MOLECULAR AND CELLULAR BIOLOGY, Sept. 1996, p. 5156–5168 Vol. 16, No. 9
0270-7306/96/$04.00+0

Surprising Deficiency of CENP-B Binding Sites in African Green Monkey α-Satellite DNA: Implications for CENP-B Function at Centromeres

ILYA G. GOLDBERG,† HANITA SAWHNEY, ANN F. PLUTA,‡
PETER E. WARBURTON,† AND WILLIAM C. EARNSHAW*

Department of Cell Biology and Anatomy, Johns Hopkins School of Medicine, Baltimore, Maryland 21205

Received 15 February 1996/Returned for modification 3 April 1996/Accepted 20 May 1996

Centromeres of mammalian chromosomes are rich in repetitive DNAs that are packaged into specialized nucleoprotein structures called heterochromatin. In humans, the major centromeric repetitive DNA, α-satellite DNA, has been extensively sequenced and shown to contain binding sites for CENP-B, an 80-kDa centromeric autoantigen. The present report reveals that African green monkey (AGM) cells, which contain extensive α-satellite arrays at centromeres, appear to lack the well-characterized CENP-B binding site (the CENP-B box). We show that AGM cells express a functional CENP-B homolog that binds to the CENP-B box and is recognized by several independent anti-CENP-B antibodies. However, three independent assays fail to reveal CENP-B binding sites in AGM DNA. Methods used include a gel mobility shift competition assay using purified AGM α-satellite, a novel kinetic electrophoretic mobility shift assay competition protocol using bulk genomic DNA, and bulk sequencing of 76 AGM α-satellite monomers. Immunofluorescence studies reveal the presence of significant levels of CENP-B antigen dispersed diffusely throughout the nuclei of interphase cells. These experiments reveal a paradox. CENP-B is highly conserved among mammals, yet its DNA binding site is conserved in human and mouse genomes but not in the AGM genome. One interpretation of these findings is that the role of CENP-B may be in the maintenance and/or organization of centromeric satellite DNA arrays rather than a more direct involvement in centromere structure.

The way in which specific DNA sequences nucleate the formation of functional centromeres has been a challenging area of study for many years. In the yeast *Saccharomyces cerevisiae*, the discovery that a small *cis*-acting element (125 bp) was necessary and sufficient to confer centromere activity (10) ultimately led to the discovery of a number of genes involved in centromere function (8, 15, 29, 30, 34, 39). Subsequently it was shown that several of these gene products exist in a multimeric complex (CBF3) that not only recognizes this DNA element but also is essential for binding to microtubules (27). Substantial efforts have been devoted to the description of centromeres from other eukaryotes at a similar level of detail. A particular difficulty in this pursuit has been that the centromeric DNA of higher eukaryotes is much more complex than that of *S. cerevisiae*. Most higher eukaryotes, including mammals, have large arrays of repeated DNA at their centromeres. It is not known to what extent these arrays play a role in centromere function. At one extreme, it is possible that these arrays fold into a tertiary structure onto which centromeric proteins assemble without the need for recognition of specific DNA sequences per se. It is also possible that within these arrays lies a small centromere-determining element that remains undiscovered, implying a mechanism similar to that in *S. cerevisiae*. Several proteins that localize to the centromeres of mammals have been recently discovered (13, 18, 23, 42, 44–46, 53, 61, 62), but their roles in centromere structure and function are largely unknown.

CENP-B is a centromeric polypeptide that is conserved in several mammalian species (18, 54). Our original observation that CENP-B levels vary dramatically between human chromosomes initially led to the suggestion that CENP-B may bind α-satellite because the amount of this DNA also varies between chromosomes (19). Subsequently, it was shown that CENP-B has a specific DNA binding activity (36). The DNA sequence that CENP-B recognizes (the CENP-B box [36]) is located in the repeats of α-satellite in humans, minor satellite in *Mus musculus* (60), and the 79-bp satellite of *Mus caroli* (31). The DNA binding domain of CENP-B is necessary and sufficient to localize this protein to the centromere in vivo (47). CENP-B is distributed throughout the centromeric heterochromatin beneath the kinetochore plates (12). The functions that CENP-B performs in this heterochromatic region of the chromosome are not known. It has been postulated (54) that CENP-B must perform a vital role in centromere function because of its conservation between humans and mice: both the DNA binding domain of CENP-B and its CENP-B box binding site are identical in these two organisms. However, cytological and molecular analysis of both normal and aberrant chromosomes suggests that the presence of significant levels of CENP-B at centromeres may not be obligatory for centromere function and is certainly not sufficient to confer mitotic activity on a centromere. For example, CENP-B is found at inactive centromeres on stable dicentric chromosomes (17). These inactive centromeres do not make productive attachments to the spindle during mitosis. Furthermore, there are some centromeres, including the Y chromosomes of mice and humans (16),

* Corresponding author. Present address: Institute of Cell and Molecular Biology, University of Edinburgh, Swann Building, The King's Buildings, Mayfield Rd., Edinburgh EH9 3JR, Scotland, United Kingdom. Phone: 44-131-650-7101. Fax: 44-131-650-7100. Electronic mail address: bill.earnshaw@ed.ac.uk.

† Present address: Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, Scotland, United Kingdom.

‡ Present address: Department of Dermatology, Johns Hopkins School of Medicine, Baltimore, MD 21205.

to which CENP-B does not seem to localize and in which CENP-B boxes cannot be detected by in situ hybridization (38). In mice, the Y chromosome seems to lack the minor satellite that carries CENP-B boxes, while in humans, although the Y chromosome does carry α-satellite, the regions of this satellite sequenced to date (55) lack CENP-B boxes.

In the present study, we reveal a surprising paradox that raises questions about current thinking on the possible role of CENP-B at centromeres. The paradox was revealed by our studies of the binding of CENP-B to the genome of the African green monkey (AGM). Although AGM cells express a functional CENP-B protein and the AGM genome contains large amounts of α-satellite DNA, this DNA appears to lack detectable binding sites for the CENP-B protein. Why a polypeptide should be expressed in the absence of its specific binding site, and why this binding site was conserved in mice and humans but lost in AGM, poses a clear paradox. We suggest that the role of CENP-B may be tied to the maintenance and/or organization of centromeric satellite DNA arrays rather than a more direct involvement in centromere structure or function.

MATERIALS AND METHODS

DNA reagents. Plasmids p17M5 and p17M15 (M5 and M15 in this work), containing monomers of human α-satellite derived from chromosome 17, have been previously described (47). The plasmids were digested with restriction endonucleases *Bam*HI and *Eco*RI (Gibco-BRL) in order to generate fragments containing the α-satellite monomers. The α-satellite fragments were purified by fast protein liquid chromatography on a MonoQ (Pharmacia) column, using a gradient of 750 to 850 mM NaCl (in 20 mM Tris-HCl [pH 7.5], 1 mM EDTA) over 30 column volumes. Concentrations of the purified DNAs were determined by A_{260} in triplicate. Fragment M15, which contains a single CENP-B box, was used as a specific competitor for CENP-B binding. Fragment M5, which is a monomer of α-satellite that does not contain a CENP-B box, was used as a control for the specific competitor and has been previously shown not to bind CENP-B (47).

Genomic DNAs (gDNAs) were isolated as described previously (4). After digestion with proteinase K, lysates were extracted twice with phenol and chloroform. The DNA was precipitated with isopropanol, resuspended in TE (10 mM Tris-HCl [pH 8.0], 1 mM EDTA), and treated with RNase A (Sigma). Following RNase treatment, the gDNAs were again extracted two times with phenol and chloroform and precipitated twice with isopropanol to remove residual oligoribonucleotides. Following purification, the gDNAs were sonicated to an average size of 750 bp, as measured by agarose gel electrophoresis. DNA concentrations were determined by A_{260} in triplicate dilutions. The A_{260}/A_{280} ratios of all gDNA preparations used were between 1.9 and 2.0.

The size of the diploid AGM genome has been estimated to be 6 pg as determined by reassociation kinetics (52). The size of the plasmid (p17M15) carrying the single high-affinity binding site is 3,131 bp or 3.4×10^{-6} pg. By adding 5.7×10^{-6} μg of plasmid to 1 μg of gDNA, one obtains DNA with the equivalent of one additional binding site per haploid genome equivalent. In making mixed competitor DNAs, if the amount (by weight) of added plasmid exceeded 1% of the total DNA, the gDNA concentration was corrected for the amount of added plasmid.

Plasmid libraries were prepared by digesting COS gDNA with *Hin*dIII and isolating the 170- and 340-bp fragments from agarose gels by using NA45 paper. The fragments were ligated to the *Hin*dIII site of Bluescript KS+ (Stratagene) and electroporated into *Escherichia coli* Top10 (Invitrogen). Random clones containing appropriately sized inserts (170 or 340 bp) were sequenced. All inserts sequenced contained α-satellite DNA.

Sequences of human and AGM α-satellite were optimally aligned by using PILEUP (Genetics Computer Group). The degree of conservation between aligned monomers was determined by using the program DISTANCES (Genetics Computer Group). Default parameters were used for both programs.

Protein reagents. Nuclei were isolated from COS and HeLa cells as described previously (54) by osmotic shock followed by Dounce homogenization. Nuclei were washed twice in hypotonic buffer [10 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (PIPES), 10 mM KCl, 2 mM $MgCl_2$, 1 mM dithiothreitol, 10 μM cytochalasin B, protease inhibitors (pH 7.4)] and resuspended with 1 pellet volume of 2× extract buffer (1× extract buffer contains 500 mM NaCl, 20 mM Tris [pH 7.5], 2 mM EDTA, and 1 mM dithiothreitol) with protease inhibitors and incubated 1 h at 4°C. Nuclear extracts were cleared by a $15,000 \times g$ spin for 30 min. Protein concentrations were measured using the Bio-Rad protein assay reagent. In cases in which protein concentrations were standardized between two extracts, extract buffer was added to the concentrated extract until both were within 1% A_{595} of each other in the Bio-Rad protein assay.

Antibodies to human CENP-B were previously described. RaL was raised against the entire CENP-B polypeptide except the N-terminal six amino acids (12). Monoclonal antibody 2D7 was described in reference 19. Its epitope is believed to lie in the C-terminal 147 amino acids of CENP-B. Polyclonal antibody 764 (46a) was raised against the entire CENP-B polypeptide expressed in *E. coli*. All polyclonal sera were raised in rabbits.

CENP-B DNA binding assays. Radioactive probes were prepared as previously described (50). Purified fragment M15 was dephosphorylated with calf intestinal alkaline phosphatase (Boehringer Mannheim) and rephosphorylated with [γ-^{32}P]ATP and polynucleotide kinase (New England Biolabs). An oligonucleotide (5′GGATCCGAATTCTTTCGTTGGAAACGGGATAAACTGCAGAAGATCTTC3′) containing the CENP-B box (underlined) was phosphorylated with [γ-^{32}P]ATP and polynucleotide kinase (New England Biolabs) and hybridized to oligonucleotide 5′GAAGATCTTC3′. The complementary strand was extended with Klenow enzyme (New England Biolabs). The labeled, double-stranded oligonucleotide was purified by polyacrylamide gel electrophoresis.

In equilibrium competition assays, the indicated amount of competitor was mixed with probe (100,000 cpm of M15 or the B-box oligonucleotide) in binding buffer (final concentrations: 10 mM Tris-HCl, 100 mM NaCl, 1 mM dithiothreitol, 1 mM EDTA, 100 μg of bovine serum albumin [BSA], 0.1% Triton X-100, 100 μg of sonicated salmon sperm DNA per ml [pH 8.0]). Nuclear extracts were added as the last component and allowed to incubate for 12 to 16 h at 4°C. The final reaction volume was 50 μl. Loading dye (20×; 0.4% bromophenol blue, 0.4% xylene cyanol FF, 25% Ficoll 400) was added to the reaction mixtures, which were loaded on polyacrylamide gels (4% acrylamide, 0.05% bisacrylamide, 2.5% glycerol, 0.5× Tris-borate-EDTA) and electrophoresed for 6 h at 4°C at 200 V. The gels were dried prior to autoradiography.

In nonequilibrium competition assays, binding reactions were set up as described above except that the probe was omitted in the initial 12- to 16-h incubation. Probe (100,000 cpm of M15) was added after preincubation with competitor and allowed to incubate for 1 h at 26°C. The reactions were then quenched on ice and prepared for electrophoresis as described above.

When binding reactions were quantitated, dried gels were exposed in PhosphorImager cassettes and scanned by a PhosphorImager (Molecular Dynamics). The electronic images of the gels were quantitated by using Collage software (Fotodyne). Equal-size squares were used to determine the total number of counts in the CENP-B complexes. The fraction bound was determined by dividing the total counts in each square by the total counts in the square where no additional competitor was used. Backgrounds were not subtracted. Similar results were obtained using ImageQuant software (Molecular Dynamics).

Immunolocalization of CENP-B. Because of difficulties encountered in localizing AGM CENP-B to centromeres, a number of different fixation and specimen preparation procedures were used, as described below. All experiments used three distinct CENP-B antibodies (two polyclonal and one monoclonal), all of which were shown to recognize native AGM CENP-B in gel mobility shift experiments (see Fig. 2). The positive control for these experiments was a rabbit antibody to CENP-C. This antibody localized to centromeres, confirming that the secondary and tertiary detection reagents were fully functional and that centromere structure was preserved under our preparation conditions.

Immunolocalization of CENP-B on mitotic chromosome spreads of COS cell lines 4-8 and C32 containing ectopic arrays of human α-satellite DNA (25) was performed as described previously (17). Briefly, chromosome spreads were prepared by swelling cells in hypotonic medium (75 mM KCl), followed by fixation with methanol-acetic acid and spreading by blowing with a discontinuous stream of air from an aquarium pump. CENP-B was then detected by immunostaining with rabbit polyclonal antibody in TEEN buffer (1 mM triethanolamine-HCl [pH 8.5], 0.2 mM NaEDTA [pH 8.5], 25 mM NaCl). Transfected human CENP-B was tagged with a doubly iterated 25-residue moiety derived from the carboxy terminus of the avian coronavirus M glycoprotein (previously known as E1) (35) and detected with antibody against the tag as previously described (47). Endogenous AGM CENP-B was detected with polyclonal antibody 764. Subsequent labeling steps with biotinylated anti-rabbit antibody (Vector Laboratories) and streptavidin-fluorescein (Bethesda Research Laboratories) were carried out in KB buffer (10 mM Tris-HCl [pH 7.5], 150 mM NaCl, 0.1% BSA, 0.1% Triton X-100) at 37°C for 30 min. Chromosomal DNA was counterstained with propidium iodide or 4′,6-diamidino-2-phenylindole (DAPI) for imaging in a confocal or conventional microscope, respectively. In control experiments using the COS 4-8 and C32 cell lines, in situ hybridization under conditions specific for the ectopic human α-satellite DNA arrays was performed as described previously (25).

Immunolocalization of CENP-B in adherent cell cultures was also performed as described previously (13). Cells growing on coverslips were fixed briefly (5 min with 4% formaldehyde at 4°C), permeabilized with KB (three washes of 2 min each at room temperature), placed at low salt in TEEN buffer to maximize accessibility of the chromatin to the antibodies (three washes of 2 min each at room temperature), and stained with the appropriate first antibody diluted in TEEN for 1 h at 37°C. Subsequent stainings with biotinylated secondary antibody and streptavidin-Texas red were carried out as described above. The low-salt swelling step was necessary to obtain reproducible CENP-C labeling of kinetochores in these cells. Cells fixed with methanol or methanol-acetic acid were negative for CENP-C immunolocalization in these experiments.

Images were captured on three different microscopes as follows. Figure 1A was obtained by using a Bio-Rad MRC 600 confocal microscope. Figure 1B is a

two-dimensional projection through a three-dimensional data set that was obtained by using a computer-driven Olympus IX-70 fluorescence light microscope attached to a Photometrics PXL cooled charge-coupled device camera (Applied Precision, Mercer Island, Wash.). The data set was subjected to a computational deconvolution algorithm to correct for out-of-focus information (1). Images in Fig. 4 were obtained by using an Olympus Vanox microscope attached to a DAGE SIT camera and driven by a modified version of Adobe Photoshop.

RESULTS

Transfected human CENP-B does not localize to AGM centromeres. In an earlier study, we demonstrated that targeting of CENP-B to centromeres in vivo is directed by the sequence-specific binding of the amino-terminal portion of the CENP-B polypeptide to the 17-bp CENP-B box (47). We noted at the time that the published consensus sequence for AGM α-satellite DNA lacks the CENP-B box consensus and therefore decided to examine the distribution of CENP-B at centromeres of AGM chromosomes.

To examine the distribution of CENP-B on AGM chromosomes, we used a method developed earlier to determine the distribution of CENP-B and CENP-C on dicentric chromosomes by indirect immunofluorescence (17). We obtained the surprising result that AGM chromosomes prepared by this method lack detectable levels of CENP-B at their centromeres (data not shown). This result suggested several possibilities: (i) our antibodies do not recognize AGM CENP-B, (ii) AGM cells lack CENP-B protein, and (iii) AGM cells lack binding sites for CENP-B protein at centromeres.

Initial support for the third possibility came when we examined the binding of transfected human CENP-B to AGM chromosomes. To have an internal control for the activity of the transfected CENP-B, we used a COS cell line in which transfected human α-satellite DNA had been stably integrated at ectopic locations in the arms of several AGM chromosomes. When these cells were transfected with epitope-tagged human CENP-B, we could reproducibly detect the presence of human CENP-B binding to the ectopic α-satellite arrays (Fig. 1A). Interestingly, these regions often appeared to form secondary constrictions of the chromosome arms. However, we never saw binding of transfected human CENP-B to the centromeres of the AGM chromosomes. This result could not be explained as being due to the expression of abnormally low levels of CENP-B in the AGM cells. Certain transfected cells grossly overexpressed human CENP-B. In those cells, the entire chromosomes were painted with CENP-B, and there was no hint of a preference of the protein for the AGM centromeres (data not shown).

In control experiments, we also detected binding of the endogenous AGM CENP-B to the ectopic human α-satellite arrays (Fig. 1B). Thus, our indirect immunofluorescence protocol is capable of staining both human and AGM CENP-B when they are bound to α-satellite DNA. As we will argue below, AGM CENP-B binds to the ectopic α-satellite by virtue of its recognition of CENP-B box motifs in the human DNA. However, neither the endogenous AGM CENP-B nor transfected human CENP-B is detectable in significant amounts at AGM centromeres by using our antibodies and methods.

These results suggested that AGM centromeres have few, if any, accessible binding sites for CENP-B. This deficit of binding sites could reflect two possibilities: (i) a true absence of CENP-B binding sites in the sequence of AGM α-satellite DNA or (ii) methylation of all free CENP-B boxes (reported to block binding of CENP-B to the CENP-B box [35a]).

AGM cells express a protein structurally and functionally similar to human CENP-B. We next wished to determine whether AGM cells express a protein similar to CENP-B and whether this protein binds to the human CENP-B box consensus sequence. We approached these questions by establishing an electrophoretic mobility shift assay (EMSA) protocol for the study of AGM CENP-B, using the human CENP-B box as a well-characterized DNA substrate for protein binding.

For this assay, nuclear extracts from COS cells were incubated with a labeled oligonucleotide containing a CENP-B box. The resulting nucleoprotein complexes were then resolved on polyacrylamide gels and visualized by autoradiography. The specificity of each complex was determined by including in the reaction specific and nonspecific unlabeled competitor DNAs. The specific competitor DNA was a cloned α-satellite monomer containing a CENP-B box (M15), while the nonspecific competitor was a different monomer that lacked a CENP-B box (M5). The only sequence shared between the oligonucleotide that was used as the probe and the human α-satellite monomer that was used as the specific competitor was the canonical 17-bp CENP-B box. As shown in Fig. 2, two of the major complexes formed in these AGM extracts could be competed for by DNA containing a CENP-B box but not by DNA lacking a CENP-B box (lanes 7 and 6, respectively). This result suggests that AGM cells express a protein with the DNA recognition properties of CENP-B.

In control experiments, we compared this EMSA activity with that in HeLa extracts by using the α-satellite monomer containing a CENP-B box both as a specific competitor and as a probe. As shown in Fig. 3, HeLa extracts (lanes 2 and 6) produced a complex with the probe that had the same mobility as complexes produced in the COS extracts (lanes 1 and 5). In this particular experiment, the majority of the signal was due to the lower-mobility complex; however, the two complexes are clearly visible in Fig. 2 and 6. These complexes were efficiently competed for with a 100-fold excess of unlabeled probe (Fig. 3, lanes 3, 4, 7, and 8). It has been previously shown (41, 63) that the two specific CENP-B box complexes visible by EMSA are due to the dimerization activity of CENP-B. The difference in mobility is probably due to a CENP-B dimer binding one or two molecules of DNA (41, 63).

Standardization of the protein concentrations in the HeLa and COS cell extracts allowed us to estimate that the specific CENP-B box binding activity of HeLa extracts was approximately fourfold higher than that of COS extracts (Fig. 3; compare lanes 2 and 5). Similar observations were made in the accompanying report (64). We conclude from these experiments that COS cells express significant levels of a protein that specifically binds to human CENP-B boxes and most likely has the same dimerization activity as human CENP-B.

To determine whether the CENP-B box binding activity in AGM extracts corresponded to CENP-B itself, we added several different antibodies that recognize human CENP-B to the EMSA reaction. As shown in Fig. 2, two polyclonal antibodies and one monoclonal antibody raised against human CENP-B caused the CENP-B box-specific complexes to be further retarded in their mobility, while preimmune serum did not. We conclude from these experiments that AGM cells express a protein that shares both DNA binding properties and epitopes with human CENP-B. In the study described in the accompanying report (64), the authors cloned an AGM CENP-B cDNA and found that it encodes a polypeptide predicted to be 98% identical to human CENP-B. The DNA binding and dimerization domains of the two proteins are 100% identical. We will refer to this protein as AGM CENP-B for the remainder of this study.

Most AGM CENP-B is distributed diffusely throughout the interphase nucleus. The knowledge that AGM cells express a CENP-B polypeptide that is recognized by antibodies to the

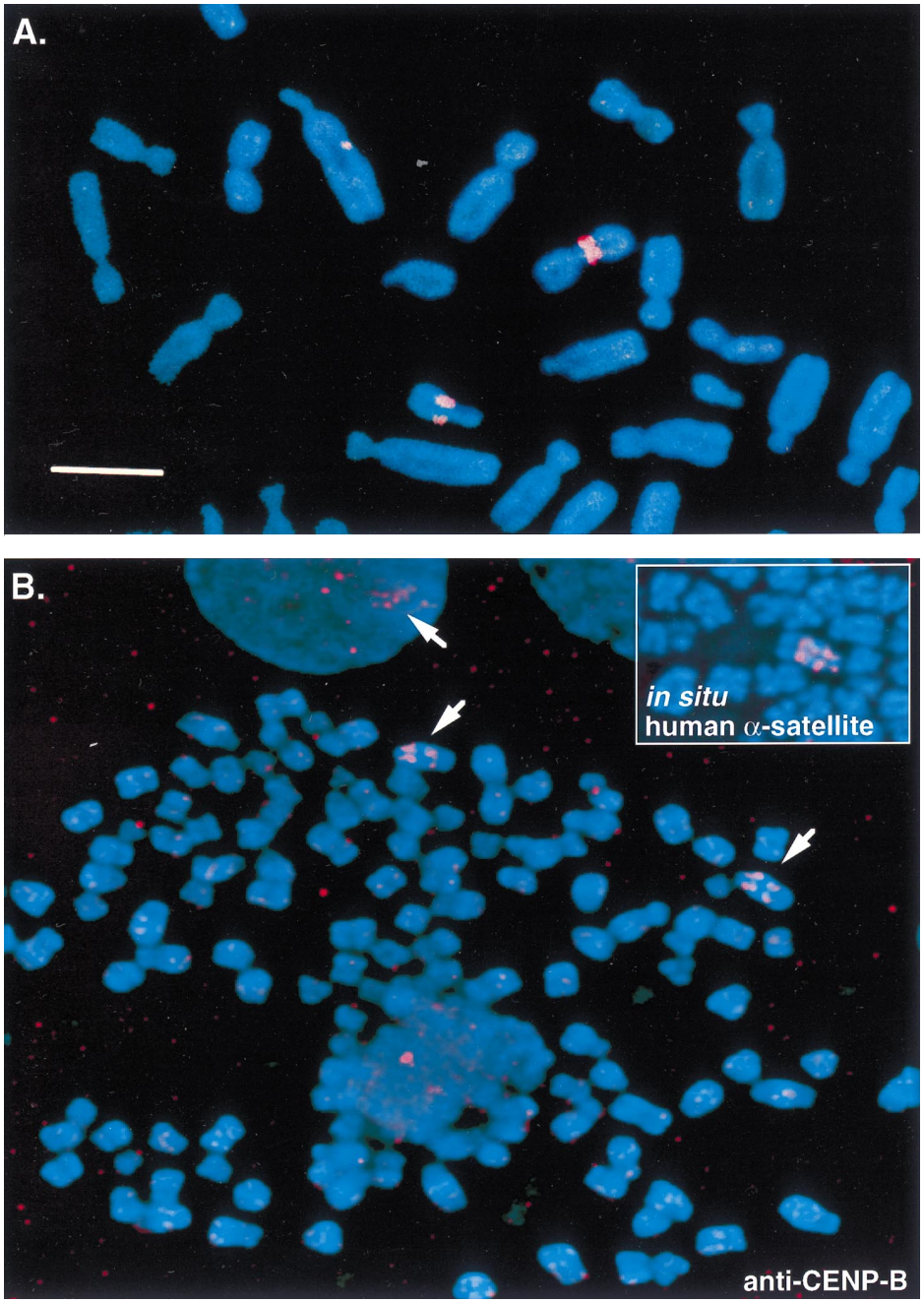

FIG. 1. (A) Human CENP-B does not localize to centromeres in transfected AGM cells. Shown is a mitotic spread of a stable COS subline (4-8) containing human α-satellite DNA integrated into the arms of several chromosomes (25). These cells were transiently transfected with a pECE vector (20) expressing human CENP-B (47). Red, antibody localization of CENP-B; blue, chromosomal DNA stained with propidium iodide. (The image shown is a false color representation of a pair of merged confocal images.) CENP-B staining is restricted to the sites of human α-satellite DNA integration on the chromosomal arms. (B) AGM CENP-B localizes to human α-satellite DNA integrated into the arms of AGM chromosomes. Shown is a mitotic spread of a stable COS subline (C32) containing an amplified region of human α-satellite DNA in one chromosome arm (25) (confirmed by in situ hybridization [inset]). The main image shows a tetraploid chromosome spread in which the binding of AGM CENP-B to two copies of this chromosome can clearly be seen (arrows). Note also the labeling in the interphase nucleus indicated with an arrow at the top. Under these conditions, in which the endogenous AGM CENP-B is clearly labeled on the human α-satellite DNA, little or no signal is seen at AGM centromeres.

human polypeptide led us to directly examine the distribution of endogenous CENP-B in cultured AGM cells. Using a variety of fixation conditions and several antibodies, including one monoclonal and three independent polyclonal antibodies to CENP-B, we were unable to reproducibly detect significant levels of CENP-B at AGM centromeres in interphase COS, Vero, or CV-1 cells. Instead, the endogenous CENP-B antigen was distributed diffusely throughout the entire nucleus, with the exception of nucleoli (Fig. 4). To control for the integrity of centromeres in these cells as well as for possible difficulties with detection of the rabbit primary antibodies, we also stained Vero cells with an antibody to CENP-C, a component of the

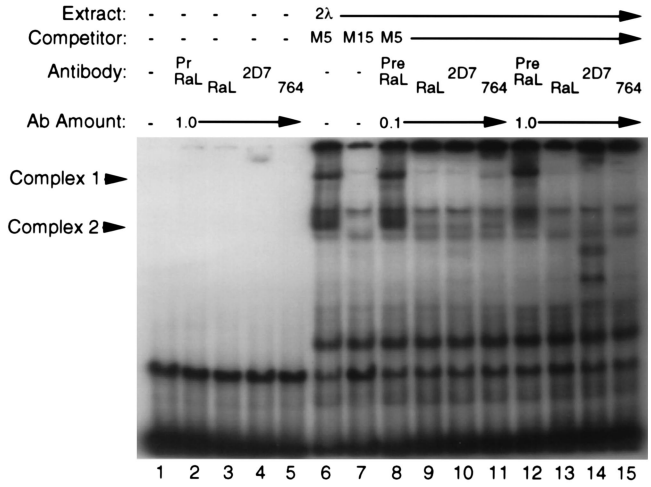

FIG. 2. COS nuclear extracts have a DNA binding activity similar to that of human CENP-B. COS nuclear extracts were incubated with a labeled oligonucleotide containing a CENP-B box. The complexes formed were electrophoresed on an acrylamide gel and visualized by autoradiography. Human α-satellite monomers were used as specific and nonspecific competitor DNAs to determine the specificity of the complexes formed. The specific competitor (M15) contained a single CENP-B box, while the nonspecific competitor (M5) did not. Antibodies (Ab) raised against human CENP-B were also added to the reactions to test if they would further retard the mobility of the complexes that are specific to the CENP-B box. RaL and 764 are polyclonal CENP-B antisera, and 2D7 is purified monoclonal immunoglobulin G from ascites fluid; 0.1 and 1.0 μl of serum and immunoglobulin G were used. We assume that the increased radioactivity in the wells to which specific antibodies were added to the COS extracts is due to CENP-B complexes retarded in their mobility by the antibodies. It is also possible that the antibodies specifically disrupt nucleoprotein complex formation.

inner kinetochore plate (49). CENP-C showed a normal punctate distribution in these cells, indicating that the cells do have properly assembled centromeres during interphase and that the protocols used in immunostaining did not disrupt centromere structure.

In some instances, faint CENP-B staining could be detected at the centromeres of mitotic chromosomes in these cells (Fig. 4). Even in these cells, however, it appeared that the bulk of CENP-B was dispersed throughout the cytoplasm, in contrast to the indirect immunofluorescence results presented in the accompanying report (64). In that study, under different fixation conditions and with a different antibody, a weak CENP-B signal could be observed in the outer regions of AGM centromeres outside the kinetochore. (In human chromosomes, CENP-B is located beneath the kinetochores in the centromeric heterochromatin [12].) It is difficult to reconcile the results of the two studies. There may be a small amount of CENP-B localized to the centromeres of AGM, which could be detected by the authors of the accompanying report but not by us (if, for example, their protocol resulted in the solubilization or extraction of much of the nonchromosomal CENP-B). Below we present evidence that the number of binding sites for this protein in AGM centromeres is so few that unless this protein has an alternate mechanism of centromere localization, the number of CENP-B molecules at each AGM centromere should be several hundred-fold less than at each human centromere, or equivalent to one or two dimers of CENP-B.

An enriched AGM α-satellite fraction is deficient in high-affinity CENP-B binding sites. Since most endogenous AGM CENP-B fails to localize to centromeres even though it retains the ability to bind CENP-B boxes, we proceeded to determine whether CENP-B boxes could be found in AGM centromeric DNA. It has been shown (22) that when AGM gDNA is digested with the restriction endonuclease *Hin*dIII, a ladder of bands with a 170-bp spacing is observed by agarose gel electrophoresis (Fig. 5). The 170- and 340-bp bands are composed largely, if not entirely, of AGM α-satellite (48). These fractions have been shown to hybridize to AGM centromeres (33, 51). We wanted to compare the abundance of CENP-B boxes in this AGM α-satellite fraction with their abundance in human α-satellite. Unfortunately, it is not possible to use restriction endonuclease digestion to obtain a comparable bulk α-satellite fraction from human gDNA. However, it is possible to predict the abundance of CENP-B boxes in humans by examining the extensive database of sequenced human α-satellite sequences. A search of GenBank for human α-satellite clones produced 255 entries containing a total of 137,606 bp, or 804 171-bp monomers. A search for the canonical CENP-B box (YTTCG TTGGAARCGGGA) in these entries produced 165 matches, or an average of one CENP-B box in five monomers. A less stringent search with only the most conserved bases of the CENP-B box (TTCGNNNNANNCGGG) (37) produced 206 matches, or an average of one CENP-B box in four monomers.

Since we could obtain large amounts of the gDNA *Hin*dIII fraction from COS cells, we used this fraction as a competitor in the EMSA. In this assay, we used a monomer of human α-satellite that contained a single CENP-B box both as a probe

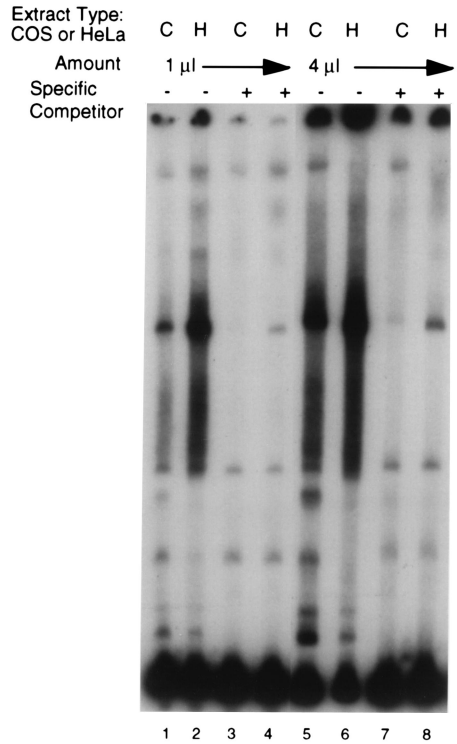

FIG. 3. COS and HeLa nuclear extracts form similar complexes with the CENP-B box. Nuclear extracts were prepared from COS (C) and HeLa (H) cells and standardized for their protein concentrations by adding extract buffer. The extracts were incubated with labeled human α-satellite monomer M15 that contained a single CENP-B box. The complexes were resolved on acrylamide gels and autoradiographed as before. Unlabeled probe in 100-fold excess was added to the lanes marked +. One and four microliters of the standardized extracts were used. The majority of the probe is shifted to the position of the low-mobility CENP-B complex, which is presumably composed of a CENP-B dimer and two molecules of DNA. The high-mobility specific complex is difficult to visualize because of a high background but is easily visible in Fig. 6. There are several minor complexes that seem to exhibit some degree of specificity; we presume these to be due to degradation products of CENP-B because their appearance is not consistent between experiments.

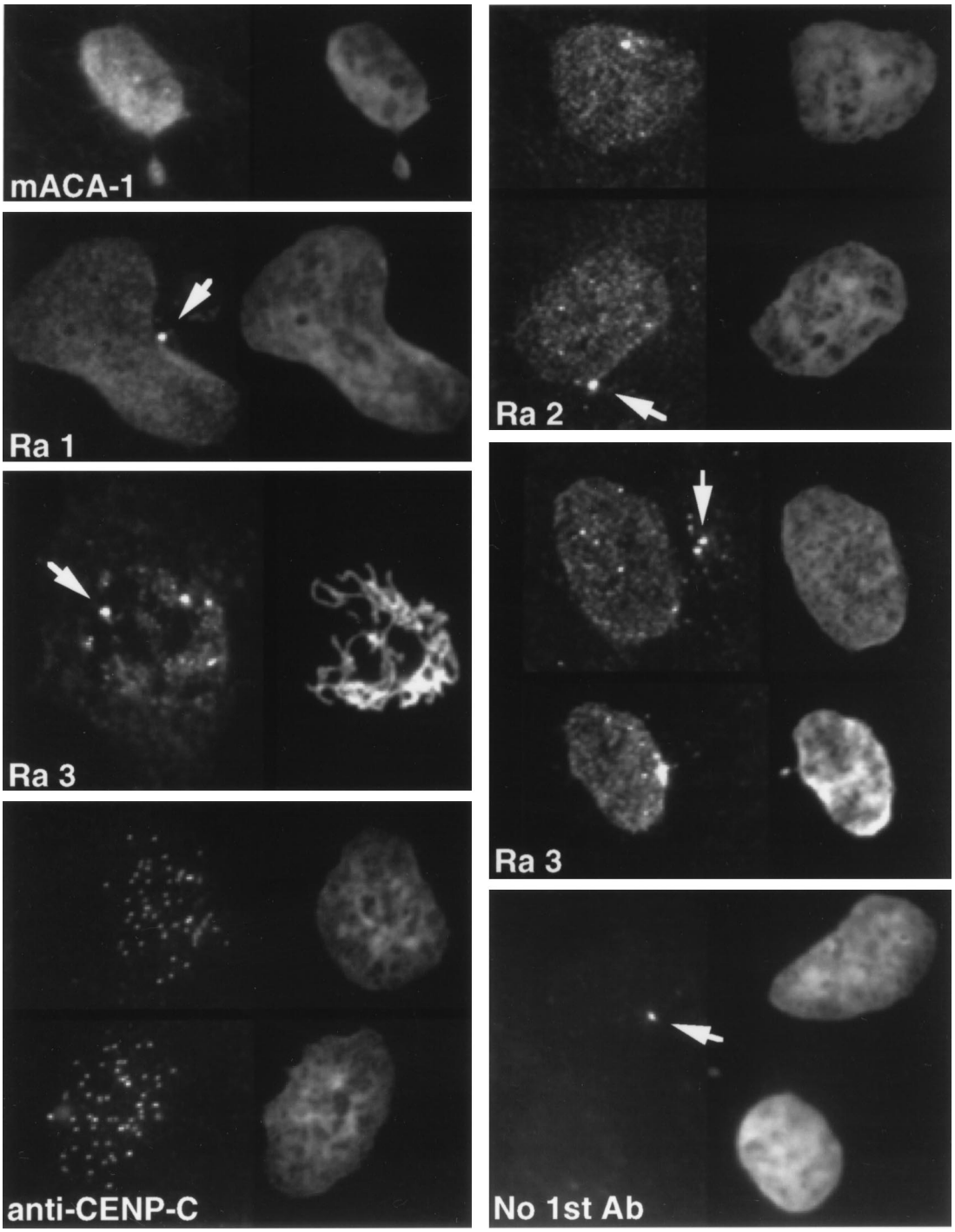

FIG. 4. Most CENP-B is distributed diffusely throughout the nucleus in AGM cells. Vero cells in culture were stained with the following antibodies (Ab) specific for human CENP-B or CENP-C: mACA-1, monoclonal anti-CENP-B elicited by immunization with a β-galactosidase fusion protein encoding the C-terminal 146 amino acids (aa) of CENP-B (19); Ra1, rabbit anti-CENP-B (ra-ACA1) elicited by immunization with a β-galactosidase fusion protein encoding the C-terminal 146 aa of CENP-B (19); Ra2, rabbit anti-CENP-B (ra-ACA2) elicited by immunization with a TrpE fusion protein encoding the C-terminal 593 aa of CENP-B (12); Ra3, rabbit anti-CENP-B elicited by immunization with full-length CENP-B expressed in *E. coli*; anti-CENP-C, rabbit anti-CENP-C CNPC3 elicited by immunization with a fusion protein encoding the C-terminal 647 aa of CENP-C expressed in *E. coli* (49). The arrowheads indicate nonspecific reactivity of the secondary antibody with centrosomes. The left panel of each pair is the antibody stain, while the right panel is DNA stained with DAPI.

and as a specific competitor. In the experiment shown in Fig. 6, we compared the competition activity of the AGM *Hin*dIII fraction with that of the human α-satellite monomer. We found that 5 ng of the human monomer competed more effectively than 5 μg of the AGM *Hin*dIII fraction (lanes 4 and 10, respectively). This result suggests that the *Hin*dIII fraction of AGM α-satellite contains less than one CENP-B binding site in 1,000 monomers, as opposed to a predicted average of

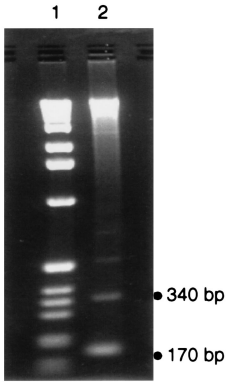

FIG. 5. COS genomic DNA digested with *Hin*dIII. Lane 1, size markers; lane 2, COS gDNA digested with *Hin*dIII. The 170- and 340-bp α-satellite bands used in making plasmid libraries are indicated.

one in 4 to 5 for human α-satellite. These results are consistent with the apparent absence of CENP-B localization to centromeres in AGM nuclei.

A novel kinetic EMSA reveals that AGM gDNA contains 3 orders of magnitude fewer CENP-B specific binding sites than human DNA. Although competition experiments with the AGM *Hin*dIII fraction suggested that CENP-B binding sites are very rare in AGM α-satellite, the possibility remained that the *Hin*dIII fraction is not representative of all α-satellite in the AGM genome and that AGM could have binding sites for CENP-B in α-satellite domains with no *Hin*dIII restriction sites. Such segregation of human α-satellite into large domains with and without CENP-B binding sites has recently been described for chromosome 21 (28). Furthermore, we wanted to address the possibility of CENP-B binding sites occurring in other regions of the genome. We therefore expanded our comparison of CENP-B binding sites between AGM and humans to cover the entire genomes of both organisms.

We designed a novel EMSA protocol that uses the full complexity of genomic DNA as a competitor in a reaction involving AGM CENP-B binding to a labeled CENP-B box. This assay should detect CENP-B binding sites wherever they might occur throughout the genome. To do this required the development of a new competition protocol for the following reasons. There are two types of binding sites in gDNA: high-affinity sites (CENP-B boxes), and low-affinity (or nonspecific) sites that constitute the remainder of the gDNA. When one considers the entire genomic complement as a competitor, the number of low-affinity sites is extremely large compared with the predicted number of high-affinity sites for any sequence-specific DNA-binding protein. In equilibrium competition assays, this multitude of nonspecific sites has the potential to bind substantial amounts of CENP-B and thus mask competition by specific sites. Since we were interested in determining the number of specific high-affinity sites throughout the entire genome, we needed to devise a strategy to minimize the effects of nonspecific competition. To do this, we performed the EMSA reaction under nonequilibrium conditions (a kinetic EMSA). COS nuclear extracts were preincubated with the gDNA competitor for an extended period of time (12 to 16 h at 4°C) in order to reach equilibrium. Probe was then added for a brief incubation (1 h at 26°C), and CENP-B–DNA complexes were resolved as usual. This protocol relies on the fact that the difference in binding constants between specific and nonspecific interactions is primarily due to a difference in off rates rather than a difference in on rates (which are usually diffusion limited). Therefore, protein that is associated nonspecifically with DNA is much more likely to dissociate and rebind to the probe in a short time than protein that is tightly associated with its specific binding site. With this protocol, we were able to increase the sensitivity of the competition assay by 10- to 100-fold (data not shown). In the experiment shown in Fig. 7, we directly compared the competition activities of HeLa and COS gDNA in the kinetic EMSA. It took ~100-fold more AGM gDNA than human (HeLa) gDNA to compete for the CENP-B complex. The slopes of the competition profiles, which are a measure of the competition activity per microgram of gDNA, also showed a 100-fold difference between humans and AGM. It is important to note that this does not necessarily mean that AGMs have precisely 100-fold fewer binding sites than humans. In fact, if the number of binding sites in the AGM genome were too small to be detected by this assay, the competition by COS gDNA could be due entirely to nonspecific binding.

To test if COS gDNA could be distinguished from a gDNA with no specific CENP-B binding sites, we compared the competition activity of COS gDNA with that of *E. coli* gDNA. We were surprised to find that *E. coli* gDNA competed more strongly for AGM CENP-B than did AGM gDNA. We obtained similar results with gDNA from two other species (salmon and the moth *Spodoptera frugiperda*), which were poorer competitors than *E. coli* but still better than AGM (data not shown). Although it is formally possible that these three control species have cryptic CENP-B binding sites, a more likely explanation is that the number of CENP-B binding sites in *E. coli*, salmon, moth, and AGM cells is too low to be detected in this assay.

Since the number of CENP-B binding sites in COS gDNA is evidently below the sensitivity range of the kinetic EMSA, we wished to determine the minimum number of sites per genome that we could detect in order to calibrate the sensitivity of the assay. To do this, we supplemented COS gDNA with a known number of additional high-affinity binding sites. The high-affinity binding sites were provided by a plasmid clone of the human α-satellite monomer (M15) used in the assay described above. We prepared DNAs containing 50 and 250 additional

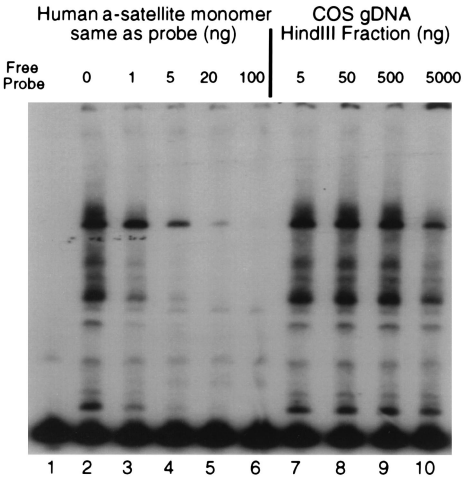

FIG. 6. The *Hin*dIII fraction of AGM gDNA competes poorly for AGM CENP-B binding. A human α-satellite monomer containing a CENP-B box was compared with the COS *Hin*dIII fraction for its ability to compete the two AGM CENP-B complexes in an EMSA. The *Hin*dIII fraction is at least 1,000 times less effective at competing for the upper CENP-B complex and at least 5,000 times less effective at competing for the lower complex.

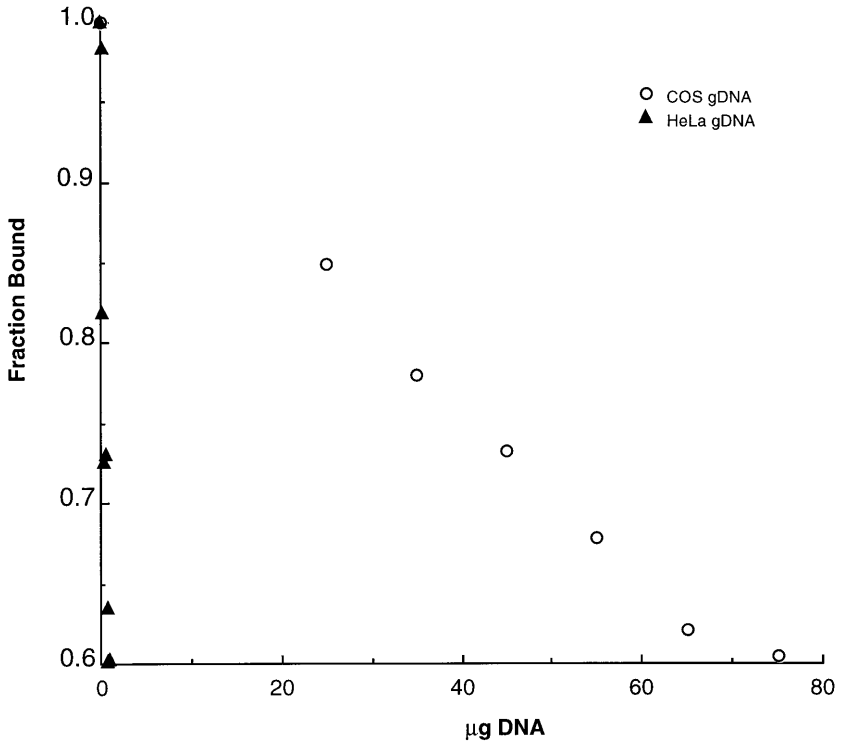

FIG. 7. COS gDNA is at least 100 times less effective than HeLa gDNA in competing for AGM CENP-B binding to a human CENP-B box. Human and COS gDNA were used as competitors in a kinetic EMSA as described in the text. Fraction bound is the counts retained in the AGM CENP-B complex in the presence of competitor divided by the counts in the AGM CENP-B complex when no competitor is added. The competition plateau is reached at 0.6, at which point a CENP-B complex is no longer observed. The reason fraction bound does not reach 0.0 is that background was not subtracted (see Materials and Methods).

binding sites per haploid genome and compared their competition profiles with that for COS gDNA. We reproducibly detected a difference in competition activity between COS gDNA alone and a mixture with as few as 50 additional binding sites per genome equivalent (Fig. 8). This result confirms the sensitivity of the kinetic EMSA and suggests that COS gDNA may have fewer than 50 binding sites per haploid genome, assuming that there are no binding sites in *E. coli*. This estimate is consistent with the low frequency of CENP-B boxes detected by the highly sensitive PCR amplification protocol used in the study reported in the accompanying paper (64).

Our results with AGM DNA are in marked contrast to those obtained when we used the kinetic EMSA to determine the frequency of CENP-B boxes in the human genome. We can predict the total number of CENP-B boxes in the human genome on the basis of our database analysis, suggesting that there is one CENP-B box for every four monomers. If 5% of the 4×10^9-bp human genome is composed of α-satellite (58), then there would be 10^6 170-bp monomers and 2.5×10^5 CENP-B boxes. We used our kinetic EMSA to test this estimate experimentally. As shown in Fig. 7, the AGM CENP-B complex can be completely competed for with as little as 1 μg of HeLa gDNA, a range in which COS gDNA gives little competition. In Fig. 9, we compared the competition activity of HeLa gDNA with that of COS gDNA to which additional CENP-B boxes had been added. We found that the concentration of CENP-B boxes in human α-satellite lies between the values seen when COS gDNA is supplemented with between 8,000 and 40,000 CENP-B boxes per genome. This experimental value could be slightly lower than the predicted value for several reasons. For example, some fraction of human CENP-B boxes may be unavailable for CENP-B binding as a result of epigenetic effects such as DNA methylation. Alternatively, α-satellite arrays rich in CENP-B boxes may be underrepresented in the genome relative to the database, thereby inflating our predicted number of boxes. Nevertheless, the experimental results are within an order of magnitude of the predicted value and show that there is a 1,000-fold difference in the number of CENP-B binding sites between AGM and human gDNAs.

The absence of functional CENP-B binding sites correlates with an absence of CENP-B box sequences. We wanted to determine if the apparent absence of CENP-B binding sites in AGM as measured by competition experiments correlates with an absence of CENP-B box sequences, or if the competition results are due to epigenetic effects. Since we could obtain purified α-satellite DNA from AGM, we prepared plasmid libraries of the 170- and 340-bp fractions (Fig. 5) and determined the DNA sequences of a number of clones chosen at random. If CENP-B boxes are as abundant in AGM as in human α-satellite, we would expect them to be easily found by this approach. This approach has the added merit that little is known about the sequences of individual AGM α-satellite monomers; a recent search of GenBank produced only nine examples. We were therefore interested in examining how closely related each individual monomer is to the consensus determined by Rosenberg et al. (48) and how closely the monomers are related to each other. In our sequences of 76 individual AGM α-satellite monomers, we were unable to find any canonical 17-bp CENP-B box sequences, nor were we able to find any 15-bp sequences that contained the nine conserved bases necessary for CENP-B binding (37). Among these 76 monomers, we would expect to find roughly 19 CENP-B boxes if the motif was as abundant in AGM as it is in humans. The accompanying paper (64) reports that of 24 randomly se-

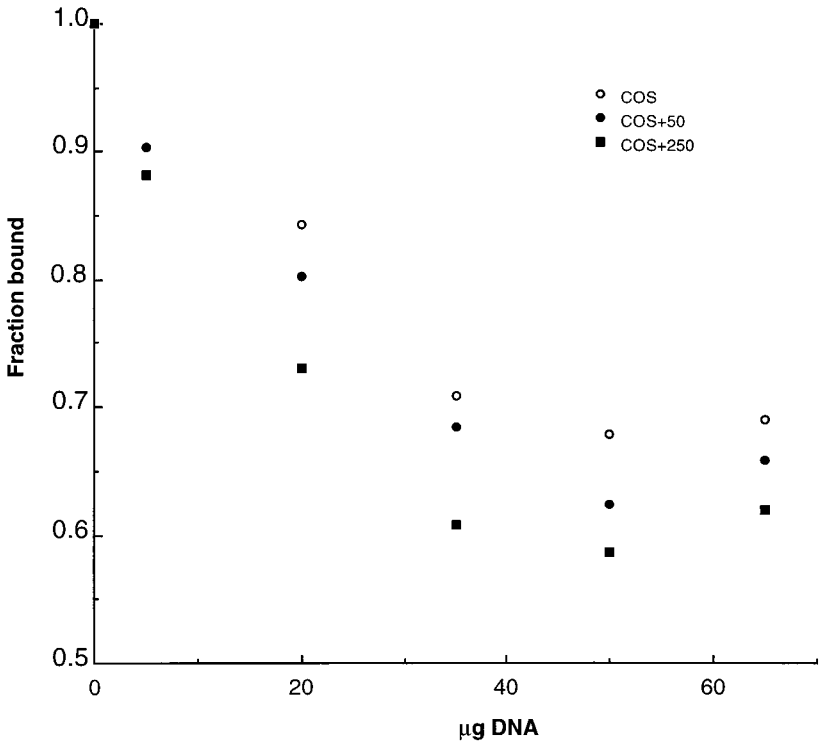

FIG. 8. Competition assay can discriminate between COS genomic DNA and COS DNA with 50 added binding sites per genome. COS gDNA was mixed with a plasmid containing a CENP-B box in order to obtain gDNAs with a known number of binding sites per genome. Mixtures containing 0, 50, and 250 CENP-B boxes per haploid genome were used as competitors for AGM CENP-B binding to a human CENP-B box in the kinetic EMSA. Fraction bound reflects the fraction of counts retained in the CENP-B complex relative to the number of counts in the CENP-B complex with no added competitor.

quenced monomers from a similar fraction of α-satellite DNA, none had the CENP-B box sequence. These results confirm directly that the abundance of CENP-B boxes in α-satellite is at least 20-fold less in AGM than it is in humans and suggest that the absence of binding sites as observed by gDNA competition is probably not solely due to epigenetic effects.

Our sequencing results also show that AGM α-satellite monomers differ significantly from their human counterparts in that they are highly homologous to each other. An optimal alignment of the 76 monomers was made and used to determine the number of mismatched bases in each monomer relative to the others. On average, each AGM α-satellite monomer was 95% identical to every other monomer. The two most divergent monomers were 84% identical, while the two most conserved were 99% identical. A similar analysis was performed with 76 randomly selected human α-satellite monomers. On average, each human monomer was 75% identical to the other human monomers, with a minimum of 60% identity and a maximum of 99%. The high degree of conservation of AGM α-satellite monomers from this gDNA fraction was also confirmed independently in the accompanying report (64).

DISCUSSION

AGM chromosomes have a greatly reduced number of specific CENP-B DNA binding sites. Although in many ways CENP-B is the best-characterized mammalian centromere protein, the biological role of this protein remains obscure. It is widely believed that CENP-B is important for the structure of the centromeric heterochromatin. For example, binding of this very acidic protein near the linker region of the nucleosome could significantly alter the structure of the 30-nm chromatin fiber (47). Alternatively, the ability of CENP-B to dimerize via a motif near its carboxy terminus (63) raises the possibility that the protein can cross-link adjacent chromatin strands (36). In this way, CENP-B might contribute to the unusually compact organization of centromeric chromatin and to the production of the primary constriction of mitotic chromosomes. Models that imply a structural role for CENP-B at centromeres are supported by the distribution and abundance of this protein and its binding sites in a variety of mammals. CENP-B is expressed at a level of 2×10^4 to 5×10^4 monomers per HeLa cell (5, 40), which corresponds to ~300 molecules per sister chromatid. In the present study, the number of CENP-B binding sites in human α-satellite DNA was predicted to be 4×10^5 per cell on the basis of a statistical analysis of the genome database and determined to be somewhat less than 8×10^4 per cell on the basis of EMSA experiments. Thus, the number of CENP-B binding sites almost certainly exceeds the number of CENP-B molecules in human cells. Given this ratio of binding sites to protein, the majority of the CENP-B protein in human cells would be predicted to be bound to its specific sites in α-satellite DNA.

In the present study, we demonstrate by three independent approaches that the AGM genome has ~100- to 1,000-fold fewer CENP-B binding sites than the HeLa genome. Yet, AGM chromosomes appear to have a normal centromeric structure, with condensed heterochromatin, a primary constriction, and normal levels of other centromeric proteins, such as CENP-C. The accompanying report (64) shows that AGM CENP-B closely resembles human CENP-B and that low levels of the protein can be detected at the periphery of AGM centromeres under certain fixation conditions. However, our data indicate that few if any AGM CENP-B dimers could target to

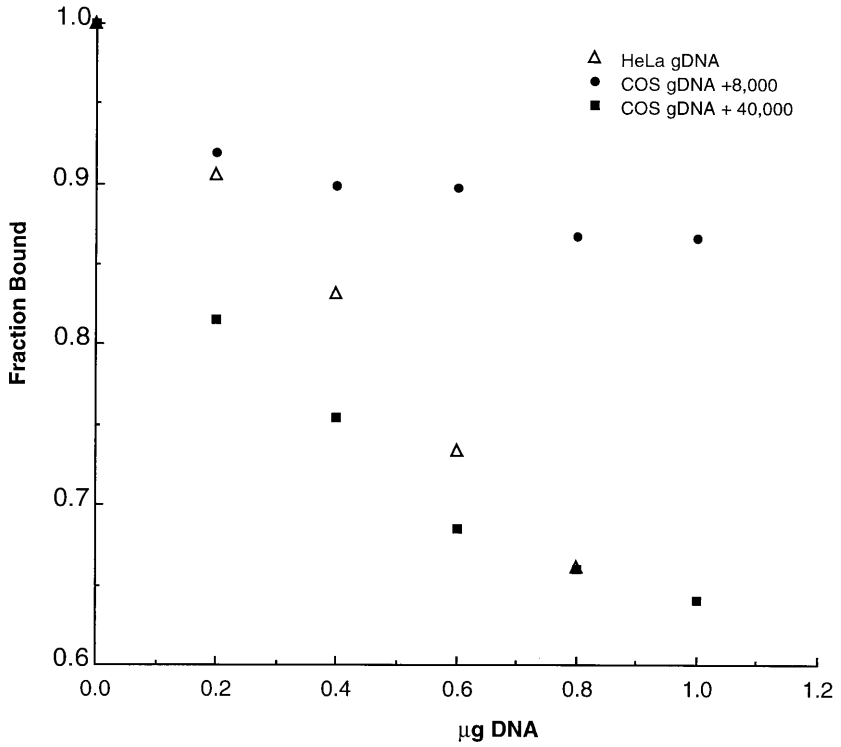

FIG. 9. HeLa gDNA contains between 8,000 and 40,000 CENP-B binding sites per genome. Kinetic EMSA competition experiments were performed as described in the text. COS gDNA was doped with a plasmid containing a CENP-B box in order to obtain DNA mixtures containing a known number of binding sites per genome. The COS standard mixtures were compared with HeLa gDNA for their activity in competing for AGM CENP-B binding to a human CENP-B box.

AGM centromeres in vivo by recognizing CENP-B boxes as previously shown for the targeting of this protein in human cells (47). We have shown in a novel kinetic EMSA that *E. coli* DNA, which presumably lacks high-affinity binding sites for CENP-B, competes slightly better than AGM genomic DNA for binding to AGM CENP-B. This assay can reproducibly detect the addition of as few as 50 CENP-B binding sites to the AGM competitor total genomic DNA. This finding confirms that we are within the range of sensitivity of the kinetic EMSA and suggests that AGM DNA most likely has on the order of $\leq$50 high-affinity CENP-B binding sites, i.e., $\leq$1 for each of the 92 chromatid centromeres. This number compares with ~300 dimers present at human centromeres (5, 37).

AGMs do not provide the only example of mitotically stable chromosomes with severely lowered levels of CENP-B. Human Y chromosome α-satellite sequences presently available in the databases lack CENP-B boxes (55), and where examined, CENP-B and CENP-B box localization to the Y chromosome could not be detected (16, 38). Also several abnormal human marker chromosomes lacking detectable α-satellite DNA altogether have been described (6, 43, 56). We have examined one such stable chromosome (derived from chromosome 11) for the presence of CENP-B and found that levels of the antigen are below detectable limits (19a).

Our results are supported by a recent study which found that CENP-B boxes may be absent from the centromeres of many primate species (24). In that study, a fluorescence in situ hybridization (FISH) assay using a CENP-B box oligonucleotide was used to probe for CENP-B boxes in several simian species. CENP-B box sequences were detected at the centromeres of humans, chimpanzees, pygmy chimpanzees, gorillas, and orangutans but not at the centromeres of gibbons, Old and New World monkeys, and prosimians (24). These results should be interpreted with caution, as sequences that are not completely homologous to the canonical CENP-B box might still have CENP-B binding activity (37). For example, this was recently noted for the 79-bp satellite of *M. caroli* (31). However, our results confirm that in the AGM, absence of signal by FISH does correlate with a substantial reduction in the number of functional CENP-B binding sites. It thus appears that the centromeres of gibbons, Old and New World monkeys, and prosimians function with a greatly reduced number of functional CENP-B binding sites compared with that found in the great apes (and in mice).

The CENP-B paradox. Our results reveal an intriguing paradox that challenges present views of CENP-B function at centromeres. cDNA cloning studies reveal that CENP-B polypeptide is highly conserved in human, mouse, and monkey genomes (19, 54, 64). This clearly implies that the polypeptide serves a role of sufficient importance for its sequence to be under positive selection. Previous to the present study, a similar claim could be made for the CENP-B DNA binding site. Conservation of this sequence between human and mouse genomes in the context of centromeric satellite DNAs that are otherwise unrelated clearly argues for the operation of positive selection pressure and thereby for an important functional role (36). However we now show, in agreement with the FISH studies of monkeys and apes (24), that a species much closer to humans than to mice appears to have a severe deficit of CENP-B boxes. This would normally be taken as evidence for the functional dispensability of the box and, by extension, of CENP-B.

How can this paradox be resolved? We have no resolution at present, but several possibilities suggest themselves.

(i) Mammals may express a second functional homolog of CENP-B that serves a similar role at centromeres but binds to

a distinct sequence in α-satellite DNA. In AGM, this homolog may substitute entirely for CENP-B, whereas in humans, this homolog could function primarily on the Y chromosome. However, two caveats must be considered with respect to this model. First, the putative homolog must be sufficiently distinct from CENP-B to avoid detection by a variety of CENP-B-specific antibodies or by Southern or Northern (RNA) hybridization. Second, if such a homolog exists, then why is CENP-B itself still expressed with such a high degree of conservation in human, mouse, and monkey genomes?

(ii) CENP-B may serve a function that is necessary in humans and mice but is not needed in AGMs. For example, the important function of CENP-B might be to impart a characteristic bend to the centromeric DNA. If such a bend were an intrinsic property of the AGM α-satellite, then CENP-B function might be superfluous (see below). However the second caveat from the preceding point would then apply.

(iii) CENP-B might have a backup mechanism for centromere targeting. This might have become the principal pathway of targeting to AGM centromeres. The mechanism would likely be based on protein-protein interactions, though low-affinity recognition of a secondary structure of the DNA cannot be excluded. The former notion fits with the observation in the accompanying report that the centromeric AGM CENP-B appears to be located peripheral to the kinetochore as defined by anticentromere autoantibodies (64). We have previously shown that these peripheral regions of the kinetochore contain little, if any, DNA (11). Our previous studies also indicated that whatever the centromere-targeting mechanism used by CENP-B in vivo, it must require the amino-terminal domain of the protein. This is also the CENP-B DNA binding domain (47, 63).

(iv) The essential function of CENP-B at centromeres may be nonstructural and may require the presence of only one or two CENP-B dimers. We discuss below the possibility that CENP-B functions in the maintenance of α-satellite DNA arrays.

(v) CENP-B may have an essential function that is distinct from its ability to associate with centromeres. This could explain the continued presence of CENP-B protein in the AGM nucleus. It is likely that such a function will be revealed only by future genetic analysis of CENP-B null mutants.

Likely roles for CENP-B in human and AGM chromosomes. To gain some insight into the possible roles of CENP-B in vivo, it may be useful to consider the differences between satellites that have CENP-B boxes and those that lack them. Human α-satellite repeats are arranged in a hierarchy: divergent tandem monomers of 171 bp are grouped in distinct linear arrangements to form higher-order repeat units that are then nearly perfectly conserved across megabases of DNA (59). Clones of these higher-order arrays hybridize either to specific chromosomes or to a small subset of the chromosome complement. Some aspects of this hierarchical structure may be conserved in *M. musculus* (32). Two divergent forms of mouse minor satellite that hybridize specifically to chromosome 2 (26) and chromosome 4 (7) have been cloned. This minor satellite, which is otherwise unrelated to α-satellite, has a CENP-B binding motif (36).

In contrast, AGM α-satellite is composed of highly homogeneous 170-bp repeats, which appear to possess a monomeric rather than hierarchical organization (3). The original consensus sequence for AGM was determined by bulk sequencing a purified *Hin*dIII fraction. The fact that specific bands were observed on the sequencing gels indicated a priori that each position in the 170-bp monomer must be relatively conserved. This hypothesis is confirmed by the sequences of 100 AGM monomers from a similar fraction of the AGM genome reported in this and the accompanying report (64). Inspection of the base frequencies in several regions of the consensus sequence of human α-satellite DNA (9), most notably in the region occupied by the CENP-B box, indicates that considerable heterogeneities exist in human α-satellite sequences. These would preclude unambiguous identification of bases in a bulk-sequencing experiment. Thus, human α-satellite is composed of more divergent monomers than is AGM α-satellite.

It has been proposed that all primate α-satellite belongs to two distinct homology groupings, which probably arose from two distinct ancestral monomers (2). Monomers from group B are abundant on all human chromosomes but have never been found on the Y chromosome. Monomers from group A have been found on only a subset of the human chromosomes but are abundant on the Y chromosome. This distribution of A and B monomers is similar in gorilla, orangutan, and chimpanzee cells (3). However, it appears that AGM monomers all belong to the A group (3). The main sequence differences between A and B monomers cluster in the region occupied by the CENP-B box. Consensus sequences of B monomers have at most one mismatch to the CENP-B box, while consensus sequences of A monomers have at least five mismatches (3). If AGM did inherit only one of the two ancestral monomers (the one without a CENP-B box), then this would explain why the α-satellite in this organism is much more homogeneous than it is in other primates and why it seems to lack CENP-B boxes. One role for CENP-B boxes may be to recruit CENP-B in order to somehow compensate for these heterogeneities in the centromeric repetitive DNA of humans and mice, a role that would be unnecessary in AGM α-satellite.

One such role for CENP-B might be the induction or modulation of DNA bends in centromeric satellite DNA. It is possible that the primary constriction is formed not because specific DNA-binding proteins are recruited to high-affinity sites on the centromeric DNA but because formation of a primary constriction is an intrinsic property of the repeats that make up centromeric DNA. In a recent study (21), computer analysis was used to predict the bending potential of the DNA in several types of satellite DNA. This analysis showed that even though the sequences of the repeats and the repeat lengths were not conserved, the spacing of the bending potential in the repeats was nonetheless conserved. The conserved pattern consisted of two 50- to 60-bp bending elements separated by a 20- to 30-bp region of low curvature. The spacing of the bends is just long enough to accommodate a nucleosome core and a small linker. If regions of repeated DNA sequence prone to bending have a conserved spacing just large enough to accommodate a nucleosome, a large cooperative force would result to phase the nucleosomes along this DNA. The combination of tightly and evenly spaced phased nucleosomes may be the driving force for formation of primary constrictions. One role of CENP-B might be to enforce a particular ordered nucleosome phasing on the human alphoid array, with its relatively variable underlying sequence. Perhaps the homogeneous α-satellite AGM DNA array can adopt a sufficiently ordered nucleosome phase on its own, without the need for CENP-B. Alternatively, CENP-B may directly cause the human α-satellite DNA array to adopt a suitable frequency of bends. If the homogeneous AGM α-satellite DNA array can adopt a suitable bend on its own, then this role of CENP-B might be dispensable.

A second possible role of CENP-B is in the remodeling of satellite arrays at centromeres. In a study of the molecular basis of recombination events found in 13 variant α-satellite higher-order repeat units (57), it was found that these events

cluster in a 20- to 25-bp region that borders the CENP-B box. This finding suggests that dimerized CENP-B bridging CENP-B boxes might have a role either in the unequal crossing over or subsequent recombination events that lead to conversion within satellite arrays. This observation is especially interesting because one protein with sequence homology to CENP-B is an open reading frame in the *Drosophila* transposon pogo. Although this open reading frame has not been shown to produce a transposase, it does have the D35E motif present in many transposases (14). Thus, it is possible that in AGM, in which the α-satellite arrays apparently lack CENP-B boxes, the protein remains essentially as an evolutionary vestige. It will be informative in the future to determine whether the α-satellite rearrangements that occur in humans continue to occur in AGM.

We are left with a paradox: If the function of CENP-B at centromeres is expendable, then why are the protein and its binding sites conserved between human and mouse genomes? If the function of CENP-B at centromeres is essential, then why are its binding sites lost in African green monkeys? A resolution to this paradox can come only from further studies of the nature of CENP-B and the nature of centromeric satellites.

ACKNOWLEDGMENTS

We acknowledge Rebecca Bernat for first suggesting that CENP-B might have a role in the maintenance of α-satellite DNA arrays and the members of the laboratory (plus one anonymous referee) for constructive criticisms of the manuscript.

This work was supported by grant GM35212 from the NIH and a Principal Research Fellowship from the Wellcome Trust.

ADDENDUM IN PROOF

A recent study of DNA transposons in the human genome has identified a new class of elements related to *Drosophila pogo* (A. F. A. Smit and A. D. Riggs, Proc. Natl. Acad. Sci. USA **93:**1443–1448, 1996). These elements, *Tigger1* and *-2*, have an open reading frame that resembles CENP-B. Furthermore, the terminal inverted repeats of *Tigger2* contain a sequence similar to that of the minimal CENP-B box. This study further supports the possibilities that CENP-B is derived from a transposase and that it transferred laterally between diverse species.

REFERENCES

1. **Agard, D. A., Y. Hiraoka, P. J. Shaw, and J. W. Sedat.** 1989. Fluorescence microscopy in three dimensions. Methods Cell Biol. **30:**353–377.
2. **Alexandrov, I. A., T. D. Mashkova, T. A. Akopian, L. I. Medvedev, L. L. Kisselev, S. P. Mitkevich, and Y. B. Yurov.** 1991. Chromosome-specific alpha satellites: two distinct families on human chromosome 18. Genomics **11:**15–23.
3. **Alexandrov, I. A., L. I. Medvedev, L. L. Kisselev, L. Y. Romanova, and Y. B. Yurov.** 1993. Definition of a new alpha satellite suprachromosomal family characterized by monomeric organization. Nucleic Acids Res. **21:**2209–2215.
4. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1991. Current protocols in molecular biology. John Wiley & Sons, New York.
5. **Bernat, R. L.** Ph.D. thesis. Johns Hopkins University School of Medicine, Baltimore, Md.
6. **Blennow, E., H. Telenius, D. de Vos, C. Larsson, P. Henriksson, Ö. Johansson, N. P. Carter, and M. Nordenskjöld.** 1994. Tetrasomy 15q: two marker chromosomes with no detectable alpha-satellite DNA. Am. J. Hum. Genet. **54:**877–883.
7. **Broccoli, D., K. T. Trevor, O. J. Miller, and D. A. Miller.** 1991. Isolation of a variant family of mouse minor satellite DNA that hybridizes preferentially to chromosome 4. Genomics **10:**68–74.
8. **Cai, M., and R. W. Davis.** 1990. Yeast centromere binding protein CBF1, of the helix-loop-helix protein family, is required for chromosome stability and methionine protrophy. Cell **61:**437–446.
9. **Choo, K. H., B. Vissel, A. Nagy, E. Earle, and P. Kalitsis.** 1991. A survey of the genomic distribution of alpha satellite DNA on all human chromosomes, and derivation of a new consensus sequence. Nucleic Acids Res. **19:**1179–1182.
10. **Clarke, L., and J. Carbon.** 1980. Isolation of a yeast centromere and construction of functional small circular chromosomes. Nature (London) **287:**504–509.
11. **Cooke, C. A., D. P. Bazett-Jones, W. C. Earnshaw, and J. B. Rattner.** 1993. Mapping DNA within the mammalian kinetochore. J. Cell Biol. **120:**1083–1091.
12. **Cooke, C. A., R. L. Bernat, and W. C. Earnshaw.** 1990. CENP-B: a major human centromere protein located beneath the kinetochore. J. Cell Biol. **110:**1475–1488.
13. **Cooke, C. A., M. M. S. Heck, and W. C. Earnshaw.** 1987. The INCENP antigens: movement from the inner centromere to the midbody during mitosis. J. Cell Biol. **105:**2053–2067.
14. **Doak, T. G., F. P. Doerder, C. L. Jahn, and G. Herrick.** 1994. A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. Proc. Natl. Acad. Sci. USA **91:**942–946.
15. **Doheny, K. F., P. K. Sorger, A. A. Hyman, S. Tugendreich, F. Spencer, and P. Hieter.** 1993. Identification of essential components of the S. cerevisiae kinetochore. Cell **73:**761–774.
16. **Earnshaw, W. C., R. L. Bernat, C. A. Cooke, and N. F. Rothfield.** 1991. Role of the centromere/kinetochore in cell cycle control. Cold Spring Harbor Symp. Quant. Biol. **56:**675–685.
17. **Earnshaw, W. C., H. Ratrie, and G. Stetten.** 1989. Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. Chromosoma **98:**1–12.
18. **Earnshaw, W. C., and N. Rothfield.** 1985. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. Chromosoma **91:**313–321.
19. **Earnshaw, W. C., K. F. Sullivan, P. S. Machlin, C. A. Cooke, D. A. Kaiser, T. D. Pollard, N. F. Rothfield, and D. W. Cleveland.** 1987. Molecular cloning of cDNA for CENP-B, the major human centromere autoantigen. J. Cell Biol. **104:**817–829.

19a. **Earnshaw, W. C., and J. Zackowski.** Unpublished data.

20. **Ellis, L., E. Clauser, D. O. Morgan, M. Edery, R. A. Roth, and W. J. Rutter.** 1986. Replacement of insulin receptor tyrosine residues 1162 and 1163 compromises insulin-stimulated kinase activity and uptake of 2-deoxyglucose. Cell **45:**721–732.
21. **Fitzgerald, D. J., G. L. Dryden, E. C. Bronson, J. S. Williams, and J. N. Anderson.** 1994. Conserved patterns of bending in satellite and nucleosome positioning DNA. J. Biol. Chem. **269:**21303–21314.
22. **Gruss, P., and G. Sauer.** 1975. Repetitive primate DNA containing the recognition sequences for two restriction endonucleases which generate cohesive ends. FEBS Lett. **60:**85–88.
23. **Guldner, H. H., H.-J. Lakomek, and F. A. Bautz.** 1984. Human anti-centromere sera recognise a 19.5 kD non-histone chromosomal protein from HeLa cells. Clin. Exp. Immunol. **58:**13–20.
24. **Haaf, T., A. G. Matera, J. Wienberg, and D. C. Ward.** 1995. Presence and abundance of CENP-B box sequences in great ape subsets of primate-specific alpha-satellite DNA. J. Mol. Evol. **41:**487–491.
25. **Haaf, T., P. E. Warburton, and H. F. Willard.** 1992. Integration of human α-satellite DNA into simian chromosomes: centromere protein binding and disruption of normal chromosome segregation. Cell **70:**1–20.
26. **Hayashi, T., H. Ohtsuka, K. Kuwabara, Y. Mafune, N. Miyashita, K. Moriwaki, Y. Takahashi, and R. Kominami.** 1993. A variant family of mouse minor satellite located on the centromeric region of chromosome 2. Genomics **17:**490–492.
27. **Hyman, A. A., K. Middleton, M. Centola, T. J. Mitchison, and J. Carbon.** 1992. Microtubule-motor activity of a yeast centromere-binding protein complex. Nature (London) **359:**533–536.
28. **Ikeno, M., H. Masumoto, and T. Okazaki.** 1994. Distribution of CENP-B boxes reflected in CREST centromere antigenic sites on long-range alpha-satellite DNA arrays of human chromosome 21. Hum. Mol. Genet. **3:**1245–1257.
29. **Jiang, W., J. Lechner, and J. Carbon.** 1993. Isolation and characterization of a gene (*CBF2*) specifying a protein component of the budding yeast kinetochore. J. Cell Biol. **121:**513–519.
30. **Jiang, W., K. Middleton, H.-J. Yoon, C. Fouquet, and J. Carbon.** 1993. An essential yeast protein, CBF5p, binds in vitro to centromeres and microtubules. Mol. Cell. Biol. **13:**4884–4893.
31. **Kipling, D., A. R. Mitchell, H. Masumoto, H. E. Wilson, L. Nicol, and H. J. Cooke.** 1995. CENP-B binds a novel centromeric sequence in the Asian mouse *Mus caroli*. Mol. Cell. Biol. **15:**4009–4020.
32. **Kipling, D., H. E. Wilson, A. R. Mitchell, B. A. Taylor, and H. J. Cooke.** 1994. Mouse centromere mapping using oligonucleotide probes that detect variants of the minor satellite. Chromosomal **103:**46–55.
33. **Kurnit, D. M., and J. J. Maio.** 1973. Subnuclear redistribution of DNA species in confluent and growing mammalian cells. Chromosoma **42:**23–36.
34. **Lechner, J., and J. Carbon.** 1991. A 240 kd multisubunit protein complex, CBF3, is a major component of the budding yeast centromere. Cell **64:**717–725.
35. **Machamer, C. E., and J. K. Rose.** 1987. A specific transmembrane domain of

a coronavirus E1 glycoprotein is required for its retention in the Golgi region. J. Cell Biol. **105:**1205–1214.

35a. **Masumoto, H.** Personal communication.

36. **Masumoto, H., H. Masukata, Y. Muro, N. Nozaki, and T. Okazaki.** 1989. A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. J. Cell Biol. **109:** 1963–1973.

37. **Masumoto, H., K. Yoda, M. Ikeno, K. Kitagawa, Y. Muro, and T. Okazaki.** 1993. Properties of CENP-B and its target sequence in α satellite DNA. NATO ASI Ser. Ser. H **72:**31–43.

38. **Matera, A. G., and D. C. Ward.** 1992. Oligonucleotide probes for the analysis of specific repetitive DNA sequences by fluorescence in situ hybridization. Hum. Mol. Genet. **1:**535–539.

39. **Middleton, K., and J. Carbon.** 1994. *KAR3*-encoded kinesin is a minus-end-directed motor that functions with centromere binding proteins (CBF3) on an in vitro yeast kinetochore. Proc. Natl. Acad. Sci. USA **91:**7212–7216.

40. **Muro, Y., H. Masumoto, T. Okazaki, and M. Ohashi.** 1991. Purification of a human centromere antigen (CENP-B) and application of DNA immunoprecipitation to quantitative assay for anti-CENP-B antibodies. J. Invest. Dermatol. **97:**378–380.

41. **Muro, Y., H. Masumoto, K. Yoda, N. Nozaki, M. Ohashi, and T. Okazaki.** 1992. Centromere protein B assembles human centromeric α-satellite DNA at the 17-bp sequence, CENP-B box. J. Cell Biol. **116:**585–596.

42. **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 1 of Drosophila. Chromosome Res. **2:**245–253.

43. **Ohashi, H., K. Wakui, K. Ogawa, T. Okano, N. Niikawa, and Y. Fukushima.** 1994. A stable acentric chromosome: possible existence of an intercalary ancient centromere at distal 8p. Am. J. Hum. Genet. **55:**1202–1208.

44. **Palmer, D. K., and R. L. Margolis.** 1985. Kinetochore components recognized by human autoantibodies are present on mononucleosomes. Mol. Cell. Biol. **5:**173–186.

45. **Pankov, R., M. Lemieux, and R. Hancock.** 1990. An antigen located in the kinetochore region in metaphase and on polar microtubule ends in the midbody region in anaphase, characterized using a monoclonal antibody. Chromosoma **99:**95–101.

46. **Pfarr, C. M., M. Coue, P. M. Grissom, T. S. Hays, M. E. Porter, and J. R. McIntosh.** 1990. Cytoplasmic dynein is localized to kinetochores during mitosis. Nature (London) **345:**263–265.

46a. **Pluta, A. F.** Unpublished data.

47. **Pluta, A. F., N. Saitoh, I. Goldberg, and W. C. Earnshaw.** 1992. Identification of a subdomain of CENP-B that is necessary and sufficient for targeting to the human centromere. J. Cell Biol. **116:**1081–1093.

48. **Rosenberg, H., M. Singer, and M. Rosenberg.** 1978. Highly reiterated sequences of SIMIANSIMIANSIMIANSIMIANSIMIAN. Science **200:**394–402.

49. **Saitoh, H., J. E. Tomkiel, C. A. Cooke, H. R. Ratrie, M. Maurer, N. F. Rothfield, and W. C. Earnshaw.** 1992. CENP-C, an autoantigen in scleroderma, is a component of the human inner kinetochore plate. Cell **70:**115–125.

50. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

51. **Segal, S., M. Garner, M. F. Singer, and M. Rosenberg.** 1976. In situ hybridization of repetitive monkey genome sequences isolated from defective simian virus 40 DNA. Cell **9:**247–257.

52. **Singer, D. S.** 1979. Arrangement of a highly repeated DNA sequence in the genome and chromatin of the African green monkey. J. Biol. Chem. **254:** 5506–5514.

53. **Steuer, E. R., L. Wordeman, T. A. Schroer, and M. P. Sheetz.** 1990. Localization of cytoplasmic dynein to mitotic spindles and kinetochores. Nature (London) **345:**266–268.

54. **Sullivan, K. F., and C. A. Glass.** 1991. CENP-B is a highly conserved mammalian centromere protein with homology to the helix-loop-helix family of proteins. Chromosoma **100:**360–370.

55. **Tyler-Smith, C., and W. R. A. Brown.** 1987. Structure of the major block of alphoid satellite DNA on the human Y chromosome. J. Mol. Biol. **195:**457–470.

56. **Voullaire, L. E., H. R. Slater, V. Petrovic, and K. H. A. Choo.** 1993. A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? Am. J. Hum. Genet. **52:**1153–1163.

57. **Warburton, P. E., J. S. Waye, and H. S. Willard.** 1993. Nonrandom localization of recombination events in human alpha satellite repeat unit variants: implications for higher-order structural characteristics within centromeric heterochromatin. Mol. Cell. Biol. **13:**6520–6529.

58. **Waye, J. S., and H. Willard.** 1986. Structure, organization, and sequence of alpha satellite DNA from human chromosome 17: evidence for evolution by unequal crossing-over and an ancestor pentamer repeat shared with the human X chromosome. Mol. Cell. Biol. **6:**3156–3165.

59. **Willard, H. F.** 1990. Centromeres of mammalian chromosomes. Trends Genet. **6:**410–416.

60. **Wong, A. K. C., and J. B. Rattner.** 1988. Sequence organization and cytological localization of the minor satellite of mouse. Nucleic Acids Res. **16:** 11645–11661.

61. **Wordeman, L., and T. J. Mitchison.** 1995. Identification and partial characterization of mitotic centromere-associated kinesin, a kinesin-related protein that associates with centromeres during mitosis. J. Cell Biol. **128:**95–105.

62. **Yen, T. J., D. A. Compton, W. C. Earnshaw, and D. W. Cleveland.** 1991. CENP-E, a human centromere associated protein released from chromosomes at the onset of anaphase. EMBO J. **10:**1245–1991.

63. **Yoda, K., K. Kitagawa, H. Masumoto, Y. Muro, and T. Okazaki.** 1992. A human centromere protein, CENP-B, has a DNA binding domain containing four potential α helices at the NH2 terminus, which is separable from dimerizing activity. J. Cell Biol. **119:**1413–1427.

64. **Yoda, K., T. Nakamura, H. Masumoto, N. Suzuki, K. Kitagawa, M. Nakano, A. Shinjo, and T. Okazaki.** 1996. Centromere protein B of African green monkey cells: gene structure, cellular expression, and centromeric localization. Mol. Cell. Biol. **16:**5169–5177.