

# Identification of Overlapping DNA-Binding and Centromere-Targeting Domains in the Human Kinetochores Protein CENP-C

CHARLES H. YANG,<sup>1†</sup> JOHN TOMKIEL,<sup>1‡</sup> HISATO SAITOH,<sup>1§</sup> DANIEL H. JOHNSON,<sup>2||</sup>  
AND WILLIAM C. EARNSHAW<sup>1\*</sup>

*Department of Cell Biology and Anatomy, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205,<sup>1</sup>  
and Department of Biochemistry and Molecular Biology, University of Miami School of Medicine,  
Miami, Florida 33136-1019<sup>2</sup>*

Received 15 February 1996/Accepted 18 April 1996

**The kinetochore in eukaryotes serves as the chromosomal site of attachment for microtubules of the mitotic spindle and directs the movements necessary for proper chromosome segregation. In mammalian cells, the kinetochore is a highly differentiated trilaminar structure situated at the surface of the centromeric heterochromatin. CENP-C is a basic, DNA-binding protein that localizes to the inner kinetochore plate, the region that abuts the heterochromatin. Microinjection experiments using antibodies specific for CENP-C have demonstrated that this protein is required for the assembly and/or stability of the kinetochore as well as for a timely transition through mitosis. From these observations, it has been suggested that CENP-C is a structural protein that is involved in the organization of the kinetochore. In this report, we wished to identify and map the functional domains of CENP-C. Analysis of CENP-C truncation mutants expressed *in vivo* demonstrated that CENP-C possesses an autonomous centromere-targeting domain situated at the central region of the CENP-C polypeptide. Similarly, *in vitro* assays revealed that a region of CENP-C with the ability to bind DNA is also located at the center of the CENP-C molecule, where it overlaps the centromere-targeting domain.**

The centromere of eukaryotic chromosomes is essential for the proper segregation of chromosomes during mitosis and meiosis (8, 55). In addition to serving as the site of attachment for microtubules, centromeres regulate the generation of force that is responsible for the various chromosomal movements along microtubules. Centromeres also mediate the attachment of sister chromatids during early mitosis and regulate their disengagement at the metaphase-anaphase transition. Centromere function in mitosis is monitored by the metaphase cell cycle control checkpoint.

The centromere of budding *Saccharomyces cerevisiae* is a nucleoprotein complex in which specific *cis*-acting centromeric DNA sequences (the *CEN* locus) serve as binding sites for proteins that interact with one another, with microtubules, and with motor molecules (36). This *CEN* locus occupies a 125-bp DNA fragment that confers mitotic stability after it is inserted into plasmid vectors (5, 15, 43). A protein complex, *cbf3*, that binds to a conserved 25-bp element within the *CEN* locus can mediate the binding of microtubules to the DNA (18, 45) and, under some circumstances, promote translocation along microtubules *in vitro* (16, 27).

A comparable functional understanding of the mammalian

centromere at the molecular level has yet to be achieved because of, in part, the considerably greater size and complexity of this structure compared with that of the yeast centromere. Mammalian centromeres typically encompass several million bases of DNA (53, 55) and are composed predominantly of repetitive satellite DNAs that are packaged into a specialized type of chromatin known as heterochromatin. Situated at the surface of the centromeric heterochromatin is the kinetochore, a structure that serves as the site of attachment for microtubules of the mitotic spindle (39). Ultrastructural analyses have shown that the kinetochore is a trilaminar disc composed of electron-dense inner and outer plates that are separated by an electron-translucent interzone (25, 39).

It has been proposed that all three layers of the kinetochore are composed of specialized chromatin fibers that loop between different levels of this structure (25, 38, 59). However, more recent experiments have shown that while DNA can be detected at the inner plate of the kinetochore, it is apparently absent from the outer plate (6). This result suggests that the kinetochore outer plate may be predominantly a proteinaceous structure tethered atop the centromeric heterochromatin.

A number of proteins that resemble known microtubule motor molecules of the cytoplasmic dynein and kinesin super-families have been shown to localize, at least transiently, in the vicinity of a fibrous corona coating the outer surface of the kinetochore outer plate (35, 46, 56, 57). Thus, the kinetochore periphery may be involved in tethering spindle microtubules to the chromosome and in generating mitotic forces necessary for chromosome movements. In contrast, although very little is known about the role of the inner plate, antibody injection experiments (see below) suggest that this structure may be important in kinetochore assembly.

Despite concerted efforts to identify centromere components in recent years, very few intrinsic structural components of the kinetochores have been identified. The best known of

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

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

‡ Present address: Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI 48202.

§ Present address: Laboratory of Molecular Embryology, National Institute of Child Health and Human Development, Bethesda, MD 20892-5430.

|| Present address: 9391 Southwest 212th Terr., Miami, FL 33189.

these are centromere proteins A, B, and C (CENP-A, B, and C) (10), which were identified with centromere-specific autoantibodies from sera of patients afflicted with scleroderma spectrum diseases (10, 29, 31, 42). All three CENP antigens bind DNA, and they may therefore be involved in the organization of the centromeric heterochromatin that underlies the kinetochore.

CENP-A ( $M_r$ , 17,000) is a divergent H3-like histone that is present in nucleosome-like complexes. CENP-A may be involved in establishing the differentiated chromatin structure at the centromere (33, 34, 49). CENP-B ( $M_r$ , 80,000), the most extensively studied centromere protein to date (11), is a sequence-specific DNA-binding protein (23) that localizes to the centromeric heterochromatin beneath the kinetochore (7). The DNA-binding activity of CENP-B maps to its amino terminus (48, 58). A detailed analysis of CENP-B-DNA binding suggests that the protein likely interacts with DNA in a novel manner (58). This sequence-specific DNA-binding activity of CENP-B has been shown to be necessary and sufficient to target the protein to centromeres *in vivo* (37). Because CENP-B binds specifically to  $\alpha$ -satellite DNA, the major DNA component of human centromeres, it has been suggested that this protein plays a role in the higher-order organization of the centromeric heterochromatin (30, 37).

CENP-C ( $M_r$ , 140,000) is a basic protein that is localized in the inner kinetochore plate (41). Antibody microinjection experiments revealed that CENP-C is required for determination of the diameter of the kinetochore disk and possibly for the overall assembly and/or stability of the kinetochore (51). CENP-C function is also required for a timely progression through mitosis. Cells injected with anti-CENP-C antibodies are delayed extensively in metaphase, although they do eventually enter anaphase (51). CENP-C shares three regions of amino acid similarity with an *S. cerevisiae* protein, MIF2p, that has been shown to be essential for the completion of mitosis in the budding yeast cell (2, 3, 26). It has been proposed that MIF2p may be a yeast centromere protein (26). In support of this view, MIF2 mutants display a number of defects in chromosome segregation and genetically interact with mutants having mutations in several of the confirmed centromere proteins.

While these observations are consistent with CENP-C being a structural protein that is important for proper chromosome segregation, nothing is known about the biochemical mechanisms by which this protein accomplishes its functions. Early observations with cloned CENP-C expressed in *Escherichia coli* led us to suggest that this protein binds DNA (41). This supposition has since been confirmed, and the DNA-binding domain roughly mapped by deletion analysis to a region at the center of the molecule (47). Given the location of CENP-C at the inner kinetochore plate (the outermost region of the kinetochore in which DNA can be detected [6]) and its role in kinetochore structure, we suggested that CENP-C may be important in organizing the interface between the outer kinetochore and underlying chromatin (41).

Here, we address the mechanism by which CENP-C is localized to centromeres *in vivo*. Utilizing a transient transfection assay to investigate the expression and localization patterns of various truncated CENP-C polypeptides, we show that a 60-amino-acid domain in the central portion of the molecule is necessary and sufficient to direct this protein to human centromeres. We have also investigated the DNA-binding activity of various truncated CENP-C derivatives by Southern (DNA-protein) analyses. We identify a putative DNA-binding domain that has approximately 88 residues and that is also located in the central region of the CENP-C molecule. The minimal *in vivo* centromere localization and *in vitro* DNA-

binding motifs overlap one another, suggesting that the targeting of CENP-C to the centromere-kinetochore complex might be directed by its DNA-binding activity. However, three independent approaches have failed to identify specific DNA sequences with which CENP-C interacts preferentially. This result clearly gives impetus to future studies aimed at determining whether CENP-C binds specifically to a mammalian *CEN* DNA locus or whether its ability to target to centromeres is in part determined by protein-protein interactions.

## MATERIALS AND METHODS

**Construction of expression vectors.** The *EcoRI* fragment of pCENPC<sub>B</sub> (41) was subcloned into modified expression vector pECE/SK (12) upstream of DNA sequences encoding two tandem copies of the avian coronavirus E1 glycoprotein epitope tag (37). The *EcoRI* site at the 5' end of the CENP-C cDNA was destroyed by a partial cleavage reaction with *EcoRI*, blunting with DNA polymerase I (Klenow fragment), and resealing with T4 ligase. The resulting construct was digested with *EcoRI* and *KpnI*, and a nested series of deletions of CENP-C carboxy-terminal sequences was generated by exonuclease III digestion with the Erase-a-Base system (Promega, Madison, Wis.). Constructs encoding in-frame fusions of CENP-C polypeptides to the E1 tags were identified by dideoxy sequencing. pCENPC(1-868), pCENPC(1-829), pCENPC(1-537), pCENPC(1-410), and pCENPC(1-318) were created in this manner. The amino acids of CENP-C coded for by each construct are indicated within the parentheses.

Expression constructs encoding E1-tagged amino-terminal truncations of CENP-C were generated by exonuclease III digestion of pCENPC(1-829). The plasmid was first cleaved with the restriction enzyme *HindIII* and filled in with thiophosphorylated nucleotides, using Klenow fragment to confer resistance to exonuclease III. The plasmid was then digested with *EcoRV* and subjected to unidirectional exonuclease III digestion. Subsequent analyses of the nested deletion series identified constructs pCENPC(192-829), pCENPC(426-829), pCENPC(462-829), and pCENPC(478-829) which had been digested to nucleotide positions +525, +1261, +1351, and +1417 of the CENP-C coding region, respectively. These constructs made use of internal in-frame ATG sites (corresponding to amino acid positions 192, 426, 462, and 478, respectively) for translational initiation.

Additional expression constructs that encoded internal regions of CENP-C were made. pCENPC(638-829) was generated by direct cloning procedures in the following manner. pCENPC(1-829) was digested with *BglII* and *XbaI* (to release the E1-tagged carboxy-terminal region of CENP-C corresponding to residues 638 to 829), blunted, and subcloned into the *SmaI* site of pT7-7 (50), in frame with the vector ATG translation initiation site. Subsequently, the CENPC::E1 moiety, with its new ATG initiation site, was cleaved from pT7-7 with *XbaI* and subcloned into the corresponding site of pECE/SK.

Expression constructs for the central region of CENP-C were generated by standard cloning procedures to combine portions of CENP-C amino-terminal truncation constructs with carboxy-terminal truncations. More specifically, the 3' sequences of pCENPC(1-726)::E1 and pCENPC(1-537)::E1 were isolated as *BstXI* fragments and used to replace the corresponding *BstXI* fragments of pCENPC(192-829)::E1, pCENPC(426-829)::E1, and pCENPC(478-829). Expression constructs pCENPC(192-726)::E1, pCENPC(426-726)::E1, pCENPC(478-726)::E1, pCENPC(192-537)::E1, pCENPC(426-537)::E1, and pCENPC(478-537)::E1 were generated in this manner.

Expression of various subregions of CENP-C in *E. coli* for DNA-binding assays was performed with a modified T7 RNA polymerase-based pT7-7 vector system (50) in which a polylinker containing unique restriction enzyme sites was inserted downstream of the T7 gene 10 translational start site (41a). pT7CENPC(23-943) (originally designated pTCNPC<sub>α</sub> in reference 41) expressing CENP-C residues 23 to 943 was constructed as previously described (41). To generate carboxy-terminal CENP-C truncations, pT7CENPC(23-943) was first cleaved with *BamHI* and *PstI* and then subjected to exonuclease III digestion from the *BamHI* site with the Erase-a-Base system. T7CENPC(23-831), T7CENPC(23-649), T7CENPC(23-576), T7CENPC(23-475), T7CENPC(23-440), T7CENPC(23-324), and T7CENPC(23-273) were derived in this way.

Constructs expressing the central and carboxy-terminal regions of CENP-C were generated from restriction digest fragments derived from CNPC3, an *EcoRI* cDNA insert contained in one of the original CENP-C  $\lambda$ gt 11 isolates (41). T7CENPC(297-429) was generated from an *EcoRI*-*AluI* fragment from CNPC3. In addition, T7CENPC(297-520) was derived from an *EcoRI*-*TaqI* fragment, T7CENPC(433-638) was derived from an *AluI* fragment, T7CENPC(297-789) was derived from an *EcoRI*-*HincII* fragment, and T7CENPC(638-943) was generated from an *AluI*-*EcoRI* fragment. The DNA fragments were isolated following digestion with the restriction enzymes, blunted with DNA polymerase I, and ligated into the appropriate pT7-7 vector. T7CENPC(459-943) was generated directly from an *EcoRI* cDNA insert from one of the original CENP-C  $\lambda$ gt 11 isolates (41). For all of the pT7 constructs discussed above, the CENP-C polypeptides were expressed as fusion proteins containing up to 15 N-terminal and 14 C-terminal residues derived from the vector.

For the expression of T7CENPC(475-537), the CENP-C coding sequence was isolated from pCENPC(478-537)::E1 as a *Bst*Y1 fragment without the E1 tag and subcloned into the *Bam*HI site of pT7-7. T7CENPC(475-537) was expressed as a fusion protein with 96 amino acids (63 residues of CENP-C from positions 475 to 537 [including 3 residues upstream of Met-478] plus 15 N-terminal and 18 C-terminal residues from the vector). Likewise, to express T7CENPC(422-537), CENP-C coding sequence was also released from pCENPC(426-537) as a *Bst*Y1 fragment without the E1 tag and ligated into pT7-7. T7CENPC(422-537) was expressed as a 149-amino-acid polypeptide (116 residues of CENP-C from positions 422 to 537 [including 4 residues upstream of Met-426] plus 15 N-terminal and 18 C-terminal residues from the vector).

**DNA sequence analysis.** All DNA sequencing of expression constructs was performed with the Sequenase kit (U.S. Biochemicals, Cleveland, Ohio). Protein secondary structure prediction analysis for CENP-C was performed with MacVector software (IBI, New Haven, Conn.).

**Induction of CENP-C expression in *E. coli*.** Overnight cultures were diluted 1:100 in Luria broth containing ampicillin. The cells were grown to an optical density at 600 nm of 0.5 and then were induced with 0.5 mM IPTG (isopropyl- $\beta$ -D-thiogalactopyranoside). The cultures were grown for an additional 5 h and harvested. The cells were then centrifuged and resuspended in a 1/20 volume of Tris-buffered saline (10 mM Tris-HCl [pH 7.4], 100 mM NaCl, 1 mM EDTA, 15% glycerol, and protease inhibitors [1 mM phenylmethylsulfonyl fluoride, 1  $\mu$ g of chymostatin per ml, 1  $\mu$ g of leupeptin per ml, 1  $\mu$ g of antipain per ml, and 1  $\mu$ g of pepstatin A per ml]), lysed by sonication, and spun at top speed in an Eppendorf microcentrifuge for 30 min at 4°C. The cell lysate supernatant was transferred to a fresh tube and boiled for 3 min with sample buffer.

**Southwestern blotting.** DNA-binding activity of CENP-C was assayed essentially as described previously (52). Briefly, *E. coli* cell lysates containing CENP-C were separated by standard sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). To renature the proteins, the acrylamide gels were then washed twice for 1 h each time in renaturation buffer (50 mM NaCl, 10 mM Tris-HCl [pH 7.4], 20 mM EDTA, 0.1 mM dithiothreitol, and 4 M urea) to remove SDS. Following the urea washes, the proteins were transferred onto nitrocellulose by electroblotting with 25 mM Tris-192 mM glycine-5% (vol/vol) methanol (pH 8.3). The filters were subsequently blocked in binding buffer (10 mM Tris-HCl [pH 7.4], 50 mM NaCl, 1 mM EDTA, and 5% [wt/vol] nonfat dry milk) for 2 h.

T7CENPC(475-537) and T7CENPC(422-537) were processed for Southwestern analysis by an alternative protocol (47). *E. coli* cell lysates were separated by standard SDS-PAGE. Proteins were then electroblotted onto a polyvinylidene difluoride membrane (Bio-Rad Laboratories, Richmond, Calif.) with 25 mM Tris-192 mM glycine (pH 8.3). The proteins were then renatured on the membrane by first incubating the membrane in HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid) buffer (25 mM HEPES [pH 7.5], 25 mM NaCl, 5 mM MgCl<sub>2</sub>) containing 6 M guanidine hydrochloride for 30 min, then in HEPES buffer plus 3 M guanidine hydrochloride, then sequentially in HEPES buffer containing 1.5, 0.75, and 0.37 M guanidine hydrochloride, respectively, and finally in HEPES buffer only. The membrane was then incubated in standard blocking buffer for 2 h and probed.

As probe DNA, we used a PCR library derived from a naturally occurring human minichromosome, designated *CX*, which we estimate to contain approximately 2 to 5 Mb of DNA. Cells containing this chromosome were provided by Gail Stetter (Johns Hopkins University). The *CX* minichromosome is enriched for centromeric DNA (see below). This mitotically stable chromosome is C-band negative but binds anti-centromere autoantibodies (recognizing CENP-A to -C) at a level comparable to that for the other chromosomes (7a). Microdissection of the *CX* minichromosome from a mitotic spread and preparation of the *CX* DNA PCR library were performed by a modification of previous methods (17). Briefly, DNA from a needle-isolated single *CX* minichromosome was first amplified with random primers to produce a primary *CX* DNA pool. No restriction endonuclease digestion was performed, thereby eliminating possible bias arising from the absence of sites in centromeric repetitive DNA. Next, in order to provide a defined primer binding site for *in vitro* DNA replication by PCR, the primary *CX* DNA pool was ligated to a double-stranded linker corresponding to the single-stranded sequence 5' GGCGGAGGCGGG 3', which was used as a primer for subsequent PCR amplifications. This library DNA, the bulk of which runs on agarose gels as a smear between 0.2 and 0.9 kb, is propagated solely by PCR (an important advantage given the instability of certain types of repetitive DNA in *E. coli* [17]). Subsequent fluorescent *in situ* hybridization analysis, with the *CX* library DNA being used as a probe, showed that this DNA hybridized predominantly to the minichromosome and to the centromeric heterochromatin of chromosome 9 in human cells (16a). Thus, we refer to this library as centromere-enriched DNA. The conclusion that the *CX* chromosome is enriched in centromeric sequences is supported by the results of simultaneous immunofluorescence-phase-contrast microscopy, during which no chromosome arms are seen to protrude beyond the region stained by anti-centromere antibodies (7a). Molecular analysis of the sequences represented in the *CX* library was begun but has been put on hold because of the illness of the author responsible (D.H.J.). It will be important in future studies to characterize the composition of this library in detail.

To perform Southwestern analyses, probe DNA was labeled with [ $\alpha$ -<sup>32</sup>P]ATP by two rounds of thermal cycling with AmpliTaq (Perkin-Elmer Cetus, Norwalk,

Conn.). Labeled DNA was separated from unincorporated nucleotides by chromatography through Sephadex G-50, added to fresh binding buffer to a concentration of  $2 \times 10^5$  cpm/ml along with a total of 5  $\mu$ g of cold unlabeled salmon sperm DNA, and incubated with the membrane for 3 to 8 h at room temperature with gentle agitation.

Following DNA incubation, the filters were washed four times for 30 min each time with binding buffer, rinsed in 10 mM Tris-HCl (pH 7.4)-100 mM NaCl-1 mM EDTA, and then air dried. Lastly, the filters were then exposed to Kodak X-ray film.

**Attempts to identify a specific target sequence for CENP-C binding.** For these experiments, two sources of potential centromeric target sequences for CENP-C were used. The nature and preparation of the *CX* minichromosome-derived PCR library are described above. A second library contained a random 21-bp sequence flanked by defined sequences for use as PCR primers. The first step in the production of this library was the chemical synthesis of degenerate 62-mer oligonucleotides that contained 21 random nucleotides flanked at the ends by the T7 and T3 promoter sequences. The oligonucleotides were then annealed with a primer that was complementary to the T7 promoter sequence and extended with DNA polymerase in the presence of deoxynucleoside triphosphates, resulting in a random library of double-stranded oligonucleotides which could be amplified by PCR with the standard T7 and T3 primers.

Identification of putative CENP-C target sequences was performed in the following manner. Approximately 50 to 100 ng of *CX* or random oligonucleotide library DNA was mixed with 50  $\mu$ l of *E. coli* cell lysates containing CENP-C [T7CENPC(23-943)] in the presence of 5  $\mu$ g of sonicated salmon sperm DNA serving as a nonspecific competitor and incubated for 2 h at room temperature with gentle agitation. CENP-C-DNA complexes were then immunoadsorbed with anti-CENP-C antibodies that were prebound to protein A agarose beads. Following several low-salt washes (in a solution containing 10 mM Tris-HCl [pH 7.4], 100 mM NaCl, and 1 mM EDTA, plus protease inhibitors), bound DNA was eluted with a high-salt solution (10 mM Tris-HCl [pH 7.4], 500 mM NaCl, 1 mM EDTA, plus protease inhibitors) and amplified by PCR. The binding and amplification steps were repeated five more times. Finally, the amplification products were separated on agarose gels and cloned into pBluescript for sequencing. To control for the specificity of interaction with CENP-C, we performed these panning experiments in parallel with either immune or preimmune anti-CENP-C serum bound to the agarose beads. No specific DNA fragments were amplified from the *CX* library in the control experiment.

**Electroporation and microinjection.** HeLa cells were transiently transfected by electroporation as follows. Adherent cells were grown to early log phase, trypsinized, rinsed, and resuspended in Opti-MEM to a concentration of  $5 \times 10^6$  cells per ml. A 0.4-ml portion of the suspended cells was then transferred to a 0.4-cm-diameter electroporation cuvette containing 20  $\mu$ g of DNA. After being mixed, the cells were subjected to a 960-mF pulse at 330 V with a Gene-Pulser (Bio-Rad Laboratories). Cells were subsequently plated onto glass coverslips in RPMI 1640 (Gibco BRL, Gaithersburg, Md.) supplemented with 5% calf serum and incubated for 12 to 18 h.

Some DNA expression constructs were also introduced into HeLa cells by nuclear microinjection. HeLa cells were seeded and grown to microcolonies on gridded coverslips (Bellco) to facilitate the tracking of injected cells. Expression vector DNA (at a concentration of 50 to 150 ng/ml) was injected into the cells with the Narishige micromanipulator and a Nikon PLI-188 microinjector apparatus. The cells were maintained at 37°C at all times during the injection with a Nikon NP-2 microscope stage incubator. The cells were injected, returned to the CO<sub>2</sub> incubator, and then fixed and stained 5 to 15 h later.

**Immunofluorescence.** Indirect immunofluorescence analysis was performed essentially as described previously (1), with the exception that cells were fixed in 4% formaldehyde. Immunostaining of the tagged CENP-Cs was performed with anti-E1 peptide antibodies (37) at a 1:1,000 dilution, then with biotinylated anti-rabbit immunoglobulin G (1:500 dilution), and then with streptavidin-Texas Red (1:1,000 dilution). The cells were double stained with the human anti-centromere serum GS (1:2,000 dilution) (10) and visualized with fluorescein isothiocyanate-conjugated goat anti-human secondary antibodies (Accurate Chemicals, Westbury, N.Y.). Images were obtained with a DAGE SIT camera with a Perceptics PixelPipeline board driven by a modified version of Adobe Photoshop (13a). unpublished).

**Nucleotide sequence accession numbers.** The sequences of the cloned fragments described in this paper have been deposited in GenBank with accession numbers U57992 to U57998.

## RESULTS

**Epitope-tagged CENP-C expressed in HeLa cells localizes to centromeres.** To express various mutant CENP-Cs for this study, we generated epitope-tagged CENP-C derivatives in the simian virus 40 promoter-based mammalian expression vector pECE (12). As our immunological tag, we used a doubly iterated 25-residue moiety derived from the carboxy terminus of the avian coronavirus M glycoprotein (previously known as E1) (21). We have previously utilized this tag to characterize the in

vivo centromeric targeting of another centromeric protein, CENP-B (37). In that study, it was shown that the E1 tag does not interfere with the proper expression or localization of CENP-B and also that it does not possess an independent centromere-targeting activity.

We created a series of deletion derivatives of these CENPC::E1 fusions in pECE (see Materials and Methods). All of our constructs produced fusion proteins that contained the E1 moiety at the carboxy terminus. The longest of our constructs was CENPC(1-868)::E1, which encodes amino acids 1 through 868 of CENP-C. This protein is missing 75 amino acid residues from the carboxy terminus of CENP-C.

To determine whether this chimeric molecule retained CENP-C function, we introduced pCENPC(1-868)::E1 into HeLa cells by electroporation or nuclear microinjection. We detected expression of CENPC::E1 chimeras by indirect immunofluorescence with anti-E1 antibodies. We observed a punctate staining pattern in all cells that reacted positively with anti-E1, i.e., >100 cells observed following electroporation and 59 cells observed following microinjection. These cells included both interphase cells (Fig. 1) and mitotic cells (data not shown). We demonstrated that this staining colocalized with centromeres by performing double immunofluorescence with highly characterized human anti-centromere antibodies (10). Immunofluorescence of uninjected cells (Fig. 1) or cells that were microinjected with vector only (data not shown) did not reveal any specific staining signal with anti-E1 antibodies in either interphase or mitotic cells. Thus, the replacement of 75 residues from the extreme carboxy terminus of CENP-C with 50 residues of the E1 tag did not interfere with the centromeric localization of the polypeptide.

**Analysis of CENP-C truncation mutant proteins identifies a 60-residue minimal centromere-targeting domain.** In addition to pCENPC(1-868)::E1, we created five other CENP-C carboxy-terminal deletion products with various lengths that were fused in frame with the E1 tag. These constructs all encoded CENP-C polypeptides that possessed at least one of the four predicted nuclear localization signals that were previously identified by their sequence similarity to nucleoplasmic-type nuclear localization signals (40, 41). These constructs were also delivered into HeLa cells by either electroporation or microinjection, and the cells were subsequently fixed and immunostained for the E1 tag and centromeres. CENP-C polypeptides expressed from all five constructs accumulated in the nucleus. A diffuse nuclear staining with anti-E1 was apparent in all five constructs (typical examples are shown in Fig. 1). Superimposed on this diffuse staining, specific staining of centromeres could also be clearly distinguished in cells expressing three of the longer polypeptides, CENPC(1-829)::E1, CENPC(1-726)::E1, and CENPC(1-537)::E1 (Fig. 1 and 2). In contrast, only the general diffuse nuclear staining pattern was observed in cells expressing the two shortest constructs, CENPC(1-410)::E1 and CENPC(1-318)::E1. These results demonstrate that an autonomous centromere-targeting domain is located within the first 537 residues of CENP-C. They also show that these chimeric molecules were not dramatically unstable after not being assembled at centromeres. As discussed below, both of these findings contrast with the results of a previous study in which human CENP-C derivatives were expressed in several nonhuman cell types (20).

To more precisely delineate the CENP-C centromere-targeting domain, we created constructs expressing five different amino-terminally truncated CENP-C polypeptides. Four of these constituted a nested set of mutant proteins generated by exonuclease III digestion—CENPC(192-829)::E1, CENPC(426-829)::E1, CENPC(462-829)::E1, and CENPC(478-829)::E1. A

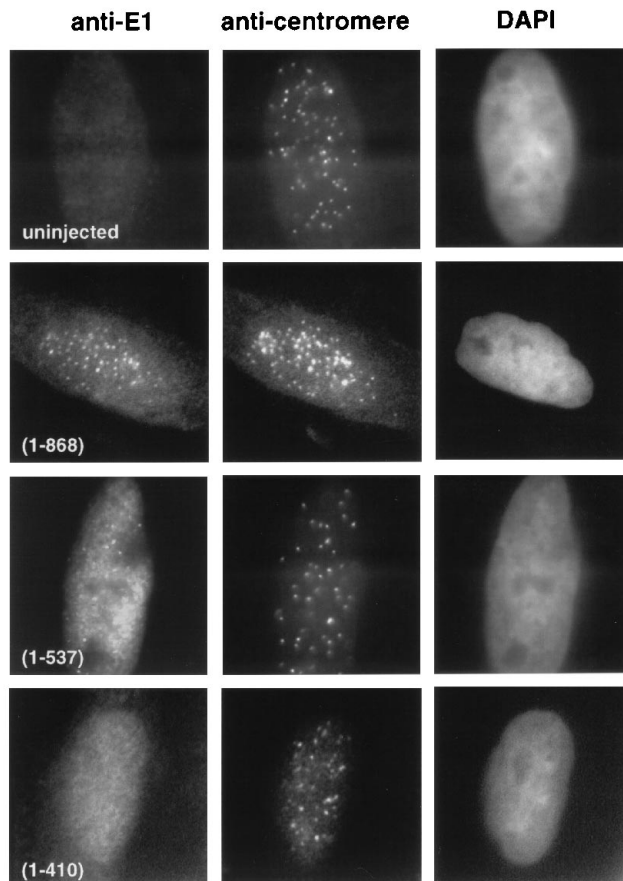


FIG. 1. A centromere-targeting signal is localized in the amino-terminal 537 amino acids of CENP-C. HeLa cells were electroporated or microinjected with a simian virus 40 promoter-based plasmid expressing CENP-C. Following incubation to allow expression of the transfected DNA (see Materials and Methods), cells were fixed and stained. Rabbit anti-E1 tag antibodies reveal the localization of the expressed chimeric CENPC::E1 protein (anti-E1). Human anti-centromere autoantibodies reveal the localization of centromeres (anti-centromere). 4',6-Diamidino-2-phenylindole (DAPI) stains the nuclear DNA (DAPI). Numbers in parentheses indicate the amino acid residues of the truncated CENP-C being expressed. Subtle differences between the images in the anti-E1 and anti-centromere panels in this figure and Fig. 3 are presumed to arise from slight differences in the focal planes for the two images.

fifth construct was generated by first subcloning the E1-tagged carboxy-terminal region of CENP-C encoding residues 638 to 829 plus the E1 tag into bacterial expression vector pT7-7 in frame with the vector ATG translation initiation site and then subcloning the fused gene with its new ATG initiation site back into pECE. In the case of the exonuclease III mutant proteins, translation of the truncated molecule was assumed to be initiated at the first internal in-frame ATG codon.

All five of these constructs contained at least one of the four putative nuclear localization signals of CENP-C, and all expressed CENP-C polypeptides that localized in the nucleus: a diffuse nuclear immunofluorescence signal with anti-E1 antibodies was apparent in all cases. In addition to being responsible for this diffuse nuclear staining, three of the chimeric CENP-C polypeptides—CENPC(192-829)::E1, CENPC(426-829)::E1, and CENPC(462-829)::E1—also localized to centromeres (Fig. 2 and 3). A fourth construct, CENPC(478-829)::E1, could also localize to centromeres (Fig. 3), although the observed centromeric signal was generally weaker than that in

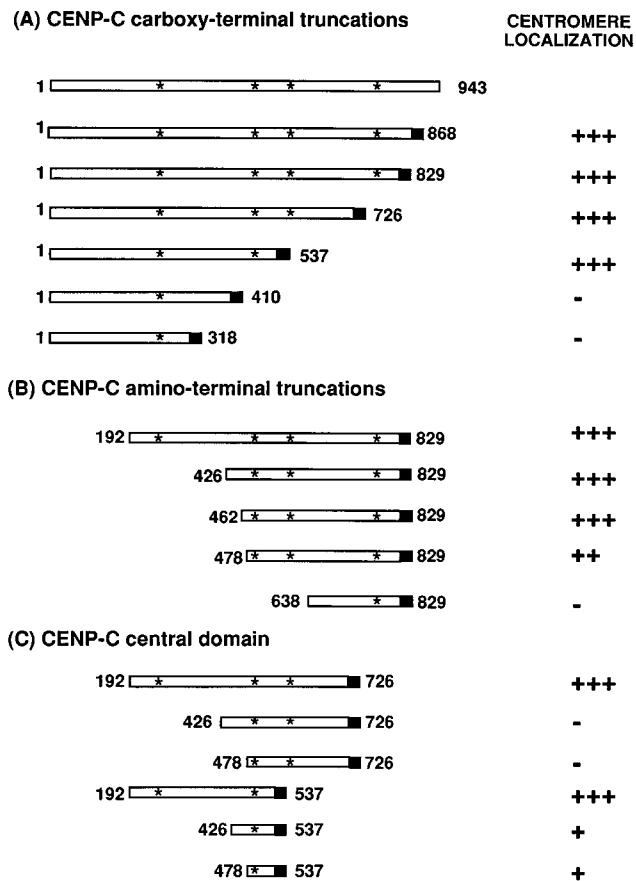


FIG. 2. Schematic representation of CENP-C::E1 constructs and summary of centromere localization results. Numbers indicate the amino acid residues of CENP-C being expressed. Three plus signs indicate that the corresponding mutant CENP-C localized to the centromere in every HeLa cell in which expression was detected. A minus sign indicates that the corresponding mutant CENP-C failed to localize at the centromere. Two plus signs indicate that the corresponding mutant CENP-C localized to the centromere in 60% of the HeLa cells in which expression was detected following nuclear microinjection of plasmid DNA. A single plus sign indicates that the corresponding mutant CENP-C localized to the centromere in 2 to 5% of HeLa cells expressing microinjected plasmid DNA. See the text for more details. Asterisks identify the locations of the putative CENP-C nuclear localization signals (35).

cells expressing larger CENP-C derivatives. Moreover, centromere staining was apparent in only a subset of cells transfected with CENP-C(478-829)::E1. Whereas centromere staining was detected in 206 of 206 cells that were scored following nuclear microinjection with the three larger CENP-C expression constructs, specific immunostaining of centromeres was apparent in only 35 of 57 cells expressing CENP-C(478-829)::E1. Only diffuse nuclear staining was discerned in the other 22 cells.

Expression of CENP-C(638-829)::E1 resulted in a surprisingly specific localization pattern in 21 of 75 microinjected cells that expressed this truncated CENP-C. In these cells, immunofluorescence with anti-E1 revealed the staining of nuclear patches that colocalized with nucleoli (Fig. 3). Thus, CENP-C(638-829)::E1 can specifically accumulate in the nucleolus. Independent biochemical data from our laboratory also suggest that the carboxy-terminal domain of CENP-C can interact with a nucleolar antigen (35a). We never observed centromeric staining with anti-E1 antibodies in cells that were transfected or microinjected with this construct. In cells in

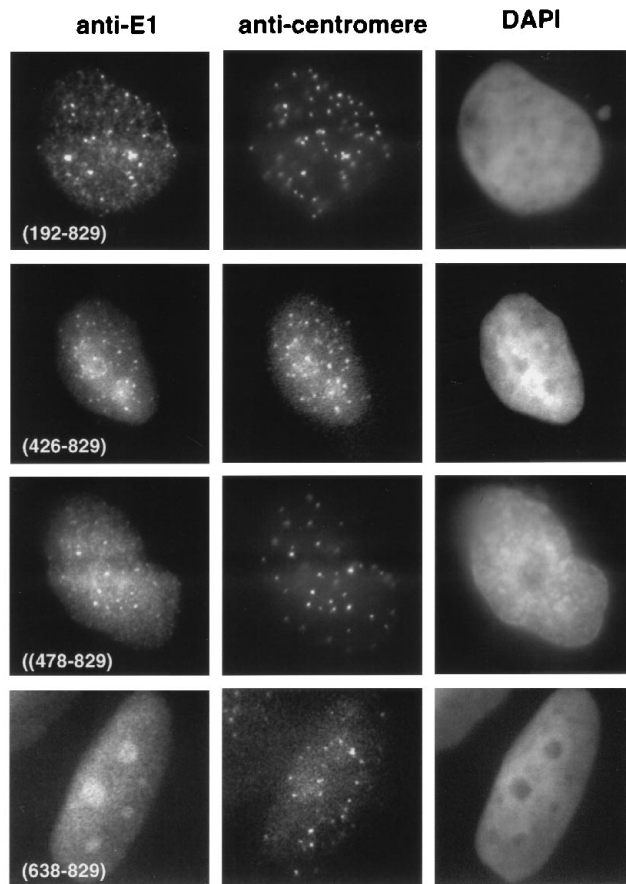


FIG. 3. CENP-C residues 1 to 478 are dispensable for centromere localization. Rabbit anti-E1 tag antibodies reveal the localization of the expressed chimeric CENP-C::E1 protein (anti-E1). Human anti-centromere autoantibodies reveal the localization of centromeres (anti-centromere). DAPI stains the nuclear DNA (DAPI). Numbers in parentheses indicate the amino acid residues of the truncated CENP-C being expressed.

which nucleoli were not stained, immunofluorescence with anti-E1 resulted in a diffuse nuclear background staining pattern.

The results obtained from the *in vivo* expression of amino- and carboxy-terminally truncated CENP-C derivatives suggested that the central region of the polypeptide chain might be involved in targeting CENP-C to the centromere. Furthermore, these studies suggested that the relevant domain might be as small as 60 residues in size and span residues 478 to 537—the area of overlap for CENP-C(1-537)::E1 and CENP-C(478-829)::E1. To directly confirm this hypothesis, we generated several smaller constructs expressing just the central region of the CENP-C polypeptide. The results of this analysis are summarized in Fig. 2. The relatively large CENP-C derivatives, CENP-C(192-726)::E1 and CENP-C(192-537)::E1, localized efficiently to centromeres. Although centromere localization by CENP-C(426-726)::E1, CENP-C(478-726)::E1, CENP-C(426-537)::E1, and CENP-C(478-537)::E1 was less consistent, we did observe weak but unmistakable centromere localization in 8 of 159 cells with CENP-C(426-537)::E1 and 3 of 129 cells with CENP-C(478-537)::E1 (Fig. 4). Thus, while centromere labeling in these transfected cells was usually obscured by the diffuse nuclear background stain, it was clear that this short 60-amino-acid region of CENP-C can be sufficient to target the E1 tag to centromeres.

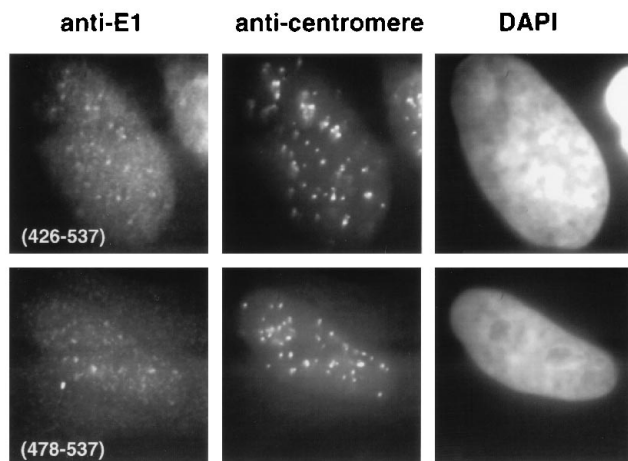


FIG. 4. The minimal centromere-targeting domain is located between residues 478 and 537 of CENP-C. Rabbit anti-E1 tag antibodies reveal the localization of the expressed chimeric CENPC::E1 protein (anti-E1). Human anti-centromere autoantibodies reveal the location of centromeres (anti-centromere). DAPI stains the nuclear DNA (DAPI). Numbers in parentheses indicate the amino acid residues of the truncated CENP-C being expressed.

**CENP-C expressed in *E. coli* binds DNA, but no preferred target sequence has yet been identified.** We previously noted that highly enriched preparations of CENP-C contained significant amounts of DNA and that treatment of these preparations with several different DNases resulted in a dramatic change to the chromatographic properties of CENP-C (40a). We therefore postulated that native CENP-C possesses DNA-binding activity (41). This supposition raised the interesting possibility that CENP-C, like CENP-B (23), might recognize a specific DNA target sequence. On the basis of our immunolocalization of CENP-C (41), such a sequence would be expected to be located within the human kinetochore. We therefore decided to attempt to identify a specific DNA target sequence for CENP-C.

We began by confirming that CENP-C does possess DNA-binding activity by using a standard Southwestern blotting protocol. As a source of DNA, we used a human genomic DNA library that is derived from the CX minichromosome as a probe (see Materials and Methods). This library is enriched for human centromere sequences. We note that this library might well contain human kinetochore DNA; however, this first experiment was performed under conditions favoring nonspecific DNA binding.

To obtain recombinant CENP-C, we expressed the protein with an inducible T7 RNA polymerase *E. coli* expression system (50). First, we tested construct T7CENPC(23-943), which expressed a near-full-length derivative of CENP-C fused at its amino terminus to 10 residues derived from the pT7 expression vector. Although this fusion protein lacked 22 residues of the extreme amino terminus of CENP-C, we chose to utilize it in our initial investigations because of its relatively high level of expression in *E. coli*. Cellular lysates containing expressed CENP-C were separated by SDS-PAGE, renatured, and subsequently transferred to nitrocellulose filters. As can be seen from Fig. 5A (lane 1), T7CENPC(23-943) was expressed well in bacteria, although it underwent a high rate of degradation, as was revealed by immunoblotting analysis with anti-CENP-C antibodies. For Southwestern analysis, a duplicate filter containing CENP-C was then probed with  $^{32}$ P-labeled human DNA. This DNA bound specifically to a major band with an  $M_r$

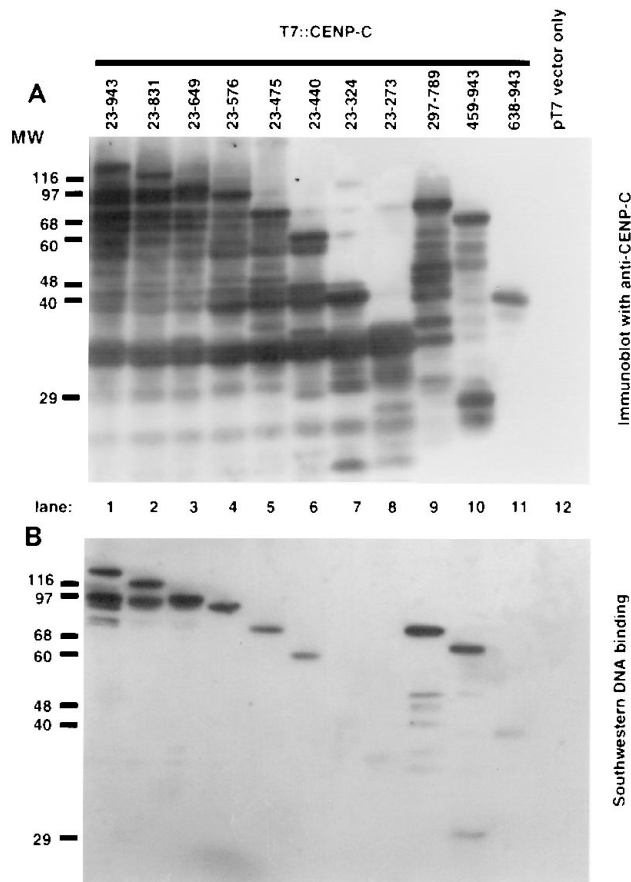


FIG. 5. Binding of DNA by full-length and truncated CENP-C polypeptides expressed in *E. coli*. CENP-C derivatives expressed in *E. coli* with the pT7-7 expression system were subjected to SDS-PAGE, transferred to duplicate nitrocellulose filters, and probed with anti-CENP-C antibodies (A) or labeled DNA (B) to assay for DNA binding. Numbers above the lanes indicate the amino acid residues of the truncated CENP-C being expressed. MW, molecular weight (in thousands).

of 130,000 (at a position similar to the position of the expressed CENP-C [Fig. 5B, lane 1]). DNA was also bound by several lower-molecular-weight species that probably corresponded to CENP-C degradation products. In contrast, no DNA bound to cellular proteins in the pT7 vector-only control lane (Fig. 5B, lane 12).

Having confirmed our earlier deduction that CENP-C possesses an intrinsic DNA-binding capability, we then attempted to identify a specific DNA target sequence for CENP-C binding. We began by asking whether CENP-C could bind preferentially to several known centromeric DNAs. In vitro electrophoretic mobility shift assays failed to reveal any preferential binding of CENP-C to satellite 3 (14) or to two different alpha-satellite DNA monomers derived from chromosome 17 (54). Oligonucleotides containing the consensus sequence motif needed for binding of CENP-B also failed to exhibit preferential binding of CENP-C (data not shown). Since these known centromeric DNAs failed to exhibit specific binding to CENP-C, we next tried two different experiments that were designed to identify preferred CENP-C-DNA binding sites without prior assumptions as to the identity of those sites.

We first used bacterially expressed CENP-C as an affinity ligand to screen the CX centromere-enriched library by the binding selection and PCR amplification strategy described in

**Materials and Methods.** After five rounds of binding, immunoprecipitation, and amplification, this approach yielded major fragments with sizes of 0.6, 0.7, 0.8, and 1.1 kb. In preliminary experiments, these DNAs were found to localize to the centromere of chromosome 9 and to the minichromosome by *in situ* hybridization (45a). The 0.6-kb clone was sequenced completely, and partial sequences were obtained for the others. No significant matches were found, with the exception of a partial long interspersed nuclear element (LINE) (170 nucleotides). A repeat of this screen yielded three fragments with sizes of 0.6, 0.75, and 0.9 kb. The 0.6-kb bands from both screens appear to be identical by both size and restriction analyses.

Although these results initially appeared to be very promising, it now appears that the binding of these DNAs to CENP-C, if specific, must be very weak. None of the cloned fragments undergoes a gel mobility shift in the presence of CENP-C or binds CENP-C at levels detected by immunoblotting (with biotinylated DNA on streptavidin beads). This result could be explained if (i) none of these fragments contains an exact match to the hypothetical CENP-C box, (ii) binding requires other components in addition to CENP-C, or (iii) binding requires a modification of CENP-C not performed in *E. coli*.

Our second attempt to identify a specific DNA target for CENP-C involved the use of PCR enrichment for CENP-C-associated sequences from a random oligonucleotide library (see Materials and Methods). After five rounds of binding, immunoprecipitation, and amplification, 25 clones were obtained and sequenced. Interestingly, 13 of these clones contain sequences (G.A/T.A/T.A/T.G) that loosely fit the consensus for human satellite III DNA (GGAAT = GAATG)<sub>n</sub> (44). However, as was described above, further experiments indicated that bacterially expressed CENP-C did not cause a specific mobility shift of <sup>32</sup>P-labeled satellite III DNA under a variety of conditions.

Although our experiments taken together demonstrate a generalized affinity of CENP-C for DNA, the question of whether CENP-C binds to a specific target sequence remains unanswered. CENP-C may be only one component of a CENP-DNA binding complex, and other factors may confer binding specificity.

**The CENP-C DNA-binding domain maps near the centromere-targeting domain.** We next sought to define the DNA-binding domain more precisely. Using an approach similar to our approach of mapping the centromere-targeting domain, we systematically examined the ability of a number of CENP-C deletion mutants to bind DNA. First, a nested series of plasmids encoding CENP-C carboxy-terminal deletion mutant proteins was generated by unidirectional exonuclease III digestion (see Materials and Methods). As can be seen from the immunoblot analyses of Fig. 5A (lanes 2 to 8), we identified several constructs that expressed truncated CENP-C polypeptides of various lengths in *E. coli*. We then tested these truncated CENP-C polypeptides for their ability to bind DNA. As can be seen from Fig. 5B (lanes 2 to 8) and as is summarized in Fig. 6, DNA-binding activity was retained by constructs expressing approximately the amino-terminal half of CENP-C [T7CENPC(23-440) and larger] but was abolished in constructs that expressed smaller portions of CENP-C [T7CENPC(23-324) and T7CENPC(23-273)].

To further define the DNA-binding domain, we also generated a set of three amino-terminal deletion mutants by direct subcloning procedures (see Materials and Methods). Two of these mutants, T7CENPC(297-789) and T7CENPC(459-943), retained the ability to bind DNA in the Southwestern assay (Fig. 5B, lanes 9 and 10). However, the DNA-binding activity

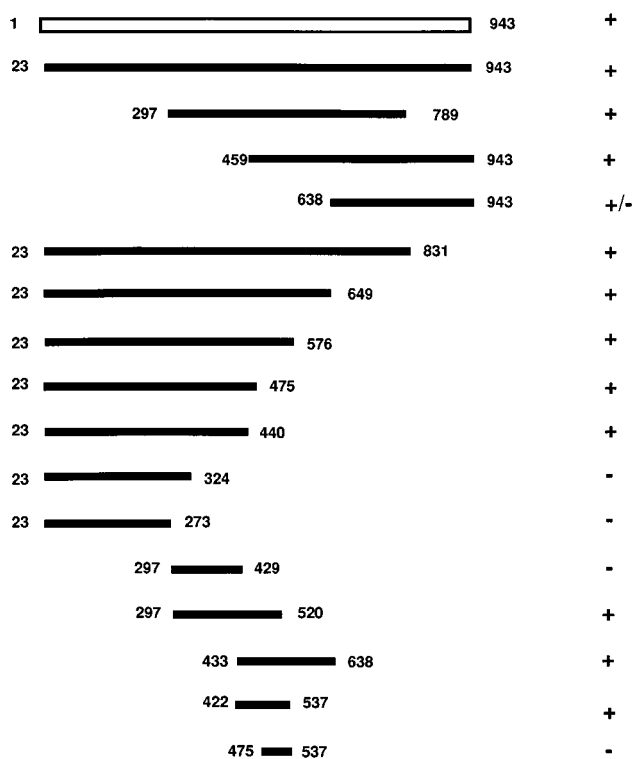


FIG. 6. Schematic representation of CENP-C constructs expressed in *E. coli* and summary of DNA-binding assay results. Numbers indicate the amino acid residues of the CENP-C being expressed. A plus sign indicates that the corresponding CENP-C derivative bound DNA in the Southwestern assay. A minus sign indicates that the corresponding CENP-C derivative did not bind DNA in this assay. A plus-or-minus sign indicates that the corresponding CENP-C bound DNA weakly, resulting in a weak band in this assay (Fig. 5B, lane 11).

of the shortest construct, T7CENPC(638-943), was greatly diminished relative to those of the other two constructs (Fig. 5B, lane 11), indicating that the primary DNA-binding activity of CENP-C was located approximately within the amino-terminal two-thirds of the CENP-C polypeptide.

In sum, our analysis of CENP-C deletion mutant proteins identified the region that falls between residues 324 and 638 and that corresponds to the central region of CENP-C polypeptide as the putative DNA-binding domain for this protein (Fig. 6). Furthermore, our analysis suggested that DNA binding by CENP-C may entail more than one binding site in this region. This outcome is suggested by the fact that two constructs expressing noncontiguous regions of CENP-C [T7CENPC(23-440) and T7CENPC(459-943)] were both able to bind DNA in the Southwestern blotting assay.

To directly confirm that the central region of CENP-C contains DNA-binding motifs, we generated several constructs expressing only this region of the polypeptide. As can be seen from Fig. 7A, T7CENPC(433-638) and T7CENPC(297-520) retained DNA-binding activity. However, T7CENPC(297-429) was no longer able to bind DNA. These results further narrowed the location of the DNA-binding domain of CENP-C to residues 433 to 520 [the region of overlap between T7CENPC(433-638) and T7CENPC(297-520)].

To test this hypothesis, we next generated construct T7CENPC(422-537), which expresses 116 residues from the center of the CENP-C molecule. As can be seen from Fig. 7B, Southwestern analysis confirmed that T7CENPC(422-537) possessed DNA-binding activity.

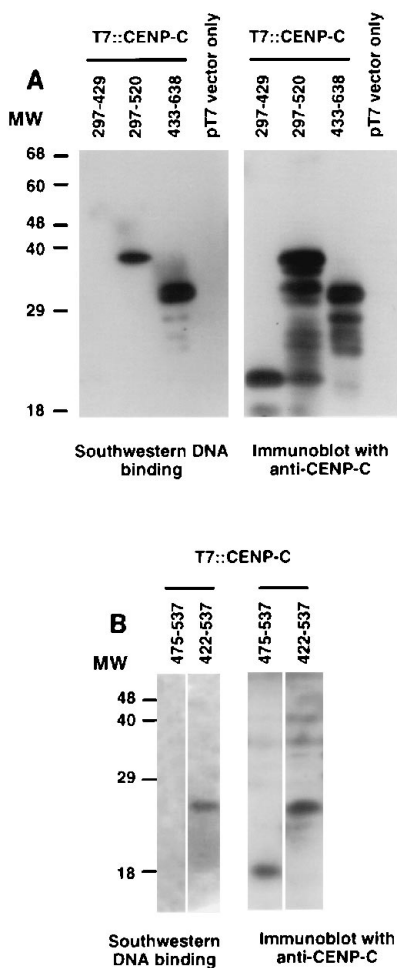


FIG. 7. The minimal DNA-binding domain is located between residues 422 and 537 of CENP-C. (A) Localization of the DNA-binding domain within the central portion of the CENP-C polypeptide. (B) Identification of the minimal DNA-binding domain. CENP-C derivatives expressed in *E. coli* with the pT7-7 expression system were subjected to SDS-PAGE, transferred to duplicate nitrocellulose filters, and probed with labeled DNA to assay for DNA binding or anti-CENP-C antibodies. Numbers above each lane indicate the amino acid residues of the truncated CENP-C being expressed. MW, molecular weight (in thousands).

Because this region closely corresponded to the putative centromere-targeting domain (residues 478 to 537) identified by our *in vivo* expression studies, we sought to test whether the small 60-residue minimal centromere-targeting moiety was also capable of binding DNA. We therefore generated construct T7CENPC(475-537) (see Materials and Methods), which expresses the relevant region of CENP-C without the E1 tag. As can be seen from Fig. 7B, this molecule lacked DNA-binding activity in our assay. This result is important, because this portion of the CENP-C molecule is highly basic. Therefore, DNA binding by CENP-C under our conditions is unlikely to arise from nonspecific electrostatic interactions.

## DISCUSSION

A variety of evidence suggests that CENP-C plays an essential role in the structure and/or assembly of the human kinetochore. (i) CENP-C is located in the inner kinetochore plate (41), the region with the outermost DNA detectable within the kinetochore by cytological means (6). (ii) Injection of mono-

specific and affinity-purified anti-CENP-C antibodies into HeLa cells causes a metaphase delay that is associated with abnormalities in kinetochore structure (51). In many cells, the chromosomes lack detectable kinetochores and appear not to be attached to the spindle. In cells examined after only brief delays in metaphase, trilaminar kinetochores are present, but these are significantly reduced in diameter relative to their counterparts in control cells injected with preimmune immunoglobulin G (51). (iii) CENP-C is not present at inactive centromeres on stable dicentric chromosomes (9, 32). Such centromeres, which do contain normal levels of other centromeric components (9), apparently do not make functional interactions with spindle microtubules during mitosis. (iv) The carboxy-terminal region of CENP-C shares several regions of limited amino acid similarity with Mif2p—the product of a gene that is believed to be important for mitotic spindle integrity during anaphase in the budding yeast *S. cerevisiae* (2, 3, 26). Mif2p possesses an HMG box DNA-binding motif, suggesting that it may bind preferentially to AT-rich DNA. It has been suggested that Mif2p may locate to the yeast centromere as a result of interactions between its DNA-binding domain and the AT-rich DNA of CDE II (26). CENP-C lacks this particular DNA-binding motif, although, as is discussed below, CENP-C does bind to DNA. The function of the Mif2p-CENP-C homology domains is currently unknown.

We presently know nothing about how CENP-C accomplishes its demonstrated role in kinetochore assembly. With the exception of the identification of several possible nuclear targeting sequences, analysis of the CENP-C cDNA sequence failed to identify amino acid motifs of known function (41). Furthermore, the deduced CENP-C sequence failed to provide any hints as to the domain structure of the polypeptide. The principle goal of the present investigation was to identify and map the centromere-targeting and DNA-binding regions of human CENP-C.

**Centromere-targeting domain.** A 60-amino-acid region of CENP-C (residues 478 to 537) carries information sufficient to specify the centromere localization of the protein. However, both CENPC(478-537)::E1 as well as the slightly larger CENPC(426-537)::E1 targeted to centromeres only inefficiently, even though both proteins could be expressed to apparently high levels and accumulated in the nucleus. Efficient centromere localization was obtained only when the 60-residue centromere-targeting region was expressed together with flanking regions, as in CENPC(192-537)::E1 and CENPC(478-829)::E1 (Fig. 2).

We have considered three probable explanations for this size effect. First, the minimal centromere-targeting domain identified in truncation experiments may not correspond to an independent structural domain of CENP-C and may experience difficulties in folding on its own. Second, although the minimal region may possess the specificity determinants to direct CENP-C to centromeres, it may bind to its ligands with low affinity. Interactions between flanking regions and other centromeric components may stabilize the binding at centromeres. Third, supplemental centromere-targeting signals may be present in the flanking regions. However, while we cannot rule out the latter possibility, the observation that either amino- or carboxy-flanking regions can enhance the efficiency of centromere targeting suggests that neither region on its own contains structural information that is essential for centromere targeting.

A recent functional dissection of CENP-C (20) used *in vivo* expression of various human CENP-C constructs in nonhuman cells to map the centromere-targeting domain to the carboxy terminus of the protein. This study failed to detect any centromere-



mere-targeting activity of a large amino-terminal CENP-C polypeptide (residues 1 to 584) in BHK cells (20). It was therefore deduced that the Mif2 homology region (2, 3, 26) contains essential centromere-targeting information, though this was not directly tested by independently expressing this region of CENP-C (residues ~728 to 943) on its own (20).

Those findings exhibit several fundamental differences from the results reported here. We observed consistent centromere localization of an amino-terminal CENP-C polypeptide (residues 1 to 537). We suggest that the published failure of CENP-C(1-584) to target to BHK centromeres may be explained either by folding difficulties with the truncated polypeptide or by a decreased affinity of the truncated human polypeptide to centromeric components in nonhuman cells. Our results demonstrate conclusively that (i) the carboxy-terminal portion of CENP-C is entirely dispensable for centromere targeting and that (ii) the Mif2 homology domains cannot serve an essential role in centromere targeting, although they could have supplementary targeting activity. We currently favor models in which the Mif2 homology domains participate in some other conserved aspect of kinetochore function (26) rather than in targeting.

The previous study identified a discrete region within the CENP-C amino terminus as the site of a novel instability domain that marks CENP-C polypeptides that fail to localize to centromeres for rapid destruction, thereby preventing the toxic accumulation of noncentromeric CENP-C in nuclei. Transfected polypeptides lacking this domain accumulated to high levels in transfected cells and resulted in aberrant chromatin organization and cell death (20).

We did not observe any specific region of CENP-C that appeared to be associated with increased proteolysis of noncentromeric CENP-C. Instead, we typically observed a wide variation in the expression levels of CENP-C from cell to cell. This result duplicates results obtained previously in our lab for other polypeptides expressed from the pECE vector in transient transfection experiments (22, 37). For all CENP-C constructs tested, we always observed cells in which CENP-C polypeptides accumulated in the nucleus, sometimes to very high levels, so that centromere staining was almost obscured [Fig. 1; compare CENP-C(1-537)::E1 with the uninjected control]. Furthermore, we failed to detect any gross aberrations in chromatin organization in cells expressing high levels of CENP-C polypeptides. It is possible that human CENP-C performs slightly differently in human and nonhuman cells. Alternatively, the instability effects observed in BHK cells may be a result of the accumulation of abnormally high levels of CENP-C as a result of the highly active cytomegalovirus promoter being used to drive CENP-C expression.

**DNA-binding domain.** We mapped a minimal DNA-binding region of CENP-C encompassing residues 433 to 520. This DNA-binding region virtually coincided with the minimal centromere-targeting region (residues 478 to 537). Further experiments revealed an unexpected redundancy in the DNA-binding region. Two nonoverlapping CENP-C derivatives [T7CENPC(23-440) and T7CENPC(459-943)], each expressing only a subset of the putative DNA-binding region, were both able to bind DNA. This result suggests that while the CENP-C DNA-binding domain maps to a single region, it may in fact comprise two independent DNA-binding sites.

A recent study used deletion analysis to delineate the borders of the CENP-C DNA-binding domain (47). That study showed that deletions into the region encompassing residues 400 to 502 of CENP-C eliminated DNA binding in a Southwestern blotting assay. Although no attempt was made to express this putative binding domain on its own, those results are

consistent with the results for the minimal region that we have identified (residues 433 to 520). From secondary structure predictions, it was suggested that this region of CENP-C might resemble a helix-loop-helix structural motif comprising four alpha helices that are necessary for DNA-binding activity, similar to the proposed models for the CENP-B DNA-binding domain (48, 58). However, our results showing that the minimal DNA-binding region apparently contains two independent DNA-binding motifs are inconsistent with this interpretation.

**Does DNA binding target CENP-C to centromeres?** At present, the best-characterized centromere-targeting domain of a constitutively centromeric protein has been that of CENP-B. Deletion analysis showed that the amino-terminal 158 amino acids of CENP-B are necessary and sufficient to target this protein to centromeres *in vivo* (37). This centromere-targeting region of CENP-B also doubles as the DNA-binding domain (37, 58), which binds specifically to a 17-bp sequence motif known as the CENP-B box (23). This motif is located in the centromeric  $\alpha$ -satellite DNA of humans as well as the minor satellite of *Mus musculus* (23) and (in its 9-bp minimal functional form [24]) the 79-bp satellite of *Mus caroli* (19).

A second characterized centromere-targeting domain is the histone fold domain of CENP-A, a histone H3-related protein which is believed to be involved in the organization of centromeric nucleosomes (49). Histones are typically thought of as sequence nonspecific DNA-binding proteins, and the participation of this region of CENP-A in centromere targeting was initially surprising. Whether in fact CENP-A is capable of recognizing specific centromeric DNA sequences or is rather directed to centromeres by interaction with another protein is not known at present.

Our observation that the minimal DNA-binding and centromere-targeting domains of CENP-C colocalize within the same region of the polypeptide suggests that DNA binding could be involved in the targeting of CENP-C to centromeres *in vivo*. However, the evidence in support of this is by no means conclusive, particularly because the putative CENP-C-binding sequence has yet to be identified. As we noted above, a concerted effort involving electrophoretic mobility shift assays for DNA binding and library panning with recombinant CENP-C did not identify specific interactions between CENP-C and any centromeric satellite DNAs. Thus, if DNA binding is important for centromere targeting, it is currently not known whether CENP-C targets centromeres via the recognition of specific DNA sequences, as is the case for CENP-B (37), or through more subtle interactions with centromeric DNA, as is presumed to be the case for CENP-A.

The region of CENP-C containing the minimal centromere-targeting and DNA-binding domains (i.e., encompassing residues ~433 to 520) is unrelated to other described centromere-targeting motifs or to other DNA-binding motifs at the level of protein primary sequence. The 60-residue minimal centromere-targeting region (residues 478 to 537) is enriched in basic residues (17 of 60 are Arg or Lys), a number of which occur in a putative bipartite nuclear localization signal (residues 484 to 499 [Fig. 8]). Interestingly, it was previously shown that the functional bipartite nuclear localization signals of transcription factors EB1 and Jun (28) are involved in the DNA-binding activity of those proteins. Whether or not the CENP-C nuclear localization signal is also involved in DNA binding in a similar manner is not known at present. However, we note that T7CENPC(475-537), which corresponds to the minimal centromere-targeting domain and includes this putative nuclear localization signal, did not bind DNA in the Southwestern assay (Fig. 7B). The failure of this highly basic polypeptide to



FIG. 8. Predicted secondary structure of the minimal centromere-targeting domain of CENP-C. Secondary structure predictions were performed with the combined Chou-Fasman (4) and Robson-Garnier (13) algorithms of the MacVector sequence analysis software package (IBI). NLS, nuclear localization signal.

bind DNA suggests that simple charge interactions between the basic regions of CENP-C and the more acidic DNA are unlikely to be solely responsible for the observed DNA-binding activities of CENP-C polypeptides.

It remains to be determined whether interactions of CENP-C with DNA play an active role in targeting the protein to centromeres. If so, it will be important to establish whether the DNA recognition properties of CENP-C are intrinsic properties of the polypeptide itself or are modulated through interactions with other centromeric components. Thus, while some of the components of human centromeres are now yielding to functional analysis, there is still much to learn about the assembly and function of these vital chromosomal structures.

#### ACKNOWLEDGMENTS

C.H.Y. and J.T. were supported by American Cancer Society Postdoctoral Fellowships. These experiments were supported by NIH grant GM35212 and a Principal Fellowship from the Wellcome Trust to W.C.E.

#### REFERENCES

- Bernat, R. L., M. R. Delannoy, N. F. Rothfield, and W. C. Earnshaw. 1991. Disruption of centromere assembly during interphase inhibits kinetochore morphogenesis and function in mitosis. *Cell* **66**:1229-1238.
- Brown, M. T. 1995. Sequence similarities between the yeast chromosome segregation protein Mif2 and the mammalian centromere protein CENP-C. *Gene* **160**:111-116.
- Brown, M. T., L. Goetsch, and L. H. Hartwell. 1993. *MIF2* is required for mitotic spindle integrity during anaphase spindle elongation in *Saccharomyces cerevisiae*. *J. Cell Biol.* **123**:387-403.
- Chou, P. Y., and G. D. Fasman. 1974. Conformational parameters for amino acids in helical, beta-sheet and random coil regions calculated from proteins. *Biochemistry* **13**:211-222.
- Clarke, L. 1990. Centromeres of budding and fission yeasts. *Trends Genet.* **6**:150-154.
- 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.
- 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.
- Earnshaw, W. C. Unpublished data.
- Earnshaw, W. C., and A. M. Mackay. 1994. The role of non-histone proteins in the chromosomal events of mitosis. *FASEB J.* **8**:947-956.
- 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 (Berlin)* **98**:1-12.
- Earnshaw, W. C., and N. Rothfield. 1985. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma (Berlin)* **91**:313-321.
- 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.
- 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.
- Garnier, J., D. J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* **120**:97-120.
- Goldberg, I. G. Unpublished data.
- Grady, D. L., R. L. Ratliff, D. L. Robinson, E. C. McCanlies, J. Meyne, and R. K. Moyzis. 1992. Highly conserved repetitive DNA sequences are present at human centromeres. *Proc. Natl. Acad. Sci. USA* **89**:1695-1699.
- Hegemann, J. H., and U. N. Fleig. 1993. The centromere of budding yeast. *Bioessays* **15**:451-460.
- 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.
- Johnson, D., G. Stetten, and W. C. Earnshaw. Unpublished data.
- Johnson, D. H., P. M. Kroisel, H. J. Klapper, and W. Rosenkranz. 1992. Microdissection of a human marker chromosome reveals its origin and a new family of centromeric repetitive DNA. *Hum. Mol. Genet.* **1**:741-747.
- Kingsbury, J., and D. Koshland. 1991. Centromere-dependent binding of yeast minichromosomes to microtubules in vitro. *Cell* **66**:483-495.
- 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.
- Lanini, L., and F. McKeon. 1995. Domains required for CENP-C assembly at the kinetochore. *Mol. Biol. Cell* **6**:1049-1059.
- 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.
- Mackay, A. M., D. M. Eckley, C. Chue, and W. C. Earnshaw. 1993. Molecular analysis of the INCENPs (inner centromere proteins): separate domains are required for association with microtubules during interphase and with the central spindle during anaphase. *J. Cell Biol.* **123**:373-385.
- 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.
- Masumoto, H., K. Yoda, M. Ikeno, K. Kitagawa, Y. Muro, and T. Okazaki. 1993. Properties of CENP-B and its target sequence in  $\alpha$  satellite DNA, p. 31-43. *In* B. K. Vig (ed.), *Chromosome segregation and aneuploidy*. Springer-Verlag, Berlin.
- McEwen, B. F., J. T. Arena, J. Frank, and C. L. Rieder. 1993. Structure of the colcemid treated PtK1 kinetochore outer plate as determined by high voltage electron microscopic tomography. *J. Cell Biol.* **120**:301-312.
- Meluh, P. B., and D. Koshland. 1995. Evidence that the *MIF2* gene of *Saccharomyces cerevisiae* encodes a centromere protein with homology to the mammalian centromere protein CENP-C. *Mol. Biol. Cell* **6**:793-807.
- 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.
- Mikaélian, I., E. Drouet, V. Marechal, G. Denoyel, J.-C. Nicolas, and A. Sergeant. 1993. The DNA-binding domain of two bZIP transcription factors, the Epstein-Barr virus switch gene product EB1 and Jun, is a bipartite nuclear targeting sequence. *J. Virol.* **67**:734-742.
- Moro, Y., C. Peebles, M. J. Fritzler, J. Steigerwald, and E. M. Tan. 1980. Autoantibody to centromere (kinetochore) in scleroderma sera. *Proc. Natl. Acad. Sci. USA* **77**:1627-1631.
- Muro, Y., H. Masumoto, K. Yoda, N. Nozaki, M. Ohashi, and T. Okazaki. 1992. Centromere protein B assembles human centromeric  $\alpha$ -satellite DNA at the 17-bp sequence, CENP-B box. *J. Cell Biol.* **116**:585-596.
- 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.
- Page, S. L., W. C. Earnshaw, K. H. A. Choo, and L. G. Shaffer. 1995. Further evidence that CENP-C is a necessary component of active centromeres: studies of a dic(X;15) with simultaneous immunofluorescence and FISH. *Hum. Mol. Genet.* **4**:289-294.
- Palmer, D. K., and R. L. Margolis. 1987. A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *J. Cell Biol.* **104**:805-815.
- Palmer, D. K., K. O'Day, H. Le Trong, H. Charbonneau, and R. L. Margolis. 1991. Purification of the centromeric protein CENP-A and demonstration that it is a centromere specific histone. *Proc. Natl. Acad. Sci. USA* **88**:3734-3738.
- 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.
- Pluta, A., and W. C. Earnshaw. Unpublished data.
- Pluta, A. F., A. M. Mackay, A. M. Ainsztein, I. G. Goldberg, and W. C. Earnshaw. 1995. The centromere: hub of chromosomal activities. *Science* **270**:1591-1594.
- 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.
- Rattner, J. B. 1986. Organization within the mammalian kinetochore. *Chromosoma (Berlin)* **93**:515-520.
- Rieder, C. L. 1982. The formation, structure and composition of the mam-

- malian kinetochore and kinetochore fiber. *Int. Rev. Cytol.* **79**:1–58.
40. Robbins, J., S. M. Dilworth, R. A. Laskey, and C. Dingwall. 1991. Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell* **64**:615–623.
  - 40a. Saitoh, H., A. Pluta, and W. C. Earnshaw. Unpublished data.
  41. 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.
  - 41a. Saitoh, N. Unpublished data.
  42. Saunders, W. S., C. Chue, M. Goebel, C. Craig, R. F. Clark, J. A. Powers, J. C. Eissenberg, S. C. Elgin, N. F. Rothfield, and W. C. Earnshaw. 1993. Molecular cloning of a human homologue of *Drosophila* heterochromatin protein HP1 using anti-centromere autoantibodies with *anti-chromo* specificity. *J. Cell Sci.* **104**:573–582.
  43. Schulman, I., and K. S. Bloom. 1991. Centromeres: an integrated protein/DNA complex required for chromosome movement. *Annu. Rev. Cell Biol.* **7**:311–336.
  44. Singer, M. F. 1982. Highly repeated sequences in mammalian genomes. *Int. Rev. Cytol.* **76**:67–112.
  45. Sorger, P. K., F. F. Severin, and A. A. Hyman. 1994. Factors required for the binding of reassembled yeast kinetochores to microtubules in vitro. *J. Cell Biol.* **127**:995–1008.
  - 45a. Stetten, G. Personal communication.
  46. 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.
  47. Sugimoto, K., H. Yata, Y. Muro, and M. Himeno. 1994. Human centromere protein C (CENP-C) is a DNA-binding protein which possesses a novel DNA-binding motif. *J. Biochem.* **116**:877–881.
  48. 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 (Berlin)* **100**:360–370.
  49. Sullivan, K. F., M. Hechenberger, and K. Masri. 1994. Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J. Cell Biol.* **127**:581–592.
  50. Tabor, S., and C. C. Richardson. 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl. Acad. Sci. USA* **82**:1074–1078.
  51. Tomkiel, J. E., C. A. Cooke, H. Saitoh, R. L. Bernat, and W. C. Earnshaw. 1994. CENP-C is required for maintaining proper kinetochore size and for a timely transition to anaphase. *J. Cell Biol.* **125**:531–545.
  52. Tully, D. B., and J. A. Cidlowski. 1993. Protein-blotting procedures to evaluate interactions of steroid receptors with DNA. *Methods Enzymol.* **218**:535–551.
  53. Tyler-Smith, C., and H. F. Willard. 1993. Mammalian chromosome structure. *Curr. Opin. Genet. Dev.* **3**:390–397.
  54. Waye, J. S., and H. F. 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.
  55. Willard, H. F. 1990. Centromeres of mammalian chromosomes. *Trends Genet.* **6**:410–416.
  56. Wordeman, L., E. Steurer, M. Sheetz, and T. Mitchison. 1991. Chemical subdomains within the kinetochore domain of isolated CHO mitotic chromosomes. *J. Cell Biol.* **114**:285–294.
  57. Yen, T. J., G. Li, B. Schaar, I. Szilak, and D. W. Cleveland. 1992. CENP-E is a putative kinetochore motor that accumulates just prior to mitosis. *Nature (London)* **359**:536–539.
  58. 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  $\alpha$  helices at the NH2 terminus, which is separable from dimerizing activity. *J. Cell Biol.* **119**:1413–1427.
  59. Zinkowski, R. P., J. Meyne, and B. R. Brinkley. 1991. The centromere-kinetochore complex: a repeat subunit model. *J. Cell Biol.* **113**:1091–1110.