Molecular Cloning of cDNA for CENP-B, the Major Human Centromere Autoantigen


Departments of * Cell Biology and Anatomy and † Biological Chemistry, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205; § Division of Rheumatic Diseases, Department of Medicine, University of Connecticut Health Center, Farmington, Connecticut 60305

Abstract. We have isolated a series of overlapping cDNA clones for ~95% of the mRNA that encodes CENP-B, the 80-kD human centromere autoantigen recognized by patients with anticentromere antibodies. The cloned sequences encode a polypeptide with an apparent molecular mass appropriate for CENP-B. This polypeptide and CENP-B share three non-overlapping epitopes. The first two are defined by monoclonal antibodies elicited by injection of cloned fusion protein. Epitope 1 corresponds to a major antigenic site recognized by the anticentromere autoantibody used to obtain the original clone. Epitope 2 is a novel one not recognized by the autoantibody. These epitopes were shown to be distinct both by competitive binding experiments and by their presence or absence on different subcloned portions of the fusion protein. The third independent epitope, recognized by a subset of anticentromere-positive patient sera, maps to a region substantially closer to the amino terminus of the fusion protein. DNA and RNA blot analyses indicate that CENP-B is unrelated to CENP-C, a 140-kD centromere antigen also recognized by these antisera. CENP-B is the product of a 2.9-kb mRNA that is encoded by a single genetic locus. This mRNA is far too short to encode a polypeptide the size of CENP-C. The carboxy terminus of CENP-B contains two long domains comprised almost entirely of glutamic and aspartic acid residues. These domains may be responsible for anomalous migration of CENP-B on SDS-polyacrylamide gels, since the true molecular mass of CENP-B is ~65 kD, 15 kD less than the apparent molecular mass deduced from gel electrophoresis. Quite unexpectedly, immunofluorescence analysis using antibodies specific for CENP-B reveals that the levels of antigen vary widely between chromosomes.

The structure and function of the centromere (CEN) regions of mitotic chromosomes have recently become the focal point of investigations by workers in three diverse disciplines. First, cell biologists have known for a number of years that the centromere is both the site of sister chromatid pairing (the mechanism of which is currently unknown) and the site of mitotic spindle attachment. The latter site, the kinetochore, has been well-characterized at the ultrastructural level (5, 40-42). The kinetochore is a trilaminar plaque structure embedded in the chromatin at the surface of the chromosome. It is solidly anchored in the chromosomal infrastructure and is found in the chromosome scaffold fraction (13).

Second, geneticists have long been interested in the role of cis-acting DNA sequences in the control of chromosomal segregation. Substantial advances have been made with lower eukaryotes such as in Saccharomyces, where DNA from 12 of the 16 CENs has been cloned (6, 15, 20, 21, 28, 36, 45). Each CEN sequence consists of ~120 base pairs (bp) of DNA organized into three subdomains and assembled into a characteristic chromatin structure (2). In the fission yeast Schizosaccharomyces, the three centromere regions have also been cloned by chromosome walking (33). These centromere regions are much larger and more complex (roughly 50 kb with a conserved subdomain of 4,000 base pairs), and as yet have not been amenable to functional analysis. We therefore know much about the DNA organization of centromeres in these two yeasts: however, the proteins that interact with the CEN sequences remain unknown.

Finally, rheumatologists became interested in centromere structure when it was observed that centromere components are the target of a highly selective autoimmune response in certain patients with rheumatic diseases (see references 16, 32, 46). Anticentromere antibodies (ACAs) occur in patients with Raynaud's phenomenon (hypersensitivity of the peripheral vasculature to cold), many of whom also have
scleroderma. This observation has proven seminal in recent studies of centromere structure in higher eukaryotes, since the antibodies are sensitive probes for both structural and biochemical analyses. We previously showed that ACA(+) patient sera recognize three immunologically related centromere antigens, CENP-A (17 kDa), CENP-B (80 kDa), and CENP-C (440 kDa) (10). Both CENP-A and CENP-B have been recognized independently by several other research groups (CENP-A: 8, 17, 34, 49, 50; CENP-B: 49). CENP-B is referred to as the major centromere antigen, since antibodies to it are present at high titer in all ACA(+) patient sera, while the titer of antibodies to CENP-A and CENP-C is often lower (11).

Very little is known about the functions of CENP-A, CENP-B, and CENP-C. Although an early study by Brenner et al. (4) suggested that some or all of the autoantigens recognized by one ACA(+) serum might be localized to the trilaminar kinetochore structure, it is not known which of the CENP antigens was thus localized. In addition, we have shown that all three of the antigens are absent from the latent (inactive) centromere of a stable dicentric X chromosome (9). Others have also found that the antigens are either absent, or present in much reduced amounts, at the latent centromeres of other stable dicentric chromosomes (31, 37). These results suggest that the antigens might be involved in microtubule binding. However, the detailed function and distribution of individual CENP antigens in vivo is currently unknown.

To begin a more thorough analysis of centromere antigens, we have now used immunological screening of a human λgt11 cDNA expression library to isolate cloned cDNA sequences encoding polypeptides recognized by anticentromere antibodies. A series of overlapping clones corresponding to >90% of the mRNA encoding CENP-B have been obtained. Five independent lines of biochemical and immunological evidence authenticate the identity of the clones. CENP-B is encoded by a single human gene whose unusual protein product contains two long stretches of residues comprised almost entirely of acidic residues.

Materials and Methods

Patient Sera

All sera come from the collection at the University of Connecticut, and are derived from individuals examined by the staff of the Division of Rheumatic Diseases. This collection currently includes sera from 54 patients with anticentromere antibodies, 32 age- and sex-matched normal controls, and 230 individuals with similar clinical manifestations but lacking anticentromere antibodies. All patient sera are characterized both by indirect immunofluorescence using colcemid-arrested HeLa cells spread by centrifugation on glow-discharged coverslips, and by immunoblotting of centromere protein preparations (29). Patient sera are preadsorbed with bacterial lysates (43) to remove antibodies nonspecifically reactive with bacterial proteins.

Immuno blotting. Antigens were identified by immunoblotting using patient sera. Nineteen ACA(+) sera were used. These sera were shown to be reactive with CENP-B in immunoblots of isolated chromosomes. This polyclonal antibody preparation was found to be useful in these studies.

Preparation of Polyclonal Anticentromere Antibodies

Fusion protein fp-βgal-CNBP1 was isolated from YO89 lysogens harboring ϕX174M13 lacZ by a simple differential sedimentation procedure (to be described in detail elsewhere). Briefly, frozen induced lysogens were disrupted with lysozyme, EDTA, sonication, and detergent, releasing a granule of fusion protein. This granule was separated from soluble and membrane proteins by three cycles of sedimentation at 10,000 g (15 min) in the presence of 1% Triton X-100 in TEN buffer (10 mM Tris, pH 7.4, 1 mM Na EDTA, 50 mM NaCl). The granule was then dissolved in 8 M urea in TEN buffer at 65°C, insoluble material removed by centrifugation at 12,500 g for 10 min, and then dialyzed overnight versus TEN buffer.

Two male rabbits (6 mo; Harlan Sprague Dawley, Inc., Indianapolis, IN) were injected at days 0, 29, 80, 122, and 125 with 70 μg of solubilized fp-βgal-CNBP1. For the first injection we used Freund's complete adjuvant; for subsequent injections we used Freund's incomplete adjuvant. Injections were given at multiple sites: scutaneous, intramuscular, and intraperitoneal. Animals were bled from the ear on days 0, 38, 92, 135, and 169 by venipuncture. One rabbit responded by producing antibodies that recognize CENP-B in immunoblots of isolated chromosomes. This polyclonal anticientromere serum is designated as ACA(+).
tants were screened by ELISA against fp-βgal-CNPB1 and β-galactosidase (Sigma Chemical Co.). Wells positive for the former and negative for the latter were designated possible anti-human candidates and were cloned in soft agar (24, 27). Wells positive on both ELISAs were assumed to be producing anti-β-galactosidase, and several of these were cloned as well.

Primary clones were screened by immunoblotting against fp-trpE-CNPB1. Positive wells were immediately recloned in soft agar. All clones that showed binding to fp-trpE-CNPB1 also bound to chromosomal CENP-B. Four independent clones (m-ACA1, m-ACA2, m-ACA3, and m-ACA4) were obtained. Competitive binding experiments (24) showed that m-ACA1 and m-ACA4 bind to overlapping determinants on fp-trpE-CNPB1. The same is true of m-ACA2 and m-ACA3.

Competitive Binding Experiments

Competitive binding experiments were performed to examine the ability of the autoantibody and ra-ACA1 to block binding of m-ACA1 and m-ACA2 to fp-trpE-CNPB2 and to chromosomal CENP-B. Bacterial lysates containing fp-trpE-CNPB2, or alternatively, isolated Hela chromosomes prepared as previously described (22, 26), were subjected to SDS PAGE. The separated proteins were subsequently transferred to nitrocellulose by electroblotting (48). The nitrocellulose filters were then cut into strips and treated as follows.

Autoantibody. Nitrocellulose strips were incubated overnight with a 1:100 dilution of human serum (either autoantibody or normal serum WE). In the morning, the antibody solutions were removed and replaced by an equal volume of hybridoma culture supernatant from either m-ACA1 or m-ACA2. This incubation proceeded for 3 h, following which the unbound antibodies were removed (five 3-min washes with GB followed by one 2-min wash with PTX). PTX-BSA was then added, and after 5 min 125I-goat anti-mouse IgG was added to 1 × 106 cpmp/ml. 2 h later, unbound probe was removed by five 3-min washes with GB. The filter strips were then dried, mounted, and autoradiographed.

ra-ACA1. Competitive binding experiments were also performed with ra-ACA1 as described above with the following exceptions: a 1:10 dilution of the ra-ACA1 rabbit serum was used for blocking, and incubation with the hybridoma culture supernatant was for 2 h.

Expression of CENP-B as Fusion Proteins in Bacteria

For expression of cDNA from CNPB1-4 as fusion proteins with the bacterial protein trpE, appropriate regions of cDNA were cloned into one of the pATH series of bacterial expression plasmids (provided by T. J. Koerner, Duke University, Durham, NC). This series of plasmids was specifically designed to allow expression of the hybrid protein under control of the lac operon. Fusion protein was induced in a fashion analogous to that for trpE proteins with the exception that the inducer was 10 μM isopropylthiogalactoside.

For expression of CNPB1, CNPB2, CNPB3, and CNPB4, each cDNA was excised with Eco RI (from the original λ clones or from pUC9 subclones) and the fragment was isolated after electrophoresis on low melting temperature agarose gels. The gel slices were melted and added to the appropriate pATH plasmid that had been linearized with Eco RI and treated with alkaline phosphatase. The mixtures were then diluted into DNA ligase buffer (New England Biolabs), brought to room temperature, and DNA ligation added for 2 h to overnight. Ligation reactions were transformed by standard methods into Escherichia coli strains DH1 (18) or CAG-456 (a temperature-sensitive strain of E. coli that contains a mutation in the lon protease gene [43]). Transformants were selected by ampicillin resistance and plasmids carrying the recombinant trpE-CNPB fusion gene were identified by colony hybridization. Restriction mapping of isolated plasmid DNA identified those plasmids into which the cDNA had been cloned in the proper orientation.

TrpE-CNPB fusion proteins were induced from overnight cultures (grown at 30°C in M9 [29] supplemented with 100 μg/ml ampicillin, and 20 μg/ml tryptophane) of CAG-456 transformants harboring the desired plasmids. The overnight culture was diluted 1:10 into the same medium without tryptophane. The bacteria were incubated at 30°C for 2 h and the tryptophane operon induced for an additional 2 h by addition of indoleacrylic acid (Sigma Chemical Co.) to 5 μg/ml. Bacteria were lysed by boiling in SDS gel sample buffer and proteins resolved by PAGE (25).

For expression of amino- and carboxyl-terminal sequences of CNPB1, a subclone of CNPB1 in pBR322 was linearized at the unique Asp 718 site within the coding portions of the cDNA (see Fig. 1). The 5' overhang produced was filled in by Klenow DNA polymerase and 1 μM dTTP, dGTP, dCTP, and dATP. A portion of the reaction was digested with Eco RI and the 318-base fragment corresponding to the more NH2-terminal sequences of CNPB1 was cloned into pATH II that had been digested with Eco RI and Sma I. For expression of the carboxy-terminal sequences of CNPB1, the remaining aliquot was digested with Hind III (at the pbr322 Hind III site, 29 bp 3' to the terminus of CNPB1 sequences), the appropriate fragment purifed by gel electrophoresis, and cloned into pATH 3 (digested at the Bam HI site in the polylinker) which was then filled in with Klenow and then digested with Hind III.

For expression of the amino-terminal portion of CNPB4, CNPB4 cDNA was double digested with Bam HI and Eco RI and the 1,050-base fragment corresponding to the 5' portion of CNPB4 (see Fig. 1) was cloned into Eco RI-Bam HI-digested pATH II.

For expression of CNPB2 as a fusion protein with a smaller bacterial moiety (consisting of 16 amino-terminal residues; four from β-galactosidase and 12 from polylinker sequences), the entire CNPB2 cDNA was subcloned into the Eco RI site of pUC9. This construction placed CNPB2 expression in frame and under control of the lac operon. Fusion protein was induced in a fashion analogous to that for trpE proteins with the exception that the inducer was 10 μM isopropylthiogalactoside.

**Table 1. Identification of CENP-B cDNA Clones**

<table>
<thead>
<tr>
<th>Round</th>
<th>Probe (base pairs)</th>
<th>cDNA library</th>
<th>Number of phage screened</th>
<th>Longest cDNA clone obtained</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Serum GS</td>
<td>Endothelial cell</td>
<td>9 × 10⁴</td>
<td>λCNPB1 (1282)</td>
</tr>
<tr>
<td>2</td>
<td>λCNPB1 Eco RI-Kpn I (320)</td>
<td>Endothelial cell</td>
<td>2.7 × 10⁴</td>
<td>1282</td>
</tr>
<tr>
<td>3</td>
<td>λCNPB1 Eco RI-Kpn I (320)</td>
<td>Placental</td>
<td>8 × 10⁴</td>
<td>λCNPB2 (2133)</td>
</tr>
<tr>
<td>4</td>
<td>λCNPB2 Eco RI-Sac II (138)</td>
<td>Placental</td>
<td>8 × 10⁴</td>
<td>λCNPB3 (2578)</td>
</tr>
<tr>
<td>5</td>
<td>λCNPB2 Eco RI-Sac II (138)</td>
<td>Placental</td>
<td>1.2 × 10⁵</td>
<td>λCNPB4 (2617)</td>
</tr>
<tr>
<td>6</td>
<td>λCNPB4 Eco RI-Xho I (223)</td>
<td>Placental</td>
<td>4 × 10⁴</td>
<td>2617</td>
</tr>
</tbody>
</table>

**Figure 1.** Schematic drawing and sequencing strategy for four overlapping cDNA clones for the mRNA encoding CENP-B. Schematic representations of the cDNA portions of four independent cDNA clones for CENP-B are shown along with a partial restriction endonuclease map. Bold lines indicate the protein coding sequences, whereas thin lines represent the 3' untranslated region of the mRNA. Beneath the schematic is the sequencing strategy used for determining the complete cDNA sequence. X, Xho I; N, Nae I; P, Pst I; S, Sac II; B, Bam HI; Sac, Sac I; A, Asp 718; Sma, Sma I; pA, site of polyadenylation.
Identification of the epitope encoded by \( \lambda \)CNPB1. Lysogens were prepared from \( \lambda \)CNPB1, production of fusion protein was induced, and autoantibodies binding to this fusion protein were affinity purified and characterized by immunoblotting and immunofluorescence. (a) Induction of fusion protein fp-\( \beta \)-gal-CNPB1 in a lysogen. Coomassie Blue staining of SDS PAGE is shown. (Lane 1) The basal polypeptides in an uninduced culture; (lane 2) polypeptides after the induction of lytic growth and exposure of the cells to IPTG. The fusion protein is indicated with an arrowhead. (b) Immunoblots of chromosomal proteins. Whole autoimmune serum GS and antibodies affinity purified from it by binding to fp-\( \beta \)-gal-CNPB1 were used to probe immunoblots of proteins present in isolated HeLa chromosomes. (Lane 1) Binding of whole serum GS (dilution 1:1,000); (lane 2) antibody affinity purified from fp-\( \beta \)-gal-CNPB1 (dilution 1:20). (As discussed in the text, the binding to histones [solid circles] is due to the presence of denatured antibody resulting from the harsh conditions used to elute the antibodies.) (c) Indirect immunofluorescence of human chromosomes. Binding of (panel 1) whole serum GS (diluted 1:5,000) and (panel 2) antibodies (diluted 1:2) affinity purified from fp-\( \beta \)-gal-CNPB1 to centromeres of human chromosomes. This and all subsequent immunofluorescence figures are simultaneous double exposures of the indirect immunofluorescence (Texas Red label) and phase-contrast images. Bar, 10 \( \mu \)m.

**Blot Analysis of RNA and DNA**

Gels for separation of RNA were poured from 1% agarose containing 2.2 M formaldehyde (3). RNA was transferred to nitrocellulose according to Thomas (47). Electrophoresis of DNA was performed in 0.8% agarose gels containing 40 mM Tris-Cl, pH 8.1, 20 mM Na acetate, and 2 mM EDTA. Gels were stained with 1 \( \mu \)g/ml ethidium bromide and transferred according to Southern (44). For both RNA and DNA blots, specific sequences were localized by hybridization with DNA probes labeled with \( \text{\textsuperscript{32}P} \) (3,000 Ci/mmol) by the random priming method (42a). Hybridization conditions were 50% formamide, 750 mM NaCl, 75 mM sodium citrate, 0.02% albumin, 0.02% polyvinylpyrrolidone, 0.02% ficoll, 500 \( \mu \)g/ml yeast RNA, and 50 \( \mu \)g/ml sonicated herring sperm DNA. Blots were washed at 53°C in 0.1% SDS, 15 mM NaCl, before visualization by exposure at -80°C to X-ray film (XAR-2, Kodak) with Lightning Plus intensifying screens (DuPont Co., Wilmington, DE).

**Results**

**Isolation of a cDNA Clone Encoding an Epitope Present on Human CENP-B**

To identify cDNA clones for human centromere autoantigens, we screened a \( \lambda \)gt11 expression library (54, 55) using serum from a patient with anticentromere antibodies (ACAs). Selection of the antisera was critical. Many antisera from these patients have high titers of ACAs, but most also have other autoantibodies that recognize a variety of cellular components. Serum GS (from a patient with Raynaud's phenomenon and scleroderma) was chosen as the most suitable of the 38 ACA(+) antisera then available to us. This serum contains 375 \( \mu \)g/ml (or 3% of the total) of IgG specific for the 80-kD centromere autoantigen CENP-B (based on a quantitative ELISA using cloned CENP-B (Rothfield, N. E., C. Marino, D. W. Cleveland, and W. C. Earnshaw, manuscript in preparation), and is negative for other common autoantibodies, including Ro, La, Sm, DNA (ss and ds), RNP, Scl-70 (topoisomerase I), actin, and intermediate filaments. Serum GS binds detectibly to CENP-B in immunoblots of isolated chromosomal proteins at dilutions of up to 1:600,000.

When serum GS was used to screen a \( \lambda \)gt11 library constructed from human endothelial cell mRNA (the generous gift of Drs. D. Littman, M. Chow, and R. Axel), immunopositive plaques were detected at a frequency of about 1/50,000. After plaque purification, Eco RI digestion of purified phage DNA from an initial isolate, \( \lambda \)-CNPB1, liberated a single cDNA fragment of 1,282 bp.
Figure 3. Indirect immunofluorescence of human chromosomes using rabbit polyclonal antibody ra-ACA1, raised against fp-β-gal-CNPB1. (a) Preimmune serum (dilution 1:1,000). (b) Immune serum, ra-ACA1 (dilution 1:1,000).

Using nucleic acid hybridization with this insert, we found no longer cDNA inserts in the original library, but a second λgt11 cDNA library (from human placenta) yielded three longer clones (λ-CNPB2 [2,133 bp], λ-CNPB3 [2,578 bp], and λ-CNPB4 [2,617 bp]—see Fig. 1 and Table I for details).

Determination That the Isolated cDNA Clones Contain Sequences Corresponding to an mRNA Encoding an Epitope on CENP-B

When bacteria lysogenic for λ-CNPB1 were induced to express the lac operon (43), they produced a 150-kD polypeptide consisting of the cloned protein fused to β-galactosidase (Fig. 2 a). Autoantibodies from GS antiserum purified by binding to this fusion protein (immobilized on nitrocellulose) bound to centromeres (Fig. 2 c) and to CENP-B on blots of human chromosomal proteins (Fig. 2 b), but not to CENP-A or CENP-C. (The affinity-purified antibodies also bound weakly to histones H1 and H3 [identified by Ponceau S staining of the nitrocellulose strips—see solid circles in Fig. 2 b]). Several observations suggest that this binding is due to nonspecific sticking of denatured antibody to the histones. [a] No antibody from whole serum bound to histones provided that blots were washed stringently [Fig. 2 a, lane 1]. [b] Antibody affinity purified from histone H1 did not rebind to H1 on blots of chromosomal proteins [10]. [c] Prolonged exposure of the autoantibody to 3 M NH₄-K-thiocyanate during elution increases binding to histones [Saunders, W., and W. C. Earnshaw, unpublished results], suggesting that histone binding may involve antibody species denatured by exposure to elevated pH. [d] Sequence analysis of clone λ-CNPB1 reveals no homology between the cloned protein and histones. There is thus no evidence that ACAs
Figure 5. Indirect immunofluorescence analysis of human chromosomes using monoclonal anti-CENP-B. HeLa chromosomes were centrifuged onto glow-discharged coverslips (12) and then prepared for indirect immunofluorescence using monoclonal antibodies (a) m-ACA1 and (b) m-ACA2 (both monoclonals were ascites fluid used at a dilution of 1:2,000). The arrow in a points out a large metacentric chromosome whose centromere has apparently split into two halves during spreading. (This effect has previously been observed using affinity-purified anti-CENP-B [10].) Bar, 10 μm.
Figure 6. Demonstration that m-ACA1 and m-ACA2 bind to separate epitopes on fp-CNPB1. The cDNA from CNPB1 was digested at the unique Asp-718 site and the two subfragments subcloned into the appropriate pATH vectors where they were expressed as fusion proteins in frame with the bacterial trpE polypeptide (see Materials and Methods for details). (A) Expression of the various fusion proteins: (lanes 1 in all panels) fp-trpE-CNPB1; (lanes 2) fp-trpE-3'CNPB1, the 3' Asp 718-Hind III fragment of λCNPB1; and (lanes 3) fp-trpE-5'CNPB1, the 5' Eco RI-Asp 718 fragment of λCNPB1. (A) Coomassie staining of SDS PAGE of the various lysates. The positions of each induced fusion protein are denoted by arrows at the left. Also noted at the left are molecular mass markers (68 kD [bovine serum albumin], 47 kD [ovalbumin], and 29 kD [carbonic anhydrase]). B-D show the results of probing blots of these fusion polypeptides with (B) autoantiserum GS; (C) m-ACA1, and (D) m-ACA2.

Figure 7. Autoantibody GS and m-ACA1 bind to a common epitope on fp-trpE-CNPB2; m-ACA2 binds to a separate determinant. A bacterial lysate containing fp-trpE-CNPB2 was separated by SDS PAGE and electrophoretically transferred to nitrocellulose. Individual strips of the resultant blot were then probed for immunoreactivity with the following sera. (A) lanes 1-3: positive controls. (Lane 1) Normal human serum WE (dilution 1:1,000); (lane 2) rabbit anti-CENP-B (dilution 1:1,000); (lane 3) autoantibody GS (dilution 1:5,000). Antibody binding was detected with 125I-protein A. Lanes 4-6: negative controls. (Lane 4) PTX-BSA buffer; (lane 5) autoantibody GS (dilution 1:5,000); (lane 6) normal serum WE (dilution 1:100). Bound human antibody was detected with 125I-labeled goat anti-mouse IgG. Note that the anti-mouse secondary antibody does cross-react weakly with the human autoantibody. (Lanes 7-10) Epitope blocking experiment. Each nitrocellulose strip was incubated overnight with autoantiserum GS (dilution 1:100); (lanes 3 and 5) normal serum WE (dilution 1:100). Bound human antibody was detected with 125I-labeled goat anti-mouse IgG. Note that the anti-mouse secondary antibody does cross-react weakly with the human autoantibody. (Lanes 7-10) Epitope blocking experiment. Each nitrocellulose strip was incubated overnight with autoantiserum GS (dilution 1:100); (lanes 3 and 5) normal serum WE (dilution 1:100); (lanes 7 and 9) m-ACA1; (lanes 8 and 10) m-ACA2. Bound mouse antibody was detected with 125I-goat anti-mouse IgG. (B) Binding of both mAbs m-ACA1 and m-ACA2 is blocked by preadsorption with rabbit anti-CENP-B (ra-ACA1). (Lanes 1 and 2) Binding of immune (lane 1) or preimmune (lane 2) rabbit serum ra-ACA1 (both diluted 1:10), detected with 125I-goat anti-mouse IgG. There is no visible cross-reaction. (Lanes 3-6) Epitope blocking experiment. Nitrocellulose strips were incubated overnight with immune serum ra-ACA1 (lanes 3 and 5) or preimmune serum (lanes 4 and 6), both at a dilution of 1:10. After washing, the strips were probed with (lanes 3 and 4) m-ACA1 or (lanes 5 and 6) m-ACA2. In both cases pretreatment with the rabbit polyclonal serum abolished subsequent binding of the monoclonal anti-CENP-B.

Earnshaw et al. Molecular Cloning of cDNA for CENP-B

823
Figure 8. A determinant from the amino-terminal portion of CNPB4 represents a third epitope shared by fp-trpE-CNPB4 and chromosomal CENP-B. (A) Indirect immunofluorescence of chromosomes using (panel 1) antibodies affinity purified against fp-trpE-5' CNPB4, (panel 2) antibodies affinity purified against fp-trpE-CNPB1, and (panel 3) whole KG serum. Affinity-purified antibodies were used at a dilution of 1:40 for immunoblots and 1:2 for immunofluorescence. Whole serum KG was used at a dilution of 1:4,000 throughout. Lysates from induced bacteria were prepared from clones expressing either fp-trpE-5' CNPB4 (B, lanes a) or fp-trpE-CNPB1 (B, lanes b). Parallel samples were subjected to SDS PAGE and three were electrophoblot to nitrocellulose, while a fourth (B 4) was stained with Coomassie Blue. (Note that the 5' fusion accumulates to a much greater level than does fp-trpE-CNPB1.) Each blot was probed as follows: (B 1) Antibodies affinity purified from serum KG by binding to frp-trpE-5' CNPB4; (B 2) antibodies affinity purified from serum KG by binding to fp-trpE-CNPB1; and (B 3) whole serum KG. (C) Immunoblotting of chromosomal proteins probed with antibodies affinity purified against fp-trpE-5' CNPB4 (lane 1), affinity purified against fp-trpE-CNPB1 (lane 2), and whole KG serum (lane 3).

COOH terminus of CNPB1 (Fig. 6, lanes 2), m-ACA2 bound to the NH2-terminal region (Fig. 6 D) but neither autoimmune serum GS nor m-ACA1 bound to either subfragment of CNPB1 (Fig. 6, B and C). Therefore, m-ACA1 and 2 bind to distinct epitopes on fp-trpE-CNPB1, a conclusion that is further supported by the observation that m-ACA2, but not m-ACA1, binds to a proteolytic fragment derived from the fusion protein (see the star in Fig. 6). The data also suggest that the autoepitope spans the junction between the two subcloned portions of CNPB1 or alternatively, that the epitope may be located close to this site and may not adopt a native conformation in the trpE fusion protein.
In competitive binding experiments on blots of fp-trpE-CNPB2, antiserum GS inhibits binding of m-ACA1 but not m-ACA2 (Fig. 7 A). Similar results were obtained using chromosomal CENP-B rather than fp-trpE-CNPB2 as antigen (not shown). In a control experiment, m-ACA1 completely inhibited the binding to fp-trpE-CNPB2 of both m-ACA1 (Fig. 7 B, lanes 3 and 4) and m-ACA2 (Fig. 7 B, lanes 5 and 6). Therefore, m-ACA1 binds at, or adjacent to, an epitope recognized by the human antiserum.

Overall, these data demonstrate that m-ACA1 and 2 bind to distinct sites on fp-trpE-CNPB1. Since both antibodies also bind to CENP-B (Fig. 4), the cloned fusion protein and CENP-B contain at least two epitopes in common.

**A Third Epitope Shared by CNPB4 and Chromosomal CENP-B**

To test whether any epitopes recognized by autoimmune sera are present in the NH2-terminal region of CNPB4, we constructed a plasmid that produces a fusion protein (fp-trpE-5'CNPB4) consisting of the 347 NH2-terminal amino acids of CNPB4. None of these residues is on the original antibody-positive fusion protein CNPB1. When 22 ACA(+) patient sera were screened for binding to fp-trpE-5'CNPB4, five sera showed significant binding (Fig. 8 B, lanes a in gel 3). Earnshaw, W. C., manuscript in preparation). Since the residues in 5'CNPB4 and CNPB1 do not overlap, these five sera must recognize a second, independent autoepitope on CNPB4.

Affinity purification demonstrates that this epitope is also present on authentic CENP-B. ACA(+) serum KG bound to the newly identified epitope as well as it did to the original epitope on CNPB1, thereby permitting us to prepare KG-anti-5' (affinity purified from fp-trpE-5'CNPB4) and KG-anti-3' (purified from frp-trpE-CNPB1). These affinity-purified antibodies are highly specific (Fig. 8 B). For example, KG-anti-5' recognizes only 5'CNPB4 and not CNPB4 in immunoblots (Fig. 8 B, lanes a and b in gel 1). Most importantly, this new 5' epitope is also present on authentic CENP-B (Fig. 8 C) and at centromeres (Fig. 8 A, panel 1). In control experiments, serum KG and KG-anti-3' also bound to CENP-B and to centromeres (Fig. 8 A and C). KG-anti-3' also bound weakly to fp-trpE-5'CNPB4 [Fig. 8 B, lanes a and b in gel 2]. We attribute this slight binding to the fact that CNPB1 is a very "sticky" protein (Earnshaw, W. C., unpublished observations) that may adsorb low levels of anti-5' antibodies. These would readily be detected, since fp-trpE-5'CNPB4 is the most abundant polypeptide present in the induced lysate (Fig. 8 B, lane a in gel 4)."

We have thus demonstrated that fp-CNPB4 and chromosomal CENP-B share at least three independent epitopes (two defined as epitopes and one defined by m-ACA2). We conclude that clone CNPB4 is derived from an mRNA that encodes CENP-B.

**Sequence Analysis of CENP-B Shows It to Be an Acidic Protein Containing Two Long Clusters Comprised Almost Entirely of Acidic Amino Acids**

The nucleotide sequence (determined according to the strategy in Fig. 1) of the 2,617 bases of cDNA in CNPB4 is shown in Fig. 9. The sequence contains a single open reading frame encoding 594 amino acid residues, followed by a 3' untranslated region of 832 bases that terminates at a short poly A sequence. The correct reading frame of CENP-B was specified both by this single long open translation unit and by the production of an immunoreactive fusion protein when the cDNA was translated in this reading frame. Since the cDNA sequence does not contain an appropriate methionine translation initiation codon, it is unlikely that CNPB4 carries the entire CENP-B coding sequence.

Judging from a blot of HeLa cell RNA, a single mRNA species 2,900 bases long encodes CENP-B (Fig. 10 A). Assuming a poly A tail of 100–200 residues, this suggests that CNPB4 is 50–150 bases (or 16–50 amino acids) short of full length.

At first glance these results seem inconsistent with the apparent size of 80 kD for authentic CENP-B, since the single open reading frame encodes a polypeptide of only 64,400 D. However, as tabulated in Table II, all fusion proteins containing CENP-B sequences migrate anomalously in SDS PAGE. For example, electrophoresis of the 47,400-D human CNPB2 polypeptide fused to a 16-amino acid leader (including four amino acids from β-galactosidase and 12 encoded by the polylinker sequence of pUC9) yielded an immunoreactive hybrid protein with an apparent molecular mass of 71,000.

This anomalous behavior may result from a region of primary sequence that contains two extended clusters exceedingly rich in acidic residues. In one stretch of 62 amino acids (in boxed areas of Fig. 9) 79% of the residues are glutamic or aspartic acid (45 glu, 4 asp). A second 31-amino acid region occurring 45 residues closer to the carboxy terminus is comprised of 85% acidic residues (12 glu, 15 asp).

The overall sequence predicts that the cloned portion of CENP-B is an acidic polypeptide with a pI of ~5.2. Two-dimensional immunoblots of chromosomal proteins confirm that authentic CENP-B has a pI of ~5.7 (not shown).

**CENP-B Is Encoded by a Single Copy Human Gene**

To determine the complexity of human DNA sequences with high homology to CENP-B, genomic DNA blot analysis was performed with cDNA from cloned λ-CNPB1. The resultant autoradiogram is shown in Fig. 10 B, lanes 1-3. Since only a single hybridizing fragment is present in any of three digests of human DNA, we conclude that CENP-B is encoded by a single copy gene.

**Discussion**

We have isolated and characterized cloned cDNAs for CENP-B, the major human centromere autoantigen. Five independent lines of evidence demonstrate that this clone encodes authentic CENP-B. First, expression of the cDNA in a bacterial vector yielded an immunoreactive fusion polypeptide that bound (and could be used to affinity purify) antibodies against CENP-B. Second, the apparent size on SDS PAGE and acidic pI of the nearly full-length cloned protein are appropriate for CENP-B. Third, polyclonal antibodies raised against the cloned fusion protein also recognize chromosomal CENP-B. Fourth, two mAbs that recognize distinct, non-overlapping epitopes on the cloned fusion protein also recognize the corresponding epitopes on chromosomal CENP-B. Fifth, autoimmune sera recognize at least two distinct sites present both on authentic CENP-B and on the polypeptide encoded by the cloned CNPB4 cDNA.
strength of this evidence, we conclude that we have indeed identified cloned sequences for CENP-B.

CENP-B is encoded by a single human chromosomal locus that is transcribed into a single 2,900-base mRNA (of which we have cloned ~90–95%). The predicted partial polypeptide sequence of 594 amino acids is most remarkable for the presence of two very highly acidic domains, together comprising almost 100 residues that are >80% glutamic or aspartic acid. The only significant homologies detected in the Los Alamos or Dayhoff sequence banks were to other proteins with extensive acidic regions.

Although the polyacidic regions are the most striking features of the primary sequence, the data of Table II indicate that the carboxy terminus of this polypeptide is unusual as well, since the carboxy-terminal 151-amino acid residues impart a 16-21-kD shift in apparent mobility in SDS PAGE relative to the true molecular mass. For example, in Fig. 6 D, lane 1, it was shown that monoclonal antibody m-ACA2 recognizes two polypeptides in bacterial extracts producing fp-trpE-CNPB1. The smaller polypeptide is apparently a proteolytic fragment derived from the larger one. These polypeptides migrate as though they contain 20 and 38 kD, respectively, of human polypeptide attached to 32 kD of bacterial trpE protein. However, the total human insert present in this clone encodes only 18 kD of human polypeptide. The reason for this abnormal migration in SDS PAGE is unknown.

**Relationship of CENP-A, CENP-B, and CENP-C**

ACAs recognize three distinct chromosomal antigens, CENP-A, CENP-B, and CENP-C (9, 10). Since affinity-purified antibodies to CENP-B obtained from serum GS cross-react with both CENP-A and CENP-C (10), the three polypeptides must share structural determinants. It is therefore important to determine whether the CENP species are the products of a single gene (as the result of either alternative splicing or proteolytic processing), a related gene family, or are derived from unrelated genes.

Our results suggest that CENP-B and CENP-C are the products of unrelated genes but either share a common posttranslational modification or one or more limited regions of conformational or amino acid homology. RNA blotting analysis rules out the possibility that both CENP-B and CENP-C are encoded by a single mRNA. Only a single poly A(+) mRNA (2,900 bases) is detected using the CENP-B cDNA as probe. This mRNA is too small to encode CENP-C (at least 3,800 bases are needed for a 140-kD protein) unless this
CENP-B is encoded by a single-copy human gene that is transcribed into a 2,900-base mature mRNA. (A) Poly A(+) RNA (1 μg) from HeLa cells was analyzed by RNA blotting after electrophoresis on a denaturing formaldehyde agarose gel. CENP-B RNAs were detected by autoradiography of a blot hybridized to a 32p-labeled probe from CNPB1. For accurate determination of CENP-B mRNA size, a collection of RNAs of known sizes (from Bethesda Research Laboratories) were run in an adjacent lane and their migration positions (marked in kilobases at the left) visualized by staining with ethidium bromide. (B) Genomic human DNA was digested with each of three restriction endonucleases and blotted to nitrocellulose after electrophoresis and denaturation. Fragments homologous to CENP-B were identified by autoradiography after hybridization to a 32p-labeled probe constructed from CNPB1. (Lane 1) Eco RI; (lane 2) Hind III; (lane 3) Bam HI.

protein migrates extremely anomalously in SDS PAGE. Therefore, CENP-B is highly unlikely to be derived from CENP-C by proteolysis. Nor are CENP-B and CENP-C the products of closely related genes, since only a single gene for CENP-B was detected in genomic DNA.

Evidence previously obtained with affinity-purified antibodies further highlights the complexity of the relationship between CENP-B and CENP-C. When anti–CENP-C antibodies were affinity purified from serum GS, they did not rebind to CENP-B (10). However, the anti-CENP-B antibodies from this serum did recognize epitopes common to CENP-B and CENP-C (10).

The relationship between CENP-A and CENP-B is less clear. CENP-A could be derived from CENP-B by proteolytic cleavage, since antibodies affinity purified from it also recognize CENP-B in immunoblots (10). However, if CENP-A were in fact derived from CENP-B, the shared epitope must reside in the extreme NH2-terminal region, since autoantibodies affinity purified from our longest CENP-B fusion protein (fp-CNPB4) do not recognize CENP-A.

Although we have screened our cDNA libraries extensively for clones that correspond to CENP-A and CENP-C, none have yet been found. Possible explanations for this are as follows.

(a) The epitopes on CENP-C may be localized near the NH2 terminus, and cDNAs of sufficient length (>3,000 bp) may be rare in our libraries.

(b) Similarly, the epitope(s) on CENP-A might originate from the very NH2 terminus of CENP-B, and these might also be present only rarely in our libraries.

(c) The epitopes common to CENPs A and B and CENPs B and C might arise from posttranslational modification of the proteins. Such epitopes cannot be cloned by the λgt11 method, since E. coli does not perform such modifications. However, if this is the case, then two distinct modifications on CENP-A and CENP-C must be involved since affinity-purified anti–CENPs A and C do not cross-react (10).

(d) CENP-A and C may be the products of independent genes and might be transcribed from mRNAs present at very low abundance in the cells from which our cDNA libraries were made.

(e) CENP-A could be a centromere-specific histone subtype, as suggested by the experiments of Palmer and Margolis (35). In this case, CENP-A could be encoded by a poly A(−) mRNA that would be absent from our cDNA libraries.

We cannot at present distinguish among these possibilities.

Significance of the Poly-acidic Domain of CENP-B

Initially we were surprised that a chromatin protein like CENP-B would have highly acidic subdomains. However, other chromatin proteins (the most notable example being HMG-I [51]) also have acidic subdomains. It seems unlikely that such acidic regions would engage in direct interactions with the DNA. Rather, they must interact with basic regions of other chromosomal proteins.

The most striking examples of basic protein domains are the NH2-terminal regions of the core histones (22). These domains are, in general, ~20–30 residues in length, and are highly lysine and arginine rich. It was initially assumed that these regions would anchor the histones to the DNA and be

<table>
<thead>
<tr>
<th>Fusion protein</th>
<th>Vector</th>
<th>True molecular mass</th>
<th>Apparent molecular mass</th>
<th>Discrepancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>fp-β-gal-CNPB1</td>
<td>λgt11</td>
<td>133</td>
<td>150</td>
<td>17</td>
</tr>
<tr>
<td>fp-trpE-CNPB1</td>
<td>pATH 11</td>
<td>50</td>
<td>68</td>
<td>18</td>
</tr>
<tr>
<td>fp-trpE-CNPB2</td>
<td>pATH 3</td>
<td>79</td>
<td>102</td>
<td>23</td>
</tr>
<tr>
<td>fp-β-gal-CNPB2</td>
<td>pUC9</td>
<td>48.5</td>
<td>71</td>
<td>22.5</td>
</tr>
<tr>
<td>fp-trpE-CNPB4</td>
<td>pATH 11</td>
<td>97</td>
<td>123</td>
<td>26</td>
</tr>
</tbody>
</table>

Table II. Anomalous Migration of Fusion Proteins Containing CENP-B
important for nucleosome structure, but subsequent analyses suggest that their function is more subtle (see review by McGhee and Felsenfeld [30]). If the NH₂-terminal domains are removed by proteolysis (52), the truncated histones retain the capability of assembly into nucleosomes (53; reviewed in reference 30). Therefore, these domains do not appear to be essential for intra-nucleosomal contacts.

Nucleosomes constructed from truncated histones do, however, differ in one significant respect from normal nucleosomes—they are not readily induced to form 30-nm higher-order structures (1, 53). Therefore the basic domain appears to participate in nucleosome packing.

We suggest that the acidic regions of certain chromatin proteins (such as CENP-B) may serve to "capture" the NH₂-terminal histone domains. This capture might then remove the domains from participating in stabilization of the higher-order chromatin fiber, causing the fiber to open into the 10-nm "beads on a string" conformation. The open conformation would then permit other proteins (or other domains of the original protein) to more readily scan the DNA for specific recognition sequences such as have been shown to specify centromere location and function in Saccharomyces (6, 15, 20, 21, 28, 36, 45).

The Amount of CENP-B Varies between Chromosomes

The micrographs of Figs. 2, 4, and 5 reveal a very unexpected result concerning the distribution of CENP-B in the various chromosomes: the level of CENP-B antigen varies widely. This variation is seen with affinity-purified antibodies from the autoantiserum (Fig. 2 c, panel 2), with rabbit polyclonal anti-CENP-B (Fig. 4 b), and, most convincingly, with monoclonal anti-CENP-B (Fig. 5).

This finding has important implications for the possible structural distribution and physiological role of CENP-B in vivo. To our knowledge, the size of the kinetochores on various chromosomes has not been rigorously compared; however, our expectation is that kinetochore size is relatively constant. In contrast, the amount of C-band material (thought to be at least in part centromeric heterochromatin—reference 7) varies widely from chromosome to chromosome. The human Y chromosome is C-band negative and also lacks detectable CENP-B (Earnshaw, W. C., and G. Stetin, manuscript in preparation). The obvious conclusion to draw from these results is that CENP-B is somehow associated with centromeric heterochromatin.

Earlier immunolabeling (4) suggested that centromeric autoantigens are restricted to the kinetochore, where they occur throughout all layers. Based on our results, we predict that CENP-B occurs in the basal layer of the kinetochore, in association with centromeric heterochromatin. We therefore predict that CENP-B is unlikely to be involved in microtubule binding.

Examination of chromosomes stained with whole autoantiserum (Fig. 2 c, panel 7) shows all centromeres stained more or less equally. Thus, we infer that other components recognized by such autoantiserum (i.e., CENP-A and CENP-C) may show a more uniform distribution and are better candidates for components of the kinetochore trilaminar plaque structure and/or microtubule binding.

We thank Drs. Moses Chow, Dan Littman, and Richard Axel for providing the endothelial cell λgt11 library. We gratefully acknowledge the efforts of R. Bernat and L. Satterwhite in helping to isolate clone λ-CNPB4.

This work has been supported by grants from the National Institutes of Health (NIH) (GM 35212 to W. C. Earnshaw and N. F. Rothfield, GM 29513 to D. W. Cleveland and GM 26132 to T. D. Pollard) and the Devil's Bag Award from the Arthritis Foundation (to W. C. Earnshaw and D. W. Cleveland). D. W. Cleveland is the recipient of a Research Career Development Award from the NIH. K. F. Sullivan was supported by a postdoctoral fellowship from the NIH.

Received for publication 10 November 1986.

References


