ELSEVIER

Biochimica et Biophysica Acta 1307 (1996) 21–25

BBA Biochimica et Biophysica Acta

Short sequence-paper

Molecular cloning of an intronless gene for the hamster centromere antigen CENP-B [1]

Luis A. Bejarano, Manuel M. Valdivia *

Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias, Universidad de Cádiz, 11510 Puerto Real, Cádiz, Spain

Received 8 November 1995; accepted 7 February 1996

Abstract

Centromere protein B (CENP-B) is a DNA-binding protein present at both active and inactive centromeres. It was first localized at the kinetochore region by human autoimmune sera from CREST patients. Using a previously identified human cDNA we have isolated a genomic clone containing the complete hamster CENP-B intronless coding sequence. At the nucleotide level it was found to possess a high degree of homology with the human and mouse CENP-B genes, being 75% and 90% respectively. This codes for 606 amino acid residues, which represent seven more than the human and mouse centromeric proteins. Hamster CENP-B protein analysis revealed at the N-terminal region a 133 amino acid fragment of 100% homology to the DNA binding motif identified previously for the human autoantigen. Expression of hamster CENP-B during the cell cycle was analyzed by using a specific antiCENP-B serum generated against the C-terminal conserved region. These data indicate that CENP-B is highly conserved and it represents a universal component of the centromere structure and function in mammals.

Keywords: Centromere protein B; Genomic clone; Intronless coding sequence; (Hamster)

The mammalian centromere is a restricted region of the chromosome that forms the primary constriction in condensed metaphase chromatin, and represents the site for sister chromatin attachment. At the surface of each centromere, a disk-shaped kinetochore structure is assembled during early mitosis for microtubules binding [2]. Structural analysis of mammalian centromeres have shown that they contain long stretches of tandemly repeated DNA sequences. The interaction between DNA and proteins at the centromere region is important because they will determine the structure and function of the centromere. Several proteins such as CENP-A, CENP-B, CENP-C, and CENP-D, were originally identified as putative centromere proteins by Western blot analysis using antisera from patients with the autoimmune disease scleroderma [4,11,12,19]. Immunofluorescence microscopic observations revealed that these proteins localized at the centromere throughout the cell cycle [5,6,14,18]. Molecular cloning and expression demonstrated later on unequivocally the centromere localization of CENP-B [5,15] and CENP-C [14] in humans and mouse chromosomes. Other centromere components, including CENP-E and CENP-F, translocate from centromere to midbody during the cell cycle [3,8,13,20,21]. In any case the role at the molecular level of these centromeric proteins remains to be elucidated. It was reported that CENP-B binds to human alphoid DNA that contains a 17 bp sequence, the 'CENP-B box' [9,10,16]. This sequence was also found in a mouse minor satellite repeat [9]. From these studies it was suggested that this CENP-B/alphoid interaction may be of significance in organizing centromeric DNA in mammalian cells. In order to deduce the function of CENP-B/alphoid DNA complexes at the centromeres, as an universal macromolecular structure in mammalian cells, we have started to study the expression of both components in a hamster cell line (CHO). In this study, taking advantage of information on CENP-B sequences in human and mouse, we have cloned and sequenced CENP-B genomic DNA from hamster. By using a human CENP-B probe we identified from a lambda FIX CHO-K1 genomic library (Stratagene) a DNA clone that contained sequences encoding an intronless gene for the hamster centromeric protein CENP-B. Several subclones were obtained in the Bluescript (pBS SK-) cloning

* Corresponding author. Fax: +34 56 834924.

[1] The sequence data reported in this paper have been deposited with the EMBL/GenBank Data Libraries under the accession number U20951.

PII S0167-4781(96)00039-5

vector after digestion with *Sma*I, *Xba*I, *Bam*HI, *Hin*dIII and *Kpn*I enzymes. DNA sequencing was carried out using Sequenase reagents and protocols (U.S. Biochemical) by using primers from BS and thereafter synthetic oligonucleotide to the sequenced strand. The sequence determined an intronless DNA for the hamster CENP-B (Fig. 1) as indicated before for the human and mouse CENP-B clones [5,15]. Restriction mapping and Southern blot indicated that it corresponded to a single copy CENP-B sequence in the hamster genome (Fig. 1). Subsequent analysis revealed that the hamster CENP-B gene is composed of 1818 codifying nucleotides with an open reading frame encoding for 606 amino acid residues with a mass of 66.4 kDa (Fig. 2). This is in concordance with that described for human and mouse CENP-B polypeptides which migrates anomalously in SDS-PAGE with an apparent mass 15 kDa larger than predicted [5,15].

Comparison of nucleotide sequences with those in computer databases revealed that it shares 75% and 90% sequence identity within the coding region with those of human and mouse CENP-B respectively. At the amino acid level the homology of hamster CENP-B with human and mouse polypeptides are 92% and 95% respectively. The hamster CENP-B polypeptide was found to contain 7 extra amino acid residues than those of the other mammalian CENP-B already described. Those extra residues were located at the C-terminal long polyacidic region of the centromeric protein. A unique feature of the three mammalian CENP-B polypeptides so far isolated is a 100% homology of a stretch of 133 amino acid residues at the N-terminal region. This part of the CENP-B protein has been implicated in the binding to the alphoid satellite DNA sequence in human and mouse [7,9,16]. This DNA was named as the CENP-B box. This finding suggests that in hamster chromosomes, CENP-B could be associated with centromeric heterochromatin underlying the kinetochore domain interacting with a satellite DNA close in sequence to the CENP-box in a similar manner to that described for the human and mouse centromeric proteins. However, the presence in the hamster genome of a homologous centromere sequence to the human alphoid remains to be elucidated.

The expression of the hamster CENP-B centromere protein was further investigated by immunofluorescence analysis on CHO culture cells with an antiserum to the C-terminal region of the molecule. In a previous study [1] we showed the feasibility of generating specific antibodies to short peptide sequences of the CENP-B molecule in spite of the great conservation of the primary structure of the protein. In this study we revealed the expression of CENP-B during the cell cycle of the CHO culture cells by an antiCENP-B-peptide serum previously described for human cells. This antibody recognizes chromosomal regions from interphase to mitotic cells (Fig. 3). The staining was characteristic of the centromeres as revealed by the intranuclear speckles pattern in interphase cells and the pairs of fluorescence stained dots on condensed chromosomes in mitosis. The C-terminal region of CENP-B has been implicated in the self association to fold centromeric DNA sequences into a heterochromatin structure [17]. The highly conserved amino acid domain at the C-terminal region of this centromere protein from human to hamster seems plausible for a conserved role of this polypeptide in a conserved function such as the condensate stage of the centromeric region during the cell cycle.

In summary, we have cloned the hamster centromeric

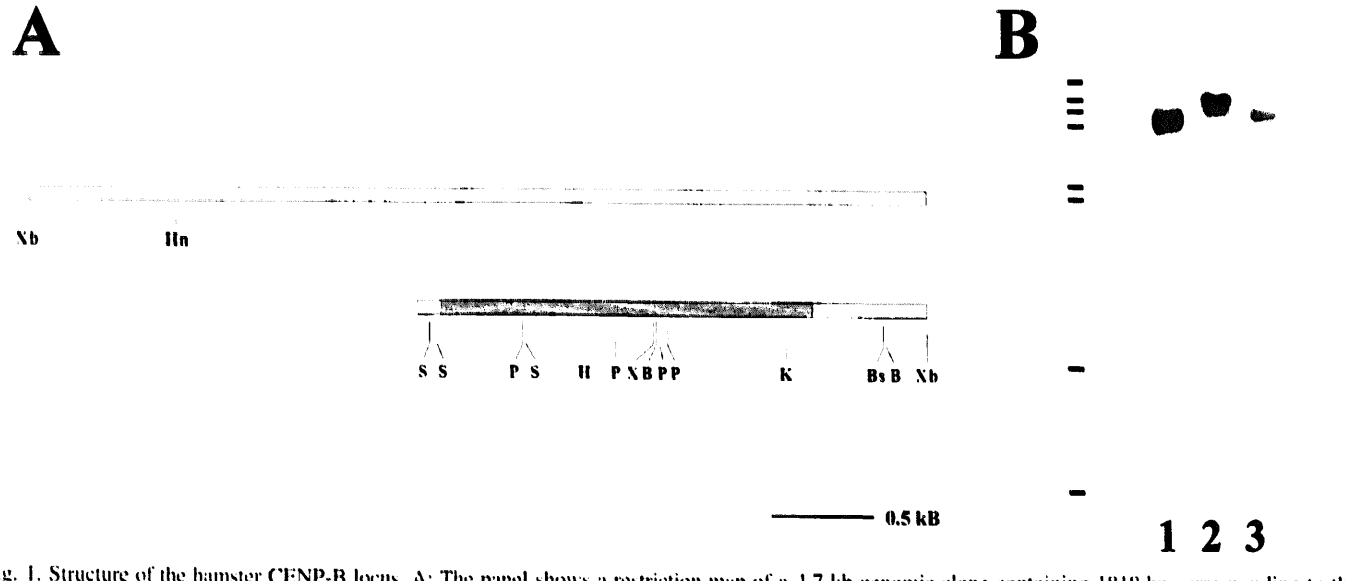

Fig. 1. Structure of the hamster CENP-B locus. A: The panel shows a restriction map of a 4.7 kb genomic clone containing 1818 bp corresponding to the intronless gene for the hamster CENP-B centromere protein. Restriction enzymes sites are marked along the length of the gene (B, *Bam*HI; Bs, *Bst*XI; H, *Hin*dIII; Hn, *Hin*cII; K, *Kpn*I; P, *Pst*I; S, *Sma*I; X, *Xho*I; Xb, *Xba*I). B: Southern blot analysis of hamster genomic DNA digested with different restriction enzymes and probed with a CENP-B clone. *Xba*I digests lane 1; *Sac*I digests lane 2; *Eco*RI digests lane 3. Markers size as shown on the left, correspond to 23, 9.4, 6.5, 4.3, 2.3, 2.0, 0.56 and 0.12 kb for λ-*Hin*dIII.

cgcgagccgctttgtctcgggcggggcgc 29

gcggagaggccgccaggtgcccccccgcgggcccgggcggcggggcggacgagcgcggtgccgggccccggggcggagggcgcgccggg 119

ATGGGCCCCAAGCGGCGGCAGCTGACGTTCCGGGAGAAGTCGCGGATCATCCAGGAGGTGGAGGAGAACCCGGACCTACGCAAGGGCGAG 209
M G P K R R Q L T F R E K S R I I Q E V E E N P D L R K G E 30

ATCGCGCGGCGCTTCAACATCCCGCCGTCCACGCTGAGCACCATCCTGAAGAACAAGCGCGCCATCCTGGCGTCGGAGCGCAAGTACGGA 299
I A R R F N I P P S T L S T I L K N K R A I L A S E R K Y G 60

GTGGCCTCCACCTGCCGTAAGACCAACAAGCTGTCCCCGTACGACAAGCTGGAGGGGCTTCTCATCGCTTGGTTCCAGCAGATCCGCGCC 389
V A S T C R K T N K L S P Y D K L E G L L I A W F Q Q I R A 90

GCGGGCCTGCCTGTCAAGGGCATCATCCTGAAAGAGAAGGCGCTACGGATAGCGGAGGAGCTGGGCATGGACGACTTCACGGCTTCCAAC 479
A G L P V K G I I L K E K A L R I A E E L G M D D F T A S N 120

GGCTGGCTGGATCGCTTCCGCCGGCGCCACGGTGTAGTGGCCTGCAGCGGCGTGACCCGCTCCCGGGCGCGAACGTCTACCCCACGGGCC 569
G W L D R F R R R H G V V A C S G V T R S R A R T S T P R A 150

CCAGCGGCACCTGCCGGCCCAGCCGCCGTGCCCTCTGAGGGCAGCGGTGGCAGTACACCAGGCTGGCGCACTCGGGAGGAGCAGCCGCCG 659
P A A P A G P A A V P S E G S G G S T P G W R T R E E Q P P 180

TCGGTGGCTGAGGGGTACGCCTCGCAGGACGTGTTCAGCGCCACCGAGACCAGCCTGTGGTACGACTTTCTGTCCGACCAGGCCTCGGGG 749
S V A E G Y A S Q D V F S A T E T S L W Y D F L S D Q A S G 210

CTGTGGGGAGGTGATGGAACGGCTCGCCAGGCCACCCAGCGTCTTAGCGTTTTGCTGTGCGCCAACCGCGATGGCAGCGAAAAGCTTCCC 839
L W G G D G T A R Q A T Q R L S V L L C A N R D G S E K L P 240

CCACTGGTTGCAGGCAAGTCCGCCAAGCCCCGTGCAAGCCAGGGTGGTCTGCCCTGCGACTACACTGCCAACTCTAAGGGTGGAGTCACC 929
P L V A G K S A K P R A S Q G G L P C D Y T A N S K G G V T 270

ACCCAGGCCCTGGCTAAGTACTTGAAAGCTCTGGACACCCGAATGGCTGCAGAATCTCGTCGGGTCCTTCTGCTTGCAGGCCGTCTGGCT 1019
T Q A L A K Y L K A L D T R M A A E S R R V L L L A G R L A 300

GCCCAGTCCTTGGACACCTCGGGCCTGCGGCACGTGCAGCTGGCCTTCTTCCCCCCCGGCACCGTGCATCCTTTGGAGCGAGGAGTGGTC 1109
A Q S L D T S G L R H V Q L A F F P P G T V H P L E R G V V 330

CAGCAGGTGAAGGGCCACTACCGCCAGGCTATGTTGCTCAAGGCCATGGCAGCACTCGAGGGCCAGGATCCCTCAGGCCTGCAGCTGGGC 1199
Q Q V K G H Y R Q A M L L K A M A A L E G Q D P S G L Q L G 360

CTAGTGGAGGCCTTACACTTTGTGGCTGCAGCCTGGCAGGCAGTGGAGCCCGCGGACATAGCAACTTGCTTTCGCGAGGCCGGTTTTGGA 1289
L V E A L H F V A A A W Q A V E P A D I A T C F R E A G F G 390

GGTGGCCTTAATGCCACTATCACCACTTCCTTCAAAAGCGAGGGAGAGGAGGAGGAGGAAGAGGAGGAGGAGGAAGAGCAGGAGGAAGAA 1379
G G L N A T I T T S F K S E G E E E E E E E E E E E E E E E 420

GAGGAGGGTGAAGGGGAAGAGGAGGAGGAGGAAGAGGAAGAAGGGGAGGAGGAAGGAGGGGAAGGAGAGGAGGTGGGAGAGGAGGAGGAG 1469
E E G E G E E E E E E E E G E E E G G E G E E V G E E E E 450

GTAGAAGAGGAGGGTGATGAGAGTGATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGAAAGCTCCTCTGAGGGCTTAGAGGCTGAA 1559
V E E E G D E S D E E E E E E E E E E E S S S E G L E A E 480

GACTGGGCACAGGGAGTAGTGGAGGCCAGTGGTGGCTTTGGGGGTTACAGTGTCCAGGAAGAGGCCCAGTGCCCAACTCTCCATTTCCTG 1649
D W A Q G V V E A S G G F G G Y S V Q E E A Q C P T L H F L 510

GAAGGTGGGGAGGACTCTGACTCGGACAGTGATGAAGAGGAGGAAGATGAAGAGGAGGATGAGGAAGATGAAGAAGATGACGATGACGAT 1739
E G G E D S D S D S D E E E E D E E E D E E D E E D D D D D 540

GAGGATGGTGATGAGGTCCCTGTGCCCAGCTTTGGGGAGGCCATGGCTTACTTTGCCATGGTCAAGAGGTACCTGACCTCCTTTCCCATT 1829
E D G D E V P V P S F G E A M A Y F A M V K R Y L T S F P I 570

GACGACCGCGTACAGAGTCACATCCTTCACTTGGAACATGATCTGGTCCATGTGACTAGGAAGAACCATGCCTGGCAGGCGGGAGTTCGG 1919
D D R V Q S H I L H L E H D L V H V T R K N H A W Q A G V R 600

GGTCTTGGACACCAAAGCTGAgctgctggacatatctgtgctccagcccagatgcgagcacctgccaaggcaggagaactccgggcagct 2009
G L G H Q S 606

gctggagacagctggagaagtcccagggccttcagcaatgctttgccagcctgagacaggccaggggttgaggtctgcctcactgctatt 2099

gcctctttctcagagtcctgtttcctccccattagtccctgggctcagggcactgggtgggaggggagctgtccggtgctaccacaccat 2189

gccatcagtgggctagaccacagcagcagccagggacgggtcctggaagctcttggccagagagtgcctctccccctgcctcccaaccag 2279

gtccttggtgggggatcccaaagccattctggaagggctccagaggaaggtccagcctaggctcccacaaaattagcagcccccctcct 2369

gcacctctaggttgtctaaagaagcacagtgtaacttagggcaggctcctgaacctgcctcttcctgctttccacctccctaaatccctt 2459

tctctggcccagtctttggcccttggttttctttctaga 2498

Fig. 2. Complete nucleotide sequence of the hamster CENP-B gene. The sequence of 2498 bases segment of the 4.7 kb genomic *Xba*I fragment is shown. Sequence was unambiguosly determined on both DNA strands. The complete predicted polypeptide sequence is shown below the DNA sequence. The hamster CENP-B polypeptide contains 606 amino acid residues in comparison with those 599 residues of human and mouse CENP-B genes. A polyadenylation signal was not found in this clone.

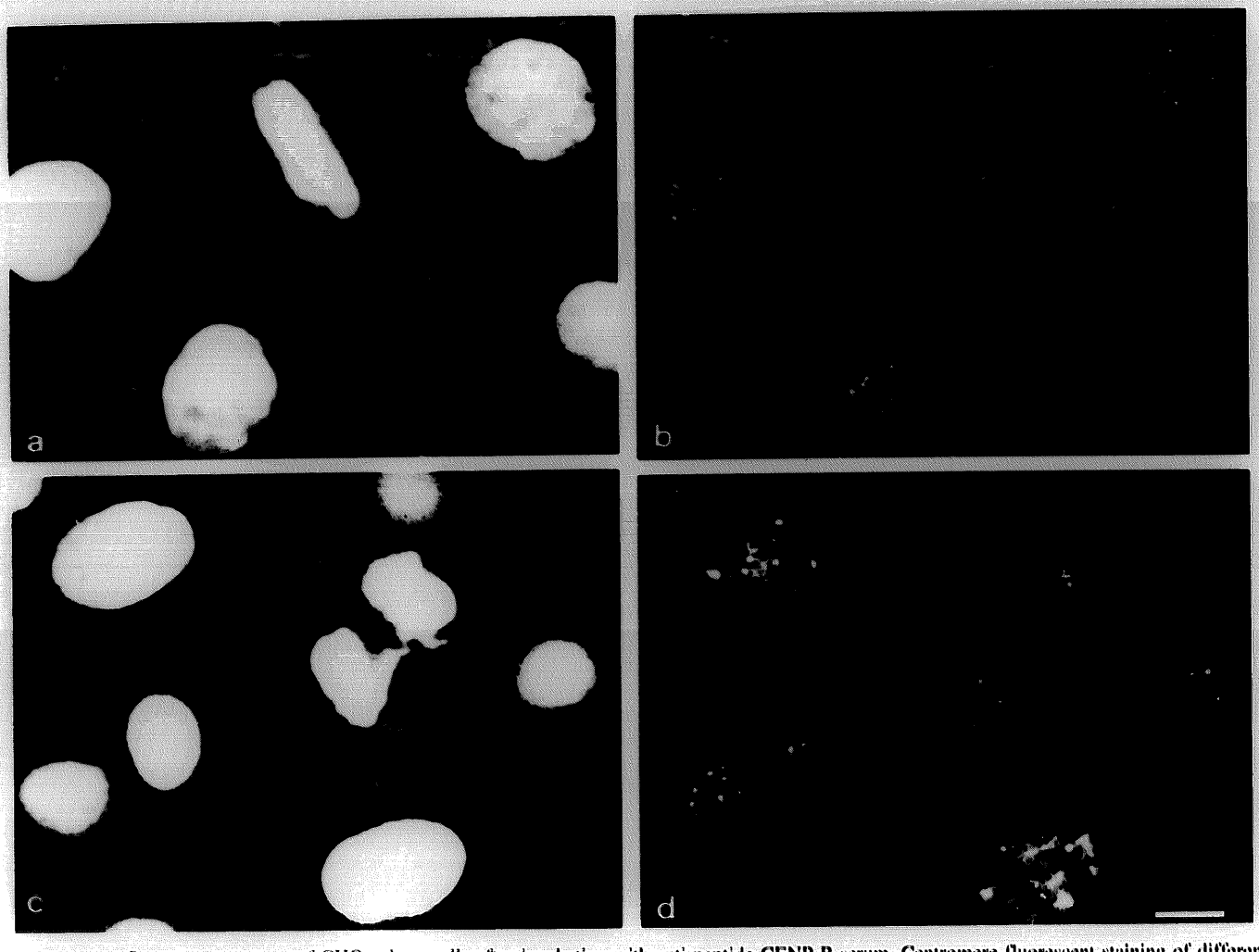

Fig. 3. Immunofluorescence pattern of CHO culture cells after incubation with anti-peptide-CENP-B serum. Centromere fluorescent staining of different stages of the CHO cell cycle is shown in b and d. Interphase, metaphase and anaphase cells are all observed with a characteristic speckled staining pattern for centromere. Same cells were stained for DNA with Hoechst 33528 (a,c). Bar = 20 μm.

protein CENP-B, and showed its expression during CHO culture cells. Further analysis of the CENP-B binding sequences in the hamster genome may hopefully elucidate how centromeric heterochromatin structure is integrated in mammals.

We thank Dr. Kevin Sullivan for kindly providing the human CENP-B probe. We also thank Concha Iglesias for maintaining cell lines. This work was supported by grants from Comisión Interministerial de Ciencia y Tecnología (PB87-0964, PB93-0726) and by Plan Andaluz de Investigación (grupo 3127).

References

[1] Bejarano, L.A. and Valdivia, M.M. (1993) Cytog. Cell Genet. 63, 54–58.
[2] Brinkley, B.R., Ouspenski, I. and Zinkowski, R.P. (1992) Trends Cell Biol. 91, 95–102.
[3] Cooke, C.A., Heck, M.M.S. and Earnshaw, W.C. (1987) J. Cell Biol. 15, 2053–2067.
[4] Earnshaw, W.C. and Rothfield, N. (1985) Chromosoma 91, 313–321.
[5] Earnshaw, W.C., Sullivan, K.F., Machlin, P.S., Cooke, C.A., Kaiser, D.A., Pollard, T.D., Rothfield, N.F. and Cleveland, D.W. (1987) J. Cell Biol. 104, 817–829.
[6] Earnshaw, W.C., Ratrie, H. and Stetten, G. (1989) Chromosoma 98, 1–12.
[7] Haaf, T., Warburton, P.E. and Willard, P.E. (1992) Cell 70, 681–696.
[8] Liao, H., Winkfein, R.J., Mack, G., Rattner, J.B. and Yen, T.J. (1995) J. Cell Biol. 130, 507–518.
[9] Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. and Okazaki, T. (1989) J. Cell Biol. 109, 1963–1973.
[10] Masumoto, H., Sugimoto, K. and Okazaki, T. (1989) Exp. Cell Res. 181, 181–196.
[11] Moroi, Y., Peebles, C., Fritzler, M.J., Steigerwald, J. and Tan, E.M. (1980) Proc. Natl. Acad. Sci. USA 77, 1627–1631.
[12] Palmer, D.K., O'Day, K., Trong, H.L., Charbonneau, H. and Margolis, R.L. (1991) Proc. Natl. Acad. Sci. USA 88, 3734–3738.
[13] Rattner, J.B., Rao, A., Fritzler, M.J., Valencia, D.W. and Yen, T.J. (1993) Cell Motil. Cytoskeleton 26, 214–226.
[14] Saitoh, H., Tomkiel, J., Cooke, C.A., Ratrie, H., Maurer, M., Rothfield, N.F. and Earnshaw, W.C. (1992) Cell 70, 115–125.
[15] Sugimoto, K., Furukawa, K. and Himeno, M. (1994) Chromosome Res. 2, 453–459.
[16] Sugimoto, K., Hagishita, Y. and Himeno, M. (1994) J. Biol. Chem. 269, 24271–24276.

[17] Sullivan, K.F. and Glass, C.A. (1991) Chromosoma 100, 360–370.
[18] Sullivan, K.F., Hechenberger, M. and Masri, K. (1994) J. Cell Biol. 127, 581–592.
[19] Valdivia, M.M. and Brinkley, B.R. (1985) J. Cell Biol. 101, 1124–1134.
[20] Yen, T.J., Compton, D.A., Wise, D., Zinkowski, R.P., Brinkley, B.R., Earnshaw, W.C. and Cleveland, D.W. (1991) EMBO J. 10, 1245–1254.
[21] Yen, T.J., Li, G., Schaar, B., Szilak, I. and Cleveland, D.W. (1992) Nature 359, 536–539.