Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 8253–8257, November 1986
Genetics

Analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe*

(repeated DNA sequence/chromosomal organization/chromosome walking)

LOUISE CLARKE, HANSPETER AMSTUTZ, BARBARA FISHEL, AND JOHN CARBON

Department of Biological Sciences, University of California, Santa Barbara, CA 93106

Contributed by John Carbon, July 1, 1986

ABSTRACT **The *Schizosaccharomyces pombe* centromere-linked genes, *LYS1* and *CYH1* on chromosome I and *TPS13* and *RAN1* on chromosome II, have been isolated. The genetic order of these markers with respect to their centromeres was determined to establish relative directionality on the genetic and physical maps. Chromosome walking toward the centromeres reveals a group of repetitive sequences that occur only in the centromere regions of chromosomes I and II and at one other specific location in the *S. pombe* genome, presumably the centromere of chromosome III. The major class of large repeated sequence elements is 6.4 kilobases (kb) long (repeat K), portions of which occur at least twice on chromosome II and in several tandemly arranged intact copies at another centromeric location. Repeat K in turn contains groups of smaller repeats. Genetic recombination is strongly suppressed in the centromere II region, which contains at least 30 kb of repeated sequences. Centromeric DNA organization is much more complex in fission yeast than has been described in budding yeast (*Saccharomyces cerevisiae*), possibly because of the larger more condensed nature of the *S. pombe* chromosomes.**

The fission yeast, *Schizosaccharomyces pombe*, has a genetic complexity roughly equivalent to that of the budding yeast, *Saccharomyces cerevisiae*. Unlike the common yeast, however, which contains 17 small chromosomes, the DNA of *S. pombe* is organized into three larger chromosomes, which condense and are visible under light microscopy (1, 2). In *S. cerevisiae* there appears to be only a single microtubule attachment site per chromosome (3), an observation consistent with the lack of repeated DNA sequences and the overall small size [150–200 base pairs (bp)] of the centromere regions (4). The number of microtubule binding sites per chromosome in *S. pombe* is unknown. It is conceivable that the larger and condensed nature of *S. pombe* chromosomes demands a correspondingly more complex kinetochore structure, more representative of chromosomes of higher eukaryotes. A DNA transformation system (5) as well as highly developed formal genetics (6) make *S. pombe* a tractable system for the study of centromere structure and chromosome segregation. Here we describe a structural analysis that reveals a group of repetitive DNA sequences that occur in the centromere regions of the *S. pombe* chromosomes. The centromere region on *S. pombe* chromosome II encompasses at least 30 kilobases (kb) of DNA, contains distinctly heterochromatin-like sequences, and is thus similar to centromeres of higher eukaryotes. Studies on repeated DNA sequences in the *S. pombe* centromere regions have also recently been reported by Nakaseko *et al.* (7).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Strains, Media, and Genetic Manipulations. *Escherichia coli* strains JA221 (*hsdR⁻ HsdM⁺ trpΔE5 leuB6 recA*) and JA226 (8) were recipients for DNA transformations. *S. pombe* strains Sp223 (h^- *leu1.32 ura4 ade6.216*) and Sp192 (h^- *ran1 ura4 ade6.216*) were gifts from David Beach (9). *S. pombe* strains cited below and in the legend to Table 1 were constructed in our laboratory from strains kindly provided by Jürg Kohli. Yeast (10) and bacterial (11) media were prepared as described. Genetic manipulations of *S. pombe* were carried out as described (6, 10).

Enzymes, Recombinant DNA Technology, Gel Electrophoresis, and DNA Sequencing. Restriction enzymes, T4 DNA ligase, and DNA polymerase were from New England Biolabs and were used according to the vendor's instructions. Unless stated otherwise, procedures for *E. coli* DNA transformation, cloning techniques, plasmid isolation, blot hybridizations, and gel electrophoresis were carried out as described by Maniatis *et al.* (11). *S. pombe* DNA was prepared according to Beach and Klar (12). The dideoxy sequencing method (13) was used on an M13mp19 clone (14) carrying the 1.0-kb *Kpn* I/*Hin*dIII fragment of pSp14. Orthogonal field alternation gel electrophoresis (15) and field inversion gel electrophoresis (16) have been described.

***S. pombe* Genomic Plasmid and Cosmid Libraries.** Construction of libraries in *E. coli* strain JA226(*recBC*) from *Sau*3A partial digests of *S. pombe* DNA (strain 972h^-; ref. 10) cloned into the *E. coli–S. pombe* shuttle vectors, PCV4 and pFL20 (17), has been described (8). The cosmid vector pHPC was constructed by inserting a 2.2-kb *Sal* I/*Xho* I restriction fragment carrying the *S. cerevisiae LEU2* gene (18) and a 1.1-kb *Eco*RI fragment carrying the *S. pombe ARS* from plasmid pFL20 (17) into the vector pJB8 (19). The *S. cerevisiae LEU2* gene complements the *leu1.32* mutation in *S. pombe* (5). The cosmid library was prepared from pHPC and *Mbo* I partial digestion products of *S. pombe* DNA according to Ish-Horowicz and Burke (19) and was propagated in *E. coli* strain JA221 (*recA*).

Transformations and Isolation of the *S. pombe* Genes *LYS1*, *CYH1*, *TPS13*, and *RAN1*. *S. pombe* DNA transformations (5) and site-directed integrations of plasmids (20) were performed as described. The *LYS1* and *TPS13* genes (6) were cloned by complementation of appropriate markers (*lys1* and *tps13 ura4*) in *S. pombe* with plasmids pSpLYS1 and pSpTPS13 (Fig. 1) from the PCV4 and pFL20 libraries, respectively. These plasmids were recovered in *E. coli* strain JA226 as described (21). Plasmids or cosmids complementing *ran1* (9) and *cyh1* (6) were subsequently identified among those isolated by overlap hybridization screening (22), using hybridization probes isolated from pSpLYS1 and pSpTPS13 DNAs.

Abbreviations: bp, base pair(s); kb, kilobase(s).

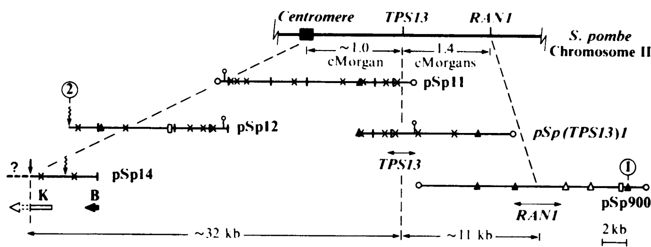

FIG. 1. Genetic and physical maps of the *S. pombe* centromere II region. Top line represents a genetic map of the centromere II region. The restriction maps show the location of all *Eco*RI (|), *Hin*dIII (×), *Bam*HI (▲), and *Sal* I (□) sites, and selected *Bgl* II (♀), *Bst*EII (⌇), *Kpn* I (↓), *Hpa* I (△), and *Sau*3A (○) sites; the location of *Hin*dIII sites was not determined for pSp900. Plasmid pSp12 was obtained as follows: The leftmost *Sau*3A/*Eco*RI fragment of pSp11 was cloned into a pBR322-*LEU2* vector. The resulting plasmid was cut at the unique *Bgl* II site and introduced by site-directed integration (20) into the homologous site in *S. pombe* strain Sp223 (*leu1.32*). Genomic DNA from the transformant was then restricted with *Bst*EII, generating a linear fragment carrying pBR322 and centromere-proximal DNA sequences, which was recovered as plasmid pSp12 in *E. coli* strain JA226. Similarly, pSp14 was recovered from *Kpn* I-restricted yeast DNA into which plasmid pSp13 (Fig. 2) had been integrated at a unique *Hin*dIII site. The locations of repeats B and K are indicated by arrows. ① and ② represent sites of *LEU2* integration for mapping studies described in Table 1.

RESULTS

Chromosome Walking in the Centromere Regions of Chromosomes I and II. The tightly centromere-linked genes *LYS1* and *CYH1* on chromosome I and *TPS13* and *RAN1* on chromosome II were isolated from *S. pombe* genomic libraries, constructed in plasmid or cosmid shuttle vectors, by complementation of the markers in appropriate *S. pombe* strains (see *Materials and Methods*; Fig. 1). The *LYS1*, *CYH1*, and *TPS13* genes have also been isolated in other laboratories (7, 23). The *lys1* and *cyh1* markers are very tightly linked genetically (Table 1; refs. 6 and 24), with a map distance of 0.5–1 cM between them. We have isolated both genes on the same 20-kb segment of DNA, and other investigators report that the loci are ≈10 kb apart (7). Mapping data indicate that *lys1* is the centromere-proximal locus and occurs ≈4 cM from centromere I (Table 1). Thus, chromosome-walking was initiated toward the centromere based on these observations. Similarly, the order of *tps13* and *ran1* on chromosome II was determined with reference to centromere II (Table 1 and Fig. 1). The two loci are physically 11 kb apart and genetically ≈1.4 cM apart. *TPS13* is the centromere-proximal gene and is located ≈1 cM from centromere II. The directionality of the physical and genetic maps on chromosome II was confirmed by site-directed integration (20) in separate strains of an *S. cerevisiae LEU2* gene at either of the locations marked ① and ② in Fig. 1. Genetic mapping of each of the integrated *LEU2* markers with respect to *tps13* and the centromere established unequivocally that the centromere was located to the left of *TPS13* on the physical map shown in Fig. 1 and must be extremely close to integration site ② (Table 1).

The physical and genetic relationships of the genetic markers and their relationship to the centromere on *S. pombe* chromosome II are summarized in Fig. 1. Representative sets of overlapping DNA segments from the region, isolated from plasmid and cosmid genomic libraries by overlap hybridization (22) or by site-directed integration (20) followed by excision of plasmid and neighboring centromere-proximal sequences, are depicted along with partial restriction maps. With the exception of the centromere-proximal regions overlined by arrows in Figs. 1 and 2, all the remaining DNA

Table 1. Genetic and physical distances in the centromere regions of *S. pombe* chromosomes I and II

Genetic cross no.	Map interval	Genetic distance, cM	Physical distance, kb	Ratio, kb/cM
	Chromosome II			
1 (106)	*tps13–ran1*	1.4	11	7.9
	tps13–cen	0.5	32	64
	ran1–cen	1.9	43	23
2 (110)	*tps13–LEU2**	1.4	18	13
	tps13–cen	1.4	32	23
	LEU2–cen*	2.8	50	18
3 (23)	*tps13–LEU2*†	4.4	30–32	—
	LEU2†*–cen*	0	0–2	—
	Chromosome I			
2 (110)	*lys1–cyh1*	0.5	ND	ND
	lys1–cen	3.6	>20	ND
	cyh1–cen	4.1	>20	ND

Genetic cross no. 1: SBP12984-19A (h^+ *cyh1 fur1 lys1 ran1*) × SBP12984-19B (h^- *tps13*); cross no. 2: SBP7185BF/301 (h^+ *ade6-210 cyh1 leu1 lys1 ura4/LEU2*) × SBP72385-60 (h^- *fur1 leu1 tps13*); cross no. 3: SBP223/13 (h^- *leu1 ura4 ade6-210/LEU2*) × SBP72385-44 (h^+ *leu1 lys1 fur1 tps13*). Strain SBP7185BF/301 was constructed by site-directed integration (20) of plasmid pSp301, which contains the 9-kb *Bam*HI fragment of pSp900 (see Fig. 1) cloned into the pBR322-*LEU2* plasmid pGT47. Strain SBP223/13 was constructed in a similar manner by site-directed integration of plasmid pSp13 (see Fig. 2*A*). Numbers in parentheses represent number of tetrads. ND, not determined.

**LEU2* integrated at site ① (Fig. 1).

†*LEU2* integrated at site ② (Fig. 1).

appears to contain only unique sequences, as determined by Southern blotting analysis (data not shown; ref. 25).

One striking observation derived from the data shown in Fig. 1 and Table 1 is that the ratio of physical to genetic distance increases considerably proceeding in the direction of the centromere on chromosome II. The *TPS13–RAN1* physical and genetic distances give a ratio of 8 kb/cM, identical to that of the overall genomic average (26). The distance from *TPS13* to the region of repeated DNA sequences (discussed in detail below) is, however, around 32 kb, but genetically the *TPS13–CEN2* distance is only ≈1 cM. This expanded physical to genetic distance ratio observed in the centromere II region of *S. pombe* is consistent with restricted recombina-

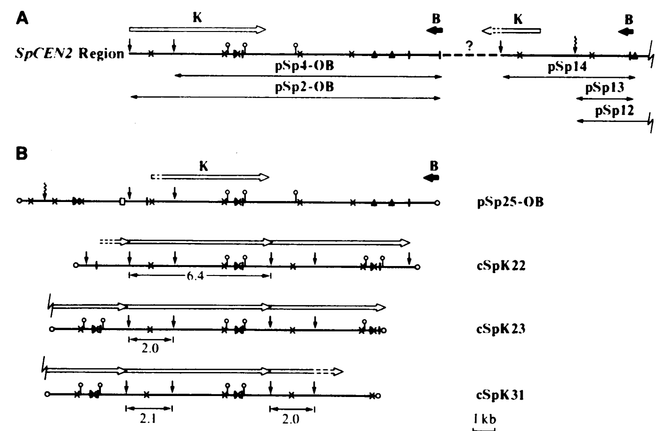

FIG. 2. (*A*) Restriction map of a portion of the *S. pombe* centromere II region showing locations of repeats B and K (indicated by arrows). The location of *Eco*RI (|), *Hin*dIII (×), *Bam*HI (▲), *Bgl* II (♀), *Kpn* I (↓), and *Bst*EII (⌇) sites are indicated. (*B*) Plasmids or cosmids carrying repeats B and/or K (designated by arrows). Sizes are in kb. Restriction enzyme sites are as indicated in *A*; *Sal* I (□) and selected *Sau*3A (○) sites are also shown.

tion phenomena observed in the centromere regions of higher eukaryotic chromosomes (27).

Two or More Classes of Large Repeated Sequence Elements Are Found in the Centromere II Region. Two restriction fragments, the 1.0-kb *Kpn* I/*Hin*dIII fragment and the 3.5-kb *Bst*EII/*Bam*HI fragment, both located near the centromere-proximal end of the region from chromosome II presented in Fig. 1, hybridize to multiple restriction fragments in the *S. pombe* genome (data not shown; see Fig. 4). The two sequences (repeats B and K in Figs. 1 and 2) do not cross-hybridize, however.

Attempts to continue the chromosome walk by overlap hybridization screening were frustrated by the extreme structural instability in *E. coli* of clones containing sequences immediately adjacent to the K repeat. Previous workers have experienced difficulty in cloning relatively large segments of DNA comprised of highly repetitive sequences, even in recombination-deficient *E. coli* hosts (28). Therefore, an alternative strategy was used to isolate DNA sequences beyond the *Kpn* I/*Hin*dIII fragment in the centromere-proximal direction. The plasmid pSp14, whose insert is delineated in Figs. 1 and 2, was obtained by site-directed integration at a unique *Hin*dIII site of another plasmid, pSp13 (Fig. 2 and Table 1, legend). Plasmid pSp13 carries the *S. cerevisiae LEU2* gene [complements the *leu1.132* mutation in *S. pombe* (5)], the 3.5-kb *Bst*EII/*Bam*HI restriction fragment containing repeat B (see Figs. 1 and 2), and also contains single *Kpn* I and *Sal* I sites within the vector sequences. The resulting integrated form of pSp13 was present in only one location in the *S. pombe* genome as determined by Southern blot analysis (data not shown). Furthermore, the introduced *LEU2* gene mapped in a standard genetic cross tightly linked to *TPS13* and *CEN2*; in a total of 23 tetrads examined, *LEU2* showed absolute linkage to *CEN2* (Table 1). Total genomic DNA from strain SBP223/13 (Table 1) was then restricted to completion with *Kpn* I or *Sal* I in order to generate linear DNA fragments carrying both pBR322 and neighboring centromere-proximal sequences; the restricted DNA was diluted, ligated, and used to transform *E. coli* strain JA226 (*recBC*) to ampicillin resistance. All transformants obtained with *Kpn* I-restricted DNA contained plasmid pSp14 (Fig. 1). No transformants were obtained with *Sal* I-restricted DNA, probably because the resulting plasmid would exceed 90 kb and contain many repeated sequences (see below). However, when total genomic DNA from *S. pombe* strain SBP223/13 was partially digested with *Kpn* I, diluted, ligated, and used to transform *E. coli*, well over one-half of the 300 clones examined contained plasmids with multiple *Kpn* I sites and inserts that extended toward *TPS13* in the centromere-distal direction, some as much as 15–20 kb. Most of the remainder were identical to pSp14.

A total of 10 *E. coli* transformants from the experiment described above with partially restricted DNA grew very poorly on ampicillin medium and contained one or the other of the plasmids whose inserts are shown in Fig. 2 and labeled pSp2-OB and pSp4-OB. Structural analysis of pSp2-OB and pSp4-OB indicates that they did not arise by a simple excision event from the genomic DNA into which pSp13 was integrated. They do not contain DNA sequences between and including the *Bst*EII site and the neighboring centromere-distal *Hin*dIII site (Fig. 2), which would be expected if they did arise in a manner similar to pSp14. The most plausible explanation for their origin is that they arose normally as partial *Kpn* I digestion products, linear fragments containing pBR322 sequences as well as centromere-proximal *S. pombe* sequences extending through several *Kpn* I sites, but upon subsequent ligation and transformation into *E. coli*, a portion of these sequences was lost by recombination between two or more B repeats that occur in the centromere region of chromosome II. Both pSp2-OB and pSp4-OB do contain sequences at their ends (rightmost in Fig. 2), and only at this location, that hybridize strongly to the B repeat region within the 3.5-kb *Bst*EII/*Bam*HI fragment. The two inserts are identical, at least with respect to arrangement of restriction sites, except that pSp2-OB contains an additional 2.0-kb *Kpn* I fragment in the centromere-proximal direction. Further structural analysis of the DNA inserts in plasmids pSp2-OB and pSp4-OB indicates that they hybridize near their leftmost ends (Fig. 2) to the repeat K sequences contained on pSp14. Thus, both repeats B and K are present at least twice in the centromere II region of *S. pombe*.

Upon probing *S. pombe* genomic plasmid and cosmid libraries with repeat K sequences, a number of additional DNA segments containing the K repeat were isolated and characterized (Fig. 2*B*). Plasmid pSp25-OB contains ≈14 kb of DNA homologous to pSp2-OB, including portions of repeats B and K, as well as an additional 7 kb of unique sequence. The cosmids cSpK22, cSpK23, and cSpK31 are of particular interest because they define the complete K repeat unit and demonstrate that this repeated sequence exists in the genome in a tandem array of at least three units. The recombinant cosmids were originally packaged *in vitro*, which presumably selects for molecules of 40–50 kb. However, cosmids cSpK22, cSpK23, and cSpK31, as recovered from *E. coli* (*recA*), are only ≈25 kb long, which may indicate that they originally contained repetitive DNA sequences that do not replicate well in *E. coli*. The size of the consensus K repeat is 6.4 kb; slightly less than a complete unit is present on pSp2-OB and pSp25-OB. Copies of repeat K can vary internally; the *Kpn* I fragment that normally measures 2.0 kb is 2.1 kb in one copy of the repeat on cSpK31 (Fig. 2*B*). This *Kpn* I fragment also shows size variation in genomic Southern blots.

Nucleotide sequence analysis has been carried out on that portion of repeat K that extends 600 bp in the centromere-distal direction from the *Kpn* I site in pSp14 (Fig. 3). The sequenced region hybridizes only to sequences surrounding the internal *Kpn* I site in the full-length K repeat (Fig. 2*B*). The first 215 bases extending from the *Kpn* I site contain two classes of short repeated elements (Fig. 3). The 12-bp sequence TGTCAGTACTAG occurs five times (with minor variations) in a tandem array beginning at the *Kpn* I site. This repeat transitions into a second 6- to 7-bp repetitive sequence element, with the consensus sequence PyACCAPy, that is present 25 times within the first 215 bp from the *Kpn* I site and occupies 70% of the sequence in this region. Other minor

```
KpnI
1                                                                    70
GGTACCATTAGTACCAGTACTAGTGTCATCTAGTGTCAGTACTAGTGTCAGTATAGGCATCAACATCATC
71                                                                  140
ACCAGTATCGGCGCCACGATCAGCACAGCACCACTACCACTACCACTTCCACTTACCACAAGCACTACCA
141                                                                 210
CCACCACCACCACTTCCACTTACCACTTCCATTACTACCACGACTACCACTAACCATCACCACTTTATCT
211                                                                 280
CCATTCTTGTCATTATTGGCTTGTTGTACGTTGTTTAAATGTTTCAACGTATGTACATTTGTGACTACGA
281                                                                 350
TTTCTTGTTAGCGATATAAACGAAGAGGATAAAGACTTTGGAGTAAATAATTTTGGCCGTGCACTAGCAT
351                                                                 420
AATGTATTTATGCGTCTAGGTATCCTGCGTCTCGGTATCCTTAGCGTTGCAAGTGGAAAGTGGCTTCACA
421                                                                 490
CTATAAATGGTTGACACAGCTAATACATTAGACATAAACTTTCTCAGCATTCACGAACAGTAAAAATACT
491                                                                 560
ATTGGTAAATGATTCATGGATATATAGTATTAGTATTTGGCATTCGTTTCGCTTATGAATTGTGATTGAA
561                                                                 630
CCGATATCGTATCATCATTAAAGTTCTGTGCACTCTACATCATTCAACATGTGAGATCTCATCATTCACA
631                   653
AATATTTCCATCGCATCGTTATT
```

FIG. 3. Nucleotide sequence of a portion of repeat K from pSp14. The sequence begins at the single *Kpn* I site in pSp14 (Fig. 1) and proceeds in the centromere-distal direction (toward *TPS13*). Various repeated sequences are indicated.

repeats present in two or three copies each are indicated in Fig. 3.

Thus, the centromere II region contains several types of repeated DNA elements, the largest of which (the 6.4-kb K repeat) occurs in part at least twice in this region and three or more times in tandem either in this or in other centromere regions (see below). Another repeat (repeat B), whose size has not been determined, is also present at least twice in the *CEN2* region. Contained within repeat K are two classes of small repeated sequences of 6 and 12 bp, which bear a striking resemblance to satellite DNAs found in higher eukaryotes.

Repeats B and K Are Present Only at Three Specific Genomic Locations, at Least Two of Which Are Centromeric. The repeated sequences discussed above have all been found at the centromere region of the chromosome II but are also located elsewhere in the *S. pombe* genome. If these classes of repeats are restricted only to the centromere regions of the three *S. pombe* chromosomes, this may imply a role for them in centromere function.

Restriction fragments containing portions of repeats K or B hybridize to what appear to be single large *Sal* I or *Bst*EII fragments in standard 1% agarose gel electrophoretic separations. However, orthogonal field alternation gel electrophoresis (ref. 15; Fig. 4*A*) or field inversion gel electrophoresis (ref. 16; Fig. 4*B*) reveals that these repeats actually hybridize to three very large DNA fragments in each case. Averaged over a number of orthogonal field alternation gel electrophoresis and field inversion gel electrophoresis experiments, the sizes of the *Sal* I fragments are 120, 90, and 50 kb, and the *Bst*EII fragments are 110, 70, and 30 kb. The 90-kb *Sal* I fragment hybridizes to a unique 4-kb *Sal* I/*Hin*dIII probe from the chromosome II walk (Fig. 1); thus, this fragment is from the centromere region of chromosome II. Restriction fragments containing repeat K also hybridize to DNA sequences that are located ≈20 kb from the *LYS1* gene on chromosome I. We have been unable to determine the exact physical distance between *LYS1* and repeat K on chromosome I, because our plasmids or cosmids all appear to contain deletions at or near the junction of repeat K with unique flanking sequences. Other investigators, however, have also identified a portion of repeat K (termed dgI) near *LYS1* and report a distance between them of ≈20 kb (7). Thus, one of the other two large *Sal* I fragments must contain the centromere region of chromosome I. It is likely that the third fragment includes the centromere region of chromosome III. However, this has not yet been established.

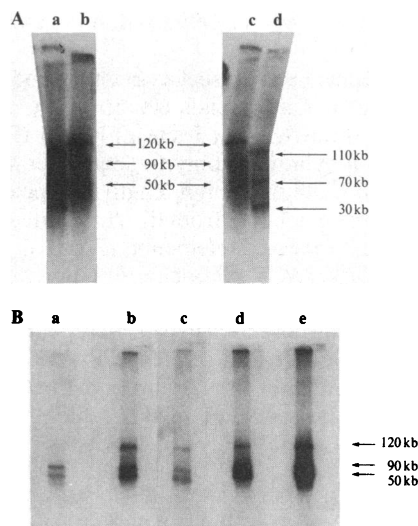

FIG. 4. The K repeat occurs only on three large *Sal* I or *Bst*EII restriction fragments in the *S. pombe* genome. *S. pombe* genomic DNA was digested with *Bst*EII (lanes a and d) or *Sal* I (lanes b and c) and subjected to orthogonal field alternation gel electrophoresis analysis (15). The gel was dried and hybridized (29) to 0.2 μg of ^{32}P-labeled 1.0-kb *Hin*dIII/*Kpn* I fragment (≈10^8 cpm/μg), which carries the arrowhead-distal portion of repeat K (Fig. 1) (lanes a and b), or to 0.2 μg of the labeled 0.7-kb *Hin*dIII fragment (10^8 cpm/μg), which carries the arrowhead-proximal portion of repeat K (Fig. 3) (lanes c and d). (*B*) Sequences from the centromere II region occur at three specific locations in the *S. pombe* genome. *Sal* I-restricted *S. pombe* genomic DNA was subjected to field inversion gel electrophoresis analysis (16). All lanes are from a single gel, which was dried, cut into individual lanes, and hybridized (29) to the following ^{32}P-labeled probes, as illustrated from right to left at the top of Fig. 2*A*: lane a, the 0.8-kb *Hin*dIII/*Bst*EII fragment; lane b, the 2.4-kb *Bst*EII/*Hin*dIII fragment; lane c, the 1.0-kb *Hin*dIII/*Kpn* I fragment; lane d, the 2.4-kb *Hin*dIII fragment; lane e, the 2.8-kb *Hin*dIII fragment. Approximately 0.2 μg of each probe (≈10^8 cpm/μg) was used for each hybridization.

Selected hybridization probes derived from DNA sequences in the centromere region of chromosome II were used in orthogonal field alternation gel electrophoresis and field inversion gel electrophoresis experiments (Fig. 4) to determine how much of this region is also present at the other two locations, as defined by the large *Sal* I fragments. Individual labeled restriction fragment probes that include essentially all the DNA from the repeat B region shown at the right of Fig. 2*A* through the repeat K region at the left hybridized to the same three large *Sal* I fragments. These probes include (from right to left in Fig. 2*A*) the 3.5-kb *Bam*HI/*Bst*EII, the 2.6-kb *Bst*EII/*Hin*dIII, the 1.0-kb *Hin*dIII/*Kpn* I, the 0.6-kb *Bam*HI, the 2.8-kb *Hin*dIII, the 2.4-kb *Hin*dIII, and the 0.7-kb *Hin*dIII fragments. Examples of field inversion gel electrophoresis hybridizations with the probes are shown in Fig. 4*B*. Probably not all sequences within a given fragment are hybridizing at each location, however. For example, the 3.5-kb *Bam*HI/*Bst*EII fragment, which includes repeat B, hybridizes more strongly to the 90-kb *Sal* I fragment (from chromosome II) than it does to the other two *Sal* I fragments (data not shown), and the 0.8-kb *Bst*EII/*Hin*dIII portion of the 3.5-kb fragment appears to hybridize only to the 90-kb and the 50-kb *Sal* I genomic fragments and extremely weakly, if at all, to the largest *Sal* I fragment (see lane a, Fig. 4*B*). Thus, the size of the region on chromosome II that contains sequences in common with the other two genomic locations is >23 kb but <90 kb. A unique 0.9-kb *Eco*RI/*Mbo* II fragment from the leftmost end of cosmid cSpK22 (Fig. 2*B*) hybridizes only to the 120-kb *Sal* I genomic fragment, associating at least three tandem copies of repeat K with this region of the genome.

DISCUSSION

In the genomes of higher eukaryotes, many classes of repeated DNA sequences have been identified, and chromosomes often contain heterochromatin in their centromeric and telomeric regions. The genome of the budding yeast *S. cerevisiae* contains a class of large (6 kb) repeats, the Ty elements, which are normally found in ≈30 copies dispersed throughout the chromosomes (30), but no repeated DNA sequences are found within the functional centromeric DNA, and actively transcribed genes occur within a few hundred base pairs of the centromeres (31).

In this study, we have encountered several types of repeated DNA sequences in the centromere regions of *S. pombe*. Repeat K, whose unit length is 6.4 kb, occurs in part at least twice in the *CEN2* region and at least three times in tandem either in this region or at another centromeric location. Repeat B is less well-defined, but is present at least twice in the *CEN2* region, and probes that encompass repeat B hybridize with various intensities to all three large *Sal* I fragments that define two, and, possibly, all three centromere regions. Furthermore, probes from the entire 23-kb region

between repeats B and K on chromosome II also hybridize only to these three large genomic *Sal* I fragments, indicating that relatively long segments are repeated within the centromere regions of the three chromosomes. Repeat K in turn contains several classes of short internal repeats. Both on plasmids and on cosmids isolated in *recA* or *recBC E. coli* hosts, repeat K appears to be adjacent to a region, possibly of additional repeated sequences, that replicates poorly in *E. coli* and deletes with high frequency.

Nakaseko *et al.* (7) have recently identified two classes of large repeated sequences in the *S. pombe* centromere II region. Their repeat yn1 appears, in part, to be analogous to repeat B. Their repeated sequence, dg, is analogous by nucleotide sequence analysis to a portion of repeat K, although dg is reported to be 3.8 kb long, whereas the consensus K repeat measures 6.4 kb. Difficulties in maintaining all of these repeated sequences intact during replication in *E. coli* may explain discrepancies in our data with that of Nakaseko *et al.* (7), who have also analyzed nucleotide sequences within the repeat K unit. For example, our 653-bp sequence from repeat K (Fig. 3) is nearly identical to a portion of their 3.9-kb dgIIa repeat, which occurs in a similar chromosomal location, except dgIIa contains an additional two copies of the 12-bp TGTCAGTACTAG (consensus) repeat, and, after nucleotide 613 (Fig. 3), the sequences diverge for 25–30 bp.

In *S. cerevisiae*, a single microtubule attaches to each centromere, directly to the 20-nm chromatin fiber, and there are no visible differentiated kinetochore structures (3). The outer layer of the highly differentiated mammalian kinetochores, however, is comprised of an arrangement of chromatin stacked in hairpin loops, and bundles of microtubules are seen directly attached to this chromatin (32). Presumably, the arrangement of centromeric DNA sequences in organisms with highly differentiated kinetochores would reflect both the multiplicity of spindle binding sites and the presence of repeated chromatin loops. Therefore, it would not be surprising to find DNA sequences that are repeated at regular intervals in the centromere regions. In *S. pombe*, repeats B and/or K may participate in forming a specialized chromatin structure that differentiates the kinetochore from the chromosomal arms. Furthermore, recombinational suppression, which is well-documented for centromere regions of *Drosophila* (33, 34), might be important for organisms with complex kinetochore structures, where an optimal number of repeated DNA sequences per kinetochore must be maintained (27). In *S. pombe*, the suppression of recombination in the centromere II region might be necessary to maintain a complex arrangement of essential repeated sequences. Alternatively, the repeats may have merely accumulated in the recombination-deficient centromere regions and serve no useful function (27).

We have previously described a DNA segment from plasmid pSp11 (Fig. 1) that contains a sequence highly homologous to the *S. cerevisiae CEN* element III consensus sequence (4) and that controls the copy number of *ARS* plasmids to 1 or 2 per cell in *S. pombe* (35). However, deletion of this sequence from the *S. pombe* genome does not destabilize chromosome II mitotically; thus, it is unlikely that this region is important in centromere function. Neither repeat B nor K hybridizes to *S. cerevisiae* genomic DNA in standard Southern blots or orthogonal field alternation gel electrophoresis analysis.

In *S. cerevisiae*, the presence of a functional centromere (*CEN*) sequence on an *ARS* plasmid stabilizes the plasmid mitotically and enables it to segregate properly through both meiotic divisions (36). We have not detected such behavior, either in *S. pombe* or in *S. cerevisiae* host cells, with circular plasmids containing any of the segments of centromeric DNA isolated in this study. However, we have not as yet obtained an entire centromere region, as defined by the region of repeated DNA sequences, on a single cloned DNA segment.

We thank David Beach and Jürg Kohli for providing *S. pombe* strains. We also gratefully acknowledge the excellent technical assistance of Virginia Kuga. This research was supported by postdoctoral fellowships from the National Institutes of Health (B.F.) and the Swiss National Foundation (H.A.), and by a research grant from the National Institutes of Health (GM-33783).

1. Robinow, E. F. (1977) *Genetics* **87**, 491–497.
2. Umesono, K., Toda, T., Hayashi, S. & Yanagida, M. (1983) *J. Mol. Biol.* **168**, 271–284.
3. Peterson, J. B. & Ris, H. (1976) *J. Cell Sci.* **22**, 219–242.
4. Clarke, L. & Carbon, J. (1985) *Annu. Rev. Genet.* **19**, 29–56.
5. Beach, D. & Nurse, P. (1981) *Nature (London)* **290**, 140–142.
6. Kohli, J., Hottinger, H., Munz, P., Strauss, A. & Thuriaux, P. (1977) *Genetics* **87**, 471–489.
7. Nakaseko, Y., Adachi, Y., Funahashi, S., Niwa, O. & Yanagida, M. (1986) *EMBO J.* **5**, 1011–1021.
8. Elliott, S., Chang, C.-W., Schweingruber, M. E., Schaller, J., Rickli, E. E. & Carbon, J. (1986) *J. Biol. Chem.* **261**, 2936–2941.
9. Beach, D., Rodgers, L. & Gould, J. (1985) *Curr. Genet.* **10**, 297–311.
10. Gutz, H., Heslot, H., Leupold, U. & Loprieno, N. (1974) in *Handbook of Genetics*, ed. King, R. C. (Plenum, NY), Vol. 1, pp. 395–446.
11. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
12. Beach, D. H. & Klar, A. J. S. (1984) *EMBO J.* **3**, 603–610.
13. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
14. Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* **9**, 309–321.
15. Carle, G. F. & Olsen, M. V. (1984) *Proc. Natl. Acad. Sci. USA* **82**, 3756–3760.
16. Carle, G. F., Frank, M. & Olson, M. V. (1986) *Science* **232**, 65–68.
17. Losson, R. & Lacroute, F. (1983) *Cell* **32**, 371–377.
18. Ratzkin, B. & Carbon, J. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 487–491.
19. Ish-Horowicz, D. & Burke, J. F. (1981) *Nucleic Acids Res.* **9**, 2989–2998.
20. Orr-Weaver, T. L., Szostak, J. W. & Rothstein, R. J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6354–6358.
21. Devenish, R. J. & Newlon, C. S. (1982) *Gene* **18**, 227–288.
22. Chinault, A. C. & Carbon, J. (1979) *Gene* **5**, 111–126.
23. Beach, D., Piper, M. & Nurse, P. (1982) *Mol. Gen. Genet.* **187**, 326–329.
24. Nakaseko, Y., Niwa, O. & Yanagida, M. (1984) *J. Bacteriol.* **157**, 334–336.
25. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
26. Gygax, A. & Thuriaux, P. (1984) *Curr. Genet.* **81**, 85–92.
27. Charlesworth, B., Langley, C. H. & Stephan, W. (1986) *Genetics* **112**, 947–962.
28. Lohe, A. R. & Brutlag, D. L. (1985) *Proc. Natl. Acad. Sci. USA* **83**, 696–700.
29. Tsao, S. G. S., Brunk, C. F. & Pearlman, R. E. (1983) *Anal. Biochem.* **131**, 355–372.
30. Cameron, J. R., Loh, E. Y. & Davis, R. W. (1979) *Cell* **16**, 739–751.
31. Yeh, E., Carbon, J. & Bloom, K. (1986) *Mol. Cell. Biol.* **6**, 158–167.
32. Ris, H. & Witt, P. L. (1981) *Chromosoma* **82**, 153–170.
33. Lindsey, D. L. & Sandler, L. (1977) *Philos. Trans. R. Soc. London Ser. B* **277**, 295–312.
34. Yamamoto, M. & Miklos, G. L. G. (1978) *Chromosoma* **66**, 71–98.
35. Clarke, L., Fishel, B. R., Amstutz, H. & Carbon, J. (1985) *J. Cell. Biochem.* Suppl. **9C**, 132.
36. Clarke, L. & Carbon, J. (1980) *Nature (London)* **287**, 504–509.