Volume 12 Number 18 1984 Nucleic Acids Research

Subtelomeric regions of yeast chromosomes contain a 36 base-pair tandemly repeated sequence

Heidi Horowitz and James E.Haber

Rosenstiel Basic Medical Sciences Research Center and Department of Biology, Brandeis University, Waltham, MA 02254, USA

Received 11 June 1984; Revised and Accepted 21 August 1984

ABSTRACT

We have determined the nucleotide sequence of a region of DNA derived from the end of one chromosome of the yeast, Saccharomyces cerevisiae. Inspection of the sequence reveals the presence of 12 tandem direct repeats, each 36 nucleotides long and having nearly identical sequence. Each 36 base-pair repeat can be further subdivided into three tandem sub-repeats of a similar 12 base-pair sequence. Analysis of total genomic yeast DNA from several strains by Southern hybridization suggests that the number of tandem 36 base-pair repeat units may vary from approximately 8 to 25 among different telomeric regions. Differences in the number of repeats may have arisen by unequal crossing over between them. Furthermore, the finding that the pattern of bases at multiple variable positions within the repeat unit is not random suggests that these regions may undergo gene conversion events that render them homogeneous.

INTRODUCTION

The ends of eucaryotic chromosomes, telomeres, are specialized structures which permit complete replication of the linear DNA molecule and lend stability to the chromosome. In contrast, the ends of broken chromosomes (which lack normal telomeres) are highly recombinogenic (1-6) and are frequently lost from the cell during cell division.

Recently, Szostak and Blackburn (7) cloned one telomere from the yeast Saccharomyces cerevisiae onto the end of a linear plasmid vector. The resulting linear plasmid was capable of replicating in yeast and was maintained with virtually the same structure. A portion of the linear plasmid was then circularized so that it could be transformed into, and maintained in E. coli. The circular derivative, pSZ220, has since been used as a convenient source of telomere-specific probe in Southern blots of total genomic yeast DNA. Experiments of this type have indicated that the DNA at the ends of many other yeast chromosomes shares a large degree of homology with the specific telomere cloned.

Experiments performed by Chan and Tye (8) have suggested a basic

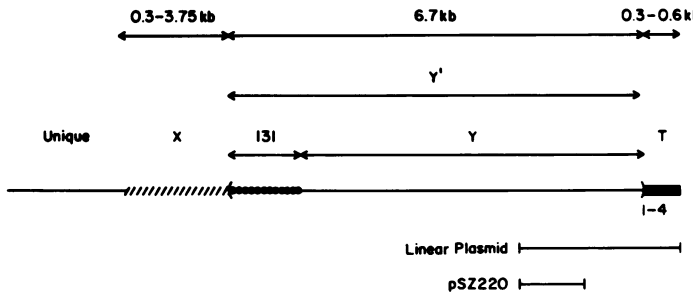

Fig. 1. General structure for a Y'-containing yeast telomere. a) The structure presented is from Chan and Tye (8). The very end of the telomere (T) is heterogeneous in length (7,9,10). Centromere-proximal to T is a highly conserved 6.7 kb sequence, Y', which consists of a Y region and an associated repetitive element, 131. Y' may be present as a single copy or tandemly repeated up to 4 times. Centromere-proximal to the Y' region is a less highly conserved sequence, X, and adjacent to that is unique chromosomal DNA. b) The position of the cloned telomere DNA (7) present in both the linear plasmid and its circular derivative (pSZ220) is indicated.

structure for one class of yeast telomeres as indicated in Fig. 1. The DNA at the very end of the chromosome (T) is heterogeneous in length and is composed of a DNA sequence in which one strand is rich in C+A (7,9-12). Centromere-proximal to the T sequence is the very highly conserved Y' region which, in some cases, is arranged in a tandem array of up to 4 copies. Centromere-proximal to the Y' regions is the heterogeneous and less highly conserved X region, and proximal to that is unique chromosomal DNA. The yeast telomere cloned on the linear plasmid by Szostak and Blackburn (7) is approximately 2.5 kb long and corresponds to a fragment of DNA containing part of a Y region and a T region, as defined by Chan and Tye (8 and see Fig. 1). There is also evidence for the existence of telomeres lacking sequences homologous to Y' (11-13). The nature of the sequences present at these telomeres has not yet been elucidated, however.

Recently, we have described an unusual degree of polymorphism among telomeric regions of different yeast strains (13). As an initial attempt to better understand these strain-specific differences, we have analyzed the specific cloned telomere in greater detail by sequencing the 1 kb region of this clone present in the plasmid pSZ220. We have found that a 36 base-pair sequence is tandemly repeated 12 times in this DNA. Hybridization studies of total genomic yeast DNA have revealed that, for the most part, these tandem repeats are found embedded in similar sequences of DNA; the inferred number of repeats in different Y regions ranges from approximately 8 to 25 in the

strains examined. In this paper we discuss various aspects of the DNA sequence of part of the Y region from one telomere. Specifically, we examine the potential significance of the repeats for the recombination and homogenization of telomeric regions.

EXPERIMENTAL PROCEDURES

Materials. All restriction enzymes, T4 DNA ligase and E. coli DNA polymerase were purchased from New England Biolabs. The Klenow fragment of DNA polymerase was purchased from Boehringer-Mannheim; the specific M13 primer d(GTAAAACGACGGCCAGT) was purchased from Collaborative Research Inc. [α-^{32}P]dTTP was purchased from New England Nuclear. The plasmid pSZ220 (7), which contains part of the Y region of a specific cloned yeast telomere, was kindly provided by J. Szostak.

DNA Sequence Analysis. The DNA sequence of the Y region cloned in pSZ220 was determined by the dideoxy chain termination method (14). Restriction fragments were cloned into the vector M13mp9 (15). Recombinant phage were analyzed by polymerase extension of the primer:single-stranded DNA hybrid with radioactive nucleotides, followed by restriction analysis. Clones possessing the desired constructions were then subjected to sequence analysis as described (14).

Strains. The yeast strains used for these studies were MCY130 (MATa SUC7 ade2 suc2°: the product of a strain which has been crossed 10 consecutive times to S288C [Marion Carlson, personal communication]), U184 (HO HMLα MATα-inc HMRα ura3 thr4 his4 lys5 ade1 cry1), and HH2, which is the diploid obtained from mating MCY130 and U184.

Analysis of Yeast Genomic DNA by Southern Hybridization. Yeast DNA was prepared as previously described (16). 0.5 µg of DNA was digested with restriction enzyme and subjected to electrophoresis on agarose gels. The DNA in the gel was denatured, transferred to nitrocellulose filter paper (Millipore) and then hybridized to the specific nick-translated DNA fragments indicated.

RESULTS

Nucleotide Sequence.

The sequencing strategy is diagramed in Fig. 2. The heavy line represents the Y region present in the plasmid pSZ220. The thin lines correspond to vector-specific sequences (see Figure Legend for details). DNA sequence was determined from a number of overlapping clones. Both strands of

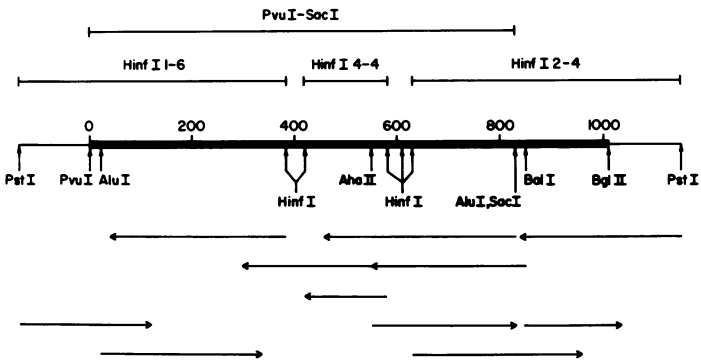

Fig. 2. Sequencing strategy for the Y region present in pSZ220. A map with relevant restriction sites is shown. The thick line indicates Y sequence. The thin line on the left corresponds to pBR322 DNA, and on the right to HIS3 DNA present in the pSZ220 plasmid. Numbering initiates at the PvuI site used in the cloning of the telomere (7). Arrows indicate the sequence determined: the base of the arrow corresponds to the end of the DNA fragment cloned closest to the site of M13 primer hybridization. The head of the arrow indicates the extent to which the sequence was determined. The labelled bars (⟨———⟩) above the map indicate the clones used as specific probes in Southern analyses.

DNA were sequenced for approximately 75% of the sequence reported. For those regions in which sequence was determined from only one strand, multiple rounds of sequencing on one or more clones were performed.

The Y region cloned in pSZ220 consists of approximately 1 kb of DNA, the sequence of which is presented in Fig. 3. Numbering begins at the PvuI site used to clone the yeast telomere onto one end of a linear plasmid vector (7), and ends at a BglII site which is approximately 1.5 kb proximal to the end of the cloned telomere (7).

Presence of a Repeated Sequence.

We have identified a 36 base-pair sequence which is tandemly repeated 12 times from residues 106 to 537. The repeats are not identical, though they are very highly conserved (see Fig. 4a). The first and last of the 12 repeats, however, are somewhat less conserved than the second through tenth. In particular, the last five base-pairs of the 12th repeat, TAGTG, differ from those found in any of the other repeats. This same 5 base-pair sequence is found directly proximal to the repeat units, in positions 101-105. Thus, the repeat units are bracketed by the pentamer. In addition, some remnants of the repeat sequence may be identified in yet another 36 base-pair interval adjacent to the 12th repeat. A consensus for the 36 base-pair repeat unit is shown in Fig. 4b.

```
       10        20        30        40        50        60        70        80
CGATCGTTGCGTTACCGTCTAGCTTCCAGGAGAGCAATAGCAGTGACAGGTGCAGAAAGTATTGCAGCAGTGATGAGGAC

       90       100       110       120       130       140       150       160
AGCGACACGTGCATTCATGGTAGTGCTAATGCCAGTACCAATGCGACTACCAACTCCAGCACTAATGCTACTACCACTGC

      170       180       190       200       210       220       230       240
CAGCACCAACGTCAGGACTAGTGCTACTACCACTGCCAGCATCAACGTCAGGACTAGTGCGACTACCACTGAAAGTACCA

      250       260       270       280       290       300       310       320
ACTCCAGCACTAATGCTACTACCACTGCCAGCACCAACGTCAGGACTAGTGCTACTACCACTGCCAGCATCAACGTCAGG

      330       340       350       360       370       380       390       400
ACTAGTGCGACTACCACTGAAAGTACCAACTCCAACACTAGTGCTACTACCACCGAAAGTACCGACTCCAACACTAGTGC

      410       420       430       440       450       460       470       480
TACTACCACCGAAAGTACCGACTCCAACACTAGTGCTACTACCACTGCTAGCACCAACTCCAGCACTAATGCCACTACCA

      490       500       510       520       530       540       550       560
CTGCTAGCACCAACTCCAGCACTAATGCCACTACCACTGAAAGTACCAACGCTAGTGCCAAGGAGGACGCCAATAAAGAT

      570       580       590       600       610       620       630       640
GGCAATGCTGAGGATAATAGATTCCATCCAGTCACCGACATTAACAAAGAGTCGTATAAGCGGAAAGGGAGTCAAATGGT

      650       660       670       680       690       700       710       720
TTTGCTAGAGAGAAAGAAACTGAAAGCACAATTTCCCAATACTTCCGAGAATATGAATGTCTTACAGTTTCTTGGATTTC

      730       740       750       760       770       780       790       800
GGTCTGACGAAATTAAACATCTTTTCCTCTATGGTATTGACGTATACTTCTGCCCAGAGGGAGTATTCACACAATACGGA

      810       820       830       840       850       860       870       880
TTATGCAAGGGCTGTCAAAAGATGTTCGAGCTCTGTGTCTGTTGGGCTGGCCAGAAAGTATCGTATCGGAGGATGGCTTG

      890       900       910       920       930       940       950       960
GGAAGCACTAGCTGTGGAGAGAATGCTGCGAAATGACGAGGAATACAAAGAATACTTGGAAGACATCGAGCCATATCATG

      970       980       990      1000      1010
GGGACCCTGTAGGATATTTGAAATATTTTAGCGTAAAAAGGGGAGAGATCT
```

Fig. 3. Nucleotide sequence of the Y region present in pSZ220. The sequence shown extends from the PvuI site used in cloning the telomere (7) to the BglII site used to circularize the linear plasmid and form pSZ220.

A striking pattern has emerged from inspection of the repeated region. Several positions in the sequence appear to be occupied by two alternative base-pairs, as shown in the consensus sequence in Figure 4b. Remarkably, a specific variant at one position is highly correlated with specific bases at the other variable sites. The variants displayed above or below the line for positions 19, 20, 23, 30, 31, and 35 in Figure 4b are almost always found together within the same repeat. For example, four of the internal repeats have a G at position 35. All four also have CC at positions 19 and 20, a C at position 23, and a GT at positions 30 and 31. Similarly, of six internal repeats having a C at position 35, all of them have a TC at positions 30 and 31; four have a T at position 23 and AA at positions 19 and 20. The probability of finding such exclusive groupings at six positions within the sequence if the base selection were random is extremely low.

```
                1   6   12     18   24     30   36
                .   .   .      .    .      .    .
a.          106-CTAATGCCAGTA CCAATGCGACTA CCAACTCCAGCA
                CTAATGCTACTA CCACTGCCAGCA CCAACGTCAGGA
                CTAGTGCTACTA CCACTGCCAGCA TCAACGTCAGGA
                CTAGTGCGACTA CCACTGAAAGTA CCAACTCCAGCA
                CTAATGCTACTA CCACTGCCAGCA CCAACGTCAGGA
                CTAGTGCTACTA CCACTGCCAGCA TCAACGTCAGGA
                CTAGTGCGACTA CCACTGAAAGTA CCAACTCCAACA
                CTAGTGCTACTA CCACCGAAAGTA CCGACTCCAACA
                CTAGTGCTACTA CCACCGAAAGTA CCGACTCCAACA
                CTAGTGCTACTA CCACTGCTAGCA CCAACTCCAGCA
                CTAATGCCACTA CCACTGCTAGCA CCAACTCCAGCA
                CTAATGCCACTA CCACTGAAAGTA CCAACGCTAGTG-537

                  A              AA T      TC  C
b.          CTA TGCTACTA CCACTG  AG A CCAAC  CAG A
                  G              CC C      GT  G

                        A                      G
c.          C  Py  A    Py  G  C  Py  A      Py  A
                        C                      C
            28 30  28 23  30 24 22  24 30 27  26 30
```

Fig. 4. Analysis of the tandemly repeated 36 base-pair sequence. a) The DNA sequence harboring the tandem repeats is aligned to facilitate comparisons between them. The twelve repeats are shown. Underscored numbers refer to the positions in the nucleotide sequence presented in Fig. 3. b) A consensus sequence for the 36 base-pair repeat is shown. Only the second through tenth repeats were considered since less homology exists for the first and last repeats. Bases which are invariant in all 10 repeats are indicated by an underscored upper-case letter; those present in at least 7 repeats are indicated by an upper-case letter. In those cases where two bases predominate, they are both shown. c) A consensus sequence for the 12 base-pair sub-repeat is shown. Beneath each position of the consensus dodecamer is shown the number of sub-repeats (out of a possible 30) which have the base(s) indicated. Py, pyrimidine.

The 36 base-pair repeat unit itself can be subdivided into three sub-repeats of a 12 base-pair sequence which is much more variable than the 36 base-pair repeat described above. Nevertheless, a definite pattern can be discerned. Some positions are completely invariant. For example, the 9th and 12th positions of the 12 base-pair sub-repeat are represented by an A in each of the 30 dodecamer units in the region from 142 to 501 of the nucleotide sequence. In contrast, the base selected at some sites is more strongly determined by the relative position in the 36 base-pair repeat of

the dodecamer in which it resides. For example, the second base of every second and third dodecamer in the 36 base-pair repeat units is a C, while in every first dodecamer, it is a T. A consensus sequence for the 12 base-pair sub-repeat is shown in Fig. 4c.

The base composition of the repeated region is significantly different from that of the flanking DNA sequences. The top strand of DNA in both flanking regions is AT-rich (57-58%), and 22-25% G, but noticeably low in its C content (17-21%). In contrast, the top strand of the repeated region is very rich in C (35.5%) as it is in A (31.5%), but is very low in G (15.4%) and T (17.5%). Thus, the repeated region displays a very large difference in base content of the two DNA strands (the upper strand is 67% A+C, while the lower strand is 67% G+T).

Presence of Inverted Repeats.

We also searched the sequence for the presence of inverted repeats. No perfect inverted repeats of substantial length were found, although a small imperfect inverted repeat extending from position 3 to 17 could be identified in each 36 base-pair direct repeat. The best of these inverted repeats have the structure:

AGTGCTACTACCACT
·
TCACGATGATGGTGA

The dot indicates the center of symmetry (position 10 of the designated 36 base-pair repeat unit); over- and underlined positions indicate symmetrical bases.

Organization of the 36 Base-Pair Repeats in the Yeast Genome.

We wished to obtain additional information regarding the organization of Y regions in the yeast genome. Experiments in which the plasmid pSZ220 was used as a probe of total yeast DNA digested with a variety of different enzymes have shown that a large number of chromosome ends are homologous to the cloned DNA (6-13). We wanted to determine if the repeated sequence is present at most of these ends, and whether or not it is always embedded in the same surrounding sequences. DNA from three separate strains was digested with the restriction enzyme KpnI, which produces many fragments of different sizes that hybridize to pSZ220 DNA (6,7,13). Most of these fragments are presumably generated by one cut distal to the PvuI-BglII region and the other within the proximal flanking unique DNA, and thus represent different chromosome ends. DNA from each digest was divided into three aliquots which were run in parallel on an agarose gel, transferred to nitrocellulose filters and probed with DNA fragments HinfI 1-6, 4-4 or 2-4, the locations of which are indicated in Fig. 2. HinfI 1-6 contains approximately 8 repeat units, as

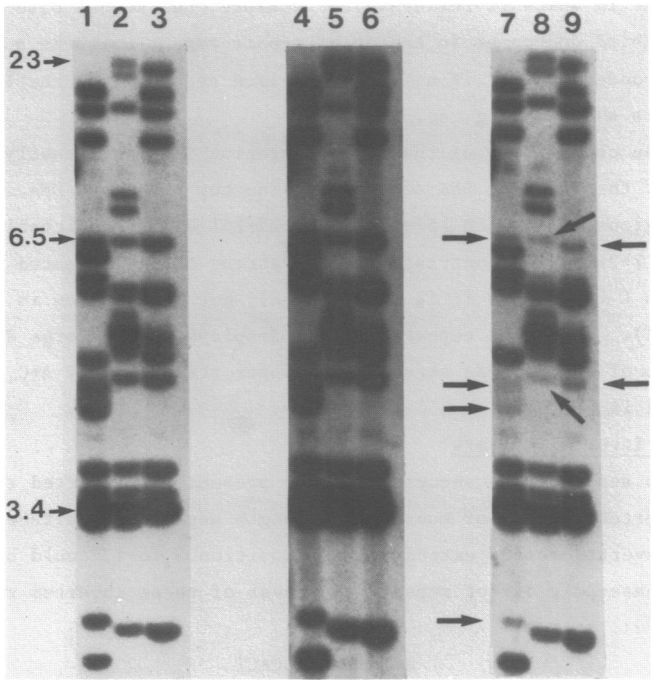

Fig. 5. The 36 base-pair repeats are located in sequences sharing homology with one another. KpnI digests of DNA from three strains were each divided into three aliquots and electrophoresed in parallel on an agarose gel. The gel was cut into three portions, the DNA transferred to nitrocellulose filter paper and then probed with the indicated DNA fragment. Lanes 1,4, and 7 are MCY130 DNA; lanes 2,5, and 8 are U184 DNA; lanes 3,6 and 9 are HH2-12D DNA. Lanes 1-3 were probed with HinfI 1-6, lanes 4-6 with HinfI 4-4, and lanes 7-9 with HinfI 2-4 (see Fig. 2 for positions of probes). The arrows indicate bands which hybridize much less intensely with the 2-4 probe than with 1-6 or 4-4. The positions of size markers are indicated to the left of the Figure (sizes in kb).

well as 105 base-pairs of proximal DNA. HinfI 4-4 contains approximately 3 good repeat units as well as 40 base-pairs of distal DNA which shares some homology to the repeat unit. HinfI 2-4 contains only sequences distal to the repeats. It is clear from Fig. 5 that the KpnI digests of the three separate strains show identical hybridization patterns when probed with HinfI 1-6 or 4-4. However, when probed with HinfI 2-4, several bands are much lighter in intensity (see arrows in Fig. 5). This difference in hybridization patterns may be explained in two ways: (1) the regions downstream from some repeated sequences are unlike that in the cloned telomere of pSZ220; (2) the fragments that hybridize with reduced efficiency to the 2-4 probe contain sequences

similar to those in pSZ220 upstream from the probe. However, they have only a very small amount of DNA from the leftmost part of HinfI 2-4 due to the presence of a KpnI site close to the left end of the probe.

In summary, it appears that for the most part, the repeats in the genome are flanked on the telomere-proximal side by DNA sequences similar to those found in the pSZ220 clone since almost all bands that hybridize to HinfI 1-6 or 4-4 also hybridize to 2-4. Several repeats, however, may not be associated with sequences sharing homology with the HinfI 2-4 probe.

The Y Regions of Different Telomeres May Have Variable Numbers of 36 Base-Pair Repeats.

We wished to determine if the number of 36 base-pair repeats varies at different telomeres. To do so, we performed Southern blots using a variety of digests with enzymes which we knew would cut on one side or the other of the repeated region in the pSZ220 clone. The linear plasmid constructed by Szostak and Blackburn (7) possesses a PstI site approximately 150 base-pairs distal to the BglII site (position 1006 in Fig. 3) which they used to circularize the linear plasmid to form pSZ220. If this PstI site were highly conserved within Y sequences, and if there were also a conserved PstI site proximal to the repeated sequences, then digestion with PstI should produce fragments containing the repeats. Figure 6 presents a Southern blot in which DNA from several different strains was digested with PstI (lanes 1-8) and probed with the PvuI-SacI fragment from pSZ220. It is clear that a very regular pattern of bands is produced. Indeed, it is surprising that the patterns are so similar in different strains, given that these strains produce very different hybridization patterns when the enzyme KpnI is used (compare with lanes 1, 2 & 3 in Fig. 5). The bands in the pattern (designated A-G in Fig. 6) range from 1520 to 2140 base-pairs in length.

The most strongly hybridizing band in the pattern is E, which may represent Y regions from many different telomeres. In contrast, band A most likely represents a fragment derived from a single Y-containing region: the strains MCY130 (which has A) and U184 (which lacks the band) were mated to form the diploid HH2 and the diploid was sporulated. The four segregants obtained from each of two separate tetrads show 2:0 segregation of the A band, which is indicative of a single Mendelian locus (see lanes 4-7).

We hypothesized that the ladder of bands might be due to the presence of variable numbers of the repeat unit. If so, we should be able to truncate the PstI fragments at either or both ends, and obtain the same regular pattern of bands, shifted to a lower position on the gel by virtue of their

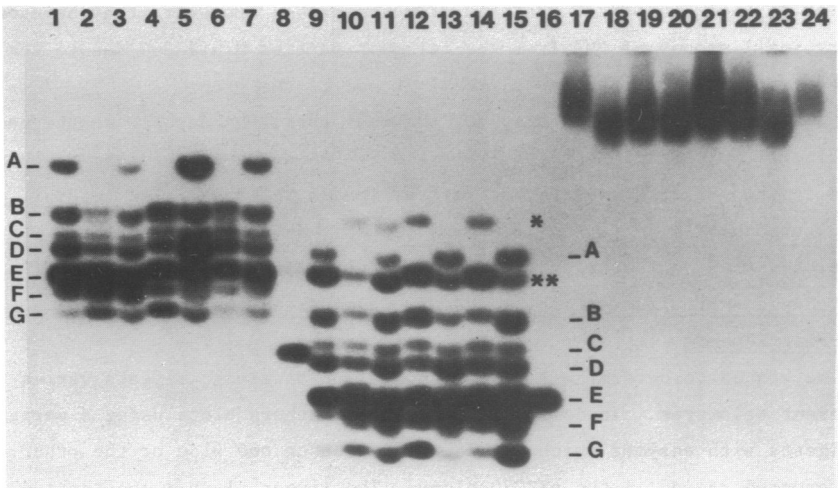

Fig. 6. Y regions contain variable numbers of 36 base-pair repeats. The strains MCY130 and U184 were crossed to form the diploid HH2, which was then sporulated. DNA prepared from the parent strains, the diploid and one tetrad was digested with the restriction enzyme(s) indicated, and the Southern blot probed with the PvuI-SacI fragment (see Fig. 2). MCY130 (lanes 1,9,17), U184 (lanes 2,10,18), HH2 (lanes 3,11,19), HH2-11A (lanes 4,12,20), HH2-11B (lanes 5,13,21), HH2-11C (lanes 6,14,22), HH2-11D (lanes 7,15,23), linear plasmid (7) (lanes 8,16,24). Lanes 1-8 were digested with PstI; lanes 9-16 with PstI and PvuI; lanes 17-24 with PvuI. The bands discussed in the text are labelled A-G. Band B in the PstI digest of U184 is a doublet: one of the bands is not cleaved in double digests with PvuI (indicated with an *). In addition, not all of the component bands comprising band E in PstI digests are cut with PvuI in the double digests. These residual bands are indicated with **. The size of the PstI band obtained from the linear yeast plasmid cannot be directly compared to the bands from genomic yeast DNA since it is composed partially of pBR322 sequences. However, the comparison can be made in the PstI-PvuI digest (lane 16). The band obtained corresponds to band E in the other strains shown. The bands in lanes 17-24 are derived from fragments that extend from the PvuI site to the end of the chromosome. The "fuzzy" nature of these bands is due to differences in size among telomeres of different chromosomes and/or variation in length at any particular telomere.

smaller size. Lanes 9-16 of Fig. 6 show the Southern blot of DNA digested with PvuI and PstI. Virtually the same pattern as was observed for the PstI digests alone was obtained, except that it was shifted down by approximately 450 base-pairs. (In addition, some extra bands are present; see Figure Legend for a detailed explanation). The finding that all of the PstI bands are shifted by equal amounts indicates that these fragments are all derived from similar regions, that both the PstI and PvuI sites from each region are basically equivalent, and that the source of the size differences is confined

to the region between the PvuI and the telomere-proximal PstI site. As our sequence analysis revealed that the repeats lie between these two sites in the cloned Y region, it seemed likely that a variable number of repeats might be responsible for creating the observed pattern. The linear plasmid of Szostak and Blackburn was also subjected to digestion with PstI and PvuI (see lane 16 of Fig. 6). It is clear that the cloned telomere has a PvuI-PstI fragment corresponding to band E, the most strongly hybridizing band of the series. Although the greater degree of hybridization to E in the genomic DNA digests could be due to increased homology with the probe, it more likely reflects a greater number of Y regions possessing that particular length PvuI-PstI fragment, and, by inference, that number (twelve) of repeat units. Thus, the cloned telomere is indeed a prototypic chromosome end by this criterion.

To perform a preliminary test of the hypothesis that the PstI and PstI-PvuI digestion patterns reflect differences in the number of repeats, rather than a peculiarity of digestion with PstI, we performed a number of other digests on two of the strains analyzed in Fig. 6. Figure 7a shows the Southern blot of restriction digests of DNA from strains MCY130 and U184 probed with the HinfI 1-6 clone.

Cutting of DNA fragments at any position predicted to be outside the repeats produced the same characteristic pattern of bands A-G as was obtained in digestions with PstI, though the relative size ranges varied (see Fig. 7a). We attempted to localize the source of the phenomenon responsible for generating the observed pattern of bands to a small segment of DNA by using the enzyme DdeI. One site for this enzyme is located 31 base-pairs beyond the last 36 base-pair repeat in the pSZ220 clone. As shown in lanes 17 and 18, digestion with DdeI results in the same pattern of bands observed in the other digests. In conjunction with the information obtained from other digests, we determined the source of band size variability to be localized between the PvuI and DdeI sites (see restriction map, Fig. 7b). The PvuI-DdeI fragment of the cloned telomere is 565 base-pairs long and is composed entirely of repeat units except for 102 base-pairs at one end and 31 base-pairs at the other. The production of the characteristic band pattern in DNA digests with these two enzymes, then, suggests that the pattern may be derived from Y regions having variable numbers of repeat units.

Characterization of the fragment sizes in all of the digests of Fig. 7a reveals that the size range of bands A-G varies over 600 base-pairs. The resolution of the gel does not permit one to unambiguously determine whether

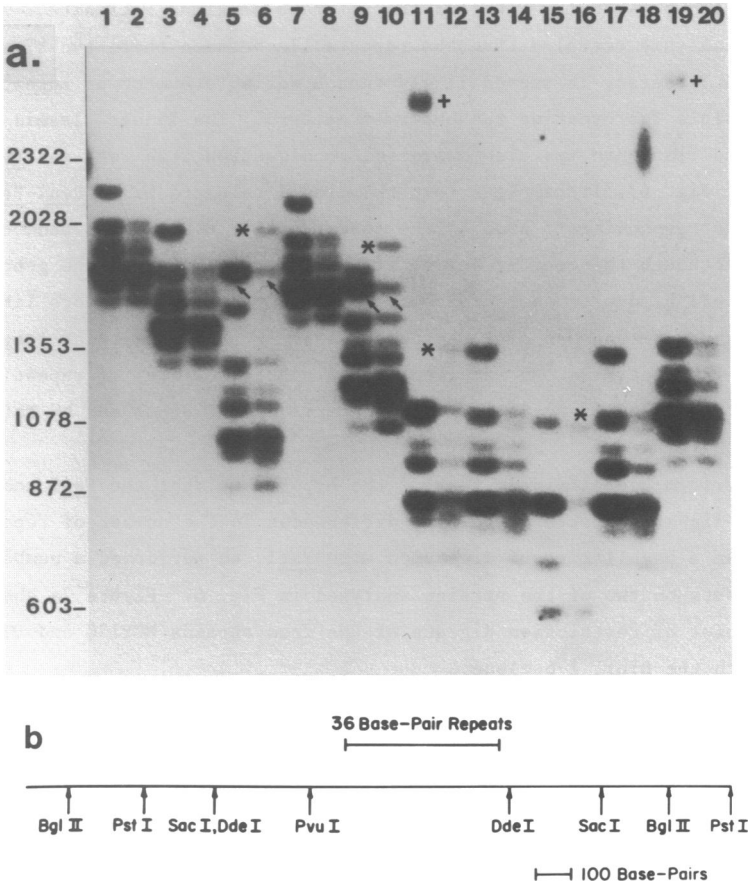

Fig. 7. Southern blots illustrating the repeat pattern in a variety of restriction digests. a) The Southern blot presented (probed with HinfI 1-6) shows that the pattern of bands A-G is obtained upon digestion of DNA with a variety of enzymes. Odd number lanes, MCY130 DNA; even number lanes, U184 DNA. Digestion was with the indicated enzyme(s): BglII (lanes 1,2), PstI-BglII (lanes 3,4), PvuI-BglII (lanes 5,6), PstI (lanes 7,8), PstI-PvuI (lanes 9,10), PvuI-SacI (lanes 11,12), DdeI-PstI (lanes 13,14), PvuI-DdeI (lanes 15, 16), DdeI (lanes 17, 18), SacI (lanes 19,20). In lanes 15 and 16, hybridization to bands F and G is poor due to inefficient transfer of these very small fragments. (Longer exposures of the autoradiogram reveal that they are indeed present). The absence of band A in the SacI digest of MCY130 DNA may be due to the lack of the usual site for this enzyme on one side or the other of the repeats in the relevant Y region. In fact, a larger band is found in digests of MCY130 DNA with SacI (lanes 11 and 19, indicated with a +). Not all E bands are cut with PvuI. In PstI-PvuI and PvuI-BglII digests, the uncut PstI or BglII E fragments appear as extra bands in the pattern (denoted with arrows). For the PvuI-SacI and PvuI-DdeI digests, the uncut SacI and DdeI E fragments migrate to the same extent as the B fragments of the respective double digests, resulting in increased intensity of these bands. One of the bands in the B doublet of U184 DNA is not cut by PvuI.

Thus, in double digests with PvuI, this band remains (indicated with an *). Positions of size markers are shown to the left of the figure. b) A restriction map illustrating the known and inferred positions of restriction sites based on information obtained from the Southern blot shown in a is presented. The position of the repeats is indicated.

or not the bands differ by integral numbers of repeat units. Nevertheless, based on the positions of fragments relative to band E (which has 12 repeats), it can be inferred that the Y regions contain between 8 and 25 repeats. The bands we have identified most probably have the following numbers of repeat units: A:25, B:20, C:17, D:15, E:12, F:10, G:8.

We have also examined the pattern of bands obtained in PstI and DdeI digests in five other laboratory strains of S. cerevisiae. For four of the five strains, the patterns were very similar to those presented in Fig. 7a. In the fifth strain, Y55 (which is more distantly related), a somewhat different pattern of bands was observed, but the size range of these bands was similar to that observed for all the other strains. Thus, it appears that, for the most part, the pattern of 36 base-pair repeat sequences is extremely well-conserved.

DISCUSSION

Recent efforts to characterize the structure and function of telomeres in numerous organisms has revealed that these elements are far more complicated than originally presumed. Sequence analysis of the extreme ends of telomeres from a variety of ciliates, flagellates, slime molds and yeast has revealed the presence of short, simple repeated sequences having the general structure: $5'-[C_{1-8}(^{A}_{T})_{1-4}]-3'$ (10-12; 17-25). The telomeres of S. cerevisiae possess an irregular repeat having the structure $5'-[C_{2-3}A(CA)_{1-3}]-3'$ (10-12). The sequence similarity of the repeats in very different organisms lends support to the notion that these specific sequences were conserved in order to provide the necessary telomeric functions. Indeed, Szostak and Blackburn (7) found that the ends of the Tetrahymena rDNA were capable of functioning as telomeres in yeast. However, they were modified somewhat to become more "yeast-like" (11).

In addition to the simple repeats described above, some organisms possess more complex subtelomeric regions as well. Proximal to the C_nT satellite found at the ends of Dictyostelium rDNA is a region composed of four nearly perfect tandem repeats of a 29 base-pair sequence (25). Similarly, centromere-proximal to the $5'-C_3TA_2-3'$ repeats at the ends of telomeres in Trypanosoma brucei there exist numerous perfect direct repeats

of a 29 base-pair C_3TA_2-derived sequence (23). Proximal to these repeats is a complex AT-rich sequence that contains many short homopolymeric runs and several C_3TA_2 repeats (26).

Szostak and Blackburn (7) showed that, in addition to the repeated sequences near the very ends of the yeast termini, many chromosomes share at least 3-4 kb of homology near their ends. Chan and Tye (8) mapped a family of ARS-containing yeast DNA sequences to the telomeres and further characterized these regions in a number of clones. Their work extended the estimate of the homology shared by some yeast telomeres to at least 6.7 kb (8).

We have sequenced 1 kb of Y DNA from the pSZ220 clone (7) in an attempt to characterize this region in better detail. The sequence presented here suggests that the Y region itself is rather complex: embedded within the 1 kb of Y DNA we sequenced is a region containing 12 tandem imperfect repeats of a 36 base-pair sequence (Fig. 4a). The moderate bias in the base composition of the two strands (67% A+C on the top versus 67% G+T on the bottom) of the repeat unit DNA is reminiscent of the extreme bias in base composition (100% AC or GT) observed at the very ends of the chromosome. The C+A-rich stretches present in the repeated region have sequences ACCAC(C), (C)ACCAAC and CCAACA. These sequences, though similar to, are not identical to the C+A-rich sequences identified at yeast telomeres (11,12), nor are they tandemly repeated. Nevertheless, they may provide a possibly related function, especially since a large number of these C+A-rich sequences are concentrated in one region by virtue of the tandem repetition of the 36 base-pair sequence. Indeed, the relationship between the 36 base-pair repeats of the Y regions and the simple repeats at the extreme ends of the yeast chromosome may be analogous to that between the 29 base-pair C_3TA_2-derived subtelomeric and the C_3TA_2 telomeric repeats of T. brucei (23). However, it should be noted that the CA- vs GT-richness of the two strands of DNA in the 36 base-pair yeast repeat is inverted with respect to that found at the extreme ends of the telomeres.

The work presented here, when considered along with recent discoveries, contributes to the emergence of a new picture of the Y'-containing class of yeast telomeres. As thus far elucidated, the yeast telomere is basically a set of nested repeat units terminated by a C+A-rich highly repeated sequence at the extreme end of the chromosome. The nested repeats consist of (1) the 6.7 kb Y' sequence, which may be repeated up to four times in a tandem array, (2) the tandem array of possibly eight to twenty-five 36 base-pair repeats

centered approximately 3.3 kb from the telomere distal end of the Y sequence, and (3) the three tandem 12 base-pair sub-repeats within the 36 base-pair repeat.

Indeed, it is likely that the telomeric sequences are even more complicated than suggested by this picture. There is evidence that the telomeres of some chromosomes lack sequences with homology to Y' (11-13). Thus, there may be other, possibly repeated, sequences residing at these telomeres.

The evolution and potential functions of the repeated 36 base-pair sequence we have identified are of considerable interest. Recent work in our laboratory has indicated that rearrangements in yeast telomeric regions during mitosis and meiosis are quite common (13). It is likely that the extensive homology shared by the chromosome ends plays an important role in the generation of these rearrangements through gene conversion and/or recombination events. Moreover, it is possible that the presence of the tandemly repeated 36 base-pair sequences we describe here may be especially important for increasing the recombination frequency among ends. Smith (27) has proposed that the presence of closely spaced, short similar sequences could promote unequal crossover, resulting in duplication or deletion of the intervening DNA sequences. The 36 base-pair repeats are very good candidates for involvement in unequal crossover events which could serve to increase or decrease the number of repeat units at any particular chromosome end. The ramifications of changes in repeat number are unknown. We plan to examine these questions in greater detail in experiments which delete, alter or disrupt the repeat units, to ask if the new sequences are still capable of functioning as telomeres on linear plasmids in yeast.

While the subtelomeric region we have sequenced contains twelve 36 base-pair repeats, the ladder of Y-homologous restriction fragments observed in genomic Southern blots (Figs. 6 and 7) is highly suggestive of the presence of variable numbers of these repeats in different telomeres. We are currently examining other cloned yeast telomeres (28) by sequence analysis in order to rigorously test this hypothesis. If our interpretation is correct, then the ladder may in fact reflect an equilibrium state of unequal crossover events in the cell. Precedence for a phenomenon of this type may be found in the amplification of the CUP1 locus of yeast. In response to elevated levels of copper in the growth medium, the copy number of the CUP1 locus is increased in tandemly iterated units (29). Because amplification is dependent upon the presence of at least two copies of the

gene, it has been suggested that the amplification is mediated by unequal sister chromatid exchange (30). Unequal sister chromatid exchange has also been well-documented in the tandemly repeated rDNA genes of yeast (31,32). In this case, unequal crossovers occur in the absence of any applied selective pressure.

Preliminary experiments have indicated that the number of repeat units on the linear plasmid is stable when the cells are grown mitotically: no ladder of bands is generated when cells harboring the linear plasmid are grown for more than 40 generations (see Fig. 6). It is possible, however, that by virtue of its small size or lack of a centromere, the linear plasmid is incapable of participating in unequal crossover events. Alternatively, it might be necessary for the cells to undergo meiosis for these events to occur. We are currently exploring these possibilities.

Recombination between repeated sequences may also serve as a mechanism to maintain the homogeneity of these sequences (27). The identification of exclusive groupings of bases at six positions in the 36 base-pair repeat provides strong evidence that these sequences experience a form of "homogenization", most likely arising by gene conversion events, either between repeats in the same chromosome or between repeats on different chromosomes. The finding of so many variants that are apparently co-converted provides an unusual opportunity to examine gene conversion and mismatch repair in great detail.

Acknowledgements

We are very grateful to Jack Szostak for providing the pSZ220 clone. We are indebted to Michael Wormington, Norah Rudin, Bernadette Connolly and John McCusker for providing helpful comments on this manuscript. Thanks also to Edwina Hertach, Ruth Ladd and Phyllis Bishop for typing this manuscript. This research was supported by NSF grant PCM-8110633. H.H. is a Damon Runyon-Walter Winchell postdoctoral fellow.

REFERENCES

1. McClintock, B. (1941). Genetics 26, 234-282.
2. McClintock, B. (1942). Proc. Nat. Acad. Sci. USA 28, 458-463.
3. Resnick, M.A. and Martin, P. (1976). Mol. Gen. Genet. 143, 119-129.
4. Orr-Weaver, T.L., Szostak, J.W., and Rothstein, R.J. (1981). Proc. Nat. Acad. Sci. USA 78, 6354-6358.
5. McCusker, J.H. and Haber, J.E. (1981). Genetics 99, 383-403.
6. Haber, J.E. and Thorburn, P.C. (1984). Genetics 106, 207-226.
7. Szostak, J.W., and Blackburn, E.H. (1982). Cell 29, 245-255.
8. Chan, C.S.M., and Tye, B.-K. (1983). Cell 33, 563-573.

9. Walmsley, R.M., Szostak, J.W., and Petes, T.D. (1983). Nature 302, 84-86.
10. Blackburn, E.H. (1984). Cell 37, 7-8.
11. Shampay, J., Szostak, J.W. and Blackburn, E.H. (1984). Nature 310, 154-157.
12. Walmsley, R.W., Chan, C.S.M., Tye, B.-K. and Petes, T.D. (1984). Nature 310, 157-160.
13. Horowitz, H., Thorburn, P. and Haber, J.E. (1984). Mol. Cell Biol., in press.
14. Sanger, F., Nicklon, S. and Coulson, A.R. (1977). Proc. Nat. Acad. Sci. USA 74, 5463-5467.
15. Messing, J. and Viera, J. (1982). Gene 19, 269-276.
16. Sherman, F., Fink, G. and Lawrence, C. (1979). Methods in Yeast Genetics, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.
17. Blackburn, E.H. and Gall, J.G. (1978). J. Mol. Biol., 120, 33-53.
18. Katzen, A.L., Cann, G.M. and Blackburn, E.H. (1981). Cell 24, 313-320.
19. Yao, M.-C., and Yao, C.H. (1981). Proc. Nat. Acad. Sci. USA 78, 7436-7439.
20. Oka, Y., Shiota, S., Nakai, S., Nishida, Y., and Okubo, S. (1980). Gene 10, 301-306.
21. Klobutcher, L.A., Swanton, M.A., Donini, P., and Prescott, D.M. (1981). Proc. Nat. Acad. Sci. USA 78, 3015-3019.
22. Johnson, E.M. (1980). Cell 22, 875-886.
23. Van der Ploeg, L.H.T., Liu, A.Y.C., and Borst, P. (1984). Cell 36, 459-468.
24. Blackburn, E.H. and Challoner, P.B. (1984). Cell 36, 447-457.
25. Emery, H.S. and Weiner, A.M. (1981). Cell 26, 411-419.
26. De Lange, T., Kooter, J.M., Michels, P.A.M., and Borst, P. (1983). Nucl. Acids Res. 11, 8149-8165.
27. Smith, G.P. (1976). Science 191, 528-535.
28. Dunn, B., Szauter, P., Pardue, M.L. and Szostak, J.W. (1984). Cell, in press.
29. Fogel, S., Welch, J.W., Cathala, G. and Karin, M. (1983). Curr. Genet. 7, 1-9.
30. Karin, M., Najarian, R., Haslinger, A., Valenzuela, P. Welch, J. and Fogel, S. (1984). Proc. Nat. Acad. Sci. USA 81, 337-341.
31. Petes, T.D. (1979). J. Bacteriol. 138, 185-192.
32. Szostak, J.W. and Wu, R. (1980). Nature 284, 426-430.