SUPPLEMENTAL DATA

An AT-rich Sequence in Human Common Fragile Site FRA16D that Causes Fork Stalling and Chromosome Breakage in *S. cerevisiae*

Haihua Zhang and Catherine H. Freudenreich

Supplemental Results

Correlation of Percentage of FOA^R with Chromosome Breakage Events.

DNA from FOA^R colonies was purified and chromosome structure analyzed. For the short YACs, Southern blot analysis to characterize the healed structures of YACs from FOA^R colonies was done as in Callahan et. al, 2003. While creating the *LEU2·URA3* short YACs, we accidentally included two 60 bp homologous sequences flanking the FRA16D region, which resulted in the loop-out of the region in some cells, detected by Southern blot. We observed a higher percentage of loop-out events in YACs containing Flex1 compared to the Flex4, Flex5-p, and control YACs, suggesting that there is a hot spot for breakage in Flex1. To eliminate these loop-out events, the genetic assay was conducted in a recombination-deficient *rad52Δ* background. For the *LEU2·URA3* YAC in the *rad52Δ* background, the majority had healed at the C_4A_4 site, indicating a breakage event. We then replaced one of the homologous sequences with a *HIS5* marker to create the *LEU2·HIS5·URA3* YAC, allowing the assay to be conducted in a wild-type background. For this YAC, all of the *HIS5*$^-$ FOA^R colonies showed loss of the right end of the YAC and all were healed at the C_4A_4 site whereas all of the *HIS5*$^+$ FOA^R colonies retained the entire YAC (data not shown).

The 801B6 and 972D6 YACs were analyzed by high resolution PFGE and Southern blot with a *LEU2* probe. All the YACs were shorter than the original YAC, consistent with a breakage event. The expected size of a YAC healed by new telomere addition at C_4A_4 is 8 kb. We observed healed YACs of multiple sizes between 8 and 195 Kb. Since a breakage event within the FRA16D region of the 801B6 YAC would require chromosome degradation of ~ 980 Kb to reach the C_4A_4 sequence, we conclude that the majority of time telomere addition happened at other sequences with enough telomere similarity to act as a seed for *de novo* telomere addition. Previous studies have shown that GT-rich sequences 11bp or shorter can act as telomere seeds in yeast (Callahan et al., 2003; Kramer and Haber, 1993; Schulz and Zakian, 1994). Consistent with this interpretation, some of the healed products appeared to be more favored than others. In addition, some of the broken YACs likely healed by intrachromosomal recombination (see examples in Figure S1A, D). Irregardless of the exact location and method of healing, our data indicate that the percentage of FOA^R colonies is a good indicator of the frequency of double strand breaks and that the chromosomal region containing FRA16D is a natural site of *in vivo* breakage.

Physical Analysis of FRA16D Breakage by Pulsed Field Gel Electrophoresis (PFGE)

To physically visualize sites of chromosome breakage, yeast cells containing either the 801B6 or 972D3 YAC were grown to early log phase and whole chromosomes were purified and separated by PFGE followed by probing with *TRP1* or *URA3* sequences present on each end of the YAC (Figure S1B). To facilitate detection of broken intermediates, this experiment was done in a strain lacking the Rad50 protein (*rad50*Δ) which has been shown to exhibit very slow degradation of broken ends. In this strain background, we observed chromosomal deletions

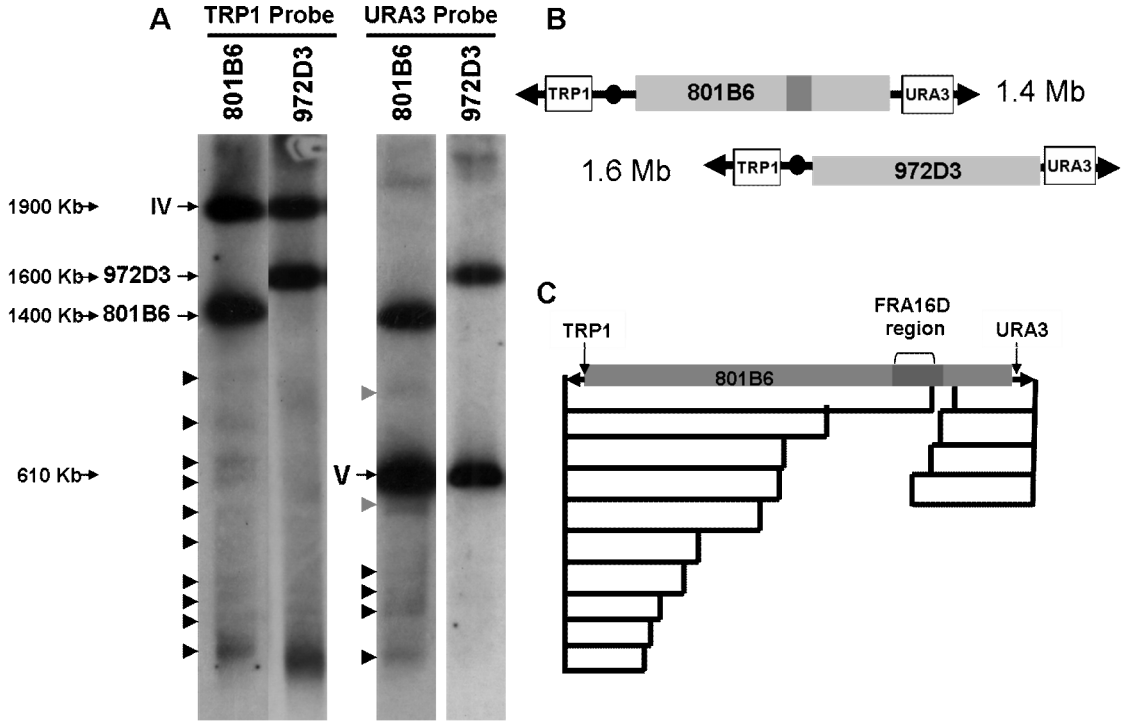
specific to the 801B6 YAC, not observed in the 972D3 control, indicating the presence of unstable elements prone to deletion or translocation. Thus we needed to screen multiple 801B6-containing colonies to find ones containing full-length YACs to start the experiment.

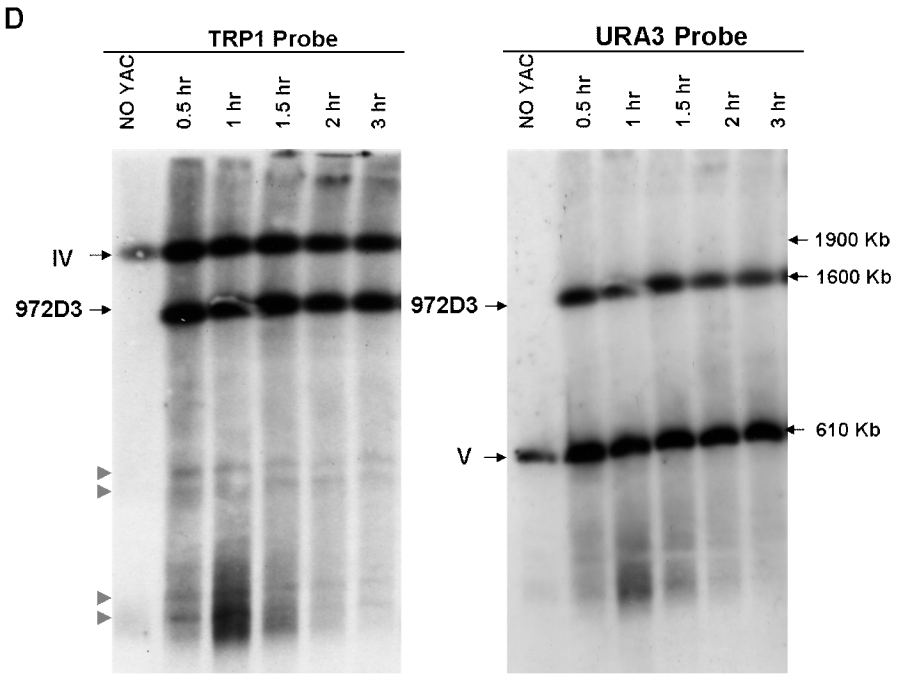
In the presence of FRA16D (801B6 YAC), ~10 broken intermediates (indicated by black arrows) were observed using a *TRP1* probe and ~4 broken intermediates hybridized to the *URA3* probe (Figure S1A). The two largest intermediates (indicated by grey arrows) shown by the URA3 probe are most likely stable recombination or healed breakage products, since these were frequently observed in stably shortened 801B6 YACs (data not shown and Figure 1D). In the absence of FRA16D (972D3 YAC), significantly fewer and less distinct broken products were observed. This result indicates that there is one or more preferred fragile regions in the 801B6 YAC compared to the 972D6 YAC, and that the regions where breaks occurred in the 972D3 YAC were more random. Mapping of the bands suggests that breakage most likely initiates within the FRA16D region, since the edge of the longest products co-localized with the FRA16D region (Figure S1C). The shorter intermediates may represent degraded or healed products. Taken together, the data show that sequences in the FRA16D region are highly unstable and prone to breakage, and are consistent with breakage initiating in the area defined by cytogenetics and containing the 6 peaks of DNA flexibility.

To further investigate whether distinct breakage regions were present in the 972D3 YAC, *rad50Δ* cells with YAC 972D3 were arrested at G1 phase, released to YC-Ura-Trp media, and time points collected. When probed with a TRP1 probe, bands that are smaller than the intact 972D3 YAC could be observed at the starting 0.5 time point, but these were present and stable throughout the rest of the 2.5 hours, suggesting that they are shorter recombinant YACs (grey arrows). No transient breakage intermediates could be observed using either the TRP1 or URA3

probe. Therefore no preferred breakage sites could be detected in YAC 972D3 as they were for YAC 801B6 which contains FRA16D.

Supplemental Figure S1





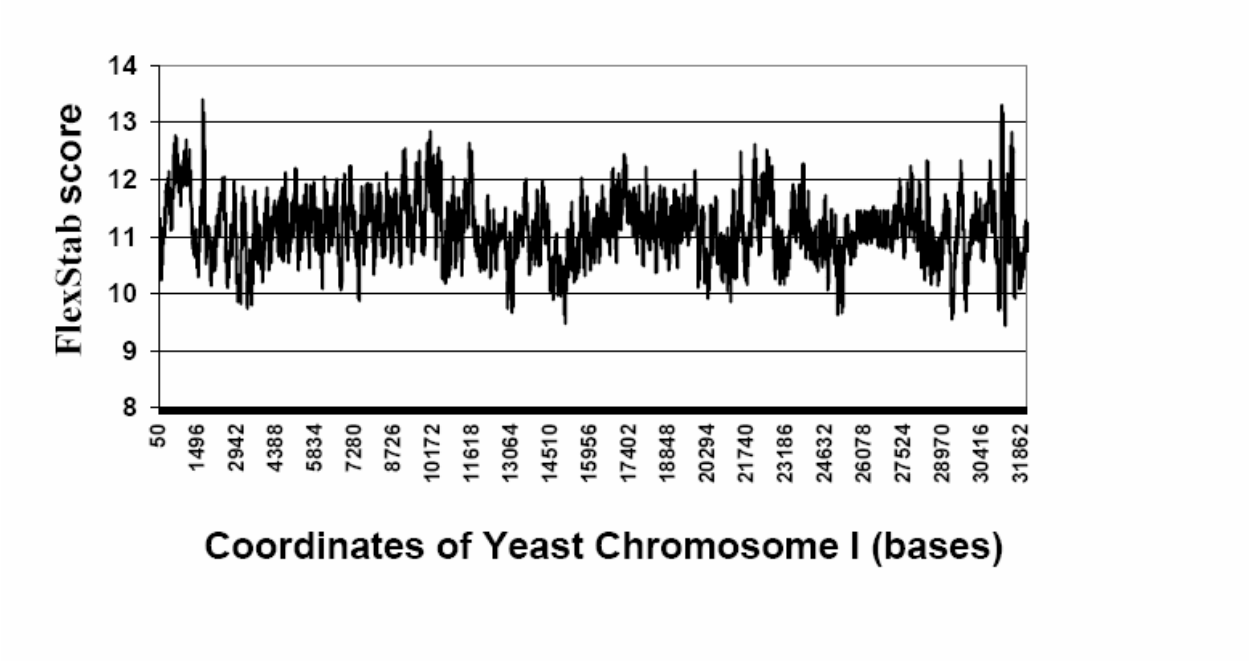
Supplemental Figure S1. The FRA16D-containing YAC exhibits a higher level of breakage compared to a control. **(A)** PFGE to detect chromosome breakage intermediates. The 801B6 and 972D3 YACs from the CEPH YAC library were grown in a *rad50Δ* strain background. Chromosomes were prepared and separated using PFGE, blotted, and hybridized to either a TRP1 or a URA3 probe to the left or right arm of the YACs, respectively. The endogenous chromosome IV (Containing *TRP1*) and V (containing *URA3*) and full length YAC 801B6 and 972D3 are indicated by long arrows. Putative breakage intermediates and degradation products are indicated by black arrowheads. Likely recombination products are indicated by grey arrowheads. **(B)** A diagram of the 801B6 and 972D3 YACs used in (A), aligned according to their coordinates in the human genome. **(C)** Mapping of the breakage intermediates and degradation products. A diagram of YAC 801B6 is shown approximately to scale. Each band shown by a black arrow head in (A) was mapped according to the size and the probe that was used. Relative sizes of the bands compared to the full length YAC 801B6 are diagrammed. All maps are based on the Whitehead YAC library. **(D)** Rad50Δ cells with YAC 972D3 were arrested at G1 phase and released to YC-Ura-Trp media. Cells were collected at the indicated time points and chromosomes prepared, separated by PFGE, and probed as in (A). Likely recombination products are indicated by grey arrowheads.

Analysis of Predicted Helix Flexibility of Yeast Chromosome I and the Flex1 YAC

The yeast genome is comprised of a slightly higher percentage of AT sequences than the human genome. In order to verify that high flexibility peaks were still significantly different from this potentially higher background, we used an algorithm based on the FlexStab program to predict the helix flexibility (twist angle) of a portion of yeast chromosome I and the Flex1(AT)34 YAC

(Figure 2C and S2). The average flexibility score of the yeast genome is around 11, which is very similar to human sequences (Mishmar et al., 1998; Ried et al., 2000). A flexibility peak in humans is arbitrarily defined as a score greater than 13.7 (Zlotorynski et al., 2003), and we did not observe any yeast background score higher than 13.7 (Figure S2). Thus human sequences that are defined as flexibility peaks should show the same characteristics in the context of the yeast genome. The background flexibility of the Flex1 YAC was also ~ 11, the C_4A_4 sequence exhibits a score of 8, which is less than the background and is consistent with the fact that CG-rich sequences are low in helix twist angle (Sarai et al., 1989) (Figure 2C). Sequences with a low flexibility score were also observed in the FRA16D region, (Figure 2B).

Supplemental Figure S2



Supplemental Figure S2. Analysis of 32 Kb of yeast chromosome I using a program based on FlexStab to predict helix flexibility (Mishmar et al., 1998). The Flexibility score for 100 bp intervals is shown.

Supplemental Experimental Procedures

Yeast strains

Genotypes of the strains used in this study are as follows:

AB103: MATα *ura3-52 can1-100 lys2-1 trp1 ura3 his5* (CEPH library) (Albertsen et al., 1990).

BY4705: MATα *ade2Δ::hisG his3Δ200 leu2Δ0* l*ys2Δ0 met15Δ0 trp1Δ63 ura3Δ0.*

VPS105: MATα *ade2 can1 leu2-3,112 trp1Δ ade3 ura3Δ lys2-891*.

Construction of the Short YACs

The Flex4, Flex5-p, Flex1, and the FRA16D 1320 bp control sequences were amplified by PCR and cloned into plasmid vectors (Finnis and Richards, unpublished data). An 871 bp PvuII-NsiI fragment from pVS20 containing the C_4A_4 sequence was blunted and inserted into the pUC19-based Flex4 and Flex5-p plasmids at the BamHI site. A 742 bp HindIII fragment from YEP24 containing the *URA3* gene was inserted into the modified plasmids at the HindIII site to create pHZ-Flex4 and pHZ-Flex5-p. A restriction fragment containing the Flex1 or 1320 bp control sequence was subcloned into pRS306 at the EcoRI or BamHI sites, respectively. An 871 bp PvuII-NsiI fragment containing the C_4A_4 sequence from pVS20 was blunted and inserted into the pRS306-based Flex1 and control plasmids at the PvuII site to create pHZ-Flex1 and pHZ-1320 control. The pHZ-400 control was derived from the pHZ-1320 control by removing 920 bp of the FRA16D DNA. All clones were verified by sequencing. Plasmids were linearized and transformed into yeast cells containing YAC VS5 (Schulz and Zakian, 1994), and correct integrants were selected and then verified by Southern blot to create YAC-Flex4, YAC-Flex5-p, YAC-Flex1 (4 versions) and YAC-control (2 versions). To eliminate loop-out events, one of the

homologous sequences distal to the Flex1 sequence (or 400 bp control) was replaced by a *HIS5* marker amplified by PCR using pFA-His3MX4 as a template (verified by Southern analysis).

Construction of the Long YACs for the Breakage Assay

A 2252 bp LEU2 PCR fragment was inserted into the SfoI site of pOTCA-1 (V. A. Zakian lab) to create pOTCA-LEU2. A 2800 bp PCR product, including the C_4A_4 sequence and the *LEU2* gene, was amplified with primers containing 60 bp sequences homologous to a region proximal to FRA16D and distal to the centromere. This PCR fragment was transformed into yeast cells containing the 801B6 or 972D3 YAC from the CEPH library to create new 801B6 and 972D3 YACs containing the C_4A_4 telomere seed. Correct integration events were confirmed by PCR.

Supplemental References

Albertsen, H. M., Abderrahim, H., Cann, H. M., Dausset, J., Le Paslier, D., and Cohen, D. (1990). Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. Proc Natl Acad Sci U S A *87*, 4256-4260.

Callahan, J. L., Andrews, K. J., Zakian, V. A., and Freudenreich, C. H. (2003). Mutations in yeast replication proteins that increase CAG/CTG expansions also increase repeat fragility. Mol Cell Biol *23*, 7849-7860.

Kramer, K. M., and Haber, J. E. (1993). New telomeres in yeast are initiated with a highly selected subset of TG1-3 repeats. Genes Dev *7*, 2345-2356.

Mishmar, D., Rahat, A., Scherer, S. W., Nyakatura, G., Hinzmann, B., Kohwi, Y., Mandel-Gutfroind, Y., Lee, J. R., Drescher, B., Sas, D. E*., et al.* (1998). Molecular characterization of a common fragile site (FRA7H) on human chromosome 7 by the cloning of a simian virus 40 integration site. Proc Natl Acad Sci U S A *95*, 8141-8146.

Ried, K., Finnis, M., Hobson, L., Mangelsdorf, M., Dayan, S., Nancarrow, J. K., Woollatt, E., Kremmidiotis, G., Gardner, A., Venter, D*., et al.* (2000). Common chromosomal fragile site FRA16D sequence: identification of the FOR gene spanning FRA16D and homozygous deletions and translocation breakpoints in cancer cells. Hum Mol Genet *9*, 1651-1663.

Sarai, A., Mazur, J., Nussinov, R., and Jernigan, R. L. (1989). Sequence dependence of DNA conformational flexibility. Biochemistry *28*, 7842-7849.

Schulz, V. P., and Zakian, V. A. (1994). The saccharomyces PIF1 DNA helicase inhibits telomere elongation and de novo telomere formation. Cell *76*, 145-155.

Zlotorynski, E., Rahat, A., Skaug, J., Ben-Porat, N., Ozeri, E., Hershberg, R., Levi, A., Scherer, S. W., Margalit, H., and Kerem, B. (2003). Molecular Basis for Expression of Common and Rare Fragile Sites. Mol Cell Biol *23*, 7143-7151.