Intrastrand Annealing Leads to the Formation of a Large DNA Palindrome and Determines the Boundaries of Genomic Amplification in Human Cancer

Hisashi Tanaka,1,2,4 Yi Cao,2 Donald A. Bergstrom,6‡ Charles Kooperberg,3 Stephen J. Tapscott,2 and Meng-Chao Yao1,5

Divisions of Basic Sciences,1 Human Biology,2 and Public Health Sciences,3 Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., Seattle, Washington 98109-1024; Department of Molecular Genetics, Cleveland Clinic Foundation, Cleveland, Ohio 44195; Institute of Molecular Biology, Academia Sinica, 128 Sec. 2 Academia Rd., Nankang, Taipei, Taiwan; and Department of Molecular Medicine, University of Washington, Seattle, Washington

Received 18 July 2006/Returned for modification 1 November 2006/Accepted 2 January 2007

Amplification of large chromosomal regions (gene amplification) is a common somatic alteration in human cancer cells and often is associated with advanced disease. A critical event initiating gene amplification is a DNA double-strand break (DSB), which is immediately followed by the formation of a large DNA palindrome. Large DNA palindromes are frequent and nonrandomly distributed in the genomes of cancer cells and facilitate a further increase in copy number. Although the importance of the formation of large DNA palindromes as a very early event in gene amplification is widely recognized, it is not known how a DSB is resolved to form a large DNA palindrome and whether any local DNA structure determines the location of large DNA palindromes. We show here that intrastrand annealing following a DNA double-strand break leads to the formation of large DNA palindromes and that DNA inverted repeats in the genome determine the efficiency of this event. Furthermore, in human Colo320DM cancer cells, a DNA inverted repeat in the genome marks the border between amplified and nonamplified DNA. Therefore, an early step of gene amplification is a regulated process that is facilitated by DNA inverted repeats in the genome.

Structural aberrations of chromosomes are a hallmark of cancer and result in either a gain or a loss of chromosomal regions (1, 15, 21). Studies of gene amplification in cancer cells often focus on the increased dosage of cellular oncogenes and its association with advanced stages of diseases, whereas in most cases the “gene” amplification involves large chromosomal regions that span several megabases of DNA (32, 38). Recent advances in the technology for genomewide copy-number surveys have identified a number of chromosomal regions in different cancer types as candidates for loci driving tumor progression. However, the molecular mechanisms and DNA contexts that initiate regional amplification and establish boundaries of amplicons remain largely unknown.

Mammalian cells treated with either radiation, DNA replication inhibitors, or other DNA-damaging agents show increased frequencies of gene amplification, suggesting chromosome breaks as initiating lesions (4, 12, 24, 39). In fact, either a chromosome break at a fragile site or a site-specific chromosomal DNA double-strand break (DSB) induced by I-SceI endonuclease leads to gene amplification by initiating the formation of large palindromic chromosomes (9, 37). Large palindromic dicentric chromosomes generate further chromosome breaks during chromosome segregation (see Fig. 1a). These breaks are again resolved into large palindromic chromosomes (breakage-fusion-bridge [BFB] cycle), establishing intrachromosomal gene amplification (9, 19, 29, 40). Thus, the formation of large palindromes occurs at a very early stage of gene amplification as a consequence of illegitimate repair of DSBs. How a DSB is processed to form a large palindrome remains unknown. A potential mechanism includes the joining of broken DNA ends of sister chromatids by nonhomologous end joining (NHEJ) (see Fig. 1a, right), a mechanism of chromosome end fusion in critically short telomeres (30). However, palindromic gene amplification still occurs in mice deficient in NHEJ (44), suggesting another, yet-unknown mechanism for large DNA palindrome formation in mammalian cells (see Fig. 1a, left).

Gene amplification in simple eukaryotes also involves the formation of large palindromes at a very early stage. In Tetrahymena, amplification of the rRNA (rDNA) gene is initiated by the formation of large palindromic chromosomes with the developmentally regulated induction of a site-specific DSB next to a short DNA inverted repeat (DNA IR) in the genome (42). DSBs initiate intramolecular recombination (intrastrand annealing) at the site of the short DNA IR, creating a large hairpin molecule prior to DNA replication. This hairpin molecule is resolved into a large palindrome following DNA replication, with the original DNA IR at a boundary between amplified and nonamplified DNA (6, 43). A short DNA IR is required for this process in Tetrahymena, as well as in Saccharomyces cerevisiae (5). Thus, a DNA IR in the genome is a critical cis-acting element for the formation of large palin-
dromes by an intramolecular recombination pathway and sets a boundary of the amplicons in lower eukaryotes.

Large palindromes are widespread in cancer cells and provide a platform for subsequent gene amplification (36). In a prior study, we showed that a short DNA IR adjacent to a DSB could catalyze the formation of a large palindrome and subsequent gene amplification in mammalian cells (37). Thus, gene amplification in cancer cells might proceed through the same intramolecular pathway as the developmentally regulated gene amplification in single-cell eukaryotes. To test this emerging model of mammalian gene amplification, it is necessary to show the following: (i) that a DNA IR in the genome facilitates large palindrome formation through a process consistent with intrasstrand annealing of broken ends and (ii) that a DNA IR in the genome marks the boundary between amplified and nonamplified DNA in spontaneous palindrome-associated gene amplification in human cancer cells.

We have now developed a system for measuring the efficiency of large palindrome formation after a DSB in the presence or absence of a short DNA IR at the same integration site in mammalian cells. Analysis of multiple amplicons at very early stages of amplification indicates that large palindrome formation after DSB is mediated by intrasstrand annealing and that a DNA IR next to a DSB significantly facilitates this process. We further determine that in human colon cancer cells, the center of the large palindrome sets the boundary between amplified and nonamplified DNA, and a 26-kb DNA IR preexisting in the human genome marks the region of recombination at the center of the palindrome. Therefore, a newly discovered mechanism of mammalian gene amplification occurs through a process similar to the intramolecular recombination used by simple eukaryotes, resulting in an amplicon bordered by a DNA IR in the genome.

MATERIALS AND METHODS

Constructs. To construct pD229IRe1sceox2, primer pairs, either XbaIR and XhoIoxIR or XhoCenPt and PstloxIR, were used to amplify each one of the two parts of the 229-base-pair inverted repeat (229IR) with pD229IRsce as a template (37). PCR products were cloned into pCRII-TOPO (Invitrogen) using the TOPO TA cloning kit to generate pTOPOXbaIR and pTOPOPstIR. An XbaI-Xhol fragment of pTOPOXbaIR was ligated into XbaI and Xhol sites of pTOPOPstIR to generate pTOPO229IRlox2. An XbaI-PstI fragment of pTOPO229IRlox2 replaced an XbaI-PstI fragment of pD79IRsce to generate pD229IRlox2sce. DNA sequences of the PCR primers are available upon request.

Cell cultures, transfections, and DNA analysis. Dihydrofolate reductase (DHFR)-deficient CHO cells were transformed with pD229IRlox2sce as described previously (37). Single-copy transformants were determined by Southern analysis and were subjected to transfection with 5 µg of pCMV3xI-SceI (16) for DSB induction using Superfect (Qiagen, Chatsworth, CA). As a control, pCMVnoSce, a vector without the I-SceI coding sequence, was transfected. Forty-eight hours after transfection, 1 × 10⁷ cells were plated in the medium with 0.2 µM methotrexate (MTX). Resistant colonies were scored after 10 to 12 days. Colonies were picked and expanded in 24-well plates for further genomic analysis. Genomic DNA extraction and Southern analysis were done as described previously (37). Briefly, 2 µg of high-molecular-weight genomic DNA was digested with a restriction enzyme, run on either a 0.8% or a 0.6% agarose gel, and blotted to a nylon membrane. Snap-back DNA was prepared as follows: 2 µg of genomic DNA in 50 µl water with 100 mM NaCl was boiled for 7 min and immediately transferred on ice. DNA was precipitated by ethanol and digested with a restriction enzyme. A 2.5-kb molecular ruler (Bio-Rad), a 14-kb DNA ladder (New England Biolabs), or a 200-bp DNA stepladder (Promega) was used as a size marker. To generate a probe for Southern analysis, human genomic DNA was amplified by PCR, and the fragment was cloned by using the TOPO TA cloning kit (Invitrogen). For sequence analysis of the centers of palindromes, either bisulfite-treated DNA or genomic DNA from each MTX-resistant clone was amplified using nested PCR. Bisulfite treatment of genomic DNA was done following published procedures (23). DNA sequences of the PCR primers are available upon request.

GAP. The genome-wide analysis of palindrome formation (GAP) procedure was performed as described previously (36) with modifications. After snapback treatment, 6 µl of S1 nuclease buffer, 4 µl of 3 M NaCl, and 100 µl of S1 nuclease (Invitrogen) were added to the DNA and incubated at 37°C for 1 h. S1 nuclease was inactivated by addition of 10 mM EDTA and phenol-chloroform extraction. DNA was precipitated with ethanol, dissolved in water, and digested with 40 µl of MspI, Tacl, or Msel (New England Biolab) for 16 h. DNA was precipitated, dissolved in 21 µl of water, and ligated to a MspI-, Tacl- or Msel-specific linker by adding 5 µl of a 20 mM linker, 3 µl of T4 DNA ligase buffer, and 400 U of T4 DNA ligase (New England Biolabs) at 16°C for 16 h. DNA was precipitated and dissolved in 200 µl Tris-EDTA, followed by application on a Microcon YM-50 spin column (Millipore) to remove excess linker. DNA was recovered in 20 µl H₂O. Thus, for each Colo320DM and normal human foreskin fibroblast (HFF2) cell culture, templates with three different linkers were prepared. For PCR, 2 µl of DNA, 0.5 µl of Faststart Taq DNA polymerase (Roche), 2.5 µl of 2 mM deoxynucleoside triphosphate, 5 µl of 10× PCR buffer with MgCl₂ (for Faststart Taq DNA polymerase; Roche), and 2 µl of a linker-specific primer were mixed with H₂O for a total reaction volume of 50 µl. For MspI-linker-ligated DNA, 5 µl of GC-rich solution (for Faststart Taq DNA polymerase; Roche) was also included in the reaction mixture. PCR was performed at 96°C for 6 min followed by 25 cycles of 96°C for 30 s, 55°C for 30 s, and 72°C for 30 s on a GeneAmp PCR system 9700 thermocycler (Perkin-Elmer). For PCR using the same template with three different linker-specific primers were concentrated using a Microcon YM-50 spin column. DNA was recovered in 25 µl H₂O and fragmented using DNase I (New England Biolabs). After heat inactivation of DNase I, DNA was labeled with biotin-11-dATP using terminal transferase (Roche). The procedure was performed in triplicate to produce three independent preparations of DNA for hybridization on GeneChip Human Genome U133A arrays (Affymetrix). After hybridization, the microarrays were washed and then scanned with GENECHIP software (Affymetrix, Santa Clara, CA). Statistical analysis was done as described previously (8). The perfect-match probe intensities were corrected by robust multichip analysis, normalized by quantile normalization, and summarized by Tukey’s median polish procedure using the Affy package of Bioconductor. The comparison of GAP profiles between Colo320DM and normal human foreskin fibroblasts was done using the LIMMA package of Bioconductor. The differentially expressed genes were determined to be significant according to a false discovery rate (FDR) of <0.05.

Microarray data accession number. All microarray data have been deposited in the Gene Expression Omnibus database at NCBI (accession no. GSE6274).

RESULTS

Significant increase of MTX-resistant clones in response to DSBs. To determine whether the formation of large palindromes is facilitated by the presence of DNA IRs in mammalian cells, we created a system to measure the efficiency of large palindrome formation in the presence or absence of a 229IR at the same integration site. In the construct, pD229IRlox2sce, one of the two parts of the repeat is flanked by a pair of loxP sites (Fig. 1b). This construct was randomly integrated into the genome of dhfr-deficient Chinese hamster ovary cells (CHO dhfr⁻/⁻). For each single-copy transformant (229IRlox2sce and -14 229IR transformants), we generated two independent subtransformants in which one repeat element was excised by Cre recombinase (noIR subtransformants) (see Fig. S1 in the supplemental material). Transformants were transfected with either pCMV3xI-SceI (16) for DSB induction at the I-SceI site or pCMVnoSce, a control vector without I-SceI coding sequence, followed by MTX selection (0.2 µM) for 10 to 12 days before the number of colonies was determined. Compared to the 229IR transformants, the numbers of MTX-resistant colonies induced after DSBs were significantly decreased for the noIR subtransformants, indicating that the 229IR pro-
motes the generation of MTX-resistant colonies (Fig. 1c). The IR-mediated increase in colony formation was 3.9-fold for D229IRlox2-35 (93.6 ± 21.0 for 229IR and 24.0 ± 4.0 for noIR-1) and 12-fold for D229IRlox2-14 (15.7 ± 1.2 for 229IR and 1.3 ± 0.6 for noIR-2). It should be noted, however, that I-SceI-induced DSBs generated a significant increase in MTX-resistant colonies for both noIR subtransformants of D229IRlox2-35 (24.0 ± 4.0 with DSB and 3.0 ± 1.7 without DSB for noIR-1; 16.3 ± 4.9 with DSB and 0.3 ± 0.6 without DSB for noIR-2 [Fig. 1c, left]).
FIG. 2. Biased processing of the DSB ends at the junctions of palindromes without a short IR. a. Southern analysis for the genomic DNA (top) and snap-back genomic DNA (bottom) was performed using the probe indicated. A DSB at the I-SceI site and palindrome formation eliminated the KpnI site that was located upstream of the I-SceI site. If two DHFR cassettes are joined to generate a palindrome after I-SceI-induced DSB, a 3.8-kb AflII fragment is expected. Most of the MTX-resistant clones showed a fragment that is a little smaller than 3.8 kb (arrowhead), and palindromes in these fragments were confirmed by Southern analysis using snap-back DNA (bottom). A 1-kb DNA ladder (NEB) mixed with a 200-bp DNA stepladder (Promega) is shown as a size marker. b. Schematic drawing of the centers of palindromes from MTX-resistant clones. Solid gray bars represents DNA sequences present in the palindromes, and dotted bars represent insertions of foreign sequences at the junctions. Each numbers in brackets represents the number of nucleotides missing at the junction. Gray triangles indicate the sites of microhomology. The locations of primers used for PCR are shown on the top. c. Sequence analysis of the junctions of palindromes. A part of the I-SceI recognition sequence (actual cutting site indicated with a line) and surrounding sequence are shown at the top. Each numbers in brackets represents the number of nucleotides missing at the junction. Gray triangles indicate the sites of microhomology. The locations of primers used for PCR are shown on the top. c. Sequence analysis of the junctions of palindromes. A part of the I-SceI recognition sequence (actual cutting site indicated with a line) and surrounding sequence are shown at the top. Each numbers in brackets represents the number of nucleotides missing at the junction. Gray triangles indicate the sites of microhomology. The locations of primers used for PCR are shown on the top. c. Sequence analysis of the junctions of palindromes. A part of the I-SceI recognition sequence (actual cutting site indicated with a line) and surrounding sequence are shown at the top. Each numbers in brackets represents the number of nucleotides missing at the junction. Gray triangles indicate the sites of microhomology. The locations of primers used for PCR are shown on the top.
Intrastrand annealing for palindrome formation is facilitated by a 229IR. Southern analysis was used to analyze DSB-induced DNA rearrangements in MTX-resistant clones from the noIR subtransformants. In contrast to a 1.9-kb KpnI-AflII fragment in parental cells, MTX-resistant clones gave fragments of about 3.8 kb or a little smaller (Fig. 2a). The fragment in each MTX-resistant clone was converted to half-size after denaturing and rapid cooling (snap-back) (36), indicating that a DSB alone, without an adjacent DNA IR, is sufficient to generate large palindromes. Heterogeneous sizes of palindromic fragments suggest variable processing of I-SceI-induced DSBs. The DNA sequences at the centers of large palindromes hold a key to identifying molecular mechanisms. To determine the DNA sequences at the centers of palindromes, bisulphite modification of genomic DNA was applied (25, 37). Bisulphite treatment converted cytosines to uracils, which disrupted the palindromes to allow the centers of palindromes to be amplified by PCR for subsequent sequencing analysis. Two clones (8 and 12) had an almost perfect palindrome with very small deletions (less than 25 bp) from the I-SceI cutting site on both arms (Fig. 2b and c). This very small loss of sequences is similar to the end of an I-SceI-induced DSB repaired by NHEJ (16, 17). Thus, these large palindromes were probably generated by an intermolecular mechanism in which two sister chromatids

**FIG. 3.** A 229-bp DNA IR facilitates palindrome formation. a. Palindrome formation mediated by a 229IR. Southern analysis for the genomic DNA (right, top) and snap-back genomic DNA (right, bottom) was performed using the probe indicated (left). A DSB at the I-SceI site and palindrome formation eliminated the KpnI site that was located upstream of the I-SceI site and generated a 3.8-kb fragment in most of the clones. A 1.9-kb fragment of lower intensity was also seen, which might represent a half-sized fragment formed in vitro by intramolecular pairing of a large 3.8-kb palindrome (20, 36). Palindrome formation was confirmed by the snap-back procedure on genomic DNA, which generated a single 1.9-kb fragment (indicated by an arrowhead) in most of the clones. A 1-kb DNA ladder (NEB) is shown as a size marker. b. Intramolecular recombination model. After 5'-to-3' resection of the DSB end, intrastrand pairing occurs either between short IR (left) (5, 6, 37) or by microhomology (right). This initially generates a hairpin molecule, and a large DNA palindrome forms after DNA replication. Microhomology serves as a very small version of IR.
FIG. 4. The center of a palindrome marks the boundary of the amplicon at 1q21 in the Colo320DM cell line. 

a. Array-CGH analysis identified the regional amplification at 1q21 in Colo320DM cells (36). The log2 ratio of CGH signal intensity of Colo320DM relative to that of human foreskin fibroblast (HFF2) cells on genes within the 20-Mb region at 1q21 is shown. A spotted human cDNA array was used for this array-CGH analysis. Genes with increased intensity in Colo320DM cluster within a 1-Mb region (solid bar). Note that the CTSK and HIST2H4 genes are located near the telomeric and centromeric boundaries, respectively, of this amplification. MB, million base pairs. 

b. Schematic drawing of the 50-kb region at 1q21, covering three genes (CTSS, CTSK, and ARNT). Exons of these genes are represented by gray rectangles. Probes used for Southern analysis (Cen and Tel) are indicated. Restriction sites for EcoRV (EV) are also shown. 

c. Palindrome formation at the boundary of the amplicon. EcoRV-digested DNA is hybridized by the probes indicated in panel b. A 16-kb amplified fragment was seen in Colo320DM (Colo) cells using the Cen probe (black arrowhead), and this fragment was converted into a half-sized one after snap-back (gray arrowhead), indicating palindromic organization of this fragment. A probe in the ARNT gene (Tel) hybridized to a single fragment of similar intensity in both human foreskin fibroblast (HFF) and Colo cells, and this fragment was also seen in Colo cells using the probe Cen, implying a normal allele was present in Colo320DM cells. SB, snap-back. A 2.5-kb molecular ruler (Bio-Rad) is shown as a DNA size marker. 

d. DNA sequence at the center of the 1998 TANAKA ET AL. MOL. CELL. BIOL.
A boundary of gene amplification in cancer cells is defined by the center of a palindrome. In order to investigate whether intrastrand annealing and recombination sets a boundary of amplicons in cancer cells, GAPF was used (22, 36). GAPF identifies regions of the genome at or near centers of DNA palindromes. Genomic DNA from Colo320DM and HFF2 (a human foreskin fibroblast cell strain) cells were processed to enrich for DNA palindromes and hybridized to the GeneChip Human Genome U133A array. Among the 42 features that were significantly increased in Colo320DM DNA (FDR [34] of <0.05) (see Table S1 in the supplemental material), four genes were located at 1q21, which is an amplified locus in Colo320DM DNA (13, 36). CTSK was among the most significantly increased genes (FDR = 0.00889). Digesting Colo320DM DNA with EcoRV and hybridizing with a probe covering the CTSK gene (Fig. 4b) yielded an amplified 16-kb fragment that was distinct from the normal fragment of 34 kb, indicating a DNA rearrangement and gene amplification (Fig. 4c). This 16-kb fragment was converted to a half-size (8-kb) fragment after snap-back, confirming its palindromic structure. The palindrome extended in the direction of the centromere (see Fig. S2 in the supplemental material). DNA sequence analysis of the center of the palindrome indicates that the palindrome has a small asymmetry of 1.8 kb (Fig. 4d). Identical stretches of sequence homology were not found. However, the genomic sequences at the proximal and distal ends of the nonpalindromic region are highly AT rich (72% and 73%, respectively). The distribution of adenine and thymine has a strong strand bias (Fig. 4d), with 47% (28/60) adenine in the proximal end of the nonpalindromic spacer (“A-rich” region) and 45% (27/60) thymine in the distal end (“T-rich” region). The AT richness and strand bias distribution of nucleotides might have promoted intrastrand annealing during the formation of this large palindrome. The region covering the ARNT gene was not amplified (Fig. 4c), indicating that the center of the palindrome marks the telomeric boundary between amplified and nonamplified regions.

A naturally existing DNA inverted repeat forms the centromeric boundary of the amplicon. Another GAPF-positive gene at 1q21 in Colo320DM DNA was the HIST2H4 gene, which is in a naturally existing 26.1-kb large DNA IR (Fig. 5a). The sequence identity of the arms is 99.7%, and there is a 6.7-kb spacer between arms (41). To determine whether this DNA IR is involved in gene amplification, DNA copy numbers were compared in the regions centromeric to, within, and telomeric to the DNA IR. Although the DNA region 7 kb centromeric to the IR showed similar copy number intensities (2N) between HFF2 and Colo320DM DNA (probe a), the regions within and telomeric to the IR (probes b and c) showed amplification. However, the degree
of amplification was different between the nonpalindromic spacer (probe b) and the telomeric region (probe c): the telomeric region showed a 4N copy number (an increase of 2N), whereas the nonpalindromic spacer showed a 3N copy number (an increase of 1N). This twofold difference in copy number increase is consistent with palindromic amplification of the entire region, with the nonpalindromic spacer at the center, in which the palindromic arms are amplified twice as many times as the nonpalindromic spacer. Thus, a naturally existing IR in the genome sets the centromeric boundary of the amplicon (Fig. 5b). Therefore, the large palindrome formed in cancer cells seems to have a structure consistent with those generated by DSB-induced intrastrand annealing in our model system of DHFR amplification.
We show here that intrastrand annealing is a prevalent mechanism for palindromic gene amplification in mammalian cells and that a DNA IR significantly facilitates this process. Furthermore, the boundaries of genomic amplification in cancer cells are consistent with palindromic formation by intrastrand annealing, and a DNA IR in the human genome marks the boundary between amplified and nonamplified DNA. These findings suggest that human gene amplification might occur through a conserved process also used for developmental gene amplification in *Tetrahymena*. A DNA IR in the genome is involved in this process, also indicating that local sequence homologies, at least in some cases, determine the regions most susceptible to amplification.

In our model system of *DHFR* amplification, a 229-bp IR significantly increases the efficiency of palindromic formation when inserted at a chromosomal locus. Furthermore, in Colo320DM cells, a 26-kb IR at 1q21 marks the boundary of an amplicon. Thus, some regions of genomic amplification are likely to be determined by the local DNA homology. The extent of the IR homology facilitates palindromic formation, which in turn determines the location of the amplicon boundaries. Several lines of evidence suggest that IR-mediated palindrome formation is a major pathway for generating DNA palindromes in cancer cells. First, large IRs with very high identity between the arms preexist in the human genome. A study by Warburton et al. identified a number of DNA IRs in the human genome with more than 8 kb in each arm length and 99% identity between the arms (41). In addition, Alu sequences often form DNA IRs with a high degree of homology (>80%) and substantial alignment (>275 bp) when present next to each other in the genome (33). Second, a DNA IR also serves either as an “at-risk-motif” or a “fragile site” to regulate large palindromic formation in *S. cerevisiae* when DNA replication, repair, and checkpoint function are compromised (14, 18). Furthermore, IR sequences preexisting in the genome are found at the centers of palindromes in *Leishmania* and *Schizosaccharomyces pombe* (2, 11). This DNA IR-facilitated model of gene amplification predicts that palindrome formation is in part regulated by the presence of DNA IRs and occurs nonrandomly in the genome, which may provide a mechanistic basis for the nonrandom distribution of DNA palindromes in human cancer cells (36).

It is also important to note that a DSB alone is sufficient to initiate large palindrome formation in mammalian cells. Even in these cases, however, many clones show biased processing of the ends of the I-SceI-induced DSB, which is likely to be derived from intrasstrand annealing. Excision (5’ to 3’) of an I-SceI-induced DSB end generates a 3’ single-strand overhang that forms a hairpin molecule prior to DNA replication through intrasstrand base pairing, resulting in a large palindrome after DNA replication (Fig. 3b). This model is analogous to the intramolecular mechanism of palindrome formation mediated by the short DNA IR in *Tetrahymena* and to the small IR-assisted intramolecular mechanism for palindrome formation in yeast (Fig. 3b, left) (6, 20, 26). Microhomology represents a very short DNA IR (Fig. 3b, right), and the length of homology determines the efficiency of intrasstrand annealing. This process is likely to be independent of NHEJ, since an intramolecular mechanism using a short stretch of homology can occur in the absence of RAD50, YKU70, and LIG4 in *S. cerevisiae* (20, 26).

The amplicon at 1q21 in Colo320DM DNA covers a 1-Mb genomic region, and the copy number increase at some loci in this region is more than fourfold (36). Regional amplification of 1q21 is often seen in primary cancer tissues, including those of osteosarcoma and ovarian cancer (3, 35), suggesting an involvement of 1q21 amplification in tumor progression. The CTSK gene is located at the telomeric end of this regional amplification. An amplicon extending centromeric to the CTSK gene suggests that a DSB at this region initiated palindrome formation and generated a dicentric chromosome (Fig. 5b). Subsequent BFB cycles probably resulted in another chromosome break near the 27-kb DNA IR at the *HIST2H4* locus, again generating a large palindrome with the DNA IR at the center for this regional amplification. The presence of one large palindrome along with the duplication of the normal allele generates a copy number profile that is consistent with the copy number increases presented here. In both cases, the center of the large palindrome sets the boundary of an amplicon.

Using both a model experimental system with CHO cells and DNA analysis with cancer cells, we show here an important molecular process that regulates the development and progression of cancer. An initial chromosome break followed by palindrome formation triggers BFB cycles, which further generate genetic diversity in premalignant and cancer cell populations (10, 28). With a genetic deficiency, such as the loss of p53 function, which leads to permissivity for gene amplification (31), local genomic sequences, such as DNA IRs, might determine the incidence and location of gene amplification. IRs might not be accurately represented in the current human genome sequences, since DNA IR sequences are well known to be unstable upon cloning in *Escherichia coli*. DNA IRs may also exist as a form of genetic polymorphism among individuals (7). Further mapping of the epicenters of large palindromes should determine genomic and genetic attributes that lead to initial palindrome formation in cancer cells.

**ACKNOWLEDGMENTS**

We thank J. Delrow and the FHCRC Genomics facility for helpful advice on the array studies and M. Marotta for technical assistance. We thank G. Stark, B. Trask, D. Gottschling, R. Monnat, M. Thayer, and members of the Yao lab for suggestions and comments on the manuscript.

This work was supported by National Institutes of Health grants R01GM 26210 (M.-C.Y.) and R01AR 045113, R01AR 045113, R01AR 045203, and POINS 46788 (S.J.T) and by grants from the Avon Breast Cancer Crusade Opportunity Fund (to H.T.). H.T. is a recipient of an Interdisciplinary Dual Mentor Fellowship from the Fred Hutchinson Cancer Research Center.

**REFERENCES**

1. **Albertson, D. G., C. Collins, F. McCormick, and J. W. Gray.** 2003. Chromosome aberrations in solid tumors. Nat. Genet. 34:369–376.
2. **Albertrecht, E. B., A. B. Hunyady, G. R. Stark, and T. E. Patterson.** 2000. Mechanisms of sod2 gene amplification in *Schizosaccharomyces pombe*. Mol. Biol. Cell 11:873–886.
3. **Atiye, J., M. Wolf, S. Kaur, O. Monni, T. Bohling, A. Kivioja, E. Tas, M. Serra, M. Tarkkanen, and S. Knuttila.** 2005. Gene amplifications in osteosarcoma-CGH microarray analysis. Genes Chromosomes Cancer 42:158–163.
4. **Brown, P. C., T. D. Tlsty, and R. T. Schimke.** 1983. Enhancement of metho-
treats resistance and dihydrofolate reductase gene amplification by treatment of mouse 3T3 cells with hydroxyurea. Mol. Cell. Biol. 3:1097–1107.

5. Butler, D. K., L. E. Yasuda, and M. C. Yao. 1996. Induction of large DNA palindrome formation in yeast: implications for gene amplification and genomic stability in eukaryotes. Cell 87:1115–1122.

6. Butler, D. K., L. E. Yasuda, and M. C. Yao. 1995. An intramolecular recombinational mechanism for the formation of the rRNA gene palindrome of \textit{Tetrahymena thermophila}. Mol. Cell. Biol. 15:7117–7126.

7. Cabaret, S., C. Conti, C. Schurra, R. Lebofsky, S. J. Edelstein, and A. Lemoine. 2000. Human ribosomal RNA gene arrays display a broad range of chromosomal structures. Genome Res. 10:1079–1085.

8. Cao, Y., R. M. Kumar, B. H. Penn, C. A. Berkes, C. Kooperberg, L. A. Boyer, R. A. Young, and S. J. Tapscott. 2006. Global and gene-specific analyses show distinct roles for Myod and Myog at a common set of promoters. EMBO J. 25:502–511.

9. Coquelle, A., E. Pipiras, F. Toledo, G. Bittlin, and M. Debatisse. 1997. Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amorphs. Cell 89:215–225.

10. Gisselsson, D., L. Pettersson, M. Hoglund, M. Heidenblad, L. Gorunova, J. Lobachev, K. S., D. A. Gordenin, and M. A. Resnick. 2002. DNA ligase IV-dependent NHEJ of deprotected mammalian DNA repair of DNA double-strand breaks in mammalian chromosomes. Mol. Cell. Biol. 16:3587–3595.

11. Hahn, P., B. Nevaldine, and W. F. Morgan. 1990. X-ray induction of methotrexate resistance due to dhfr gene amplification. Somat. Cell Mol. Genet. 16:413–423.

12. Ishikawa, A. S., A. A. Mallof, S. K. Watson, R. J. DeLeuwe, B. Chi, B. P. Cook, A. Snijders, D. G. Albertson, D. Pinkel, M. A. Marra, V. Ling, C. MacAulay, and W. L. Lam. 2004. A tiling resolution DNA microarray with complete coverage of the human genome. Nat. Genet. 36:299–303.

13. Lemoine, F. J., N. P. Degtyareva, K. Lobachev, and T. D. Petes. 2005. Chromosomal breakage-fusion-bridge events cause genetic intraturatom heterogeneity. Proc. Natl. Acad. Sci. USA 97:5357–5362.

14. Grondin, K., G. Roy, and M. Ouellette. 1996. Formation of extrachromosomal circular amplicons with direct or inverted duplications in drug-resistant \textit{Leishmania amazonica}. Mol. Cell. Biol. 16:3557–3561.

15. Hahn, P., B. Nevaldine, and W. F. Morgan. 1990. X-ray induction of methotrexate resistance due to dhfr gene amplification. Somat. Cell Mol. Genet. 16:413–423.

16. Lengauer, C., K. W. Kinzler, and B. Vogelstein. 1998. Genetic instabilities in human cancers. Nature 396:643–649.

17. Liang, F., M. Han, P. J. Romanenko, and M. Jasim. 1998. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. Proc. Natl. Acad. Sci. USA 95:5172–5177.

18. Lin, Y., T. Lukacsovich, and A. S. Waldman. 1999. Multiple pathways for repair of DNA double-strand breaks in mammalian chromosomes. Mol. Cell. Biol. 19:8353–8360.

19. Lobachev, K. S., D. A. Gordenin, and M. A. Resnick. 2002. The mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. Cell 108:183–193.

20. Ma, C., S. Martin, B. Trask, and J. L. Hamlin. 2000. Chromosomal breakage-fusion-bridge events cause genetic intraturatom heterogeneity. Proc. Natl. Acad. Sci. USA 97:5357–5362.

21. Li, H., and M. C. Yao. 2002. DNA ligase IV-dependent NHEJ of deprotected mammalian DNA repair of DNA double-strand breaks in mammalian chromosomes. Mol. Cell. Biol. 16:3587–3595.

22. Tanaka, H., S. J. Tapscott, R. J. Trask, and M. C. Yao. 2002. Short inverted repeats initiate gene amplification through the formation of a large DNA palindrome in mammalian cells. Proc. Natl. Acad. Sci. USA 99:8772–8777.

23. Tlsty, T. D. 1997. Genomic instability and its role in neoplasia. Curr. Top. Microbiol. Immunol. 221:37–46.

24. Tlsty, T. D., P. C. Brown, and R. T. Schimke. 1984. UV radiation facilitates methotrexate resistance and amplification of the dihydrofolate reductase gene in cultured 3T3 mouse cells. Mol. Cell. Biol. 4:1050–1056.

25. Polo, F., D. Le Roscuet, G. Bittlin, and M. Debatisse. 1992. Co-amplified markers alternate in megabase long chromosomal inverted repeats and cluster independently in interphase nuclei at early steps of mammalian gene amplification. EMBO J. 11:2665–2673.

26. Warburton, P. E., J. Giordano, F. Cheung, Y. Gelfand, and G. Benson. 2004. Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. Genome Res. 14:1861–1869.

27. Yao, M. C., C. H. Yao, and B. Monks. 1990. The controlling sequence for site-specific chromosome breakage in \textit{Tetrahymena}. Cell 63:763–772.

28. Yasuda, L. F., and M. C. Yao. 1991. Short inverted repeats at a free end signal large palindromic DNA formation in \textit{Tetrahymena}. Cell 67:505–516.

29. Zhu, C., K. D. Mills, D. O. Ferguson, C. Lee, J. Manis, J. Fleming, Y. Gao, C. C. Morton, and F. W. Alt. 2002. Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations. Cell 109:811–821.