The preference for GT-rich DNA by the yeast Rad51 protein defines a set of universal pairing sequences

Robert B. Tracy,[1] Jason K. Baumohl,[2] and Stephen C. Kowalczykowski[1,2,3]

Division of Biological Sciences, Sections of Microbiology and of Molecular and Cellular Biology, [1]Microbiology Graduate Group, [2]Genetics Graduate Group, University of California, Davis, California 95616 USA

The Rad51 protein of *Saccharomyces cerevisiae* is a eukaryotic homolog of the RecA protein, the prototypic DNA strand-exchange protein of *Escherichia coli*. *RAD51* gene function is required for efficient genetic recombination and for DNA double-strand break repair. Recently, we demonstrated that RecA protein has a preferential affinity for GT-rich DNA sequences—several of which exhibit enhanced RecA protein-promoted homologous pairing activity. The fundamental similarity between the RecA and Rad51 proteins suggests that Rad51 might display an analogous bias. Using in vitro selection, here we show that the yeast Rad51 protein shares the same preference for GT-rich sequences as its prokaryotic counterpart. This bias is also manifest as an increased ability of Rad51 protein to promote the invasion of supercoiled DNA by homologous GT-rich single-stranded DNA, an activity not previously described for the eukaryotic pairing protein. We propose that the preferred utilization of GT-rich sequences is a conserved feature among all homologs of RecA protein, and that GT-rich regions are loci for increased genetic exchange in both prokaryotes and eukaryotes.

[*Key Words:* Rad51 protein; in vitro selection; genetic recombination; homologous pairing; genetic instability]

Received August 5, 1997; revised version accepted October 6, 1997.

In the yeast *Saccharomyces cerevisiae*, genes belonging to the *RAD52* epistasis group are required for homologous recombination and for DNA double-strand break repair (for reviews, see Resnick 1987; Petes et al. 1991; Game 1993). The *RAD52* group includes the *RAD50, RAD51, RAD52, RAD54, RAD55, RAD57, MRE11*, and *XRS2* genes. Genetic characterization of this group has identified *RAD51* as one of the most important members of this group. Mutations in the *RAD51* gene result in a myriad of defects including sensitivity to ionizing radiation and alkylating agents, a deficiency in mating-type switching, and deficiencies in mitotic and meiotic recombination (Resnick 1987; Petes et al. 1991; Shinohara et al. 1992; Game 1993). The Rad51 protein shares significant sequence as well as structural homology with the prototypic and most well-studied DNA strand exchange protein RecA from *Escherichia coli* (Aboussekhra et al. 1992; Basile et al. 1992; Shinohara et al. 1992). The RecA protein is absolutely essential for homologous genetic recombination and DNA repair in *E. coli* (for reviews, see Clark 1973; Radding 1988; West 1992; Cox 1993; Kowalczykowski et al. 1994).

The recent purification and biochemical characterization of the yeast Rad51 protein has provided an opportunity to compare it more thoroughly to RecA protein. Similar to RecA protein, Rad51 protein has DNA-dependent ATPase activity (Sung 1994; Sugiyama et al. 1997), and possesses the ability to promote homologous pairing and DNA strand exchange in vitro (Sung 1994; Sung and Robberson 1995; Sugiyama et al. 1997). Also, Rad51 protein forms a nucleoprotein filament that is indistinguishable from the filament formed by the RecA protein (Ogawa et al. 1993b; Sung and Robberson 1995). However, although these activities establish Rad51 protein as the yeast analog of the RecA protein, some differences do exist. First, the Rad51 protein catalyzes the hydrolysis of ATP at a rate that is ~35- to 40-fold lower than for RecA protein (Sung 1994; Sugiyama et al. 1997). Second, the Rad51 protein does not require ATP hydrolysis at any step during DNA strand exchange (Sung and Stratton 1996). Finally, the Rad51 protein catalyzes DNA strand exchange with a polarity ($3' \rightarrow 5'$), which is opposite to that of RecA protein (Sung and Robberson 1995).

These DNA strand exchange proteins are characterized by their ability to bind DNA without regard to sequence and to pair any homologous DNA sequence to its partner (for reviews, see West 1992; Ogawa et al. 1993a; Kowalczykowski and Eggleston 1994; Kowalczykowski et al. 1994; Camerini-Otero and Hsieh 1995). This permits genetic recombination to occur anywhere along the length of homologous chromosomes. However, although binding and pairing of DNA are generally regarded as being nonspecific, loci exist within both prokaryotes and

[3]**Corresponding author.**
E-MAIL sckowalczykowski@ucdavis.edu; FAX (916) 752-5939.

eukaryotes that are either enhanced or depressed for homologous recombination. Previously, we showed that a component of this variance could be attributed directly to the DNA strand exchange process. Using in vitro selection and homologous pairing assays, we showed that (1) the *E. coli* RecA protein has a preferential affinity for GT-rich DNA sequences; (2) these selected sequences display enhanced RecA protein-dependent homologous pairing activity; and (3) that these sequences are found at genomic loci within prokaryotes and eukaryotes that display enhanced recombinational activity, such as the *E. coli* recombination hot spot Chi (χ; 5′-GCTGGTGG-3′) (Lam et al. 1974; Dixon and Kowalczykowski 1991), microsatellite DNA from humans, the constant region of heavy chains of immunoglobulins from several mammals, and *Alu* repetitive elements from humans (Tracy and Kowalczykowski 1996). These findings suggested that the genetic instability observed at these sequences is attributable, at least in part, to the recombinational activity of the respective DNA strand exchange protein. This suggestion proved accurate for the χ sequence, as DNA substrates containing χ exhibited elevated levels of homologous pairing in RecA protein-dependent joint molecule formation assays (Tracy and Kowalczykowski 1996).

Because of the genetic and biochemical similarities between the RecA and Rad51 proteins, we wanted to determine whether Rad51 protein displays an analogous preference for binding and homologously pairing GT-rich DNA. Using in vitro selection, we show that the yeast Rad51 protein has a similar bias for GT-rich DNA sequences. This preference is also displayed as an increased rate and extent of homologous pairing activity. Confirming our supposition above, many of the sequences share significant homology with eukaryotic genomic sequences that are genetically unstable. Therefore, we suggest that such GT-rich DNA sequences are universally important loci for genetic exchange.

Results

In vitro selection of GT-rich DNA sequences by Rad51 protein

To determine whether the Rad51 protein has a preference for a certain DNA composition, we performed in vitro selection starting with a pool of 6×10^{13} 54-mers (SKBT18; Table 1). The 54-mers are composed of a random internal region of 18 nucleotides, and two flanking defined regions of 18 nucleotides (Tracy and Kowalczykowski 1996). To ensure significant competition between all of the 54-mers, we used limiting concentrations of purified Rad51 protein during five cycles of selection and amplification. The selected 54-mers were cloned and 40 randomly chosen clones were sequenced. Figure 1 depicts the sequences of the internal 18-mer regions that were selected by Rad51 protein. The average base composition of these sequences is highly over-represented for G and T residues (45% and 31%, respectively), and is significantly under-represented for A and C residues (10% and 15%, respectively). In addition to the 32 different sequences shown, the sequence displayed for clone 1 (selected sequence 1) appeared an additional eight times (which accounts for the missing clones in Fig. 1). Clone 39 is identical to clone 1, except for single-base changes at positions 2 (C → G) and 18 (C → G). Remarkably, the sequence of clone 1 is exactly the same sequence that appeared 16 times in the RecA protein selection experiments (Tracy and Kowalczykowski 1996). The frequent selection of this sequence is not a consequence of it being over-represented in the initial pool, as sequencing results showed that the pool was completely random (data not shown).

Similarly, control reactions established that the observed preference does not arise from the in vitro selection methodology (Tracy and Kowalczykowski 1996), demonstrating that the bias shown by RecA protein and Rad51 protein is intrinsic to these proteins. These results demonstrate that Rad51 protein has a preference for DNA sequences that are GT-rich, and that it has a particular bias for the sequence found in clone 1. Strikingly, this is nearly identical to the results obtained with RecA protein, except that Rad51 protein appears to have an even greater preference for G's and a slightly decreased bias for T's.

Significant over-representation of GT-rich dinucleotides and trinucleotides in the selected DNA sequences

Because both the Rad51 and RecA proteins bind two to three nucleotides per protein monomer (Kowalczykowski 1991; Shinohara et al. 1992; Ogawa et al. 1993b), the distribution of dinucleotide and trinucleotide frequencies in the protein-selected sequences serves to

Table 1. *Oligodeoxyribonucleotides*

Designation and description	Length	Sequence
SKBT16—Contains selected sequence 1	54	5′-AGCTTGCATGCCTGCAGGGGCGTGTGTGGTGGTGTGCTAGGATCCCCGGGTACC-3′
SKBT17—Complement of SKBT16	54	5′-GGTACCCGGGGATCCTAGCACACCACCACACACGCCCCTGCAGGCATGCAAGCT-3′
SKBT18—Random pool of 54-mers	54	5′-AGGCAAGCTTGGCTGCAG(N_{18})TAGGATCCGAGTCCTCCT-3′
SKBT19—Homologous to pBT54CN1 —opposite side of plasmid	54	5′-TGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTA-3′
SKBT20—Homologous to pBT54CN1 —adjacent to SKBT16	54	5′-TCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTTGCATGCCTGCAG-3′

		%G	%A	%T	%C
Clone 1:	5'-CCGTGTGTGGTGGTGTGC-3'	56	0	33	11
Clone 2:	5'-GGTACGTGTGTGGTGTGG-3'	56	6	33	6
Clone 3:	5'-GGGAGTTGTGGGCTGCGG-3'	61	6	22	11
Clone 4:	5'-GGGTGGTCTATTGTTGTG-3'	44	6	44	6
Clone 5:	5'-GAGTAGTATTGGTCGTGG-3'	44	17	33	6
Clone 6:	5'-TCGGCGTGATTATGTTGC-3'	33	11	39	17
Clone 8:	5'-GTGGTATGTGTTGTTGGC-3'	44	6	44	6
Clone 9:	5'-GTGGCATATCTTTGTGTG-3'	33	11	44	11
Clone 10:	5'-TCCAATGAGGTGTGTTGC-3'	33	17	33	17
Clone 12:	5'-GGGTGTCATTAAGTGTCC-3'	33	17	33	17
Clone 13:	5'-GCGGATGGTGCATGGGGC-3'	61	11	17	11
Clone 14:	5'-GGGTAGGGTGGTATTGGC-3'	56	11	28	6
Clone 15:	5'-CACCATCCCTGCGTTGTC-3'	17	11	28	44
Clone 18:	5'-GGGCAGACGGTCTAGGGC-3'	50	17	11	22
Clone 19:	5'-GGGTCGAAGTCGTGTTGG-3'	50	11	28	11
Clone 20:	5'-GGAGGTTATAGTTGTGTG-3'	44	17	39	0
Clone 21:	5'-GTTGTCGACATCTGGTGG-3'	39	11	33	17
Clone 22:	5'-GTACGGTGCGTGGTCGTG-3'	50	6	28	17
Clone 25:	5'-TGGGGCGTTTCCGCTGAG-3'	56	6	28	11
Clone 26:	5'-GGGCTCATGTGTGTGATA-3'	39	17	33	11
Clone 27:	5'-CACGTCAGTCCGCTCCCC-3'	17	11	17	56
Clone 28:	5'-GGGGGGTTATCTGGGGTC-3'	56	6	28	11
Clone 29:	5'-GGTCGCCCATACGCCTGG-3'	44	11	17	28
Clone 30:	5'-GCGCTTGGACACTGTTGG-3'	39	11	28	22
Clone 32:	5'-GTACCGGCTGTTGTGTTG-3'	39	6	39	17
Clone 33:	5'-GCGGTGGTGACTCTGGGC-3'	50	6	22	22
Clone 34:	5'-GCTTGGTAGATGCCGTGC-3'	44	11	28	17
Clone 36:	5'-GGATAGCTGTTGTTGGTG-3'	44	11	39	6
Clone 37:	5'-GCATGTGGTAGTATGGTG-3'	50	17	33	0
Clone 38:	5'-GGGTGGGTAATCTTGGGC-3'	50	11	28	11
Clone 39:	5'-GGGTGTGTGGTGGTGTGC-3'	67	0	33	0
Clone 40:	5'-CATGCTACATTGTCTGGG-3'	28	17	33	22
	Average:	**44.6%**	**10.4%**	**30.5%**	**14.7%**

Figure 1. Rad51 protein selects for GT-rich DNA sequences from a random pool of oligonucleotides. In vitro selection was performed using the 54-mers containing a random internal region of 18 nucleotides. The sequences of the 18-nucleotide region after five cycles of selection and amplification are shown. In addition, the frequency of occurrence for each base is shown, along with the average base composition for all of the sequences.

quantify the intrinsic pattern of preferences displayed by a protein monomer for a given sequence. Table 2 shows that a group of trinucleotides (GTG, TGT, TGG, GGT, GGG, and TTG) are significantly over-represented. Except for one case (GGG), all of these triplets were also highly over-represented in the RecA protein-selected sequences (Tracy and Kowalczykowski 1996). Also, as observed for RecA protein, many triplets containing A and C residues are significantly under-represented. An analysis of the dinucleotide frequencies showed that three (GT, TG, and GG) are significantly over-represented relative to the expected frequency and to the other dinucleotide frequencies (Table 2). As observed for the trinucleotides, dinucleotides containing A and C residues are significantly avoided; the RecA protein-selected sequences displayed an identical bias. A χ^2 test demonstrated that the frequency pattern for both the trinucleotide and dinucleotide distributions is not the result of random chance (data not shown). Interestingly, the most over-represented triplet in the *S. cerevisiae* genome is TGG/CCA, and the second most over-represented dinucleotide is TG/CA (Burge et al. 1992), two of the most highly over-represented in the selected sequences. A similar correlation was observed between the RecA protein-selected sequences and the *E. coli* genome (Tracy and Kowalczykowski 1996).

Table 2. *Frequency of trinucleotide and dinucleotide distribution in the 18-nucleotide sequences selected by Rad51 protein*

Trinucleotides	Frequency (%)[a]	Trinucleotides	Frequency (%)[a]
GTG	10.4	CAC	0.6
TGT	7.4	CTA	0.6
TGG	7.2	CTT	0.6
GGT	7.0	CTC	0.6
GGG	5.6	CCG	0.6
TTG	4.5	CCA	0.6
GTT	3.7	CCC	0.6
GGC	2.7	AGA	0.4
GTC	2.7	AAG	0.4
GTA	2.3	AAT	0.4
CGT	2.1	ACT	0.4
TGC	2.0	ACC	0.4
ATG	2.0	TAA	0.4
CTG	2.0	TTT	0.4
TAT	1.8	CGA	0.4
CAT	1.8	CAG	0.4
AGT	1.6	CCT	0.4
GCG	1.6	GAA	0.2
TCT	1.6	AGC	0.2
GCT	1.4	TTC	0.2
TAG	1.4	CAA	0.2
TCG	1.4	AAA	0.0
CGG	1.4	AAC	0.0
GGA	1.2		
GAT	1.2	Dinucleotides	Frequency (%)[b]
ATT	1.2	GT	18.0
GAG	1.0	TG	17.8
ATA	1.0	GG	17.5
ATC	1.0	GC	5.7
ACG	1.0	TT	5.5
TGA	1.0	AT	4.8
TAC	1.0	TC	4.6
TCC	1.0	TA	4.4
GAC	0.8	CG	4.2
AGG	0.8	CT	3.5
TTA	0.8	GA	2.9
GCA	0.6	AG	2.9
GCC	0.6	CA	2.8
ACA	0.6	CC	2.4
TCA	0.6	AC	2.2
CGC	0.6	AA	0.7

[a]Frequency was calculated by determining the number of times a certain trinucleotide occurred in all 32 sequences and dividing this number by the total number of trinucleotide occurences (512). The expected frequency for a trinucleotide is 1.6% ($1/4^3$ = 1/64).

[b]Frequency was calculated by determining the number of times a certain dinucleotide occurred in all 32 sequences and dividing this number by the total number of dinucleotide occurrences (544). The expected frequency for a dinucleotide is 6.3% ($1/4^2$ = 1/16).

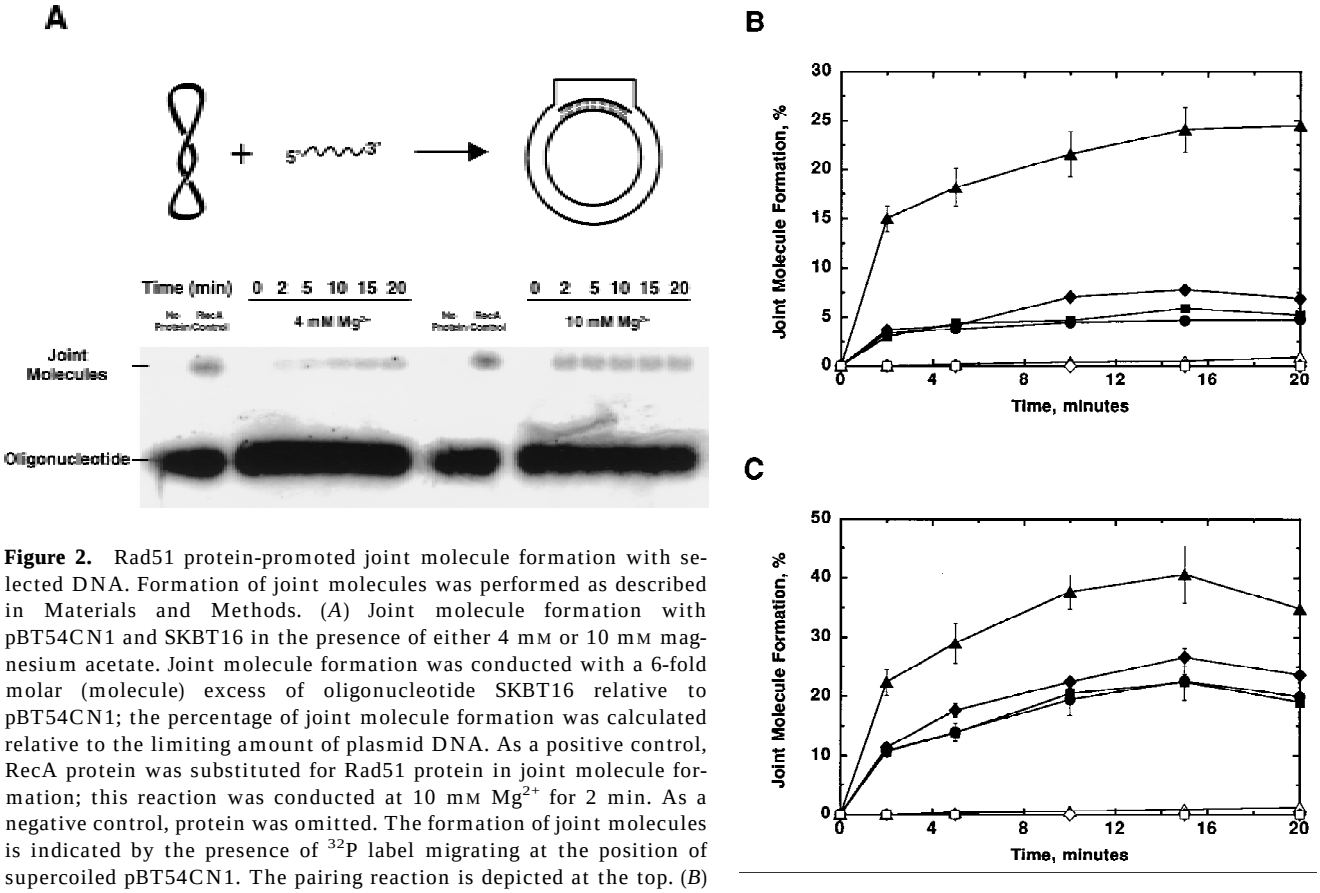

Figure 2. Rad51 protein-promoted joint molecule formation with selected DNA. Formation of joint molecules was performed as described in Materials and Methods. (*A*) Joint molecule formation with pBT54CN1 and SKBT16 in the presence of either 4 mM or 10 mM magnesium acetate. Joint molecule formation was conducted with a 6-fold molar (molecule) excess of oligonucleotide SKBT16 relative to pBT54CN1; the percentage of joint molecule formation was calculated relative to the limiting amount of plasmid DNA. As a positive control, RecA protein was substituted for Rad51 protein in joint molecule formation; this reaction was conducted at 10 mM Mg^{2+} for 2 min. As a negative control, protein was omitted. The formation of joint molecules is indicated by the presence of ^{32}P label migrating at the position of supercoiled pBT54CN1. The pairing reaction is depicted at the top. (*B*) Joint molecule formation with pBT54CN1 and either SKBT16 (▲, contains selected sequence 1); SKBT17 (●, complement of SKBT16); SKBT19 (♦, opposite side of plasmid); or SKBT20 (■, adjacent to SKBT16). Control reactions included absence of Rad51 protein (△); use of a heterologous plasmid (pBR322) (◇); or absence of pBT54CN1 (□). Reactions were performed in the presence of 4 mM magnesium acetate. (*C*) Joint molecule formation in the presence of 10 mM magnesium acetate. In both *B* and *C*, error bars indicate the standard deviation of three experiments.

Selected sequence 1 displays enhanced Rad51 protein-dependent joint molecule formation

Although Rad51 protein can promote DNA strand exchange between circular single-stranded DNA (ssDNA) and linear double-stranded DNA (dsDNA), its ability to promote invasion of supercoiled DNA by homologous ssDNA is unknown. Hence, we examined this aspect of Rad51 protein in the context of these selected DNA sequences. Because both RecA protein and Rad51 protein selected the sequence shown for clone 1, and as this sequence displayed an enhanced rate and extent of joint molecule formation (homologous pairing) in RecA protein-dependent pairing assays (3.5- and 3.4-fold increases, respectively) (Tracy and Kowalczykowski 1996), this sequence was used with Rad51 protein. Joint molecule formation was examined using the plasmid pBT54CN1, which contains selected sequence 1 as dsDNA in the multiple cloning site, and one of the following ssDNAs, (1) selected sequence 1 (SKBT16; Table 1); (2) the complement of SKBT16 (SKBT17; Table 1), which pairs in exactly the same region of the plasmid and thus eliminates the possibility of plasmid position-dependent effects; or (3) sequences (SKBT19 and SKBT20; Table 1) that are homologous to pBT54CN1 outside the vicinity of the selected sequence. In addition, because both the absolute level of homologous pairing and the pairing bias displayed by selected sequence 1 in the RecA protein-dependent reactions are sensitive to the magnesium ion (Mg^{2+}) concentration, and as the optimal Mg^{2+} concentration for in vitro pairing assays with Rad51 protein is unknown, we examined joint molecule formation at several different Mg^{2+} concentrations.

Figure 2 shows Rad51 protein-dependent formation of joint molecules at two different Mg^{2+} concentrations, 4 and 10 mM. Figure 2A demonstrates that Rad51 protein is able to promote the uptake of ssDNA (SKBT16) into a supercoiled molecule (pBT54CN1) at both 4 and 10 mM Mg^{2+}. At both concentrations, the initial rate and maximal extent of joint molecule formation are significantly increased for the selected sequence (SKBT16) relative to the control oligonucleotides. At 4 mM Mg^{2+}, the initial rate and maximal extent are increased on average ~4.4- and 4.0-fold, respectively (Fig. 2B). When the Mg^{2+} concentration was increased to 10 mM, the difference in the rate and extent between SKBT16 and the controls is still

significant, although slightly reduced (~2.0- and 1.7-fold, respectively; Fig. 2C). Pairing was also examined at higher Mg^{2+} concentrations (15 and 20 mM) for both SKBT16 and SKBT17. Even at these relatively high Mg^{2+} concentrations, the selected sequence still demonstrated enhanced joint molecule formation (data not shown). Control reactions in the absence of Rad51 protein, using a heterologous plasmid (pBR322), or lacking pBT54CN1 establish that this is an entirely homology-dependent and Rad51 protein-dependent pairing reaction, and for the first time, demonstrate that Rad51 protein is capable of promoting DNA strand invasion (Fig. 2). Similar to the results obtained for the RecA protein-dependent pairing assays, we find that the intrinsic difference in pairing activity between selected sequence 1 and the controls is most apparent at suboptimal in vitro reaction conditions (4 mM Mg^{2+}).

Discussion

Rad51 protein-selected sequences are similar to genetically unstable eukaryotic genomic sequences

Because the set of GT-rich sequences selected by RecA protein are similar to genetically unstable DNA sequences in both prokaryotes and eukaryotes, we wanted to determine whether the Rad51 protein-selected sequences might also represent attractive genomic targets for Rad51 protein action. When these sequences were compared to DNA sequences within the National Center for Biotechnology Information (NCBI) databases using the BLAST network service (Altschul et al. 1990), and the GenBank database using FastA, several (clones 1, 2, 3, 8, 9, 10, 13, 14, 20, 25, 26, 28, 30, 32, 33, 36, and 39) were found to share significant homology with sequences from eukaryotic organisms that are recombinationally active, such as microsatellite DNA, the constant region of the heavy chains of immunoglobulins, *Alu* repetitive elements, and telomeric repeat sequences (Table 3). Several studies showed that microsatellite sequences from mammalian genomes are hot spots for homologous recombination (Jeffreys et al. 1995; Steinmetz et al. 1986; Wahls et al. 1990). Wahls et al. (1990) indicated that their data were most consistent with a specific enzyme recognizing the consensus minisatellite sequence, and subsequently stimulating, either directly or indirectly, its ability to recombine. It is known that the constant region of the heavy chains of immunoglobulins from several mammals undergo rearrangement by homologous recombination at DNA sequences (called switch sequences), which are generally GT-rich (Singer and Berg 1991); the Rad51 protein is expressed and localized in B cells that carry out immunoglobulin heavy chain class switch recombination (Li et al. 1996). *Alu* repetitive elements are short repetitive DNA sequences (~300 bp long) that are dispersed throughout the genomes of vertebrates. They are of interest from a physiological point of view because they have been shown to cause genomic rearrangements in humans as a result of homologous recombination (Vanin et al. 1983; Henthorn et al. 1986; Hobbs et al. 1986; Nicholls et al. 1987; Rouyer et al. 1987; Rudiger et al. 1991). Finally, there is mounting evidence that both interstitial and terminal telomeric repeat sequences are hot spots for recombination in several eukaryotic organisms, including *S. cerevisiae* (Pluta and Zakian 1989), mice (Ashley et al. 1993; Ashley and Ward 1993; Nachman and Churchill 1996), and humans (Brown et al. 1990; Park et al. 1992; Rossi et al. 1993). In *Schizosaccharomyces pombe,* homologous pairing of chromosomes was shown to be most frequent near centromeres and telomeres (Scherthan et al. 1994). Considering that practically all telomere repeat sequences are GT-rich (Henderson 1995), our results can potentially explain their high recombination frequencies. When the Rad51 protein-selected sequences are compared to the telomere repeat sequences, clones 1 (the most prevalent sequence) and 39 show a close resemblance to the repeat sequence from *S. cerevisiae* [5′-TGTGGGTGTGGTGTGG-3′ (Singer and Gottschling 1994)]. Previously, it was shown that the maintenance of telomeric repeat arrays in the yeasts *S. cerevisiae* and *Kluyveromyces lactis* is strongly dependent on the *RAD52* pathway, which involves the Rad51 protein, in strains lacking an active telomerase (McEachern and Blackburn 1996). This demonstrates that homologous recombination can functionally substitute for telomerase, at least in these organisms. In addition, it was proposed that telomere–telomere recombination in *S. cerevisiae* could potentially be promoted by a RecA-like protein in a *RAD52* independent manner (Pluta and Zakian 1989); Rad51 protein is a candidate for this role.

Table 3. *Rad51 protein-selected sequences share significant homology with genetically unstable eukaryotic genomic sequences*

Eukaryotic genomic sequences	Organism	Clones that display similarity
Microsatellites	humans and bovine	1, 2, 9, 20, 26, 33, 36, and 39
Constant region of heavy chains of immunoglobulins	humans, bovine, sheep, rat, and mouse	1, 2, 8, 10, 25, 26, 28, 30, 32, 33, and 39
Alu repetitive elements	humans	1, 2, 3, 13, 14, 20, 32, and 39
Telomere repeat sequences	*S. cerevisiae*	1 and 39
Huntington's disease region	humans	1, 2, 20, 25, 28, 37, and 39
Fragile X syndrome gene	humans	2 and 38

Similarity was determined by using the VAX computer programs BLAST and FASTA (Genetics Computer Group, Inc.). Homology is considered to be significant when 14 of the 18 bases are identical.

Many of the selected sequences (clones 1, 2, 20, 25, 28, 37, 38, and 39) also share similarity with DNA segments associated with triplet expansion diseases in humans, such as Fragile X syndrome and Huntington's disease (Table 3). Fragile X syndrome is an X-linked dominant mental retardation disorder that is inherited (Richards and Sutherland 1992). The disorder occurs as a result of significant expansion of the $(CGG)_n$ trinucleotide repeat (from 2 to 50 copies in normal individuals to >200–2000 copies in affected individuals) (Kremer et al. 1991; Oberle et al. 1991; Verkerk et al. 1991). Huntington's disease is an inherited neurodegenerative disorder that is progressive and ultimately results in cognitive loss, motor disturbances, and psychiatric problems (Martin and Gusella 1986). Similar to Fragile X syndrome, Huntington's disease is characterized by expansion of a trinucleotide repeat (CAG_n). The repeat expands from 11–34 copies in the normal population to 42–66 copies in those affected by the disease (Huntington's Disease Collaborative Research Group 1993). One mechanism that was proposed to account for the variablity and instability associated with these tandem repeat sequences is unequal crossing over during homologous recombination (Smith 1974; Jeffreys et al. 1985). If Rad51 protein were targeted to regions directly adjacent to the triplet repeat sequences, then potentially this preference could offer one explanation for the observed instability of these sequences.

Although we advance the idea that increased genetic exchange by members of the Rad51 protein family contributes to the genetic instability observed at these sequences, we also wish to make it clear that other mechanisms can contribute to the observed genomic instability. In this regard, genetic variation has been attributed to slippage of DNA polymerase during replication (Levinson and Gutman 1987; Wells 1996), whereby the presence of DNA structures (e.g., hairpins) or proteins (e.g., histones) blocks the progression of DNA polymerase, causing it to continually slip. In addition, mismatch DNA repair has been implicated as a factor that contributes to the genetic variance (Strand et al. 1993; Jaworski et al. 1995; Wells 1996).

Because recombination in vivo is dependent on not only Rad51 protein, but also several other proteins from the *RAD52* epistasis group, it remains possible that these other proteins could further enhance or reduce the preferential binding and homologous pairing of GT-rich DNA sequences by Rad51 protein. Therefore, if increased genetic exchange mediated by Rad51 protein is indeed at least partially responsible for the observed genetic instability at these GT-rich genomic loci, then an additional consideration is whether other proteins of the *RAD52* epistasis group are influencing the ability of Rad51 protein to act preferentially at these sites. Because several members of this group have been purified, it is now possible to ascertain whether they affect Rad51 protein function at GT-rich sequences.

Implications

Taken together, these findings argue that the preferential binding of GT-rich sequences is a universal characteristic of a class of recombination enzymes, which is conserved from bacteria to eukaryotes, and that these GT-rich regions invite homologous recombination in their vicinity. A limitation of this argument is that we have examined only two members of this enzyme family and therefore we cannot exclude the possibility that other RecA protein homologs have preferences for DNA compositions that reflect that organism's genomic content. However, studies are in progress with a RecA protein homolog from the Archaea Domain, the RadA protein from *Sulfolobus solfataricus* (E.M. Seitz, R.B. Tracy, and S.C. Kowalczykowski, unpubl.). On the basis of these preliminary data, we favor the universality argument advanced here. The coincidence of GT-rich sequences at recombination hot spots in the chromosomes of a variety of organisms, however, indicates that these particular regions are more attractive for binding and subsequent genetic exchange by several different recombination enzymes. In *E. coli*, the recombination hot spot χ coordinates the efficient recombinational repair of double-stranded ends that form from direct damage or as a consequence of stalled replication forks (Kuzminov et al. 1994). This coordination results from the ability of χ to focus the efforts of the RecA and RecBCD proteins at ''recombination islands'' that are rich in GT composition, rather than at random dsDNA breaks (Dixon and Kowalczykowski 1991, 1995; Tracy and Kowalczykowski 1996; Anderson and Kowalczykowski 1997; Tracy et al. 1997). Consequently, this coordination permits a far more efficient maintenance of genomic integrity than any random, uncontrolled process. In eukaryotes, a similarly directed process is likely to be in place, with recombination hot spots playing a key role in both genetic diversity and genomic integrity.

Materials and methods

Chemicals

Chemicals were reagent grade, and solutions were prepared using Barnsted NANOpure water. ATP (Pharmacia LKB Biotechnology, Inc.) was dissolved as a concentrated stock solution at pH 7.5. The concentration of ATP was determined spectrophotometrically by using an extinction coefficient of 1.54×10^4 M/cm at 260 nm.

Proteins

Rad51 protein was purified from *E. coli* using a protocol that was modified from that described by Sung (1994) (E. Zaitseva, T. Sugiyama, J. New, and S.C. Kowalczykowski, unpubl.). T4 polynucleotide kinase was purchased from New England BioLabs. *Taq* DNA polymerase was purchased from Pharmacia. T7 sequenase, version 2.0, was purchased from U.S. Biochemical Corp. The enzymes were used as directed by the specific vendor.

DNA substrates

The oligodeoxyribonucleotides used and their sequences are shown in Table 1. All oligonucleotides were synthesized and purified as described (Tracy and Kowalczykowski 1996).

The plasmid pBT54CN1, which contains selected sequence 1 was constructed as described (Tracy and Kowalczykowski 1996). pBT54CN1 was purified by CsCl density gradient centrifugation followed by use of a Qiagen Maxi Kit (Qiagen, Inc., Chatsworth, CA).

In vitro selection assay

The oligodeoxyribonucleotide used for in vitro selection (SKBT18) is shown in Table 1. The random region contained an equimolar mixture of the four bases. The defined regions were used for PCR and cloning. The in vitro selection assay was performed as described (Tracy and Kowalczykowski 1996), except for the following changes. First, the radiolabeled oligonucleotides (100 μM nucleotides) were mixed with selection buffer [30 mM Tris acetate (pH 7.5), 20 mM magnesium acetate, 1 mM DTT, 50 mM potassium acetate, and 2.5 mM ATP] and incubated at 37°C for 2 min; 1 μM Rad51 protein was then added. The reaction mixture (100 μl) was incubated at 37°C for 10 min, immediately applied to a nitrocellulose filter (HAWP 0025 from Millipore Corp.), and washed as described previously (Tracy and Kowalczykowski 1996). Second, after each step of PCR, the amplified DNA was purified through a Microcon-10 column (Amicon).

Determination of dinucleotide and trinucleotide composition within the selected sequences

The VAX computer program Composition (Genetics Computer Group, Inc.) was used to determine the composition of dinucleotides or trinucleotides within the Rad51 protein-selected sequences. Once Composition had tabulated the occurrence of each dinucleotide or trinucleotide, the frequency (percentage) was determined.

Joint molecule formation assay

Oligonucleotides were 5′-end labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase. Joint molecule formation was carried out in a reaction mixture (250 μl) consisting of 30 mM Tris-acetate (pH 7.5), 1 mM DTT, 50 mM potassium acetate, 1 mM ATP, 1 μM nucleotides of oligonucleotides, 0.33 μM Rad51 protein, 18 μM nucleotides of pBT54CN1, and the indicated magnesium acetate concentration. The reactions were initiated by the addition of plasmid DNA, after preincubation of the other components for 5 min at 37°C. At 0, 2, 5, 10, 15, and 20 min, 40-μl aliquots were removed and immediately added to 5 μl of a 10% SDS/0.5 M EDTA mixture and 2 μl of proteinase K (14.4 mg/ml) (Boehringer Mannheim) to deproteinize the samples. Samples were incubated at 37°C for 10 min, followed by the addition of 5 μl of DNA loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol, and 25% Ficoll). Reactions were analyzed by electrophoresis through a 1% agarose gel at 30 V for 14 hr. The gels were dried and quantitated using a Betagen Betascope 603 radioisotopic analyzer. All control reactions were conducted with SKBT16. The percentage of joint molecules formed was determined relative to the total amount of plasmid DNA (which was limiting) present in a reaction.

Acknowledgments

We thank the following members of our laboratory for their helpful discussions and critical comments on the manuscript: Dan Anderson, Joel Brockman, Frederic Chedin, Jason Churchill, Deana Haddox, Frank Harmon, Noriko Kantake, Julie Kleiman, Sindy Law, Alex Mazin, Jim New, Erica Seitz, Tomohiko Sugiyama, and Eugene Zaitsev. This work was supported by a National Institutes of Health (NIH) grant (AI-18987) to S.C.K.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked ''advertisement'' in accordance with 18 USC section 1734 solely to indicate this fact.

References

Aboussekhra, A., R. Chanet, A. Adjiri, and F. Fabre. 1992. Semidominant suppressors of Srs2 helicase mutations of Saccharomyces cerevisiae map in the RAD51 gene, whose sequence predicts a protein with similarites to procaryotic RecA proteins. *Mol. Cell. Biol.* **12:** 3224–3234.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Anderson, D.G. and S.C. Kowalczykowski. 1997. The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a χ-regulated manner. *Cell* **90:** 77–86.

Ashley, T. and D.C. Ward. 1993. A ''hot spot'' of recombination coincides with an interstitial telomeric sequence in the Armenian hamster. *Cytogenet. Cell Genet.* **62:** 169–171.

Ashley, T., N.L. Cacheiro, L.B. Russell, and D.C. Ward. 1993. Molecular characterization of a pericentric inversion in mouse chromosome 8 implicates telomeres as promoters of meiotic recombination. *Chromosoma* **102:** 112–120.

Basile, G., M. Aker, and R.K. Mortimer. 1992. Nucleotide sequence and transcriptional regulation of the yeast recombinational repair gene RAD51. *Mol. Cell. Biol.* **12:** 3235–3246.

Brown, W.R., P.J. MacKinnon, A. Villasante, N. Spurr, V.J. Buckle, and M.J. Dobson. 1990. Structure and polymorphism of human telomere-associated DNA. *Cell* **63:** 119–132.

Burge, C., A.M. Campbell, and S. Karlin. 1992. Over- and underrepresentation of short oligonucleotides in DNA sequences. *Proc. Natl. Acad. Sci.* **89:** 1358–1362.

Camerini-Otero, R.D. and P. Hsieh. 1995. Homologous recombination proteins in prokaryotes and eukaryotes. *Annu. Rev. Genet.* **29:** 509–552.

Clark, A.J. 1973. Recombination deficient mutants of E. coli and other bacteria. *Annu. Rev. Genet.* **7:** 67–86.

Cox, M.M. 1993. Relating biochemistry to biology: How the recombinational repair function of RecA protein is manifested in its molecular properties. *BioEssays* **15:** 617–623.

Dixon, D.A. and S.C. Kowalczykowski. 1991. Homologous pairing in vitro stimulated by the recombination hotspot, Chi. *Cell* **66:** 361–371.

———. 1995. Role of the Escherichia coli recombination hotspot, chi, in RecABCD-dependent homologous pairing. *J. Biol. Chem.* **270:** 16360–16370.

Game, J.C. 1993. DNA double-strand breaks and the RAD50-RAD57 genes in Saccharomyces. *Semin. Cancer Biol.* **4:** 73–83.

Henderson, E. 1995. Telomere DNA Structure. In *Telomeres* (ed. E.H. Blackburn and C.W. Greider), pp. 11–34. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Henthorn, P.S., D.L. Mager, T.H. Huisman, and O. Smithies. 1986. A gene deletion ending within a complex array of repeated sequences 3′ to the human beta-globin gene cluster. *Proc. Natl. Acad. Sci.* **83:** 5194–5198.

Hobbs, H.H., M.S. Brown, J.L. Golstein, and D.W. Russell. 1986. Deletion of exon encoding cysteine-rich repeat of low density lipoprotein receptor alters its binding specificity in a

subject with familial hypercholesterolemia. *J. Biol. Chem.* **261:** 13114–13120.

Huntington's Disease Collaborative Research Group. 1993. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's Disease chromosomes. *Cell* **72:** 971–983.

Jeffreys, A.J., V. Wilson, and S.L. Thein. 1985. Hypervariable ''minisatellite'' regions in human DNA. *Nature* **314:** 67–73.

Jaworski, A., W.A. Rosche, R. Gellibolian, S. Kang, M. Shimizu, R.R. Sinden, and R.D. Wells. 1995. Mismatch repair in Escherichia coli enhances instability of (CTG)n triplet repeats from human hereditary diseases. *Proc. Natl. Acad. Sci.* **92:** 11019–11023.

Kowalczykowski, S.C. 1991. Biochemistry of genetic recombination: Energetics and mechanism of DNA strand exchange. *Annu. Rev. Biophys. Biophys. Chem.* **20:** 539–575.

Kowalczykowski, S.C. and A.K. Eggleston. 1994. Homologous pairing and DNA strand-exchange proteins. *Annu. Rev. Biochem.* **63:** 991–1043.

Kowalczykowski, S.C., D.A. Dixon, A.K. Eggleston, S.D. Lauder, and W.M. Rehrauer. 1994. Biochemistry of homologous recombination in *Escherichia coli. Microbiol. Rev.* **58:** 401–465.

Kremer, E.J., M. Pritchard, M. Lynch, S. Yu, K. Holman, E. Baker, S.T. Warren, D. Schlessinger, G.R. Sutherland, and R.I. Richards. 1991. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science* **252:** 1711–1714.

Kuzminov, A., E. Schabtach, and F.W. Stahl. 1994. Chi sites in combination with RecA protein increase the survival of linear DNA in Escherichia coli by inactivating exoV activity of RecBCD nuclease. *EMBO J.* **13:** 2764–2776.

Lam, S.T., M.M. Stahl, K.D. McMilin, and F.W. Stahl. 1974. Rec-mediated recombinatorial hot spot activity in bacteriophage lambda. II. A mutation which causes hot spot activity. *Genetics* **77:** 425–433.

Levinson, G. and G.A. Gutman. 1987. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4:** 203–221.

Li, M.J., M.C. Peakman, E.I. Golub, G. Reddy, D.C. Ward, C.M. Radding, and N. Maizels. 1996. Rad51 expression and localization in B cells carrying out class switch recombination. *Proc. Natl. Acad. Sci.* **93:** 10222–10227.

Martin, J.B. and J.F. Gusella. 1986. Huntington's disease. Pathogenesis and management. *N. Engl. J. Med.* **315:** 1267–1276.

McEachern, M.J. and E.H. Blackburn. 1996. Cap-prevented recombination between terminal telomeric repeat arrays (telomere CPR) maintains telomeres in *Kluyveromyces lactis* lacking telomerase. *Genes & Dev.* **10:** 1822–1834.

Nachman, M.W. and G.A. Churchill. 1996. Heterogeneity in rates of recombination across the mouse genome. *Genetics* **142:** 537–548.

Nicholls, R.D., N. Fischel-Ghodsian, and D.R. Higgs. 1987. Recombination at the human alpha-globin gene cluster: Sequence features and topological constraints. *Cell* **49:** 369–378.

Oberle, I., F. Rousseau, D. Heitz, C. Kretz, D. Devys, A. Hanauer, J. Boue, M.F. Bertheas, and J.L. Mandel. 1991. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* **252:** 1097–1102.

Ogawa, T., A. Shinohara, A. Nabetani, T. Ikeya, X. Yu, E.H. Egelman, and H. Ogawa. 1993a. RecA-like recombination proteins in eukaryotes: Function and structures of RAD51 genes. *Cold Spring Harbor Symp. Quant. Biol.* **58:** 567–576.

Ogawa, T., X. Yu, A. Shinohara, and E.H. Egelman. 1993b. Similarity of the yeast RAD51 filament to the bacterial RecA filament. *Science* **259:** 1896–1899.

Park, V.M., K.M. Gustashaw, and T.M. Wathen. 1992. The presence of interstitial telomeric sequences in constitutional chromosome abnormalities. *Am. J. Hum. Genet.* **50:** 914–923.

Petes, T.D., R.E. Malone, and L.S. Symington. 1991. Recombination in yeast. In *The molecular and cellular biology of the yeast* Saccharomyces: *Genome dynamics, protein synthesis, and energetics* (ed. J.R. Broach, E. Jones, and J. Pringle), pp. 407–521. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Pluta, A.F. and V.A. Zakian. 1989. Recombination occurs during telomere formation in yeast. *Nature* **337:** 429–433.

Radding, C.M. 1988. Homologous pairing and strand exchange promoted by Escherichia recA protein. In *Genetic recombination* (ed. R. Kucherlapati and G.R. Smith), pp. 193–229. American Society for Microbiology, Washington, DC.

Resnick, M.A. 1987. Investigating the genetic control of biochemical events in meiotic recombination. In *Meiosis* (ed. P.B. Moens), pp. 157–210. Academic Press, New York, NY.

Richards, R.I. and G.R. Sutherland. 1992. Fragile X syndrome: The molecular picture comes into focus. *Trends Genet.* **8:** 249–254.

Rossi, E., G. Floridia, M. Casali, C. Danesino, G. Chiumello, F. Bernardi, I. Magnani, L. Papi, M. Mura, and O. Zuffardi. 1993. Types, stability, and phenotypic consequences of chromosome rearrangements leading to interstitial telomeric sequences. *J. Med. Genet.* **30:** 926–931.

Rouyer, F., M.C. Simmler, D.C. Page, and J. Weissenbach. 1987. A sex chromosome rearrangement in a human XX male caused by Alu-Alu recombination. *Cell* **51:** 417–425.

Rudiger, N.S., P.S. Hansen, M. Jorgensen, O. Faergeman, L. Bolund, and N. Gregersen. 1991. Repetitive sequences involved in the recombination leading to deletion of exon 5 of the low-density-lipoprotein receptor gene in a patient with familial hypercholesterolemia. *Eur. J. Biochem.* **198:** 107–111.

Scherthan, H., J. Bahler, and J. Kohli. 1994. Dynamics of chromosome organization and pairing during meiotic prophase in fission yeast. *J. Cell. Biol.* **127:** 273–285.

Shinohara, A., H. Ogawa, and T. Ogawa. 1992. Rad51 protein involved in repair and recombination in S. cerevisiae is a RecA-like protein. *Cell* **69:** 457–470.

Singer, M. and P. Berg. 1991. *Genes & Genomes.* University Science Books, Mill Valley, CA.

Singer, M.S. and D.E. Gottschling. 1994. TLC1: Template RNA component of Saccharomyces cerevisiae telomerase [see comments]. *Science* **266:** 404–409.

Smith, G.P. 1974. Unequal crossover and the evolution of multigene families. *Cold Spring Harbor Symp. Quant. Biol.* **38:** 507–513.

Steinmetz, M., D. Stephan, and K. Fischer Lindahl. 1986. Gene organization and recombinational hotspots in the murine major histocompatibility complex. *Cell* **44:** 895–904.

Strand, M., T.A. Prolla, R.M. Liskay, and T.D. Petes. 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365:** 274–276.

Sugiyama, T., E.M. Zaitseva, and S.C. Kowalczykowski. 1997. A single-stranded DNA-binding protein is needed for efficient presynaptic complex formation by the Saccharomyces cerevisiae Rad51 protein. *J. Biol. Chem.* **272:** 7940–7945.

Sung, P. 1994. Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. *Science* **265:** 1241–1243.

Sung, P. and D.L. Robberson. 1995. DNA strand exchange mediated by a RAD51-ssDNA nucleoprotein filament with po-

larity opposite to that of RecA. *Cell* **82:** 453–461.
Sung, P. and S.A. Stratton. 1996. Yeast Rad51 recombinase mediates polar DNA strand exchange in the absence of ATP hydrolysis. *J. Biol. Chem*. **271:** 27983–27986.
Tracy, R.B. and S.C. Kowalczykowski. 1996. In vitro selection of preferred DNA pairing sequences by the *Escherichia coli* RecA protein. *Genes* & *Dev*. **10:** 1890–1903.
Tracy, R.B., F. Chedin, and S.C. Kowalczykowski. 1997. The recombination hot spot chi is embedded within islands of preferred DNA pairing sequences in the E. coli genome. *Cell* **90:** 205–206.
Vanin, E.F., P.S. Henthorn, D. Kioussis, F. Grosveld, and O. Smithies. 1983. Unexpected relationships between four large deletions in the human beta-globin gene cluster. *Cell* **35:** 701–709.
Verkerk, A.J., M. Pieretti, J.S. Sutcliffe, Y.H. Fu, D.P. Kuhl, A. Pizzuti, O. Reiner, S. Richards, M.F. Victoria, F.P. Zhang et al. 1991. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65:** 905–914.
Wahls, W.P., L.J. Wallace, and P.D. Moore. 1990. Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells. *Cell* **60:** 95–103.
Wells, R.D. 1996. Molecular basis of genetic instability of triplet repeats. *J. Biol. Chem*. **271:** 2875–2878.
West, S.C. 1992. Enzymes and molecular mechanisms of genetic recombination. *Annu. Rev. Biochem*. **61:** 603–640.