
Molecular Models for Repeat Expansions

Sergei M. Mirkin, University of Illinois at Chicago, College of Medicine

Characteristics of Repeat Expansions

Expansions of simple DNA repeats account for more than two dozen hereditary disorders in humans (for recent reviews see Lenzmeier and Freudenreich,¹ Jin and Warren,² Parniewski and Staczek,³ Ranum and Day,⁴ Brown and Brown,⁵ Cummings and Zoghbi,⁶ Bowater and Wells,⁷ and Siyanova and Mirkin⁸). Table 1 describes those diseases and their important genetic features. Although originally discovered expansions were limited to trinucleotide repeats $(CGG)_n \cdot (CCG)_n$,^{9,10} $(CAG)_n \cdot (CTG)_n$,^{11–17} and $(GAA)_n \cdot (TTC)_n$,¹⁸ it has now become clear that other repeats, including tetrameric $(CCTG)_n \cdot (CAGG)_n$,¹⁹ pentameric $(AATCT)_n \cdot (AGATT)_n$,²⁰ and even dodecameric $(C_4GC_4GCG)_n \cdot (CGCG_4CG_4)_n$,²¹ can expand as well.

For all cases listed in Table 1, expansions are limited to just one repeat at a given locus for every disease studied. Thus, these expansions are unlikely to be caused by mutations in general DNA metabolism, such as replication, repair, or recombination, which would tend to affect numerous repeats in various genetic loci. The latter is well illustrated by mutations in the mismatch repair genes in the case of hereditary nonpolyposis colon cancer that drastically increase length polymorphism of numerous genomic microsatellites (reviewed in Fishel and Kolodner²²). Thus, primary events leading to repeat expansions in all the cases should occur in *cis*.

Historically, two groups of expandable repeats, noncoding and polyglutamine coding, were found to cause human hereditary neurological disorders (shown in the upper half of Table 1). Several general features

Address correspondence and reprint requests to: Sergei M. Mirkin, Department of Biochemistry and Molecular Genetics, University of Illinois at Chicago, College of Medicine, Chicago, IL 60607; e-mail: mirkin@uic.edu.

CHEMTRACTS—BIOCHEMISTRY AND MOLECULAR BIOLOGY 17:639–662 (2004)

© 2004 Data Trace Publishing Company

CCC 1431-9268

Table 1. Human diseases caused by repeat expansions

Disease	Repeat	Gene	Chromosomal position	Protein	Repeat size		Repeat position	Inheritance	Mutation type	Transmission bias
					Norm	Disease				
Noncoding Repeats										
Fragile X syndrome	(CGG) _n	FMR1	Xq27	FMRP	<50	230–2,000	5'-UTR	X-linked dominant	LOF	Maternal
Fragile XE mental retardation	(GCC) _n	FMR2	Xq28	FMR2 protein	<42	200–900	5'-UTR	X-linked dominant	LOF	No
Myotonic dystrophy type 1	(CTG) _n	DMPK	19q13	DMPK	<30	80–1,000	3'-UTR	Autosomal dominant	GOF	Maternal
Myotonic dystrophy type 2	(CCTG) _n	ZNF9	3q21	ZNF9	<75	75–11,000	Intron	Autosomal dominant	GOF	ND
Spinocerebellar ataxia type 8	(CTG) _n	SCA8	13q21	SCA8	<90	100–127	3'-UTR	Autosomal dominant	GOF	Maternal
Spinocerebellar ataxia type 10	(AATCT) _n	SCA10	22q13	Ataxin-10	<22	800–4,500	Intron	Autosomal dominant	GOF	ND
Spinocerebellar ataxia type 12	(CAG) _n	PPP2R2B	5q31-33	PPP2A-PR55β	<45	55–78	5'-UTR	Autosomal dominant	LOF	ND
Huntington disease-like 2	(CTG) _n	JPH3	16q23-24	Junctophilin-3	<20	50–60	Spliced exon	Autosomal dominant	GOF	ND
Friedreich's ataxia	(GAA) _n	X25	9q13-21.1	Frxataxin	<35	200–1,700	Intron	Autosomal recessive	LOF	Maternal
Progressive myoclonus epilepsy type 1	(C ₄ GC ₄ GCG) _n	CSTB	21q22.3	Cystatin B	<3	30–75	Promoter	Autosomal recessive	LOF	Paternal
Coding Repeats—Polyglutamines										
Spinobulbar muscular atrophy	(CAG) _n	AR	Xq13-21	Androgen receptor	<36	38–62	ORF	X-linked recessive	LOF	ND
Huntington disease	(CAG) _n	IT15	4p16.3	Huntingtin	<34	40–121	ORF	Autosomal dominant	GOF	Paternal
Dentatorubralpallidolusian atrophy	(CAG) _n	DRPLA	12p13.31	Atrophin-1	<35	45–75	ORF	Autosomal dominant	GOF	Paternal

Spinocerebellar ataxia type 1	(CAG) _n	SCA1	6p23	Ataxin-1	<39	40–82	ORF	Autosomal dominant	GOF	Paternal
Spinocerebellar ataxia type 2	(CAG) _n	SCA2	12q24.1	Ataxin-2	<33	33–200	ORF	Autosomal dominant	GOF	Paternal
Spinocerebellar ataxia type 3	(CAG) _n	SCA3	14q32.1	Ataxin-3	<44	55–84	ORF	Autosomal dominant	GOF	Paternal
Spinocerebellar ataxia type 6	(CAG) _n	CACNA1A	19p13a1	Ca ²⁺ channel α 12.1	<19	20–29	ORF	Autosomal dominant	LOF	ND
Spinocerebellar ataxia type 7	(CAG) _n	SCA7	3p12-13	Ataxin-7	<35	37–306	ORF	Autosomal dominant	GOF	Paternal
Spinocerebellar ataxia type 17	(CAG) _n	SCA17	6q27	TBP	<42	47–53	ORF	Autosomal dominant	GOF	ND
Coding Repeats—										
Polyalanines										
Infantile spasm syndrome	(GCX) _n	ARX	Xp22	ARX	16	17–23	ORF	X-linked	LOF	ND
Cleidocranial dysplasia	(GCX) _n	RUNX2	6p21	RUNX2	17	27	ORF	Autosomal dominant	LOF	ND
BPE inversus syndrome	(GCX) _n	FOXL2	3q23	FOXL2	14	22–24	ORF	Autosomal dominant	LOF	ND
Hand-foot-genital syndrome	(GCX) _n	HOXA13	2q31	HOXA13	12,18	18–26	ORF	Autosomal dominant	ND	ND
Synpolydactyly syndrome	(GCX) _n	HOXD13	7p14	HOXD13	15	22–29	ORF	Autosomal dominant	ND	ND
Oculopharyngeal muscular dystrophy	(GCX) _n	PABP1	14q.11	PABP2	10	12,17	ORF	Autosomal dominant	GOF	ND
Holoprosencephaly	(GCX) _n	ZIC2	13q.32	ZIC2	15	25	ORF	ND	LOF	ND
Congenital central hypoventilation syndrome	(GCX) _n	PHOX2B	4p12	PHOX2B	20	25,33	ORF	Autosomal dominant	LOF	ND
XLMR + GHD	(GCX) _n	SOX3	6q26	SOX3	15,16	18–26	ORF	ND	ND	ND

characterize expansions of those repeats. First, expansions of repeats start when their length exceeds a certain threshold. Although the exact threshold varies for different diseases (Table 1), it generally falls within the range of 100–200 bps. Expandable repeats differ in their sequences and structural features. Hence, similar threshold lengths for various repeats indicate that the same DNA transaction might be involved in their expansions. Normal alleles usually contain several repeat interruptions. In expanding alleles, interruptions at one end of a repeat disappear, creating a long, homogenous run.²³

Second, the probability of expansions increases with the increase in the repeat's length beyond the threshold. This phenomenon accounts for the non-Mendelian inheritance of these diseases, characterized by an increased probability of transmission, a.k.a. Sherman's paradox,^{24,25} together with an earlier onset and greater severity, a.k.a. anticipation,^{26–29} of a disease as a mutant allele passes through generations. While expansions occur with mutation frequencies for repeats of threshold lengths, very long repeats expand with near 100% probability during intergenerational transmissions.

Third, expansions tend to be large-scale events (i.e., multiple repeats are added during a single transmission).³⁰ As a result, a repeat's length can increase up to 100-fold within a few generations, particularly for repeats situated in the noncoding repeats. This massive accumulation of repetitive DNA indicates that DNA synthesis is involved. Note that the smaller scale of expansions, characteristic for the polyglutamine-coding repeats,^{5,6} might simply be due to the counterselection against toxic polyaminoacid stretches in protein products.

Fourth, there is a strong bias for repeat expansions during intergenerational transmissions in humans and transgenic mice. This bias was not observed in simple model systems, such as *Escherichia coli* or *Saccharomyces cerevisiae* (reviewed in Lahue and Slater³¹).

These general features indicate that a universal mechanism could be responsible for expansions of noncoding and polyglutamine-coding repeats. Yet numerous reports from diverse experimental systems implicate different DNA transactions in repeat expansions. The main purpose of this review is to discuss various molecular models for repeat expansions with an emphasis on our hypothesis implicating replication fork stalling and restart in expansions.

Human diseases, caused by expansions of polyalanine-coding repeats (shown in the lower half of Table 1), were discovered later and are less well studied. Yet it has already become clear that the main characteristics of polyalanine diseases are markedly different from those for diseases caused by expansions of noncoding and polyglutamine-coding repeats (reviewed in Amiel et al.³² and Brown and Brown³³). First, genes containing polyalanine repeats encode transcription factors, and repeat expansions lead to developmental anomalies, rather than to neurological diseases. Second, polyalanine tracts are usually encoded by imperfect triplet repeats (GCX)_n, differently from all other cases of repeat expansions. Third, (GCX)_n tracts are not generally polymorphic in length, in sharp contrast with other expandable repeats. Fourth, the threshold length for expansion of polyalanine-coding repeats is the shortest of all,

and the scale of expansions is exceptionally small. Finally, alleles with expanded $(GCX)_n$ repeats are stable during intergenerational transmissions, as well as somatically. Thus, the most striking feature of expansions of the noncoding and polyglutamine-coding repeats (i.e., their progressive lengthening during intergenerational transmission) does not apply to polyalanine-coding repeats. These profound differences indicate that the mechanisms for polyalanine-repeat expansions are likely different from those responsible for other repeats. These mechanisms are briefly discussed in the section on genetic recombination.

Structural Features of Expandable Repeats

Early studies of expandable repeats suggested their unusual structural potential. In linear double-stranded DNA, chemical reactivity of $d(CNG)_n$ repeats differs from that of canonical B-DNA.³⁴ Denaturing and re-annealing of repeat-containing duplexes leads to the formation of slipped-stranded DNA.^{35,36} In this structure (Fig. 1A), complementary repeats are misaligned, leading to double-helical segments alternated with loop-outs. This conformation would be extremely thermodynamically unfavorable, unless the loop-outs were stabilized by hydrogen bonds. This seems to be the case for expandable repeats. Analysis of individual DNA strands of different trinucleotide repeats by various biochemical and biophysical approaches showed that they can fold into defined, compact structures.³⁷⁻⁴⁰ $d(CGG)_n$, $d(CCG)_n$, $d(CTG)_n$, and $d(CAG)_n$ stretches can fold into hairpin-like structures by both W-C and non-W-C base pairs (Fig. 1B).⁴¹⁻⁴⁴ Hairpins, formed by the above four repeats, differ in the nature of non-W-C base pairs. Thus, their stability varies because of a differential contribution from different mismatches in the following way: $CGG > CCG \approx CTG > CAG$.⁴² Recently, formation of folded structures by extended $d(GAA)_n$ and $d(TTC)_n$ repeats has also been postulated.⁴⁵

Individual strands of expandable repeats can fold into other DNA conformations, as well. For example, single-stranded $d(CGG)_n$ repeats fold into a stable quadruplex structure, shown in Figure 1C.⁴⁶ A totally different structure can be formed by the $(GAA)_n \cdot (TTC)_n$ repeat. This repeat is a subclass of homopurine-homopyrimidine mirror repeats that adopt triple-helical H-DNA configuration under the influence of negative supercoiling.⁴⁷ Interestingly, the exact nature of a triplex formed by the $(GAA)_n \cdot (TTC)_n$ repeat is a subject of controversy. Different labs proposed H-y (i.e., pyrimidine/purine/pyrimidine triplex),⁴⁸ H-r (i.e., pyrimidine/purine/purine triplex)⁴⁹ (Fig. 1D), or a composite triplex structure, called "sticky DNA."^{49,50} Finally, an AT-rich repeat implicated in SCA10, $(ATTCT)_n \cdot (AGAAT)_n$, apparently forms a peculiar unwound structure while under superhelical stress.⁵¹

Interestingly, for trinucleotide repeats that are not known to expand, formation of unusual DNA structures appeared to be less likely.⁴² Furthermore, stabilizing interruptions within trinucleotide repeats in normal alleles make formation of stable unusual DNA structures less likely.^{42,52} It is generally believed, therefore, that structure-forming potential of these repeats is involved in their expansions.^{35,53} Formation of unusual DNA structures by expandable repeats could obstruct various DNA transactions,

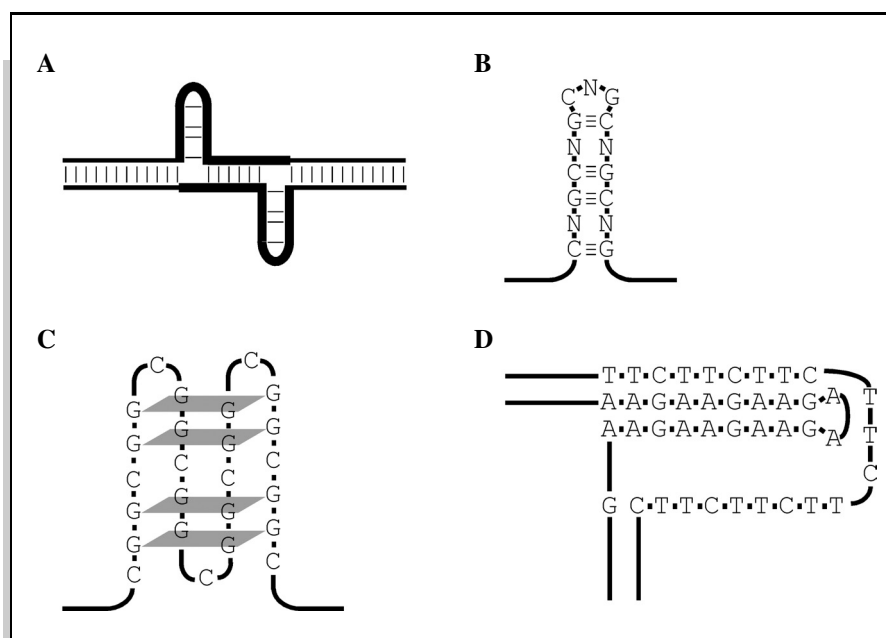


Figure 1. Unusual secondary structures of expandable repeats. **A**, S-DNA; **B**, hairpin-like structure; **C**, G-quartet; **D**, H-r DNA.

ultimately contributing to expansions. For example, it has long been known that DNA polymerases are slowed down or stopped altogether by stable hairpins,^{54–59} triplexes,^{60–63} and G-quartets.^{64–66} Unusual DNA structures were also shown to affect genetic recombination.^{67–71}

It was further suggested that the threshold length for repeat expansions might reflect the stability threshold for unusual DNA structures, formed by these repeats.⁴² This idea is challenged by two groups of data. First, unusual DNA structures formed by different repeats, such as hairpins, quadruplexes, and triplexes, have very different stabilities (reviewed in Sinden⁷²), yet the expansion threshold remains similar. Second, very stable secondary structures, capable of obstructing DNA transactions, are formed by triplet repeats that are significantly shorter than the expansion threshold.^{38,73} A link between the expansion threshold and unusual DNA structures could be that formation of all those structures requires significant DNA unwinding, or even complete separation of the two DNA strands: S-DNA is formed upon denaturing/reannealing, hairpins and quadruplexes are formed upon strand separation of a repeat, H-DNA is topologically equivalent to unwound DNA, and so forth. What factor(s) could be responsible for the separation of DNA strands of various repeats *in vivo*?

Replication Models for Repeat Expansions

The main cellular process involving DNA strand separation is replication. Figure 2 represents a basic scheme of the replication fork, in which the lagging strand is synthesized in discrete Okazaki fragments. In eukaryotes, Okazaki fragments normally range from 100 to 300 nts, comparable with the threshold length for repeat expansions.^{74,75} During the

replication fork progression, a portion of the lagging strand template must remain single-stranded to assure coordinated syntheses of leading and lagging strand syntheses and the priming of subsequent Okazaki fragments. The size of this Okazaki initiation zone (OIZ) (Fig. 2) is comparable with the length of an Okazaki fragment, that is, 200 nts long on average.^{74,75} Normally, single-stranded DNA in the OIZ is prevented from folding into stable secondary structures by two factors: binding of the RPA and primer synthesis by the Pol α -primase complex. Both RPA and Pol α -primase complex, however, are moderately sequence specific. RPA prefers pyrimidine-rich DNAs as binding substrates.⁷⁶ Pol α -primase also requires pyrimidine-rich templates, since the 5'-most nucleotide in the Okazaki primer is always purine, preferably adenine.⁷⁷ For "random" DNA templates, these sequence requirements do not pose a problem, but expandable repeats can represent a challenge. The most radical example of such challenge is the d(GAA)_n run situated on the lagging strand template,⁷⁸ which is the worst possible template for both the RPA or Pol α -primase complex. Obviously, an inefficiency of the RPA and/or Pol α -primase on single-stranded DNA repeat would facilitate its folding into unusual DNA conformations discussed above. This should become particularly plausible, when the repeat's length reaches the size of an OIZ. We hypothesize, therefore, that the threshold length for repeat expansions might reflect the average size of an Okazaki initiation zone, making it uniform for different repeats.

The above consideration, together with the fact that massive accumulation of repetitive DNA should require DNA synthesis, puts DNA replication at the forefront of repeat expansions. Indeed, replication models of repeat expansions were the first ones to emerge,^{30,79} and are supported by the largest volume of experimental data.

DNA polymerization through various trinucleotide repeats *in vitro* is known to be compromised. In double-stranded DNA templates, both pro- and eukaryotic DNA polymerases stopped at specific sites within (CTG)_n·(CAG)_n and (CGG)_n·(CCG)_n repeats, and the termination rate increased with the length of the repeat.⁸⁰ In single-stranded DNA templates, (CGG)_n runs caused a K⁺-dependent polymerization arrest, presumably due to folding into tetrahelical structures shown in Figure 1C.⁶⁶ (GAA)_n·(TTC)_n repeats represent the most potent block for DNA polymerases both in single- and double-stranded DNA templates because of the triplex formation during polymerization.⁴⁸ Thus, unusual structures of expandable repeats account for DNA synthesis blockage *in vitro*. Remarkably, this blockage could occasionally lead to the misalignment of the newly synthesized and template DNA strands.⁸¹ Reinitiation of the DNA synthesis would then lead to repeat extensions or deletions, depending on whether the newly synthesized or template DNA strand, respectively, is slipped out (Fig. 3). Note that this simple consideration is, in one or another form, employed in all replication models for repeat expansions *in vivo*.

One of the first such models was based on the data obtained in a bacterial plasmid system.⁷⁹ A feature of this system was that deletions of expandable repeats were quite common, whereas expansions were rare and could only be detected by PCR amplification. The frequencies of the two events were compared for (CTG)_n·(CAG)_n repeats cloned in

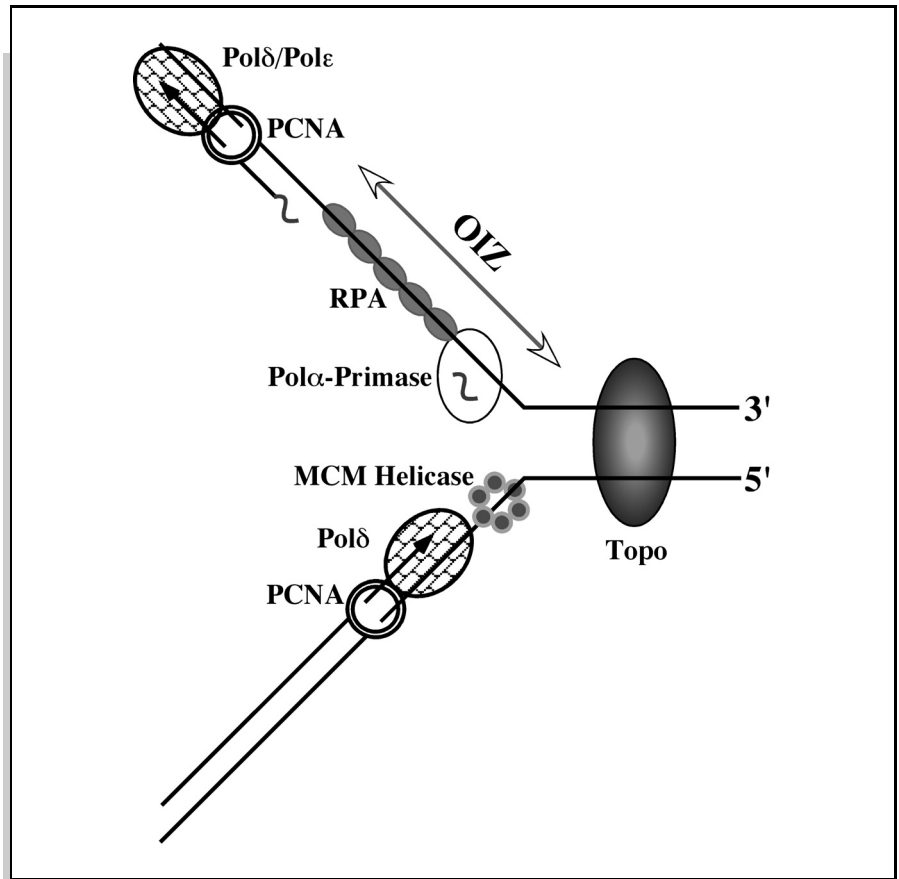


Figure 2. Eukaryotic replication fork. Black lines, DNA chains; wavy lines, RNA primers; arrows, 3'-ends of the growing DNA chains.

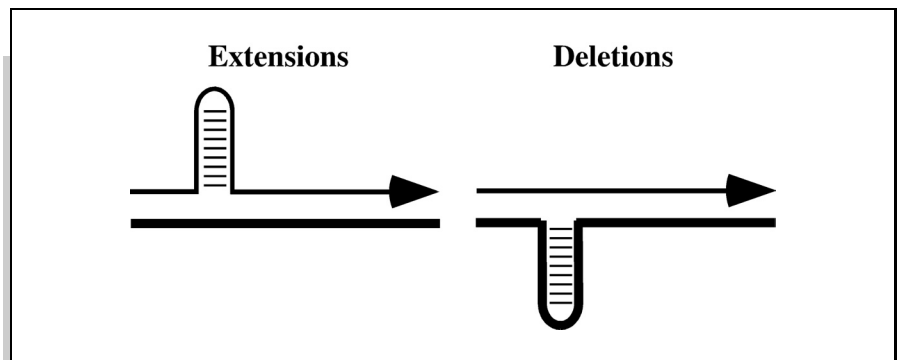


Figure 3. Slippage between repeat-containing DNA strands during DNA polymerization in vitro can lead to extensions or deletions. Thin line, nascent DNA strand; thick line, template DNA strand; arrow, 3'-end of a nascent strand.

different orientations relative to the plasmid replication *ori*. Deletions occurred more frequently and no expansions were detected when (CTG)_n runs were in the lagging strand template. When the same stretches were in the leading strand template (i.e., nascent lagging strand), deletions

occurred less frequently, while expansions became detectable. Given that the complementary strands of expandable repeats have different structure-forming capacity (see above), it was suggested that hairpin-like structures, formed by structure-prone repetitive strands, in either the lagging strand template or the newly synthesized lagging strand caused deletions or expansions, respectively (Fig. 4). This hypothesis gained significant support after expansion frequencies for various expandable repeats in bacteria, yeast, and mammalian cells were found to depend on their orientation within the replicore.^{79,82–87} Although this model implied that repeat expansions preferably occur during the lagging strand synthesis,⁷⁹ it has never been proven decisively. In fact, the same group later found that expansions can occur during the leading strand synthesis.⁸⁸

Further support for the replication model came from numerous studies showing that mutations in the replication apparatus of bacteria and yeast affect repeat expansions. These mutations included deletions of the yeast Flap-endonuclease (FEN1), missense mutations in the yeast DNA polymerase δ and PCNA, proofreading mutants of *E. coli* DNA polymerase III, and others.^{89–95} It would be fair to say that the role of FEN1 in repeat expansions generated most interest. Normally, FEN1 is involved in the removal of RNA primers from Okazaki fragments (reviewed in Liu et al.⁹⁶). At the end of an Okazaki fragment synthesis, a 5'-part of the previous Okazaki fragment is displaced, producing a composite RNA/DNA flap. The RNA portion of this flap is first removed by the Dna2 endonuclease, while its DNA part is chewed up by FEN1. The FEN1 ring loads onto the 5'-end of the flap, migrates to its junction with the duplex, and introduces a nick (Fig. 5A). It was hypothesized⁹⁷ that expandable repeats within the flap could prevent FEN1 loading by virtue of hairpin formation (Fig. 5B). Ligation of this hairpin with the next Okazaki fragment leads to expansions (Fig. 5C). Several in vitro studies indeed showed that expandable repeats strongly inhibit the cleavage activity of FEN1.^{92,98–100} Most strikingly, the rate of repeat expansions in the yeast

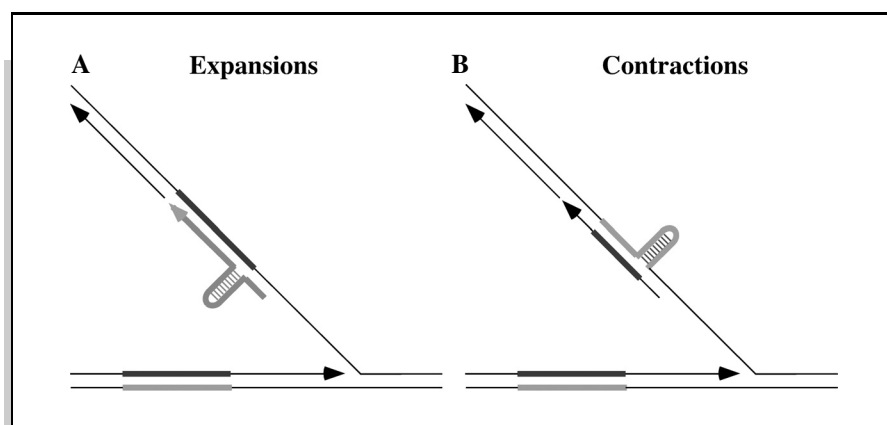


Figure 4. Replication model for repeat instability during the lagging strand synthesis (adapted from Kang et al.⁷⁹). Thin black line, flanking DNA; thick light gray line, structure-prone strand of a repeat; thick dark gray line, complementary strand of a repeat; arrows, 3'-ends of nascent DNA chains.

Rad27 mutant lacking FEN1 is increased by up to two orders of magnitude.^{90,91,93} Note, however, that the interpretation of these data is not necessarily straightforward (for a recent comprehensive review see Lahue and Slater³¹). For example, repeats, which are not known to form alternative DNA structures and/or expand in humans, efficiently expanded in the *Rad27* mutants studied.^{101,102} Furthermore, FEN1 deficiency did not affect repeat expansions in somatic and embryonic mouse cells (Bu Wieringa, personal communication). Although the final role of FEN1 in repeat expansions remains to be elucidated, the model shown in Figure 5 was the first to imply that the rate for the repeat's expansion might depend on its position within an Okazaki fragment.

All replication mutants, discussed above, increased the instability of the whole variety of DNA repeats, not only expandable ones. It was important, therefore, to search for mutations that would specifically affect expandable repeats. Presently, only two such mutations are described. An *Srs2* mutant was identified in a screen for elevated repeat expansions in yeast.¹⁰³ In this mutant, the rates for $(CTG)_n \cdot (CAG)_n$ and $(CGG)_n \cdot (CCG)_n$ repeat expansions were increased approximately 40-fold, while other repeats remained stable. It was also demonstrated that wild-type *Srs2* helicase prevents repeat expansions in conjunction with DNA polymerase δ . In a separate study, *rfc1-1* mutation was shown to increase expansions of $(CGG)_n \cdot (CCG)_n$ repeats approximately 50-fold, while having only modest

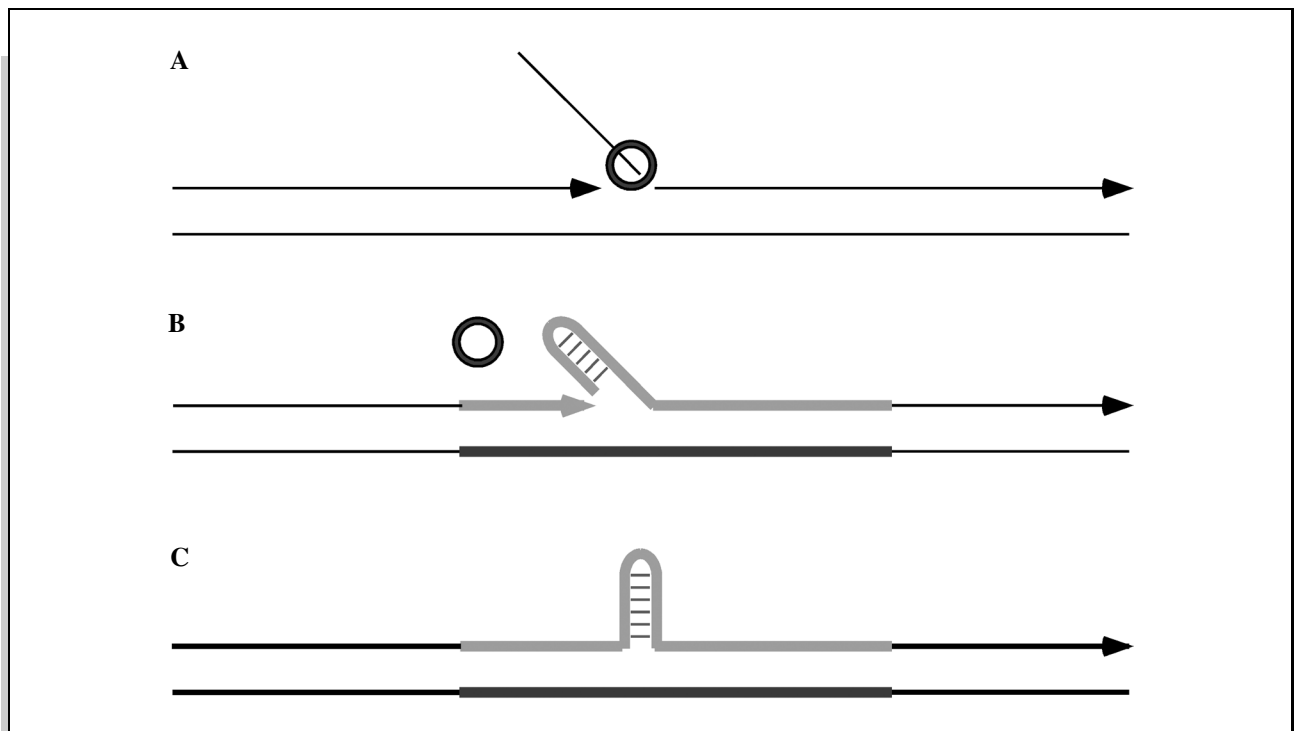


Figure 5. Flap model for repeat expansions. **A**, FEN1 processing of a nonrepetitive flap; **B**, FEN1 fails to remove repetitive flap; **C**, ligation of the repetitive flap leads to expansions. Thin black line, flanking DNA; thick light gray line, structure-prone strand of a repeat; thick dark gray line, complementary strand of a repeat; arrows, 3'-ends of nascent DNA chains. FEN1 is shown as a donut-shaped structure.

effect on other triplet repeats.¹⁰⁴ This mutation decreases ATPase activity of the large subunit of the RFC complex,¹⁰⁵ which is required for loading PCNA and, subsequently, the lagging strand polymerase onto primers for Okazaki fragments.¹⁰⁶ Noticeably, both enzymes are needed for orderly replication elongation: Srs2 helicase travels with the replication fork, helping restart stalled forks,¹⁰⁷ while the RFC complex is vital for the coordination of leading and lagging DNA strand synthesis.¹⁰⁶ Thus, mutational analysis strongly indicates that repeat stability might depend on the mode of replication fork progression through them.

To study the mode of replication fork progression through expandable repeats, our lab analyzed replication intermediates that were isolated from living cells. We looked at the effects of expandable repeats on replication of bacterial and yeast plasmids,^{78,104,108} as well as mammalian episomal vectors (unpublished results), using an approach called two-dimensional neutral/neutral gel electrophoresis. This technique, originally developed for mapping replication origins,^{109,110} appeared to be invaluable for defining replication pause sites. Using this approach, we have firmly demonstrated that $(CGG)_n \cdot (CCG)_n$, $(CAG)_n \cdot (CTG)_n$ and $(GAA)_n \cdot (TTC)_n$ repeats attenuate replication fork progression in all these systems. Replication inhibition depended on the repeat's length, becoming evident at around the expansion threshold. The strength of replication inhibition also depended on the repeat's base composition in the following order: $(CGG)_n \cdot (CCG)_n > (GAA)_n \cdot (TTC)_n > (CAG)_n \cdot (CTG)_n$. For some repeats, most explicitly $(GAA)_n \cdot (TTC)_n$ repeats, replication blockage depended on their orientation relative to the replication origin. The lagging strand in the vicinity of a repeat appeared to be underreplicated.¹⁰⁸ Finally, expansions and contractions were most prominent in the repeats' orientation, leading to the replication stalling.⁷⁸

These data, together with the replication studies discussed above, encouraged us to formulate a model for repeat expansions, which is based on the replication fork stalling and restart. This model in its universal form is presented in Figure 6; its earlier version for $(GAA)_n \cdot (TTC)_n$ repeats was described by us elsewhere.⁷⁸ When the leading strand DNA polymerase runs into an expandable repeat, an OIZ on the lagging strand template DNA becomes repetitive (Fig. 6A). If a repeat exceeds the threshold length for expansions, the whole OIZ could become a repetitive run. As discussed above, the RPA and Pol α -primase complex are relatively inefficient on certain repetitive templates, facilitating OIZ folding into an unusual conformation (Fig. 6B). This could prevent an orderly progression of the lagging strand polymerase and, consequently, leading strand polymerase, resulting in the replication fork stalling and dissociation (Fig. 6C). This model assumes that orientation dependence in the replication stalling at various repeats results from the balance between the structure-forming potential of a single-stranded repeat and its affinity for the RPA and/or Pol α -primase. For relatively short repeats, their location at the 5' or 3' end of the OIZ could have a major impact on their stability (for a comprehensive discussion, see Cleary and Pearson¹¹¹). During the replication restart, genome guardians, such as the RPA, Srs2-helicase, and others, should unravel a stable DNA structure, formed by the repeat. While this unraveling is under way, the newly synthesized leading strand and its template could dissociate and misalign by their repetitive portions

(Fig. 6D). This misalignment can occur in two ways, such that the repetitive segment on either the template or the newly synthesized DNA strand is slipped out. Resumption of the DNA synthesis would then lead to repeat contractions (Fig. 6D lower panel) or expansions (Fig. 6D upper panel), respectively.

Our model suggests that repeat expansions and contractions occur during the leading strand synthesis. This distinguishes it from other hypotheses on repeat instabilities, implicating the lagging strand synthesis.^{7,30,112} As was already mentioned, some experimental data show that expansions can occur on the leading DNA strand.⁸⁸ Clearly, however, more studies are needed to discriminate between the two DNA strands during repeat expansions. While the model in Figure 6 depicts a simple reloading of the replication fork for restart,¹¹³ this is just one opportunity. It can equally well be restarted by genetic recombination. In the latter case, repeat strand misalignment between the donor and recipient duplexes would also lead to expansions or contractions (see below).

Our model also offers an explanation for the genetic anticipation, described at the beginning of this review. When a repeat's length exceeds that of OIZ, replication stalls and restarts can happen several times. Our data for long $(GAA)_n \cdot (TTC)_n$ repeats are indeed indicative of recurrent replication stalling.⁷⁸ Both formation of stable secondary structures on the lagging

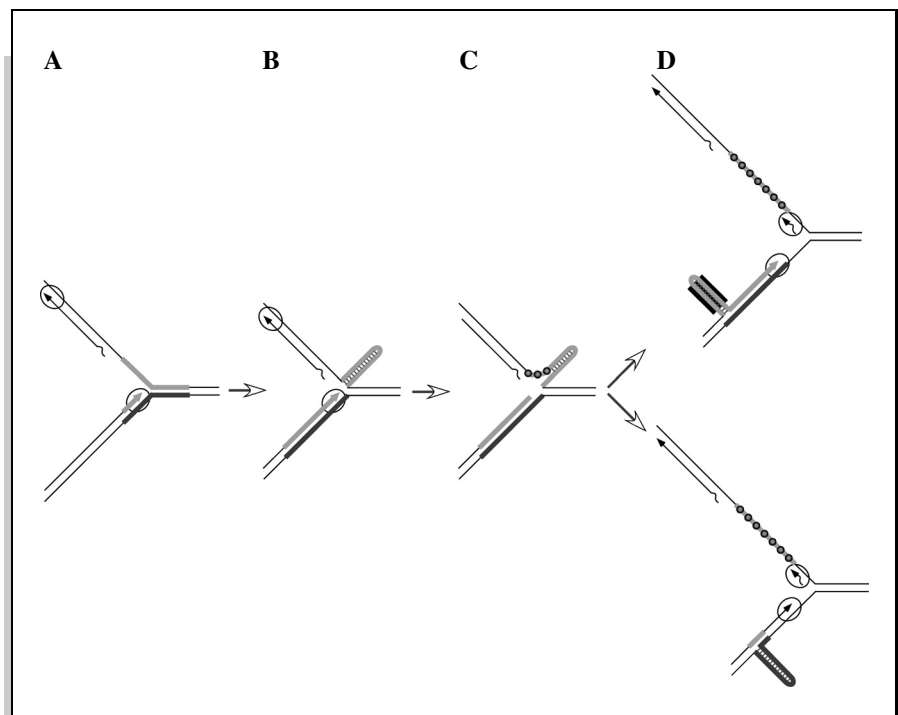


Figure 6. Replication model for repeat instability during the leading strand synthesis. **A**, entrance of the replication fork into a repeat; **B**, folding of OIZ into an unusual secondary structure; **C**, replication fork collapse; **D**, replication fork restart via polymerase reloading. Thin black line, flanking DNA; thick light gray line, structure-prone strand of a repeat; thick dark gray line, complementary strand of a repeat; wavy line, RNA primer; arrow, 3'-end of the nascent DNA chains; white oval, DNA polymerase; black circle, RPA; black rectangle, MSH2/MSH3 complex.

strand template and misalignment between the leading DNA strand and its template are fortuitous events. Thus, recurrent replication stalls and restarts within long repeats could greatly increase their probability for expansions.

Mismatch Repair and Repeat Expansions

A potential problem with our model is that it does not explain a bias toward repeat expansions in human pedigrees, since the probabilities of slippages in the newly synthesized or template strands should be similar. Recent data on the role of the mismatch repair in repeat expansions could, in our view, address this problem.

The effect of the mismatch repair on the stability of expandable repeats is a subject of substantial discussions. As was briefly discussed above, mutations in the mismatch repair in bacteria, yeast, and humans increase the length polymorphism (i.e., accumulation or deletion of small number of repeats) in various tandem repeats, including mono-, di-, tri-, and tetra-nucleotide (reviewed in Radman et al.¹¹⁴). This can be readily explained by the lack of repair of elementary replication slippages for those repeats. Expandable repeats are no different from other tandem repeats in this regard: they undergo small extensions or deletions, when maintained in mismatch repair mutants of bacteria or yeast.^{115–119} At the same time, large-scale expansions or contractions of trinucleotide repeats were not affected in yeast mismatch repair mutants during either mitosis or meiosis.^{84,93,118,120} Mutations in bacterial mismatch repair genes do not affect large-scale expansions as well, but led to rapid accumulation of large deletions.¹¹⁹ Overall, it seemed evident that unusual DNA structures formed by expandable repeats during DNA replication escaped the mismatch repair machinery in unicellular organisms.¹²⁰

In light of these conclusions, the results from studies on the role of mismatch repair in repeat expansions in transgenic mice came as a complete surprise. Several labs have demonstrated that mutational inactivation of the MSH2/MSH3 heterodimer in mice drastically decreased the frequency of repeat expansions during intergenerational transmissions, as well as in somatic cells, but had little effect on repeat contractions.^{121–124} A normal function of the MSH2/MSH3 heterodimer is to recognize and correct small insertions and deletions accumulated during DNA replication (reviewed in Modrich and Lahue¹²⁵). Thus, it was totally counterintuitive that the MSH2/MSH3 complex promoted, rather than prevented, repeat expansions. While we still do not know the mechanism for this phenomenon, some promising leads started to surface. Purified human MSH2 protein binds to DNA hairpins formed by expandable repeats.¹²⁶ Binding of the MSH2/MSH3 heterodimer to CAG-hairpins alters its conformation, abrogating the ATPase activity.¹²⁷ This led to the formulation of the so-called “hijacking” hypothesis by several labs.^{122–124} According to this hypothesis, an MSH2/MSH3 complex binds to imperfect hairpins, formed by expandable repeats, likely attracted by multiple mismatches. Its repair function is, however, lost upon this binding, possibly due to the ATPase impairment. Consequently, hairpins are stabilized, rather than repaired.

Normally, the mismatch repair machinery works in conjunction with DNA replication and is capable of distinguishing between the template and newly synthesized DNA, specifically repairing the latter (reviewed in Modrich¹²⁸). It seems plausible, therefore, that binding of the MSH2/MSH3 complex to repeat-containing hairpins on the newly synthesized, rather than template, DNA strand during the replication restart could shift the equilibrium toward repeat expansions (Fig. 6D).

It should be noted, however, that different studies of expandable repeats in transgenic mice disagreed on the exact timing for expansions. Most strikingly, a study of the HD mice concluded that expansions occurred in postmeiotic haploid germ cells in males.¹²⁴ DNA in those cells does not replicate or recombine. Thus, these data indicated that repeat expansions could occur via gap repair of nonreplicated DNA (Fig. 7). The latter model implied that the MSH2/MSH3 complex could stabilize hairpins, formed during the repair of gaps within expandable repeats. Upon repair, DNA synthesis and ligation, a strand of the repeat would extend.

The proposed timing, however, was difficult to combine with the role of the MSH2 protein in repeat expansions, observed in the same mice, because this protein is primarily expressed in dividing spermatogonia and is not detected at postmeiotic stages in mice or humans.^{129,130} Given this controversy, understanding the timing of repeat expansions during intergenerational transmissions in humans and transgenic mice became crucially important to evaluate the function of the mismatch repair (for a recent comprehensive review see Pearson¹³¹).

Early human studies pointed to either gametogenesis or early embryogenesis as time points for expansions.^{132–137} More recent and detailed studies of HD patients showed expansions of (CAG)_n runs in premeiotic male germ cells, though they may continue to occur during meiosis.^{138,139} Thus, these results significantly differed from the results on expansion timing in the HD mice, discussed above. Even more striking differences appeared to exist between the HD and DM transgenic mice. In the DM mice, expansions were limited to the premeiotic spermatogonia.¹⁴⁰ No differences were observed between spermatogonia and spermatozoa, arguing against continued expansions during postmeiotic stages. While the reasons for these striking differences remain to be elucidated, most of the data point to repeat expansions in dividing premeiotic cells during DNA replication.

Yet the gap repair model was a useful addition to the realm of expansion hypotheses. It might explain continuous repeat expansions observed in postmeiotic cells in HD patients.¹³⁹ It can also account for the progressive

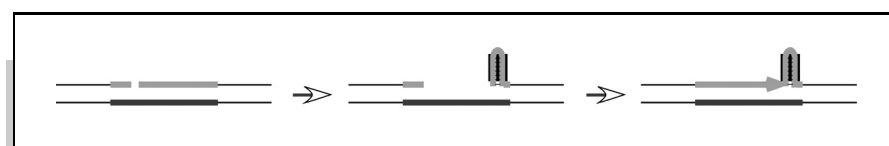


Figure 7. Gap repair model for repeat expansions (adapted from Kovtun and McMurray¹²⁴). Thin black line, flanking DNA; thick light gray line, structure-prone strand of a repeat; thick dark gray line, complementary strand of a repeat; arrow, direction of the repair synthesis; black rectangle, MSH2/MSH3 complex.

small-scale somatic expansions in aging, postmitotic tissues, such as brain and skeletal muscle, in humans^{141,142} and transgenic mice.^{143–145}

Role of Genetic Recombination

In addition to replication and repair, genetic recombination could also lead to repeat expansions. Unequal crossing-over or gene conversion between repetitive runs can a priori lead to repeat expansions and contractions (Fig. 8). Unequal crossing-over during meiosis or mitosis seems to be responsible for the expansions of polyalanine-coding repeats.¹⁴⁶ In those cases, an imperfect nature of a triplet repeat makes it easy to track various allelic changes. It appeared that practically all expanded alleles result from in-frame duplications of short $(GCX)_n$ blocks from normal alleles. Furthermore, reciprocal contractions were observed in several genes for which expansions were observed. Both groups of data are predicted by the model shown in Figure 8A.

Recombination between homologous chromosomes in meiosis is, however, an unlikely explanation for expansions of noncoding and polyglutamine-coding repeats. In humans, expansions are never accompanied by an exchange of even the closest flanking markers.¹⁴⁷ As discussed above, expansions are detected at premeiotic stages.¹³⁹ In yeast studies, the rate for expansions of $(CAG)_n \cdot (CTG)_n$ repeats was shown to increase about 5-fold, when yeast cells underwent meiosis.^{148,149} At the same time, the overall frequency of homologous recombination during meiosis in yeast increases at least three orders of magnitude.¹⁵⁰ Comparison of these two values makes meiotic recombination a minor player in repeat expansions.

For those repeats, some form of recombinational exchange, involving repeats in sister chromatids during mitosis, might be a more plausible scenario. Various triplet repeats stimulated homologous recombination in bacteria.^{151–153} $(CAG)_n \cdot (CTG)_n$ repeats were shown to undergo expansions and contractions during the double-strand break repair in yeast.^{154,155} They were also found to stimulate spontaneous unequal sister-chromatid exchange in yeast.¹⁵⁶ Finally, $(CAG)_n \cdot (CTG)_n$ repeats triggered large-scale rearrangements during recombination in cultured mammalian cells.¹⁵⁷

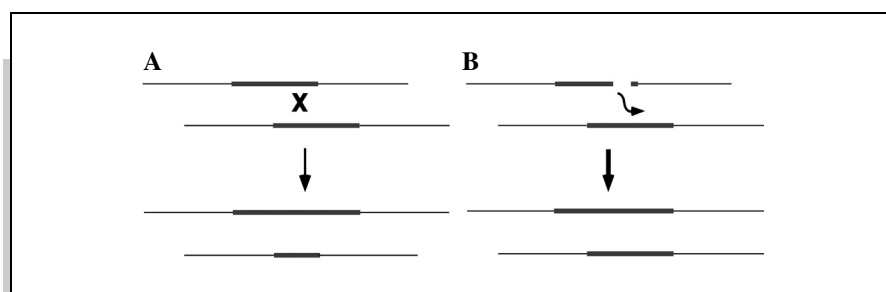


Figure 8. Recombinational models for repeat expansions. **A**, unequal crossing-over results in reciprocal appearance of an expanded and contracted repeat; **B**, gene conversion results in expansion of a repeat in one chromosome, while its counterpart on another chromosome remains intact. Repeated portions of homologous chromosomes are shown in gray.

It is generally believed that sister chromatid exchanges in mitosis are linked to the replication fork collapses and restarts (for recent reviews see McGlynn and Lloyd,¹⁵⁸ Michel et al.,¹⁵⁹ and Cox¹⁶⁰). Supporting this idea for expandable repeats, repair of double-stranded breaks induced in those repeats in *E. coli* depended on the direction of replication through them.¹⁶¹ General discussion of the fork restart via recombination is beyond the scope of this review; thus, I will briefly describe two scenarios for the recombinational restart of replication forks stalled at expandable repeats. As we have suggested above, the replication fork stalls when its leading strand runs into an expandable repeat (Fig. 9A). Collapse of the stalled replication fork commonly leads to the its reversal.^{162,163} Fork reversal in our case would lead to the formation of a four-way junction carrying a repetitive tract at the 3'-extension of the leading strand (Fig. 9B). Such single-stranded repeat can invade a parental repeat driving the replication fork restart (Fig. 9C, upper panel). Note, that invasion into the 5'-half of the parental repeat would lead to the repeat expansions upon replication restart. Alternatively, the reversed fork could be flipped backwards by the eukaryotic RuvAB homologues (Fig. 9C, lower panel). This might create a hairpin-containing structure, which is indistinguishable from that shown in Figure 6D. Replication fork restart in such structure would lead to repeat expansions.

First Events Leading to Expansions

How does the above model address the first events leading to repeat expansions? Expansions in humans can be traced to the so-called long normal

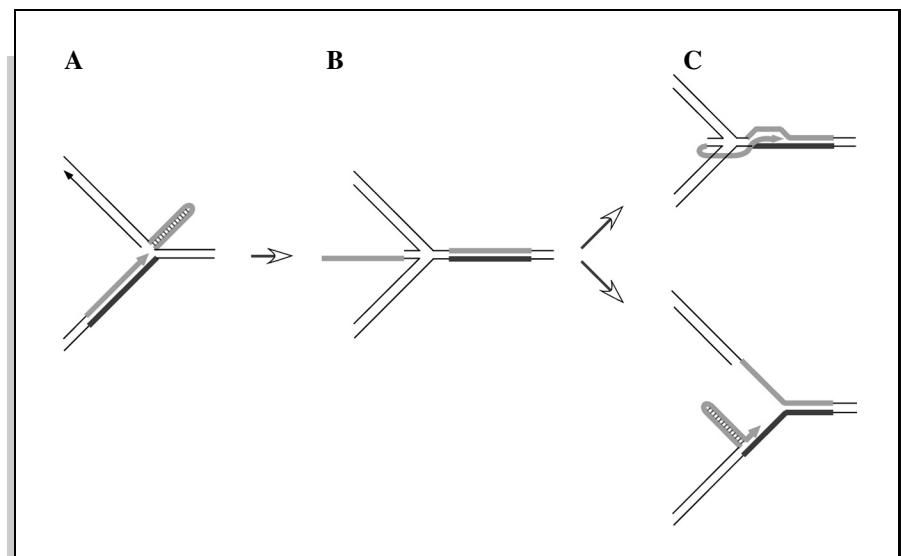


Figure 9. Recombinational restart of stalled replication forks leading to repeat expansions. **A**, replication stalling at a repetitive tract; **B**, replication fork reversal; **C**, fork restart upon strand invasion (upper panel) or by the re-enactment of the reversed fork (lower panel). Thin black line, flanking DNA; thick light gray line, structure-prone strand of a repeat; thick dark gray line, complementary strand of a repeat; arrows, 3'-ends of nascent DNA chains.

alleles, containing 30–50 repeats (reviewed in Cleary and Pearson¹⁶⁴). They are commonly interrupted by several nonrelated triplets, such as AGG triplets within the fragile X (CGG)_n runs,^{23,165,166} CAT triplets within the SCA1 (CAG)_n runs,¹⁶⁷ or GAG triplets in the Friedreich's ataxia (GAA)_n repeats.^{168,169} Familial analysis reveals that expansions in carriers and beyond occur at one end of a repeat and the expanded part lacks interrupting triplets.^{14,23} It thus seems plausible that these interruptions stabilize long normal alleles, while their loss at a repeat's end provokes expansions. Studies of repeat expansions in model systems also confirm that interruptions stabilize expandable repeats in the position-dependent manner.¹⁷⁰ Furthermore, interruptions within expandable repeats prevent the replication fork stalling.¹⁰⁸

One explanation for these phenomena could be the difference in the replication fidelity for leading and lagging DNA strands. This difference is likely due to continuous versus discontinuous modes of DNA synthesis, as well as transient single-strandedness of the lagging strand template (reviewed in Sinden et al.¹⁷¹). It has been indeed demonstrated that deletions caused by the misalignment of direct DNA repeats preferentially occurred in the lagging strand.^{172,173} Thus, the very first event leading to repeat expansions could be the loss of interrupting triplets during the lagging strand synthesis.

Several hypotheses were put forward to combine the above idea with the known position and/or orientation dependence of repeat expansions (reviewed in Cleary and Pearson¹¹¹). An ORI-SWITCH model¹⁷⁴ suggests that the first event, triggering expansions, might be inactivation (likely epigenetic) of the regular replication origin, situated on one side of the repeat, combined with an activation of a cryptic origin on the repeat's other side (Fig. 10A). Changing replication direction would reverse the point of the replication fork entry into the repeat, as well as leading and lagging strands during its replication. This could explain the loss of interruptions at one end of the repeat. This model also agrees with numerous data on orientation dependence of repeat expansions in model systems, which were described above. An ORI-SHIFT model¹⁷⁴ proposes that positioning a repeat at various parts of the OIZ could drastically affect expansion dynamics (Fig. 10B). This model is supported by the data on repeat stabilities in mammalian episomes, where the very same repeat in the same orientation prefers to expand or delete depending on its distance from the origin.⁸⁶ It is also indirectly supported by the observations that expanded alleles in myotonic dystrophy often contain an insertion of an Alu element nearby.¹⁷⁵ The most recent model, called FORK-SHIFT (Fig. 10C), proposes that repeats' propensities to expand are grounded in their intrinsic ability to position OIZ on the lagging strand template during DNA replication.¹¹¹

Neither of these models are so far corroborated by the preliminary mapping of replication origins in the vicinity of expandable repeats in humans.¹⁷⁶ Determining the direction and the mode of the replication fork progression through such repeats in human chromosomes at different stages of development could become a major direction in future studies of expansions.

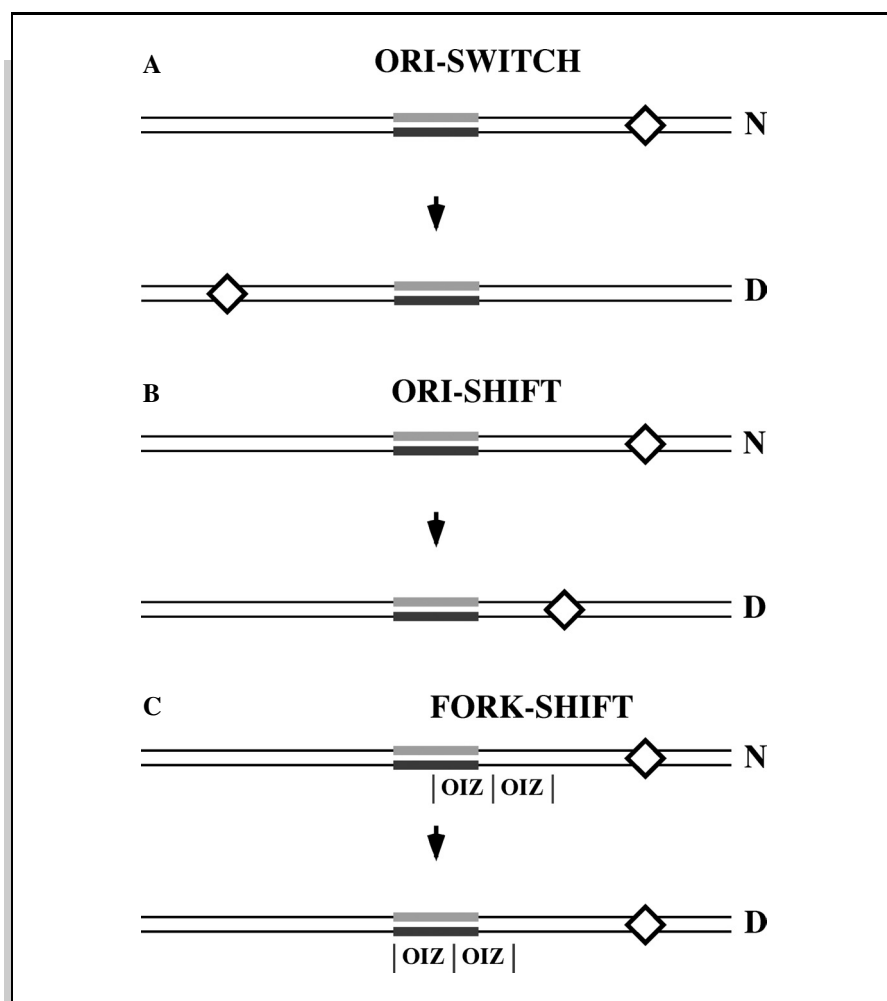


Figure 10. Replication models for the initial repeat expansions (see text for details). Thin black line, flanking DNA; thick light gray line, structure-prone strand of a repeat; thick dark gray line, complementary strand of a repeat; white diamond, replication origin.

A different idea on explaining the disappearance of repeat interruptions emerged from studies in yeast. Interruptions within expandable repeats stabilize them in all organisms, including yeast.¹⁷⁰ It appeared that this stabilization is the result of the action of the mismatch repair system.¹⁷⁷ Mutations within the mismatch repair genes led to the disappearance of interruption and ensuing expansions. To explain these results, the authors proposed the so-called coexcision model (Fig. 11). According to it, accidental slippage between the newly synthesized and template DNA strands of a repeat results in the formation of a slipped structure, similar to that shown in Figure 3. If the repeat contained interruptions, they could go out of register in both the hairpin and the duplex part of this structure (i.e., create mismatches) (Fig. 11B). Excision of the hairpin and repair of mismatches in the duplex part would restore the original composition of the repeat, maintaining interruptions and preventing expansions (Fig. 11C, left panel). Failure of the mismatch repair system to correct mismatches

in the duplex part would produce extensions at the 3'-end of the repeat, lacking interruptions (Fig. 11C, right panel).

If this hypothesis is correct, mismatch repair might play a dual role in the process of repeat expansions. At early stages, mismatch repair activity seems to be important to prevent expansions of long normal alleles by sustaining repeat interruptions. Occasional breakdowns in the mismatch repair can lead to the loss of interruptions and consequent expansions. At later stages, in contrast, mismatch repair proteins can facilitate progressive expansions, by stabilizing repetitive hairpins on the newly synthesized DNA strands during the replication fork restart.

Repeat Replication and Chromosomal Fragility

Repeat expansions seem to be associated with chromosomal fragility. Molecular mechanisms of the chromosomal fragility and the role of DNA replication were extensively discussed in several recent reviews,^{112,178,179} including the Freudenreich review in this issue. Without going into details, I would like to emphasize that fragility is usually evident for fairly long repeats. It seems plausible to speculate, therefore, that repeat expansions and chromosomal fragility could reflect the two sides of the same coin. When repeats are relatively short, replication stalls and restarts are infrequent, occasionally leading to expansions. These events become more

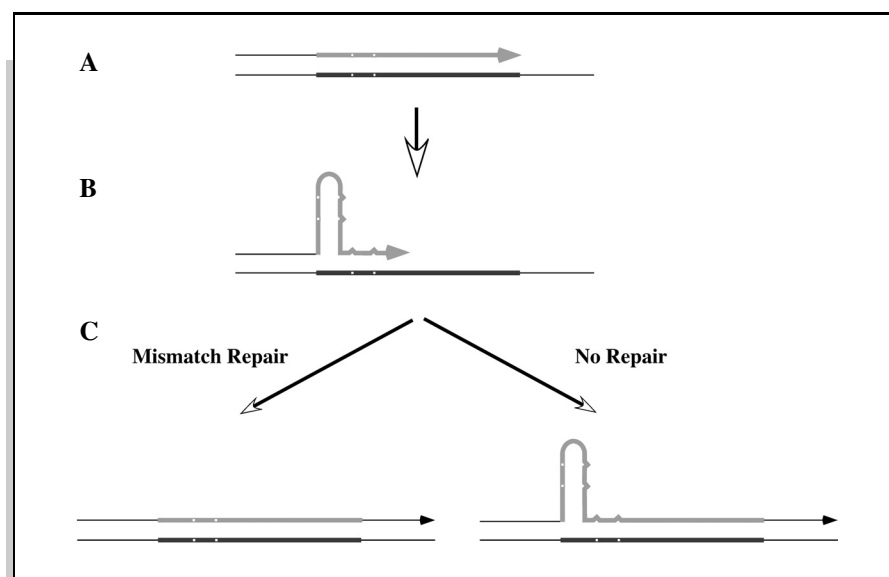


Figure 11. A coexcision model for the stabilization of long normal repeats (adapted from Rolfsmeier et al.¹⁷⁷). **A**, DNA synthesis through a long normal repeat, containing interruptions; **B**, misalignment of nascent and template DNA strands create mismatches in both the hairpin and duplex part of the slipped structure; **C**, coexcision of the hairpin and mismatches in the duplex followed by DNA synthesis restores the original repeat (left panel); breakdown of the mismatch repair leads to the expansion of the 3'-end of the repeat (right panel). Thin black line, flanking DNA; thick light gray line, structure-prone strand of a repeat; thick dark gray line, complementary strand of a repeat; small white squares, repeat interruptions; arrows, 3'-ends of nascent DNA strands.

frequent with an increase in the lengths of repeats. For the longest repeats, profuse replication stalling makes timely replication completion impossible, resulting in the chromosomal fragility in mitosis.

Long (CGG)_n repeats were indeed shown to replicate exceptionally late in the cell cycle.^{180–182} This could be explained, however, by the replication delay in the heavily heterochromatinized FRAXA region in the X chromosome, rather than the replication fork stalling within the expanded (CGG)_n run per se. A better support for the replication-fragility connection came recently from elegant yeast studies, showing that repeat-mediated chromosomal fragility increased upon inactivation of the replication checkpoint proteins, including Mec1 (ATR), Mrc1 (Claspin), or Rad53 (Chk 2).^{183,184}

Acknowledgments

I thank Mel DePamphilis for encouraging me to write this review and my wife Katya Mirkin for her support and constructive criticism. This work was supported by grant GM60987 from the NIH to S.M.M.

References

1. Lenzmeier, B.A., Freudenreich, C.H. *Cytogenet. Genome Res.* **2003**, 100.
2. Jin, P., Warren, S.T. *Hum. Mol. Genet.* **2000**, 9, 901.
3. Parniewski, P., Staczek, P. *Adv. Exp. Med. Biol.* **2002**, 516, 1.
4. Ranum, L.P., Day, J.W. *Am. J. Hum. Genet.* **2004**, 74, 793.
5. Brown, L.Y., Brown, S.A. *Trends Genet.* **2004**, 20, 51.
6. Cummings, C.J., Zoghbi, H.Y. *Annu. Rev. Genom. Hum. Genet.* **2000**, 1, 281.
7. Bowater, R.P., Wells, R.D. *Progr. Nucleic Acid Res. Mol. Biol.* **2001**, 66, 159.
8. Siyanova, E.Y., Mirkin, S.M. *Mol. Biol. (Mosc.)* **2001**, 35, 168.
9. Verkerk, A.J.M.H., Pieretti, M., Sutcliffe, J.S., et al. *Cell* **1991**, 65, 905.
10. Kremer, E.J., Pritchard, M., Lynch, M., et al. *Science* **1991**, 252, 1711.
11. The Huntington's Disease Collaborative Research Group. *Cell* **1993**, 72, 971.
12. La Spada, A.R., Wilson, E.M., Lubahn, D.B., Harding, A.E., Fischbeck, K.H. *Nature* **1994**, 352, 77.
13. Orr, H.T., Chung, M.Y., Banfi, S., et al. *Nature Genet.* **1993**, 4, 221.
14. Chung, M.Y., Ranum, L.P., Duvick, L.A., Servadio, A., Zoghbi, H.Y., Orr, H.T. *Nature Genet.* **1993**, 5, 254.
15. Koide, R., Ikeuchi, T., Onodera, O., et al. *Nature Genet.* **1994**, 6, 9.
16. Mahadevan, M., Tsilfidis, C., Sabourin, L., et al. *Science* **1992**, 255, 1253.
17. Brook, J.D., McCurrach, M.E., Harley, H.G., et al. *Cell* **1992**, 68, 799.
18. Campuzano, V., Montermini, L., Molto, M.D., et al. *Science* **1996**, 271, 1423.
19. Liquori, C.L., Ricker, K., Moseley, M.L., et al. *Science* **2001**, 293, 864.
20. Matsuura, T., Yamagata, T., Burgess, D.L., et al. *Nature Genet.* **2000**, 26, 191.
21. Lalioti, M.D., Scott, H.S., Buresi, C., et al. *Nature* **1997**, 386, 847.
22. Fishel, R., Kolodner, R.D. *Curr. Opin. Genet. Dev.* **1995**, 5, 382.
23. Kunst, C.B., Warren, S.T. *Cell* **1994**, 77, 853.
24. Sherman, S.L., Jacobs, P.A., Morton, N.E., et al. *Hum. Genet.* **1985**, 69, 289.
25. Sherman, S.L., Morton, N.E., Jacobs, P.A., Turner, G. *Ann. Hum. Genet.* **1984**, 48, 21.
26. Howeler, C.J., Busch, H.F., Geraedts, J.P., Niermeier, M.F., Staal, A. *Brain* **1989**, 112, 779.

27. Brooks, B.P., Fischbeck, K.H. *Trends Neurosci.* **1995**, *18*, 459.
28. Goldberg, Y.P., Telenius, H., Hayden, M.R. *Curr. Opin. Neurol.* **1994**, *7*, 325.
29. Pandolfo, M. *Neuromuscul. Disord.* **1998**, *8*, 409.
30. McMurray, C.T. *Chromosoma* **1995**, *104*, 2.
31. Lahue, R.S., Slater, D.L. *Front Biosci.* **2003**, *8*, 663.
32. Amiel, J., Trochet, D., Clement-Ziza, M., Munnich, A., Lyonnet, S. *Hum. Mol. Genet.* **2004**, *13*, R235.
33. Brown, L.Y., Brown, S.A. *Trends Genet.* **2004**, *20*, 51.
34. Kohwi, Y., Wang, H., Kohwi-Shigematsu, T. *Nucleic Acids Res.* **1993**, *21*, 5651.
35. Pearson, C.E., Sinden, R.R. *Curr. Opin. Struct. Biol.* **1998**, *8*, 321.
36. Pearson, C.E., Sinden, R.R. *Biochemistry* **1996**, *35*, 5041.
37. Mitchell, J.E., Newbury, S.F., McClellan, J.A. *Nucleic Acids Res.* **1995**, *23*, 1876.
38. Petruska, J., Arnheim, N., Goodman, M.F. *Nucleic Acids Res.* **1996**, *24*, 1992.
39. Yu, A., Dill, J., Wirth, S.S., Huang, G., Lee, V.H., Haworth, I.S., Mitas, M. *Nucleic Acids Res.* **1995**, *23*, 2706.
40. Yu, A., Barron, M.D., Romero, R.M., et al. *Biochemistry* **1997**, *36*, 3687.
41. Zheng, M., Huang, X., Smith, G.K., Yang, X., Gao, X. *J. Mol. Biol.* **1996**, *264*, 323.
42. Gacy, A.M., Goellner, G., Juranic, N., Macura, S., McMurray, C.T. *Cell* **1995**, *81*, 533.
43. Chen, X., Mariappan, S.V., Catasti, P., et al. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 5199.
44. Mariappan, S.V., Silks, L.A., III, Chen, X., et al. *J. Biomol. Struct. Dyn.* **1998**, *15*, 723.
45. Heidenfelder, B.L., Makhov, A.M., Topal, M.D. *J. Biol. Chem.* **2003**, *278*, 2425.
46. Usdin, K. *Nucleic Acids Res.* **1998**, *26*, 4078.
47. Mirkin, S.M., Frank-Kamenetskii, M.D. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 541.
48. Gacy, A.M., Goellner, G.M., Spiro, C., et al. *Mol. Cell* **1998**, *1*, 583.
49. Sakamoto, N., Chastain, P.D., Parniewski, P., et al. *Mol. Cell* **1999**, *3*, 465.
50. Vetcher, A.A., Napierala, M., Iyer, R.R., Chastain, P.D., Griffith, J.D., Wells, R.D. *J. Biol. Chem.* **2002**, *277*, 39217.
51. Potaman, V.N., Bissler, J.J., Hashem, V.I., et al. *J. Mol. Biol.* **2003**, *326*, 1095.
52. Pearson, C.E., Eichler, E.E., Lorenzetti, D., et al. *Biochemistry* **1998**, *37*, 2701.
53. McMurray, C.T. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 1823.
54. Bedinger, P., Munn, M., Alberts, B.M. *J. Biol. Chem.* **1989**, *264*, 16880.
55. Kaguni, L.S., Clayton, D.A. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 983.
56. Weaver, D.T., DePamphilis, M.L. *J. Biol. Chem.* **1982**, *257*, 2075.
57. Huang, C.C., Hearst, J.E. *Anal. Biochem.* **1980**, *103*, 127.
58. Chalberg, M.D., Englund, P.T. *J. Biol. Chem.* **1979**, *254*, 7820.
59. Sherman, L.A., Gefter, M.L. *J. Mol. Biol.* **1976**, *103*, 61.
60. Samadashwily, G.M., Dayn, A., Mirkin, S.M. *EMBO J.* **1993**, *12*, 4975.
61. Dayn, A., Samadashwily, G.M., Mirkin, S.M. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 11406.
62. Krasilnikov, A.S., Panyutin, I.G., Samadashwily, G.M., Cox, R., Lazurkin, Y.S., Mirkin, S.M. *Nucleic Acids Res.* **1997**, *25*, 1339.
63. Baran, N., Lapidot, A., Manor, H. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 507.
64. Woodford, K.J., Howell, R.M., Usdin, K. *J. Biol. Chem.* **1994**, *269*, 27029.
65. Weitzmann, M.N., Woodford, K.J., Usdin, K. *J. Biol. Chem.* **1997**, *272*, 9517.
66. Usdin, K., Woodford, K.J. *Nucleic Acids Res.* **1995**, *23*, 4202.
67. Kohwi, Y., Panchenko, Y. *Genes Dev.* **1993**, *7*, 1766.
68. Rooney, S.M., Moore, P.D. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 2141.
69. Wahls, W.P., Wallace, L.J., Moore, P.D. *Mol. Cell Biol.* **1990**, *10*, 785.
70. Weinreb, A., Collier, D.A., Birshtein, B.K., Wells, R.D. *J. Biol. Chem.* **1990**, *265*, 1352.

71. Wahls, W.P., Wallace, L., Moore, P.D. *Cell* **1990**, *60*, 95.
72. Sinden, R.R. *DNA Structure and Function*. Academic Press: San Diego, 1994.
73. Paiva, A.M., Sheardy, R.D. *Biochemistry* **2004**, *43*, 14218.
74. DePamphilis, M.L., Wassarman, P.M. *Annu. Rev. Biochem.* **1980**, *49*, 627.
75. DePamphilis, M.L. *Chemtracts—Biochem. Mol. Biol.* **2002**, *15*, 313.
76. Wold, M.S. *Annu. Rev. Biochem.* **1997**, *66*, 61.
77. Frick, D.N., Richardson, C.C. *Annu. Rev. Biochem.* **2001**, *70*, 39.
78. Krasilnikova, M.M., Mirkin, S.M. *Mol. Cell Biol.* **2004**, *24*, 2286.
79. Kang, S., Jaworski, A., Ohshima, K., Wells, R.D. *Nature Genet.* **1995**, *10*, 213.
80. Kang, S., Ohshima, K., Shimizu, M., Amirhaeri, S., Wells, R.D. *J. Biol. Chem.* **1995**, *270*, 27014.
81. Ohshima, K., Wells, R.D. *J. Biol. Chem.* **1997**, *272*, 16798.
82. Shimizu, M., Gellibolian, R., Oostra, B.A., Wells, R.D. *J. Mol. Biol.* **1996**, *258*, 614.
83. Freudenreich, C.H., Stavenhagen, J.B., Zakian, V.A. *Mol. Cell Biol.* **1997**, *17*, 2090.
84. Miret, J.J., Pessoa-Brandao, L., Lahue, R.S. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12438.
85. Balakumaran, B.S., Freudenreich, C.H., Zakian, V.A. *Hum. Mol. Genet.* **2000**, *9*, 93.
86. Cleary, J.D., Nichol, K., Wang, Y.H., Pearson, C.E. *Nature Genet.* **2002**, *31*, 37.
87. Panigrahi, G.B., Cleary, J.D., Pearson, C.E. *J. Biol. Chem.* **2002**, *277*, 13926.
88. Iyer, R.R., Wells, R.D. *J. Biol. Chem.* **1999**, *274*, 3865.
89. Freudenreich, C.H., Kantrow, S.M., Zakian, V.A. *Science* **1998**, *279*, 853.
90. Schweitzer, J.K., Livingston, D.M. *Hum. Mol. Genet.* **1998**, *7*, 69.
91. Schweitzer, J.K., Livingston, D.M. *Genetics* **1999**, *152*, 953.
92. Spiro, C., Pelletier, R., Rolfmeier, M.L., et al. *Mol. Cell* **1999**, *4*, 1079.
93. White, P.J., Borts, R.H., Hirst, M.C. *Mol. Cell Biol.* **1999**, *19*, 5675.
94. Ireland, M.J., Reinke, S.S., Livingston, D.M. *Genetics* **2000**, *155*, 1657.
95. Iyer, R.R., Pluciennik, A., Rosche, W.A., Sinden, R.R., Wells, R.D. *J. Biol. Chem.* **2000**, *275*, 2174.
96. Liu, Y., Kao, H.I., Bambara, R.A. *Annu. Rev. Biochem.* **2004**, *73*, 589.
97. Gordenin, D.A., Kunkel, T.A., Resnick, M.A. *Nature Genet.* **1997**, *16*, 116.
98. Lee, S., Park, M.S. *Exp. Mol. Med.* **2002**, *34*, 313.
99. Veeraraghavan, J., Rossi, M.L., Bambara, R.A. *J. Biol. Chem.* **2003**, *278*, 42854.
100. Henricksen, L.A., Tom, S., Liu, Y., Bambara, R.A. *J. Biol. Chem.* **2000**, *275*, 16420.
101. Kokoska, R.J., Stefanovic, L., Tran, H.T., Resnick, M.A., Gordenin, D.A., Petes, T.D. *Mol. Cell Biol.* **1998**, *18*, 2779.
102. Parenteau, J., Wellinger, R.J. *Mol. Cell Biol.* **1999**, *19*, 4143.
103. Bhattacharyya, S., Lahue, R.S. *Mol. Cell Biol.* **2004**, *24*, 7324.
104. Pelletier, R., Krasilnikova, M.M., Samadashwily, G.M., Lahue, R.S., Mirkin, S.M. *Mol. Cell Biol.* **2003**, *23*, 1349.
105. Beckwith, W.H., Sun, Q., Bosso, R., Gerik, K.J., Burgers, P.M.J., McAlear, M.A. *Biochemistry* **1998**, *37*, 3711.
106. Mossi, R., Hubscher, V. *Eur. J. Biochem.* **1998**, *254*, 209.
107. Barbour, L., Xiao, W. *Mutat. Res.* **2003**, *532(1–2)*, 137.
108. Samadashwily, G.M., Raca, G., Mirkin, S.M. *Nature Genet.* **1997**, *17*, 298.
109. Brewer, B.J., Fangman, W.L. *Cell* **1987**, *51*, 463.
110. Huberman, J.A., Spotila, L.D., Nawotka, K.A., El-Assouli, S.M., Davis, L.R. *Cell* **1987**, *51*, 473.
111. Cleary, J.D., Pearson, C.E. *Trends Genet.* **2005**, *272*.
112. Usdin, K., Grabczyk, E. *Cell Mol. Life Sci.* **2000**, *57*, 914.
113. Mariani, K.J. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **2004**, *359*, 71.
114. Radman, M., Matic, I., Halliday, J.A., Taddei, F. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **1995**, *347*, 97.
115. Hashem, V.I., Rosche, W.A., Sinden, R.R. *Mutat. Res.* **2002**, *502*, 25.

116. Wells, R.D., Parniewski, P., Pluciennik, A., Bacolla, A., Gellibolian, R., Jaworski, A. *J. Biol. Chem.* **1998**, 273, 19 532.
117. Parniewski, P., Jaworski, A., Wells, R.D., Bowater, R.P. *J. Mol. Biol.* **2000**, 299, 865.
118. Schweitzer, J.K., Livingston, D.M. *Hum. Mol. Genet.* **1997**, 6, 349.
119. Schmidt, K.H., Abbott, C.M., Leach, D.R. *Mol. Microbiol.* **2000**, 35, 463.
120. Moore, H., Greenwell, P.W., Liu, C.P., Arnheim, N., Petes, T.D. *Proc. Natl. Acad. Sci. USA* **1999**, 96, 1504.
121. Savouret, C., Brisson, E., Essers, J., et al. *EMBO J.* **2003**, 22, 2264.
122. Manley, K., Shirley, T.L., Flaherty, L. *Nature Genet.* **1999**, 23, 471.
123. van den Broek, W.J., Nelen, M.R., Wansink, et al. *Hum. Mol. Genet.* **2002**, 11, 191.
124. Kovtun, I.V., McMurray, C.T. *Nature Genet.* **2001**, 27, 407.
125. Modrich, P., Lahue, R. *Annu. Rev. Biochem.* **1996**, 65, 101.
126. Pearson, C.E., Ewel, A., Acharya, S., Fishel, R.A., Sinden, R.R. *Hum. Mol. Genet.* **1997**, 6, 1117.
127. Owen, B., Badger, J.D., II, Yang, Z., et al. *Nature Struct. Mol. Biol.* **2005**, in press.
128. Modrich, P. *J. Biol. Chem.* **1997**, 272, 24 727.
129. Richardson, L.L., Pedigo, C., Ann Handel, M. *Biol. Reprod.* **2000**, 62, 789.
130. Bocker, T., Barusevicius, A., Snowden, T., et al. *Cancer Res.* **1999**, 59, 816.
131. Pearson, C.E. *Trends Mol. Med.* **2003**, 9, 490.
132. Zhang, L., Leeflang, E.P., Yu, J., Arnheim, N. *Nature Genet.* **1994**, 7, 531.
133. Telenius, H., Kremer, B., Goldberg, Y.P., et al. *Nature Genet.* **1994**, 6, 409.
134. Duyao, M., Ambrose, C., Myers, R., et al. *Nature Genet.* **1993**, 4, 387.
135. Wohrle, D., Hennig, I., Vogel, W., Steinbach, P. *Nature Genet.* **1993**, 4, 140.
136. Jansen, G., Willems, P., Coerwinkel, M., et al. *Am. J. Hum. Genet.* **1994**, 54, 575.
137. MacDonald, M.E., Barnes, G., Srinidhi, J., et al. *J. Med. Genet.* **1993**, 30, 982.
138. Leeflang, E.P., Tavare, S., Marjoram, P., et al. *Hum. Mol. Genet.* **1999**, 8, 173.
139. Yoon, S.R., Dubeau, L., de Young, M., Wexler, N.S., Arnheim, N. *Proc. Natl. Acad. Sci. USA* **2003**, 100, 8834.
140. Savouret, C., Garcia-Cordier, C., Megret, J., te Riele, H., Junien, C., Gourdon, G. *Mol. Cell Biol.* **2004**, 24, 629.
141. Anvret, M., Ahlberg, G., Grandell, U., Hedberg, B., Johnson, K., Edstrom, L. *Hum. Mol. Genet.* **1993**, 2, 1397.
142. Thornton, C.A., Johnson, K., Moxley, R.T. *Ann. Neuro* **1994**, 35, 104.
143. Kennedy, L., Evans, E., Chen, C.M., et al. *Hum. Mol. Genet.* **2003**, 12, 3359.
144. Lia, A.S., Seznec, H., Hofmann-Radvanyi, H., et al. *Hum. Mol. Genet.* **1998**, 7, 1285.
145. Fortune, M.T., Vassilopoulos, C., Coolbaugh, M.I., Siciliano, M.J., Monckton, D.G. *Hum. Mol. Genet.* **2000**, 9, 439.
146. Warren, S.T. *Science* **1997**, 275, 408.
147. Richards, R.I., Holman, K., Friend, K., et al. *Nature Genet.* **1992**, 1, 257.
148. Schweitzer, J.K., Reinke, S.S., Livingston, D.M. *Genetics* **2001**, 159, 1861.
149. Jankowski, C., Nasar, F., Nag, D.K. *Proc. Natl. Acad. Sci. USA* **2000**, 97, 2134.
150. Paques, F., Haber, J.E. *Microb. Mol. Biol. Rev.* **1999**, 63, 349.
151. Napierala, M., Dere, R., Vetcher, A., Wells, R.D. *J. Biol. Chem.* **2004**, 279, 6444.
152. Jakupciak, J.P., Wells, R.D. *J. Biol. Chem.* **2000**, 275, 40 003.
153. Jakupciak, J.P., Wells, R.D. *J. Biol. Chem.* **1999**, 274, 23 468.
154. Richard, G.-F., Goellner, G.M., McMurray, C.T., Haber, J.E. *EMBO J.* **2000**, 19, 2381.
155. Richard, G.F., Cyncynatus, C., Dujon, B. *J. Mol. Biol.* **2003**, 326, 769.
156. Nag, D.K., Suri, M., Stenson, E.K. *Nucleic Acids Res.* **2004**, 32, 5677.
157. Meservy, J.L., Sargent, R.G., Iyer, R.R., et al. *Mol. Cell Biol.* **2003**, 23, 3152.
158. McGlynn, P., Lloyd, R.G. *Nature Rev. Mol. Cell Biol.* **2002**, 3, 859.
159. Michel, B., Grompone, G., Flores, M.J., Bidnenko, V. *Proc. Natl. Acad. Sci. USA* **2004**, 101, 12 783.

160. Cox, M.M. *Annu. Rev. Genet.* **2001**, 35, 53.
161. Hebert, M.L., Spitz, L.A., Wells, R.D. *J. Mol. Biol.* **2004**, 336, 655.
162. Higgins, N.P., Kato, K., Strauss, B. *J. Mol. Biol.* **1976**, 101, 417.
163. Postow, L., Ullsperger, C., Keller, R.W., Bustamante, C., Vologodskii, A.V., Cozzarelli, N.R. *J. Biol. Chem.* **2001**, 276, 2790.
164. Cleary, J.D., Pearson, C.E. *Cytogenet. Genome Res.* **2003**, 100, 25.
165. Reiss, A.L., Kazazian, H., Jr., Krebs, C.M., et al. *Hum. Mol. Genet.* **1994**, 3, 393.
166. Snow, K., Tester, D.J., Kruckeberg, K.E., Schaid, D.J., Thibodeau, S.N. *Hum. Mol. Genet.* **1994**, 3, 1543.
167. Jodice, C., Malaspina, P., Persichetti, F., et al. *Am. J. Hum. Genet.* **1994**, 54, 959.
168. Labuda, M., Labuda, D., Miranda, C., et al. *Neurology* **2000**, 54, 2322.
169. Cossee, M., Schmitt, M., Campuzano, V., et al. *Proc. Natl. Acad. Sci. USA* **1997**, 94, 7452.
170. Rolfsmeier, M.L., Lahue, R.S. *Mol. Cell Biol.* **2000**, 20, 173.
171. Sinden, R.R., Hashem, V.I., Rosche, W.A. *Ann. NY Acad. Sci.* **1999**, 870, 173.
172. Trinh, T.Q., Sinden, R.R. *Nature* **1991**, 352, 544.
173. Rosche, W.A., Trinh, T.Q., Sinden, R.R. *J. Bacteriol.* **1995**, 177, 4385.
174. Mirkin, S.M., Smirnova, E.V. *Nature Genet.* **2002**, 31, 5.
175. Neville, C.E., Mahadevan, M.S., Barcelo, J.M., Korneluk, R.G. *Hum. Mol. Genet.* **1994**, 3, 45.
176. Nenguke, T., Aladjem, M.I., Gusella, J.F., Wexler, N.S. *Hum. Mol. Genet.* **2003**, 12, 1021.
177. Rolfsmeier, M.L., Dixon, M.J., Lahue, R.S. *Mol. Cell* **2000**, 6, 1501.
178. Handt, O., Sutherland, G.R., Richards, R.I. *Mol. Genet. Metab.* **2000**, 70, 99.
179. Matsuyama, A., Croce, C.M., Huebner, K. *Eur. J. Histochem.* **2004**, 48, 29.
180. Hansen, R.S., Canfield, T.K., Lamb, M.M., Gartler, S.M., Laird, C.D. *Cell* **1993**, 73, 1403.
181. Hansen, R.S., Canfield, T.K., Fjeld, A.D., Mumm, S., Laird, C.D., Gartler, S.M. *Proc. Natl. Acad. Sci. USA* **1997**, 94, 4587.
182. Hansen, R.S., Canfield, T.K., Lamb, M.M., Gartler, S.M., Laird, C.D. *Cell* **1993**, 73, 1403.
183. Lahiri, M., Gustafson, T.L., Majors, E.R., Freudenreich, C.H. *Mol. Cell* **2004**, 15, 287.
184. Freudenreich, C.H., Lahiri, M. *Cell Cycle* **2004**, 3, 1370.