

UDC 575.2:616.056.7

Expansion of Trinucleotide Repeats

E. Yu. Siyanova[†] and S. M. Mirkin

Department of Molecular Genetics, University of Illinois at Chicago, Chicago, IL 60607, USA

Received October 2, 2000

Abstract—This review describes a novel type of genome instability, expansion of trinucleotide repeats. Originally discovered in 1991 upon cloning the gene responsible for the fragile X syndrome, it has proved to be a general phenomenon responsible for a growing number of human neurological disorders. Besides apparent medical importance, the discovery of trinucleotide repeat expansion unraveled a fundamental problem of human genetics: a non-Mendelian type of inheritance called anticipation. Understanding the mechanisms of repeat expansion and the molecular pathways leading from these expansions to human diseases became a formidable task for modern biology and one of its spectacular achievements. Here we discuss the major breakthroughs in this field made during the last decade, with an emphasis on molecular models of repeat expansion.

Key words: trinucleotide repeat, expansion, replication, recombination, gene expression, genome instability, genome polymorphism, non-Mendelian inheritance, genetic anticipation

OVERVIEW

The striking discovery of an ever-expanding world of trinucleotide repeats originally came from the studies of an inherited human disease called fragile X syndrome. This is the most common form of inherited mental retardation in humans, with incidence of approximately 1 in 5000 [1]. Clinical symptoms of this disease, which are usually not evident before mid-childhood, include learning difficulties, low IQ, macroorchidism, and characteristic elongated face with prominent ears [2]. The name for the syndrome came from cytogenetic observations of a specific chromosomal constriction Xq27.3 in a fraction of metaphases in lymphocytes of affected individuals [3]. This fragility is folate-sensitive, i.e., induced in cell culture in media depleted of folic acid, thymidine, or containing the dihydrofolate reductase inhibitor methotrexate [3]. The X chromosome fragility together with the maternal transmission of the disease [4] strongly indicated that the syndrome is caused by a mutation in the X chromosome. At the same time, the inheritance pattern of this disease was highly unusual in that the probability, onset, and severity of the disease increased as it passed through generations [5, 6]. This phenomenon, called anticipation, was inconsistent with Mendelian inheritance, and pointed to a dynamic nature of a mutation causing fragile X syndrome. This controversy was resolved upon positional cloning of the fragile X site in 1991 [7, 8]. It appeared that, in the vast majority of the cases, the disorder was caused by an expansion of the (CGG)_n repeat within the 5'-untranslated region (5'-UTR) of a gene called *FMR1*. Normal individuals have 5 to 50 repeats that are stably

transmitted through generations. Premutation carriers with 50–200 repeat copies are not affected clinically, but can, with certain probability, transmit expanded repeats to their progeny. In the following generations, expansions become more frequent so that each subsequent expansion has a higher probability than the previous one. Consequently, in individuals with the fragile X syndrome, the number of repeats easily exceeds 200 and is often as high as several thousand. Thus, the anticipation in the disease inheritance is due to the expansion of the (CGG)_n repeat.

The anticipation trait in the inheritance is not limited to the fragile X syndrome. In fact, the first observations of anticipation for another neurological disease, myotonic dystrophy, date back to 1918 [9]. It is also true for several neurodegenerative diseases including spinobulbar muscular atrophy [10], Huntington disease [11], Friedreich's ataxia [12], etc. Moreover, it was even suggested that anticipation, at least to some extent, characterize the inheritance of such complex medical genetic traits as schizophrenia [13, 14] and autism [15].

It is now totally clear that in the vast majority of cases, anticipation in disease inheritance is due to the expansion of simple DNA repeats. The table lists current examples of such diseases together with their important genetic characteristics. One can clearly see that all these cases are similar in that they are caused by an expansion of a simple DNA repeat beyond a threshold length corresponding to 60–150 bp. In most of the cases, expandable repeats are simple trinucleotide blocks. To date, three types of trinucleotide repeats have been shown to be expansion-prone: (CGG)_n·(CCG)_n, (CTG)_n·(CAG)_n, and (GAA)_n·(TTC)_n. For those runs, the threshold expansion length is

[†] Deceased.

equivalent to 20–50 repeats. It is already obvious, however, that the expansion phenomenon is not limited to trinucleotide repeats. One example is progressive myoclonus epilepsy type 1 caused by an expansion of a GC-rich dodecamer repeat in the promoter region of the cystatin B gene [16].

Beyond repeat expansion, the genetics of these diseases seem to have little in common. Genes carrying expandable repeats can be situated on either sex chromosomes or autosomes. The pattern of inheritance can be both dominant and recessive. The gender bias can be maternal, paternal, or nonexistent. Target genes have no evident functional similarities. Repeats are located in various parts of those genes, and both the loss and gain of function can result from their expansion.

Studies of triplet repeat diseases concentrated on three major questions: (i) effects of expanded trinucleotide repeats on gene expression, (ii) molecular pathogenesis of repeat-caused neurological disorders, and (iii) mechanisms of repeats expansion. *A priori* these questions might have different answers for different repeats. To give just one example, while expanded trinucleotide repeats ultimately offset carrier genes, this transpires at transcription, RNA processing, translation, or posttranslational levels. This review primarily concentrates on the mechanisms of repeat expansion. The effects of trinucleotide repeats on gene expression and consequent events leading to a disease will only briefly be described below with the best studied cases.

FROM REPEAT EXPANSION TO DISEASE

Fragile X Syndrome

The mechanism of FMR1 inactivation following (CGG)_n expansion is reasonably well understood [17]. In normal- and premutation-sized alleles, the FMR1 gene is efficiently expressed, resulting in the production of the protein product, FMRP. Interestingly, moderate expansion of CGG repeats leads to an increase in the amount of the FMR1 mRNA, but its translation efficiency is low, resulting in the regular expression level (G. Raca and E. Siyanova, unpublished results). In full mutation alleles ($n > 200$), the whole promoter region of the FMR1 gene, including, but not restricted to, the (CGG)_n repeat, is hypermethylated and transcription is shut down. In fact, methylation induced by the repeat expansion spreads for a substantial distance from the repeat resulting in the heterochromatinization of more than 1 Mb of adjacent DNA [18, 19]. Among other effects, hypermethylation leads to a very late replication of the whole FRAXA region which is largely responsible for the fragility.

The lack of FMRP apparently causes the fragile X syndrome. This notion is strongly supported by the fact that a patient with a single point mutation within

the functionally important domain of the FMRP has the most severe case of this disease [20]. FMRP is a highly evolutionarily conserved RNA-binding protein which is expressed particularly strongly in neurons and gonads [21]. It is predominantly localized in the cytoplasm and associated with the active ribosomes [22]. The protein contains two RNA-binding domains, KH and RGG, as well as nuclear localization and nuclear export signals [23]. The so-called shuttling model [24] predicts that FMRP is transported from the cytoplasm into the nucleus where it is involved in the formation of mRNP particles. Subsequently those particles are transported out of the nucleus via an active process mediated by exportin 1. In the cytoplasm, mRNP particles are transported to the polyribosomes. While potentially important in every cell, the FMRP-mediated transport should be particularly important for neurons, where mRNA is transported along the dendrite to the polyribosomes located near the synapses [24]. In accord with this idea, FMR1 knockout mice demonstrate the absence of spatial learning and flawed synapse maturation [25].

Polyglutamine Diseases

Expansion of CAG repeats situated in the coding regions of various human genes is linked to eight neurodegenerative diseases, including Huntington disease (HD) [26], spinobulbar muscular atrophy (SBMA) [27], several spinocerebellar ataxias [28, 29], and dentatorubropallidoluysian atrophy [30]. CAG expansions in all those cases do not affect transcription of target genes or translation of corresponding mRNAs, but repeat-encoded polyglutamine stretches in the respective protein products lead to their self-aggregation and aggregation with other proteins [31, 32]. Two mechanisms might be responsible for the latter phenomenon. The first one proposes that the enzyme transglutaminase crosslinks polyglutamine tracts to polypeptides containing lysyl groups [33, 34]. Consequently aggregates consisting of polyglutamine copolymers are formed. The second hypothesis states that two antiparallel β -strands of polyglutamine repeats can zip together through hydrogen bonds [35]. This so-called polar zipper can be responsible for multimerization and aggregation. Whatever the exact mode, it seems certain that polyglutamine-containing proteins form aggregates in all systems, so far studied, including cell cultures [36–39], simple model organisms [40–42], transgenic mice [38, 43–45], and human patient's neurons [46].

Aggregation of polyglutamine-containing proteins appears to induce all the above neurodegenerative disorders, except for SBMA. First, aggregate formation easily explains the toxic “gain of function” caused by polyglutamine repeat expansion, which is a prerequisite for dominant inheritance. Second, there is a clear correlation between aggregate formation in certain

Human diseases caused by expansion of simple DNA repeats

Disease	Inheritance	Gender bias	Gene	Chromosomal position	Protein	Repeat number		Repeat position	Mutation type
						normal	mutant		
Fragile X syndrome	X-linked dominant	Maternal	<i>FMR1</i>	Xq27.3	FMRP	(CGG) < 50	(CGG) > 200	5'-UTR	Loss of function
Fragile XE mental retardation	X-linked	None	<i>FMR2</i>	Xq28	FMR2 protein	(CCG) < 35	(CCG) > 200	5'-UTR	Loss of function
Myotonic dystrophy	Autosomal dominant	Maternal	<i>DMPK</i>	19q13	MD protein kinase	(CTG) < 35	(CTG) > 50	3'-UTR	Gain of function
Spinocerebellar ataxia type 8	Autosomal dominant	"	<i>SCA8</i>	13q21	None	(CTG) < 40	(CTG) > 110	Antisense RNA	Gain of function
Friedrich's ataxia	Autosomal recessive	"	<i>X25</i>	9q13-21.1	Frataxin	(GAA) < 35	(GAA) > 100	Intron 1	Loss of function
Spinobulbar muscular atrophy	X-linked recessive	None	<i>AR</i>	Xq13-21	Androgen receptor	(CAG) < 30	(CAG) > 40	Coding	Gain of function
Huntington disease	Autosomal dominant	Paternal	<i>IT15</i>	4p16.3	Huntingtin	(CAG) < 40	(CAG) > 40	"	"
Dentatorubralpallidoluysian atrophy	"	"	<i>DRPLA</i>	12p13.31	Atrophin-1	(CAG) < 35	(CAG) > 50	"	"
Spinocerebellar ataxia type 1	"	"	<i>SCA1</i>	6p23	Ataxin-1	(CAG) < 40	(CAG) > 40	"	"
Spinocerebellar ataxia type 2	"	"	<i>SCA2</i>	12q24.1	Ataxin-2	(CAG) < 30	(CAG) > 35	"	"
Spinocerebellar ataxia type 3	"	"	<i>SCA3</i>	14q32.1	Ataxin-3	(CAG) < 40	(CAG) > 40	"	"
Spinocerebellar ataxia type 7	"	"	<i>SCA7</i>	3p12-13	Ataxin-7	(CAG) < 20	(CAG) > 40	"	"
Spinocerebellar ataxia type 6	"	None	<i>CACNA1A</i>	19p13	$\alpha 1A$ voltage-dependent Ca-channel subunit	(CAG) < 20	(CAG) > 20	"	Not detected
Progressive myoclonus epilepsy type	Autosomal recessive	None	<i>CSTB</i>	21q22.3	Cystatin B	(C ₄ GC ₄ GCG) < 3	(C ₄ GC ₄ GCG) > 60	Promoter	Loss of function

groups of neurons and their vulnerability to progressive dysfunction and eventual loss characteristic for a given disease [46]. Finally, overexpression of pure polyglutamine tracts is toxic to both neurons and peripheral cells, as demonstrated in cell culture [36–38] and animal systems [38, 40–42, 44, 45]. The notable expansion here is the SBMA, the recessive disease caused by a polyglutamine repeat expansion within the androgen receptor (AR) [27]. This expansion does not increase protein aggregation, but rather it partially negates the receptor function [47].

Friedreich's Ataxia

Friedreich's ataxia, the most common form of inherited ataxias, is caused by an expansion of the GAA repeat within the first intron of the frataxin gene [48]. Expansion of $(GAA)_n$ repeats inhibits the expression of the frataxin gene in patients at the level of transcription so that this inhibition is inversely proportional to the size of an expanded repeat [49]. Cloning of $(GAA)_n \cdot (TTC)_n$ repeats of increasing length in both orientations into the intron of the reporter gene led to transcription repression in transient transfection assay [50] and in nuclear extracts [51]. The highest inhibition was achieved when the GAA repeat was in the sense strand for transcription, as in the case of the frataxin gene. It was suggested that formation of H-DNA (see below) by this homopurine–homopyrimidine repeat might be responsible for transcription elongation blockage [50]. Recent *in vitro* studies using T7 RNA polymerase generally support the idea of H-DNA formation upon transcription through $(GAA)_n \cdot (TTC)_n$ repeats [52]. This inactivation of the frataxin gene expression apparently leads to Friedreich's ataxia. This follows from observations that approximately 2% of the patients carry missense, nonsense, or splicing mutations within the frataxin gene, rather than expanded GAA repeats [53]. Interestingly, in all studied cases, affected individuals were heterozygous for mutant frataxin allele, suggesting that homozygotes for frataxin mutations are lethal.

Frataxin has no similarity with proteins of known function. Yet it is strikingly conserved in Eukarya from yeast to humans [48]. It is apparently a mitochondrial protein [49, 54]. The inactivation of yeast frataxin homolog leads to the hyper-accumulation of iron in mitochondria [55] and hypersensitivity to oxidative stress [56]. It is believed, therefore, that excess mitochondrial iron, by reacting with oxygen, causes the oxidation of vital cellular components, loss of mitochondrial DNA and ultimately irreversible cell damage. It is not implausible that anomalous iron metabolism might be responsible for Friedreich's ataxia pathogenesis in humans; however, the data are still insufficient. Iron deposits were detected in myocardial cells [57] and dentate nucleus [58] of FRDA

patients. There are also profound respiratory chain deficiencies in the heart tissue of FRDA patients [59]. Yet the loss of mitochondrial DNA in those patients was never observed. Future studies are needed to validate this very provocative hypothesis linking iron metabolism and neurodegeneration.

Myotonic Dystrophy

Myotonic dystrophy, the most common muscular dystrophy in humans, is caused by an expansion of the CTG stretch located in the 3'-UTR of the so-called myotonic dystrophy protein kinase, DMPK, gene [60, 61]. Notwithstanding the name of the gene, its actual role in the development of myotonic dystrophy is far from clear. At present, there are three conflicting hypotheses explaining molecular pathogenesis of this disease. The first hypothesis states that expansion of CTG repeat in the DMPK 3'-UTR blocks the processing of the DMPK primary transcript, resulting in the lack of the kinase and disease. This hypothesis is supported by the data that repeat-containing DMPK RNA is retained within the nuclei [62]. However, the data on the DMPK knockout mice are largely contradictory to this hypothesis: heterozygous DMPK knockouts have no muscle pathology, while homozygotes develop only mild myopathy and cardiac arrhythmia, and this only at old age [63, 64].

The second hypothesis states that expansion of CTG repeats changes the structure of the surrounding chromatin, resulting in repression of the downstream gene called SIX5/DMAHP [65, 66]. SIX5 codes for homeobox protein implicated in the regulation of muscle cell differentiation. In heterozygous and homozygous SIX5 knockout mice, frequent formation of cataracts, which is one of the characteristic features of muscular dystrophy, was detected [67, 68]. Note, however, that major disease features, such as myotonia and myopathy, were never observed in SIX5 knockouts.

The third hypothesis, which currently looks most plausible, argues that the expanded CUG repeat within the DMPK RNA sequesters certain CUG-binding proteins, resulting in altered processing or translation of muscle-specific RNAs [69]. This would lead to the toxic gain of function for the DMPK RNA, explaining dominant inheritance. At present, several proteins were shown to bind CUG repeats in RNA. One, called CUGBP1 [70], plays an important role in functioning of RNAs containing CUG repeats in their regulatory parts. It was shown to affect splicing of several muscle-specific RNAs including cardiac troponin T and DMPK itself [71]. It also affects translation of the C/EBP β [72], a transcription factor implicated in muscle differentiation. Another such protein is a kinase activated by double-stranded RNA (PKR) that binds to expanded CUG repeats due to their abil-

ity to fold into stable hairpins [73]. If correct, the third hypothesis implies that any muscle-specific RNA containing expanded CUG repeats would cause myotonic dystrophy. Quite recently this was proven to be the case: a transgenic mouse containing 250 CUG repeats in the 3'-UTR of the skeletal actin had both myopathy and myotonia characteristic for the myotonic dystrophy [74].

The case of myotonic dystrophy vividly demonstrates that expanded CUG repeats might affect quite a number of different aspects of gene functioning, including transcription, nuclear transport and splicing. Interestingly, other potential effects of CUG repeats on gene expression appear to emerge. For example, we have recently found that moderately expanded CUG repeats, when situated within the 5'-UTR of a reporter gene, downregulate its expression by blocking the scanning step of translation [75]. It is quite possible that other expandable trinucleotide repeats may also affect different stages of gene expression. Thus, searching for genes containing potentially expandable repeats in their regulatory regions may lead to revealing new human genetic disorders.

MOLECULAR MECHANISMS OF TRINUCLEOTIDE REPEAT EXPANSION

As one can see from the above discussion, the effects of expandable repeats on the expression of their carrier genes vary dramatically and, consequently, molecular pathogenesis pathways are quite different for different diseases. It is generally believed, however, that there should be a unique mechanism responsible for expansion of various repeats. This belief is grounded in two striking features characteristic for all these expansions: (i) a unique threshold of approximately 30 repeats and (ii) accumulation of a large number of a repeat's extra copies at a single step [76]. At present, however, the exact mechanism(s) of repeat expansion remains unknown and several distinct hypotheses that are most actively pursued are considered below.

The length dependency of repeat expansion might suggest that an unusual DNA secondary structure of these repeats is involved. This concept comes from studies of unusual DNA structures in supercoiled DNA, where the probability for a given DNA repeat to adopt an unusual conformation depends exponentially on its length [77]. Experimental data for trinucleotide repeats indeed show their unusual structural potential. Even in a double-stranded linear DNA, chemical probing revealed that repeated $d(CNG)_n$ differs in reactivity from the canonical B-DNA [78]. Analysis of individual DNA strands of different trinucleotide repeats by various approaches, including thermal denaturation and electrophoretic mobility, showed

that they can fold into defined, compact structures [79–82]. NMR analysis directly proved that the expandable repeats $d(CGG)_n$, $d(CCG)_n$, $d(CTG)_n$ and $d(CAG)_n$ fold into imperfect hairpins, stabilized by both WC and non-WC base pairs (Fig. 1a) [83–85]. For trinucleotide repeats which are not known to expand, hairpin formation was found to be less likely. Consequently, it was hypothesized that the threshold length for expansion in these cases may reflect the threshold energy of hairpin formation [84]. Note that hairpins formed by the above four repeats differ in the nature of non-WC base pairs. Stability of these hairpins varies due to a differential contribution from different mismatches in the following way $CGG > CCG \sim CTG > CAG$ [84].

Trinucleotide repeats can also adopt other unusual DNA conformation. Single-stranded $(CGG)_n$ repeats fold into a stable tetrahelical conformation stabilized by intertwining $G^*G^*G^*G$ and $C^*C^*C^*C$ quartets (Fig. 1b) [86]. $(GAA)_n \cdot (TTC)_n$ repeats can adopt triple-helical H-DNA conformation under the influence of negative superhelicity (Fig. 1c) [87, 88]. Note, however, the two groups responsible for the above observation made opposing claims with regard to the fine structure of the triplex: H-y structure (i.e., pyrimidine/purine/pyrimidine triplex) in one case [87], but H-r (pyrimidine/purine/purine triplex) structure in the another [88].

Formation of these unusual structures by trinucleotide repeats may obscure various DNA transactions ultimately leading to the repeat expansion. For example, it is long known that DNA polymerases can be slowed down or stopped altogether by stable hairpins [89–94], triplexes [95–98] and G-quartets [99–101]. At the same time, unusual DNA structures are believed to increase the efficiency of genetic recombination [102–106]. What is unclear, however, is how the alterations in various DNA transactions can convert into a repeat expansion.

This brings us to the second general feature of repeat expansion, its large-scale character. It was long-known that simple tandem repeats of any base composition exhibit length polymorphism. These length changes, however, are small-scale, i.e., one or two elementary repeated units. They are commonly explained by strand slippage during replication of multiply repeated DNAs [107]. In the trinucleotide repeat case, on the contrary, dozens or even hundreds of elementary repeated units are added at a single step. This rules out trivial slippage as a mechanism. A dramatic increase in repetitive DNA length can be explained *a priori* by two mechanisms: a major inaccuracy during replication of a repeat, or a form of unequal crossingover between similar repeated stretches on homologous chromosomes/sister chromatids.

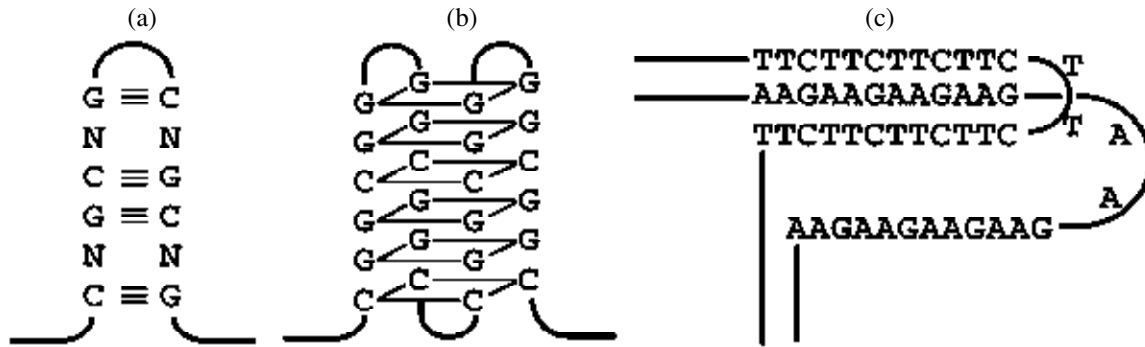


Fig. 1. Unusual DNA structures formed by trinucleotide repeats. (a) Hairpin, (b) quadruplex, (c) H-DNA.

So far most of the data points to the replication mode of expansion. DNA polymerization *in vitro* through various trinucleotide repeats was shown to be compromised. In double-stranded DNA templates, different pro- and eukaryotic DNA polymerases stopped at specific sites within $(CTG)_n \cdot (CAG)_n$ and $(CGG)_n \cdot (CCG)_n$ repeats, and the termination rate increased with the length of the repeat [108]. In single-stranded DNA templates, $(CGG)_n$ blocks caused a K^+ -dependent polymerization arrest, presumably due to quadruplex formation [109]. $(GAA)_n \cdot (TTC)_n$ repeats represent the most potent block for DNA polymerases both in single- and double-stranded DNA templates due to H-DNA formation during polymerization [87]. Therefore, it is believed that unusual DNA structures of trinucleotide repeats could account for DNA polymerization blockage. Such blockage could facilitate a misalignment between the newly synthesized and the template DNA strands [110]. Resumption of DNA polymerization might then lead to the repeat expansion or contraction depending on whether the newly-synthesized or template DNA chains, respectively, folded into a secondary structure (Fig. 2a). Two alternative models, presented in Fig. 2, suggest that expansion might occur due to the synthesis of an extra Okazaki fragment initiated at the loop of the hairpin formed by a repeat (Fig. 2b) [76], or formation of an H-DNA-like structure upon folding back the newly synthesized strand of the $(CTG)_n \cdot (CAG)_n$ repeat (Fig. 2c) [87].

All these models imply that repeats expansion occur during the lagging strand synthesis. This notion is supported by studies of trinucleotide repeat maintenance in model bacterial and yeast systems. Stability of different repeats was found to dramatically depend on their length and orientation with regard to the replication origin [111–113]. Long $(CTG)_n$ or $(CGG)_n$ stretches in the lagging strand template efficiently deleted, while the same stretches in the leading strand template tended to expand. It is believed, therefore, that formation of hairpin-like structures by structure-

prone repeats in either the lagging strand template or the newly synthesized lagging strand caused replication disorder and consequent deletion or expansion, respectively (Fig. 2a).

Additional support for this lagging strand expansion came from the studies of yeast *rad27* mutants. *Rad27* encodes the so-called “flap-endonuclease,” involved in the replacement of RNA primers in Okazaki fragments [114], required for the completion of the lagging strand synthesis. The frequency of different trinucleotide repeat expansion in *rad27* deletion mutants was found to be drastically increased [115–117]. It was suggested that the repeat-containing primer for the Okazaki fragment is not degraded in the absence of the flap endonuclease, but simply displaced by DNA polymerase. Its subsequent religation with the 3'-end of the next Okazaki fragment results in expansion (Fig. 2d). Note however, that this model can adequately explain only relatively small expansions of a size of an Okazaki primer (~20 bp). Also, an increased instability in *rad27* mutants is not limited to trinucleotide repeats but is observed for all microsatellites studied [118, 119].

In addition to model systems, a strong argument in favor of the replication model comes from the DNA sequence analysis of families with trinucleotide repeat diseases. In normal individuals from fragile X families, $(CGG)_n$ stretches are usually interrupted by several dispersed AGG triplets [120–122]. These sequences are observed to have expanded on the 3' but not on the 5' flank of the repeat in carriers and afflicted individuals. Further, the expanded part lacks the AGG interruptions [121, 122]. Similar polar and cryptic variations were reported for spinocerebellar ataxia $(CAG)_n$ repeat expansion [123]. This striking polarity can be explained by anomalous repeat replication from a single adjacent replication origin, assuming that the fidelity of the leading and lagging strand synthesis of repeated DNA is different. It is indeed known that mutations within certain DNA repeats preferentially occur in the lagging strand [124].

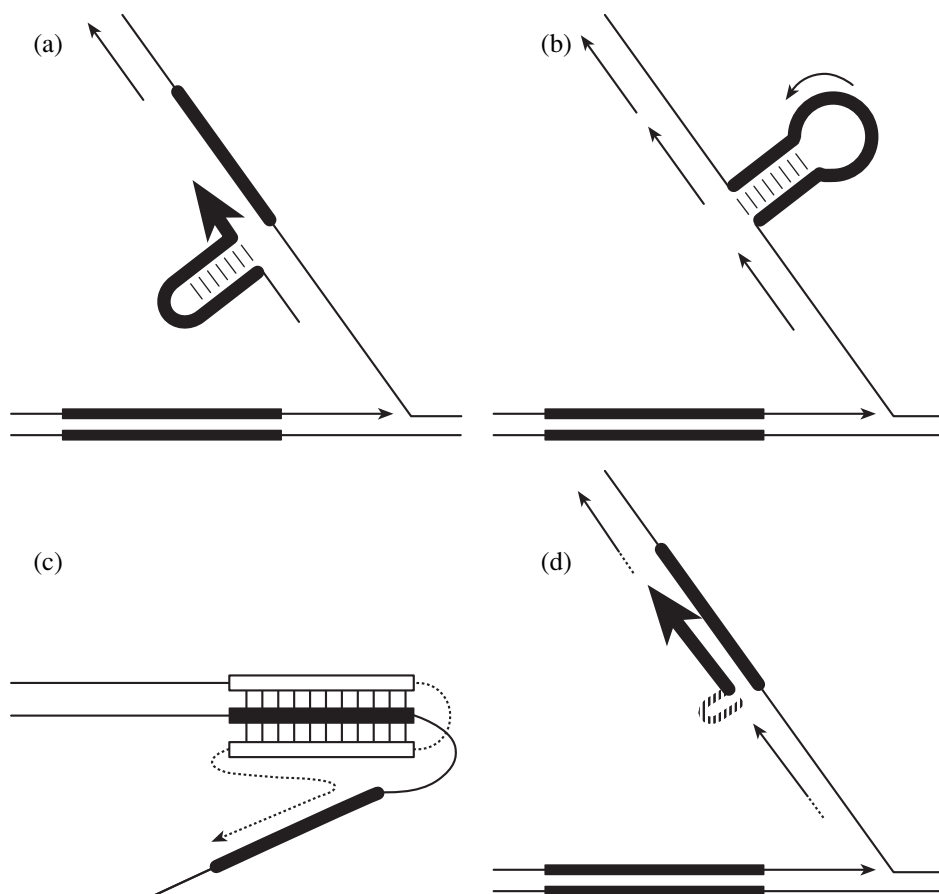


Fig. 2. Proposed mechanisms for trinucleotide repeat expansion during replication. (a) Expansion caused by the formation of a hairpin during the lagging strand synthesis. (b) Expansion due to the initiation of an extra Okazaki fragment at the hairpin loop in the lagging strand template. (c) Expansion upon triplex formation during the lagging strand synthesis. (d) Expansion upon displacement of the primer for an Okazaki fragment followed by hairpin formation. Repeated stretches are shown by black rectangles. Arrows show the direction of DNA synthesis. Hatched lines show primers of Okazaki fragments.

The replication model for expansion assumes that replication of trinucleotide repeats proceeds abnormally with a certain degree of miscoordination between the leading and lagging DNA strands. Yet direct data on the replication of trinucleotide repeats *in vivo* were strikingly scarce. This encouraged us to study the mode of replication fork progression through trinucleotide repeats *in vivo* [125]. We expected that those repeats might somewhat slow the replication fork progression. Unfortunately, this is a difficult problem to study, since the normal replication rate is very fast, ranging from 1000 bp/s in bacteria to hundreds bp/s in eukaryotes [126]. For example, given that a 100 bp-long repeat in pBR322 slows the replication fork progression 10 times, the overall plasmid replication would only be slowed from 5 to 6 s. Therefore, most conventional methods of DNA replication analysis are not applicable to this problem.

To solve this problem we analyzed the effects of different DNA repeats on the replication of bacterial plasmids *in vivo* using an approach called 2-dimen-

sional neutral/neutral electrophoresis of replication intermediates. This technique was developed for mapping of the replication origins [127, 128] but lately has become instrumental in defining replication termination sites as well [129–131]. Bacterial plasmids were chosen for two reasons: (i) they replicate unidirectionally, which unequivocally determines leading and lagging strands during DNA replication; and (ii) they replicate very efficiently, which allows for the easy isolation and analysis of replication intermediates.

The idea of electrophoretic analysis of replication intermediates applied to unidirectional replication is presented in Fig. 3. Intermediate products of plasmid replication are Q-shaped. Upon cleaving these intermediates with a restriction enzyme upstream of the replication origin, they convert into bubble-shaped molecules, where the size of the bubble correlates with the duration of replication. Bubble intermediates differ in their molecular mass (ranging from 1 to 2 plasmid masses) and shape. They are separated in two dimensions: first by mass (low percentage agar-

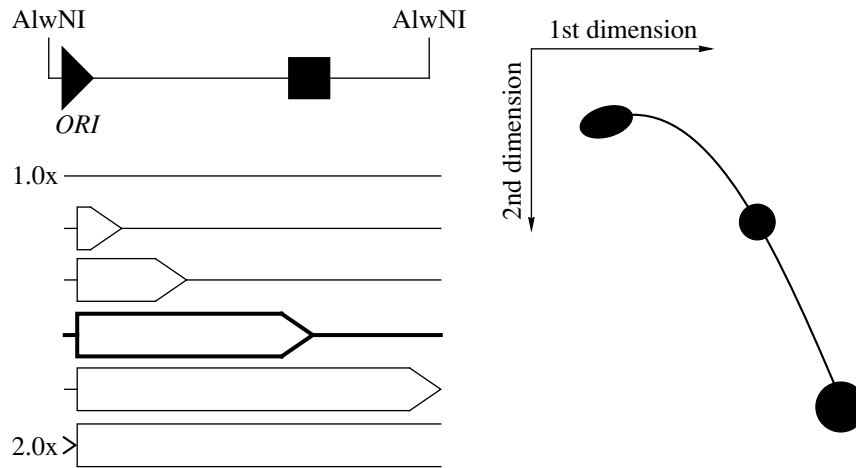


Fig. 3. Detection of repeat-caused replication blocks by 2D gel electrophoresis. Upper left panel shows the structure of the linearized plasmid DNA. The black triangle corresponds to the replication origin, the black box corresponds to cloned, repeated DNA. The lower left panel shows the shapes of different replication intermediates. The bold intermediate corresponds to the one that preferentially accumulates owing to repeat-caused replication blockage. The right panel schematically shows the bubble arc with a bulge reflecting the replication stop site.

ose) and second by mass and shape (high percentage agarose with ethidium bromide). Southern-blotting hybridization with a radioactive plasmid probe reveals a so-called “bubble arc.” If there are no roadblocks during DNA replication, this arc is smooth. Stalling of the replication fork at the $(CGG)_n$ repeat, however, leads to the accumulation of an intermediate of a given size and shape, generating a bulge on the arc. The ratio of the signal of this bulge to the signal of the corresponding area of a smooth replication arc (relative stop strength, RSS) is an index of replication fork retardation caused by the repeat.

Our data for the replication fork progression through $(CGG)_n$ $(CCG)_n$ repeats in *E. coli* cells are shown in Fig. 4a. Owing to the unidirectional character of plasmid replication, we knew precisely whether $(CGG)_n$ or $(CCG)_n$ repeats were in the lagging strand template, and the plasmids are named accordingly. One can see that lengthening $(CGG)_n \cdot (CCG)_n$ repeats lead to the appearance of prominent stop signals (shown by arrows) in the bubble arc. Quantitation of the above data is shown in Fig. 4b. One can clearly see that the efficiency of replication blockage is dependent on the repeat length and orientation relative to the replication origin. Most strikingly, the repeat length responsible for significant (5-fold) replication stalling in bacteria appeared to be similar to the threshold length for repeat expansion in humans.

To prove that the replication fork is indeed stalled within $(CGG)_n$ $(CCG)_n$ repeats, we mapped replication stop sites using a modified version of the electrophoretic analysis of replication intermediates [132]. After the first dimension of electrophoresis, replication intermediates were digested with a restriction

enzyme in the gel. The enzymes selected for this analysis cut the plasmid either upstream or downstream of the repeat. As a result, a fraction of bubble-shaped intermediates converted into identical y-shaped intermediates (Fig. 5). In the second dimension of electrophoresis, these intermediates migrate similarly and can be detected as a horizontal line upon hybridization with a probe adjacent to the replication *ori*. As is clear from Fig. 5, restriction cleavage downstream of the repeat (relative to the *ori*) would leave the bulge on the bubble-arc, while upstream cleavage shifts the bulge from the bubble-arc onto the horizontal line.

Our experimental data for the $(CGG)_{63} \cdot (CCG)_{63}$ insert are presented in Fig. 6. One can see that cleavage of replication intermediates with *EcoRI* (located downstream from the insert) leaves the replication stop on the bubble arc. By contrast, cleavage by *HindIII* shifts the stop onto the horizontal line (spot 1). It is plausible to conclude, therefore, that the replication fork is stalled within the $(CGG)_{63} \cdot (CCG)_{63}$ stretch. Notably, however, that *HindIII* cleavage also results in the appearance of an additional spot (spot 2) comigrating with spot 1 in the first dimension but migrating slower in the second dimension, i.e., the shape of this intermediate is more complex than “y” but less complex than bubble. To explain the appearance of spot 2, we speculate that it may reflect the fraction of replication intermediates in which the lagging strand around the *HindIII* site was not yet synthesized. Partial *HindIII* digestion of such intermediates would lead to the accumulation of butterfly-like DNA molecules (shown on a diagram). If true, the existence of the spot 2 indicates underreplication of the lagging strand within the repeated DNA.

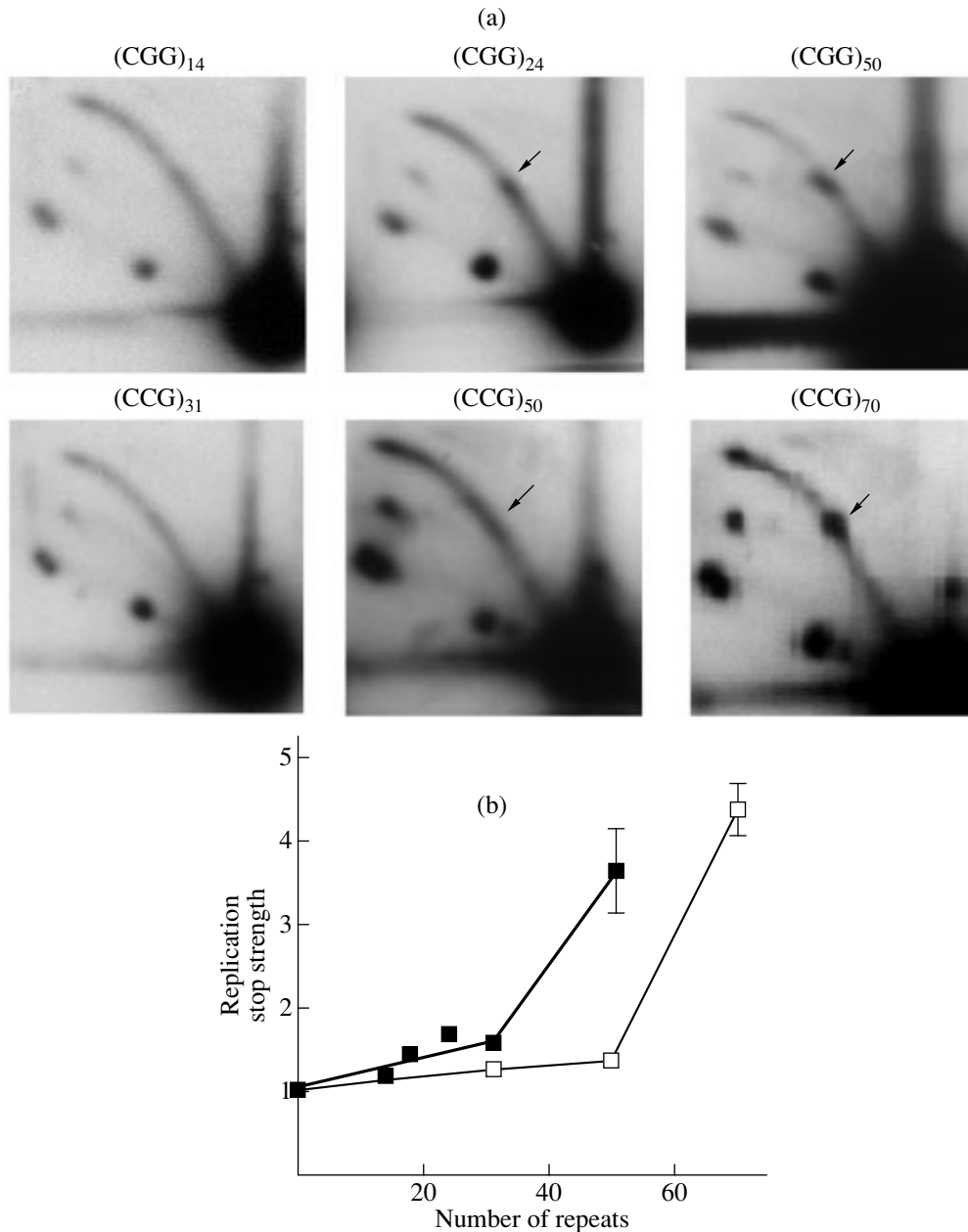


Fig. 4. Effects of $(CGG)_n \cdot (CCG)_n$ repeats on the replication fork progression *in vivo*. (a) Analysis of replication intermediates of plasmids with $(CGG)_n \cdot (CCG)_n$ inserts by 2D electrophoresis. Plasmids are named according to the sequence of the lagging strand template. Arrows show replication stop sites. (b) Quantitative analysis of the replication-stop intensity from several experiments. The strength of the stop is characterized by the ratio of the observed density of the stop signal to the expected density of the smooth arc at this position. Filled squares represent $(CGG)_n$ runs and open squares represent $(CCG)_n$ runs in the lagging strand template.

Are these results on replication in bacteria relevant to the repeat expansion in humans? As discussed above, analysis of fragile X repeats from individual human DNAs revealed that $(CGG)_n$ stretches are commonly interrupted by several AGG trinucleotides preventing expansion [121]. We, therefore, studied how AGG interruptions within the $(CGG)_n$ stretches affect their replication in our system. It appeared that AGG interruptions indeed abolish the replication arrest.

Thus, there seems to be a link between peculiarities of $(CGG)_n$ repeat replication in bacteria and their propensity to expand in humans.

We concluded that fragile X CGG repeats directly affect replication fork movement *in vivo* in a length and orientation dependent manner, being most prominent when the structure-prone strands of the repeated DNA, i.e., $(CGG)_n$, is in the lagging strand template. In view of the length-dependence, it is likely that

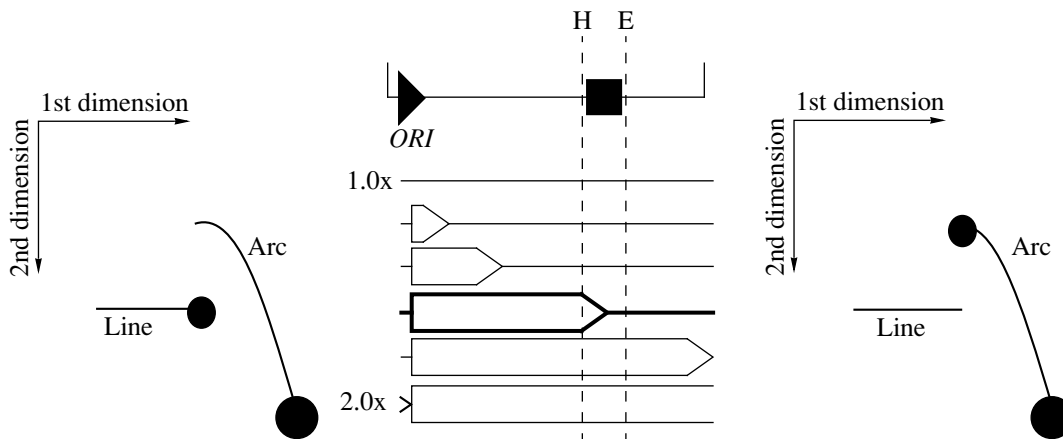


Fig. 5. Mapping of replication stop sites using restriction digestion after the first direction of 2D electrophoresis (see text for details).

some unusual structure rather than primary DNA sequence per se is responsible for replication fork arrest. Discontinuous synthesis of the lagging strand implies that a portion of the lagging strand template (of an Okazaki-fragment size) must be at least transiently single-stranded. Thus, a $(CGG)_n$ repeat has a better chance to form a secondary structure when in the lagging, but not in the leading strand template. Since synthesis of both DNA strands during replication is believed to be coordinated [126], arrest of the lagging strand synthesis would instantly resonate on the replication fork as a whole (Fig. 7).

Since the above studies were performed in bacterial cells, a legitimate question is whether trinucleotide repeats affect the eukaryotic replication fork progression as well. In order to address this question, we have recently expanded our studies of replication through trinucleotide repeats into a yeast experimental system using the same approach, electrophoretic analysis of replication intermediates. To our satisfaction, $(CGG)_n \cdot (CCG)_n$ repeats appear to stall the yeast replication fork similarly to the bacterial case (M. Krasilnikova and G. Samadashwily, unpublished results). We believe therefore, that the effects of trinucleotide repeats on DNA replication *in vivo* are principally the same in both pro- and eukaryotes.

An alternative to replication could be expansion of trinucleotide repeats via the process of recombination. First, an unequal crossingover can generate length variation for tandemly repeated sequences (Fig. 8a). However, this mechanism is not likely to account for the repeat expansion. Unequal crossingover during meiosis can be efficiently ruled out, since exchange of very close flanking markers consecutive to the expansion does not occur [133]. Unequal crossingover between the sister chromatids in mitosis can not be decisively ruled out, but it should generate equal numbers of expanded and contracted versions of a trinucleotide repeat which was never observed experimentally.

Second, a gene conversion event driven by a double strand break within or adjacent to the repeated sequence can promote its expansion (Fig. 8b). Similarly to a sister chromatid exchange, this should generate both expanded and contracted alleles. Gene conversion, however, is often biased [134] so that it can lead to a predominance of expansions over contractions. A gene conversion model has recently attracted wide attention after it was shown that a trinucleotide repeat, $(CTG)_n \cdot (CAG)_n$, is highly recombinogenic in *E. coli* [135] and likely expands via double-strand break-recombination in yeast [136].

One can think of a link between stimulation of gene conversion by the repeats with their replication abnormalities. It is well documented that the replication fork slowing or stopping at specific DNA sites can stimulate recombination [137]. This recombinational activity can, in turn, help in restarting the replication fork progression. While the mechanisms of this interplay between replication and recombination are not yet understood in fine detail, two models are most commonly considered. First, stalling of the replication fork produces a stably exposed, single-stranded piece of the lagging strand template (Fig. 9a). Nicking within this single-stranded part can generate a substrate for recombinational invasion into a different DNA duplex. Second, replication fork stalling can lead to the dissociation of both newly synthesized DNA strands and their self annealing. This would result in the so-called collapsed replication fork [138] that greatly resembles a Holliday junction possibly stabilized by RuvAB complex (Fig. 9b). Subsequent processing of this pseudo-Holliday junction by recombination enzymes can lead to repeats expansions/contractions.

It should be noted, however, that neither replication nor recombination models can satisfactorily

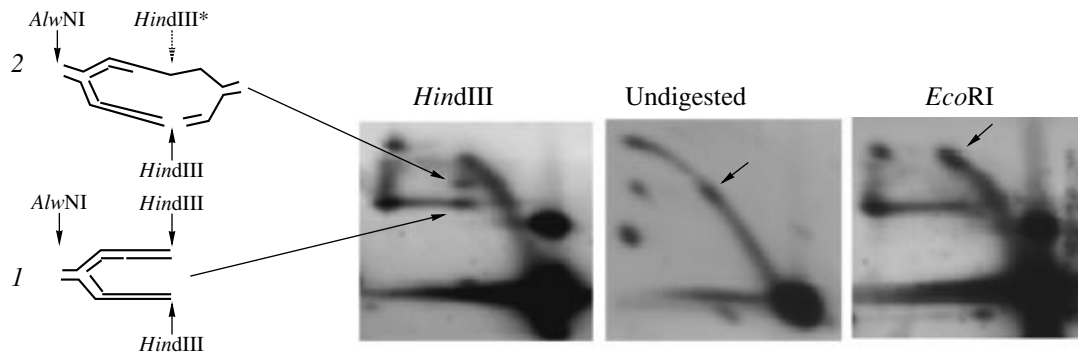


Fig. 6. Fine mapping of the replication stop site in the p(CGG)₆₃ plasmid. Central panel, no digestion; right panel, *EcoRI* digestion; left panel, *HindIII* digestion. Small arrows show stop sites on the bubble arc. Long arrows show stop sites moving towards the horizontal line. Schematic representation of the structure of the intermediates in spots 1 and 2 are presented (see text for details).

explain the striking bias towards expansion during intergenerational transmission of trinucleotide repeats in humans. This phenomenon was not so far reproduced in available experimental systems. In cultured somatic cells the length of trinucleotide repeats are fairly stably maintained. In model systems, including bacteria and yeast, the expansion frequencies for various trinucleotide repeats were relatively low and there was no bias towards expansion. In fact the opposite was true: repeat contractions were much more common than expansion. We can think of three possible explanations for this discrepancy.

First, repeat expansion may preferably occur only in specialized cell lineages and/or at specific stages of human development. For example, it is foreseeable that expansions specifically occur during gametogenesis. This hypothesis is generally supported by the single-sperm analysis of repeat length polymorphism. It was found that repeat length polymorphism is greatly increased in sperm compared to somatic cells in spinocerebellar ataxias types 1 and 7 [139, 140], spinobulbar muscular atrophy [141], and Huntington

disease [142, 143]. Moreover, there was a clear bias towards expansion. The fact that trinucleotide repeats preferably expand during spermatogenesis can easily explain the paternal pattern of disease transmission characteristic for those diseases. Maternal transmission, in turn, can be explained by preponderance for repeat expansions during oogenesis. A somewhat more complicated case of the maternal transmission provides the fragile X syndrome. In premutation males, there is a bias towards expansion during spermatogenesis [144], while in sperm of males with full mutations, repeats contract back to the premutation size [145]. It is plausible to speculate therefore that during spermatogenesis there is a selection for certain level of the FMR1 expression, e.g., against expansion beyond the premutation size. Consequently, transmission of expanded repeats from fathers to daughters does not occur.

Expansions can also happen during the very early stages of embryonic development. At these stages replication proceeds extremely rapidly [146], which may compromise its fidelity resulting in repeats expansion. In order for this hypothesis to explain the gender bias in disease transmission, one should assume imprinting of either paternal or maternal allele of a repeat-containing gene.

The second explanation comes from the data that certain mutations in the replication apparatus increase the rate of repeat expansion. One might assume that families with triplet repeat diseases carry some additional mutations in the replication apparatus that predispose them to expansion. An illustration could be the *rad27* mutation that only marginally affects the overall replication efficiency in yeast, yet drastically elevates the expansion rate for various trinucleotide repeats. This idea is not without potential pitfalls. For example, in a given family, only one type of repeats is expanding, stipulating that various "silent" replication mutations should differentially affect the expansion of different repeats. At present, the examples of such

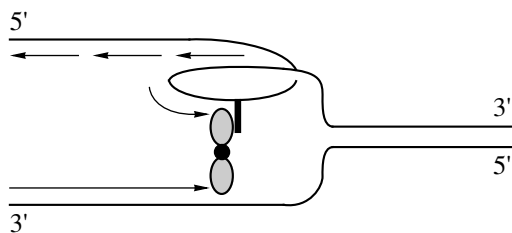


Fig. 7. The model of the replication fork blockage caused by a hairpin on the lagging strand template. Arrows show the direction of DNA synthesis. Hairpin on the lagging strand template is shown by the black rectangle. Leading and lagging strand DNA polymerases are shown by shaded ellipses, black circle stands for protein(s) linking leading and lagging strand polymerases, such as τ subunits of the DNA polymerase III holoenzyme in *E. coli*.

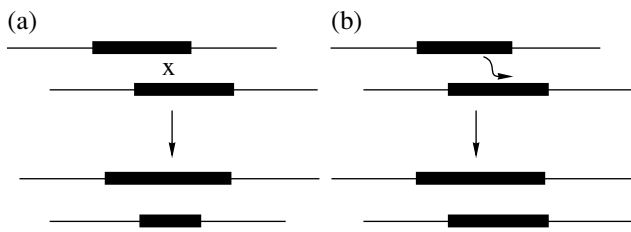


Fig. 8. Recombination models for a trinucleotide repeat expansion. (a) Via unequal crossingover, (b) via biased gene conversion. Lines correspond to double-stranded DNA molecules. Repeated areas are shown as black rectangles.

replication mutations are unknown. Note, that the first and the second explanations are not self-excluding. It is entirely possible that “silent” replication mutations preferably affect repeats replication during gametogenesis or early embryogenesis.

Finally, there might be a certain selection for an expansion of a given repeat in individuals predisposed to distinct triplet repeat diseases. Such selection could result, for example, from a mutation causing overexpression of a repeat-binding protein that have a nega-

tive effect on gametogenesis and/or embryonic development. This would give an advantage for gametes or early embryos carrying expanded alleles. Proteins that specifically bind to triplet repeats either in DNA [147, 148] or in RNA [70, 149] have been described. As described above, at least one of them, CUGBP1, might affect expression of a variety of genes at the splicing level [71]. Overexpression of repeat-binding proteins that are involved in regulation of important cellular genes is likely to be disadvantageous, but could be counterbalanced by the expansion of a corresponding repeat.

Summarizing, there exist several models explaining potential mechanisms of repeat expansion. At present, however, none of these models can satisfactorily explain why only a single trinucleotide repeat is expanding in a human pedigree, given the large number of other repeats of the same sequence in the human genome. Thus, the final molecular details of repeat expansion remain to be understood. Another crucial goal is to establish a link between the data on repeat expansion obtained either *in vitro* or in model systems and the etiology of repeat expansion in humans. Given the speed of development of this novel

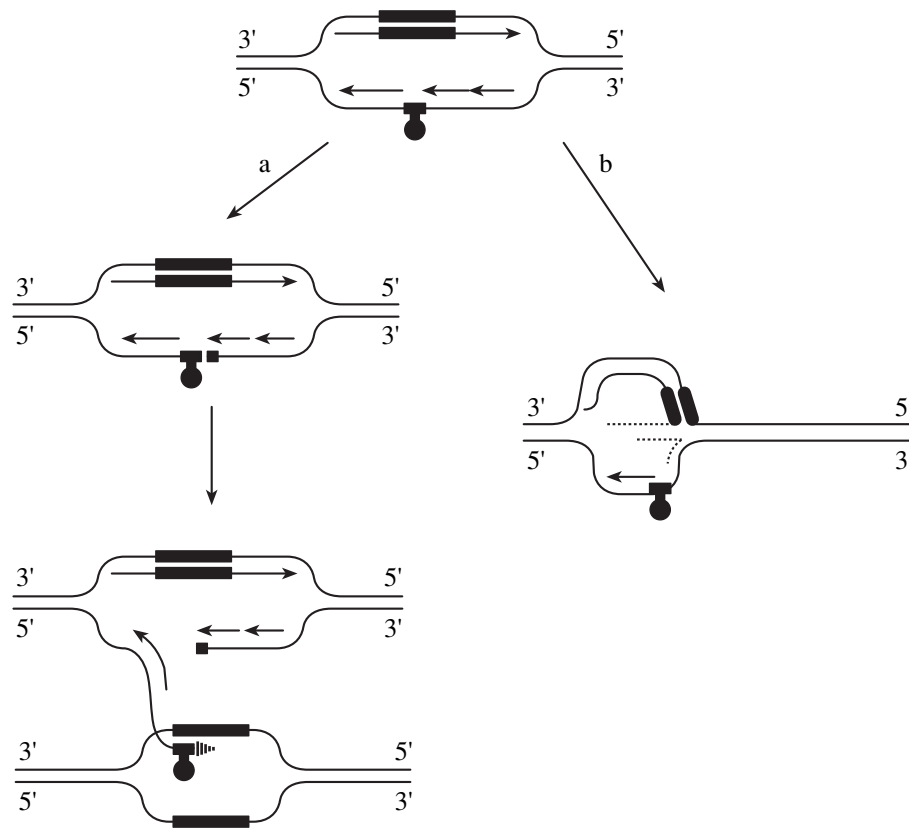


Fig. 9. Replication blockage during the lagging strand synthesis can stimulate recombination leading to repeat expansion. (a) A nick in the lagging strand template at the repeated run can initiate strand invasion in a homologous duplex followed by expansion. (b) Collapse of a stalled replication fork leads to the formation of a Holliday-like junction which can then be processed by recombination machinery. Black rectangles show trinucleotide repeats. Arrows show the direction of DNA synthesis. Newly synthesized DNA strands that annealed upon the replication fork collapse are shown by hatched lines.

field of genetics, one might fully expect major breakthroughs in both directions in the near future.

ACKNOWLEDGMENTS

We thank current and former members of our lab in Chicago: Maria Krasilnikova, Gordana Raca, and George Samadashwily for their invaluable role in our studies of trinucleotide repeats, and Gerald Buldak for his editorial help. Supported by the grant from the National Science Foundation (MCB-9723924).

REFERENCES

- Imbert, G., Feng, Y., Nelson, D.L., Warren, S.T., and Mandel, J.-L., *Genetic Instabilities and Hereditary Neurological Disorders*, Wells, R.D., and Warren, S.T., Eds., San Diego: Academic Press, 1998, pp. 27–53.
- Hagerman, R., *Genetic Instabilities and Hereditary Neurological Disorders*, Wells, R.D., and Warren, S.T., Eds., San Diego: Academic Press, 1998, pp. 3–87.
- Sutherland, G.R. and Hecht, F., *Fragile Sites on Human Chromosomes*, New York: Oxford Univ. Press, 1985.
- Van Dyke, D.L. and Weiss, L., *Am. J. Med. Genet.*, 1986, vol. 23, pp. 723–737.
- Sherman, S.L., Jacobs, P.A., Morton, N.E., *et al.*, *Hum. Genet.*, 1985, vol. 69, pp. 289–299.
- Sherman, S.L., Morton, N.E., Jacobs, P.A., and Turner, G., *Ann. Hum. Genet.*, 1984, vol. 48, pp. 21–37.
- Verkerk, A.J.M.H., Pieretti, M., Sutcliffe, J.S., *et al.*, *Cell*, 1991, vol. 65, pp. 905–914.
- Kremer, E.J., Pritchard, M., Lynch, M., *et al.*, *Science*, 1991, vol. 252, pp. 1711–1714.
- Howeler, C.J., Busch, H.F., Geraedts, J.P., Niermeier, M.F., and Staal, A., *Brain*, 1989, vol. 112, pp. 779–797.
- Brooks, B.P. and Fischbeck, K.H., *Trends Neurosci.*, 1995, vol. 18, pp. 459–461.
- Goldberg, Y.P., Telenius, H., and Hayden, M.R., *Curr. Opin. Neurol.*, 1994, vol. 7, pp. 325–332.
- Pandolfo, M., *Neuromuscul. Disord.*, 1998, vol. 8, pp. 409–415.
- Gindilis, V.M. and Shakhmatova-Pavlova, I.V., *Vest. Akad. Med. Nauk SSSR*, 1979, vol. 7, pp. 14–23.
- Bassett, A.S. and Honer, W.G., *Am. J. Hum. Genet.*, 1994, vol. 54, pp. 864–870.
- Ross, C.A., McInnis, M.G., Margolis, R.L., and Li, S.H., *Trends Neurosci.*, 1993, vol. 16, pp. 254–260.
- Lalioi, M.D., Scott, H.S., Buresi, C., *et al.*, *Nature*, 1997, vol. 386, pp. 847–851.
- Kooy, F., Willemsen, R., and Oostra, B.A., *Molecular Medicine Today*, 2000, vol. 6, pp. 193–198.
- Subramanian, P.S., Nelson, D.L., and Chinault, A.C., *Am. J. Hum. Genet.*, 1996, vol. 59, pp. 407–416.
- Hornstra, I.K., Nelson, D.L., Warren, S.T., and Yang, T.P., *Hum. Mol. Genet.*, 1993, vol. 2, pp. 1659–1665.
- De Boulle, K., Verkerk, A.J.E.R., *et al.*, *Nat. Genet.*, 1993, vol. 3, pp. 31–35.
- Devys, D., Lutz, Y., Rouyer, N., Bellocq, J.P., and Mandel, J.-L., *Nat. Genet.*, 1993, vol. 4, pp. 335–340.
- Khandjian, E.W., Corbin, F., Woerly, S., and Rousseau, F., *Nat. Genet.*, 1996, vol. 12, pp. 91–93.
- Ashley, C.J., Wilkinson, K.D., Reines, D., and Warren, S.T., *Science*, 1993, vol. 262, pp. 563–566.
- Feng, Y., Gutekunst, C.A., Eberhart, D.E., *et al.*, *J. Neurosci.*, 1997, vol. 17, pp. 1539–1547.
- Comery, T.A., Harris, J.B., Willems, P.J., *et al.*, *Proc. Natl. Acad. Sci. USA*, 1997, vol. 94, pp. 5401–5404.
- The Huntington's Disease Collaborative Research Group, *Cell*, 1993, vol. 72, pp. 971–983.
- LaSpada, A.R., Wilson, E.M., Lubahn, D.B., Harding, A.E., and Fischbeck, K.H., *Nature*, 1994, vol. 352, pp. 77–79.
- Orr, H.T., Chung, M.Y., Banfi, S., *et al.*, *Nat. Genet.*, 1993, vol. 4, pp. 221–226.
- Chung, M.Y., Ranum, L.P., Duvick, L.A., *et al.*, *Nat. Genet.*, 1993, vol. 5, pp. 254–258.
- Koide, R., Ikeuchi, T., Onodera, O., *et al.*, *Nat. Genet.*, 1994, vol. 6, pp. 9–13.
- Lunkes, A., Trottier, Y., and Mandel, J.L., *Essays Biochem.*, 1998, vol. 33, pp. 149–163.
- Bates, G., *Bioassays*, 1996, vol. 18, pp. 175–178.
- Kahlem, P., Green, H., and Djian, P., *Mol. Cell*, 1998, vol. 1, pp. 595–601.
- Green, H., *Cell*, 1993, vol. 74, pp. 955–956.
- Perutz, M.F., Johnson, T., Suzuki, M., and Finch, J.T., *Proc. Natl. Acad. Sci. USA*, 1994, vol. 91, pp. 5355–5358.
- Paulson, H.L., Perez, M.K., Trottier, Y., *et al.*, *Neuron*, 1997, vol. 19, pp. 333–334.
- Skinner, P.J., Koshy, B., Cummings, C., *et al.*, *Nature*, 1997, vol. 389, pp. 971–974.
- Ikeda, H., Yamaguchi, M., Sugai, S., *et al.*, *Nat. Genet.*, 1996, vol. 13, pp. 196–202.
- Lunkes, A. and Mandel, J.L., *Hum. Mol. Genet.*, 1998, vol. 7, pp. 1355–1361.
- Faber, P.W., Alter, J.R., MacDonald, M.E., and Hart, A.C., *Proc. Natl. Acad. Sci. USA*, 1999, vol. 96, pp. 179–184.
- Warrick, J.M., Paulson, H.L., Gray-Board, G.L., *et al.*, *Cell*, 1998, vol. 93, pp. 939–949.
- Jackson, J.R., Salecker, I., Dong, X., *et al.*, *Neuron*, 1998, vol. 21, pp. 633–642.
- Burright, E.N., Clark, H.B., Servadio, A., *et al.*, *Cell*, 1995, vol. 82, pp. 937–948.
- Klement, I.A., Skinner, P.J., Kaytor, M.D., *et al.*, *Cell*, 1998, vol. 95, pp. 41–53.
- Mangiarini, L., Sathasivam, K., Sellar, M., *et al.*, *Cell*, 1996, vol. 87, pp. 493–506.
- Zoghbi, H.Y. and Orr, H.T., *Annu. Rev. Neurosci.*, 2000, vol. 23, pp. 217–247.
- Mhatre, A.N., Trifiro, M.A., Kaufman, M., *et al.*, *Nat. Genet.*, 1993, vol. 5, pp. 185–188.
- Campuzano, V., Montermini, L., Molto, M.D., *et al.*, *Science*, 1996, vol. 271, pp. 1423–1427.
- Campuzano, V., Montermini, L., Lutz, Y., *et al.*, *Hum. Mol. Genet.*, 1997, vol. 6, pp. 1771–1780.
- Ohshima, K., Montermini, L., Wells, R.D., and Pandolfo, M., *J. Biol. Chem.*, 1998, vol. 273, pp. 14588–14595.

51. Bidichandani, S.I., Ashizawa, T., and Patel, P.I., *Am. J. Hum. Genet.*, 1998, vol. 62, pp. 111–121.
52. Grabczyk, E. and Usdin, K., *Nucleic Acids Res.*, 2000, vol. 28, pp. 2815–2822.
53. Pandolfo, M., *Arch. Neurol.*, 1999, vol. 56, pp. 1201–1208.
54. Cossee, M., Campuzano, V., Koutnikova, H., *et al.*, *Nat. Genet.*, 1997, vol. 15, pp. 337–338.
55. Babcock, M., de Silva, D., Oaks, R., *et al.*, *Science*, 1997, vol. 276, pp. 1709–1712.
56. Wilson, R.B. and Roof, D.M., *Nat. Genet.*, 1997, vol. 16, pp. 352–357.
57. Lamarche, J.B., Shapcott, D., Cote, M., and Lemieux, B., *Handbook of Cerebellar Diseases*, Lechtenberg, R., Ed., New York: Marcel Dekker, 1993, pp. 453–458.
58. Waldvogel, D., van Gelderen, P., and Hallet, M., *Ann. Neurol.*, 1999, vol. 46, pp. 123–125.
59. Rotig, A., deLonlay, P., and Chretien, D., *Nat. Genet.*, 1997, vol. 17, pp. 214–217.
60. Mahadevan, M., Tsilifidis, C., Sabourin, L., *et al.*, *Science*, 1992, vol. 255, pp. 1253–1255.
61. Brook, J.D., McCurrach, M.E., Harley, H.G., *et al.*, *Cell*, 1992, vol. 68, pp. 799–808.
62. Davis, B.M., McCurrach, M.E., Taneja, K.L., Singer, R.H., and Housman, D.E., *Proc. Natl. Acad. Sci. USA*, 1997, vol. 94, pp. 7388–7393.
63. Reddy, S., Smith, D.B.J., Rich, M.M., *et al.*, *Nat. Genet.*, 1996, vol. 13, pp. 325–335.
64. Jansen, G., Groenen, P.J.T.A., Bacher, D., *et al.*, *Nat. Genet.*, 1996, vol. 13, pp. 316–324.
65. Klesert, T.R., Otten, A.D., Bird, T.D., and Tapscott, S.J., *Nat. Genet.*, 1997, vol. 16, pp. 402–406.
66. Thornton, C.A., Wymer, J.P., Simmons, Z., McClain, C., and Moxley, R., *Nat. Genet.*, 1997, vol. 16, pp. 407–409.
67. Klessert, T.R., Cho, D.H., Clark, J.I., *et al.*, *Nat. Genet.*, 2000, vol. 25, pp. 105–109.
68. Sarkar, P.S., Appukuttan, B., Han, J., *et al.*, *Nat. Genet.*, 2000, vol. 25, pp. 110–114.
69. Timchenko, L.T., *Am. J. Hum. Genet.*, 1999, vol. 64, pp. 360–364.
70. Timchenko, L.T., Miller, J.W., Timchenko, N.A., *et al.*, *Nucleic Acids Res.*, 1996, vol. 24, pp. 4407–4414.
71. Philips, A.V., Timchenko, L.T., and Cooper, T.A., *Science*, 1998, vol. 280, pp. 737–741.
72. Timchenko, N.A., Welm, A.L., Lu, X., and Timchenko, L.T., *Nucleic Acids Res.*, 1999, vol. 27, pp. 4517–4525.
73. Tian, B., White, R.J., Xia, T., *et al.*, *RNA*, 2000, vol. 6, pp. 79–87.
74. Mankodi, A., Logigian, E., Callahan, L., *et al.*, *Science*, 2000, vol. 289, pp. 1769–1772.
75. Raca, G., Siyanova, E.Y., McMurray, C.T., and Mirkin, S.M., *Nucleic Acids Res.*, 2000, vol. 28, pp. 3943–3949.
76. McMurray, C.T., *Chromosoma*, 1995, vol. 104, pp. 2–13.
77. Vologodskii, A., *Topology and Physics of Circular DNA*, Boca Raton: CRC, 1992.
78. Kohwi, Y., Wang, H., and Kohwi-Shigematsu, T., *Nucleic Acids Res.*, 1993, vol. 21, pp. 5651–5655.
79. Mitchell, J.E., Newbury, S.F., and McClellan, J.A., *Nucleic Acids Res.*, 1995, vol. 23, pp. 1876–1881.
80. Petruska, J., Arnheim, N., and Goodman, M.F., *Nucleic Acids Res.*, 1996, vol. 24, pp. 1992–1998.
81. Yu, A., Dill, J., Wirth, S.S., Huang, G., *et al.*, *Nucleic Acids Res.*, 1995, vol. 23, pp. 2706–2714.
82. Yu, A., Barron, M.D., Romero, R.M., *et al.*, *Biochemistry*, 1997, vol. 36, pp. 3687–3699.
83. Zheng, M., Huang, X., Smith, G.K., Yang, X., and Gao, X., *J. Mol. Biol.*, 1996, vol. 264, pp. 323–336.
84. Gacy, A.M., Goellner, G., Juranic, N., Macura, S., and McMurray, C.T., *Cell*, 1995, vol. 81, pp. 533–540.
85. Chen, X., Mariappan, S.V., Catasti, P., *et al.*, *Proc. Natl. Acad. Sci. USA*, 1995, vol. 92, pp. 5199–5203.
86. Usdin, K., *Nucleic Acids Res.*, 1998, vol. 26, pp. 4078–4085.
87. Gacy, A.M., Goellner, G.M., Spiro, C., *et al.*, *Mol. Cell*, 1998, vol. 1, pp. 583–593.
88. Sakamoto, N., Chastain, P.D., Parniewski, P., *et al.*, *Mol. Cell*, 1999, vol. 3, pp. 465–475.
89. Mathern, P., Goldmuntz, E.A., Zha, H., *et al.*, *Biochem. Genet.*, 1993, vol. 31, pp. 441–448.
90. Olds, R.J., Lane, D.A., Chowdhury, V., *et al.*, *Hum. Mutat.*, 1994, vol. 4, pp. 31–41.
91. Banfi, S., Servadio, A., Chung, M.Y., *et al.*, *Nat. Genet.*, 1994, vol. 7, pp. 513–520.
92. Burke, J.R., Wingfield, M.S., Lewis, K.E., *et al.*, *Nat. Genet.*, 1994, vol. 7, pp. 521–524.
93. Zhang, L., Leeftang, E.P., Yu, J., and Arnheim, N., *Nat. Genet.*, 1994, vol. 7, pp. 531–535.
94. Subramony, S.H., *Curr. Opin. Neurol.*, 1994, vol. 7, pp. 316–322.
95. Samadashwily, G.M., Dayn, A., and Mirkin, S.M., *EMBO J.*, 1993, vol. 12, pp. 4975–4983.
96. Dayn, A., Samadashwily, G.M., and Mirkin, S.M., *Proc. Natl. Acad. Sci. USA*, 1992, vol. 89, pp. 11406–11410.
97. Krasilnikov, A.S., Panyutin, I.G., Samadashwily, G.M., Cox, R., Lazurkin, Y.S., and Mirkin, S.M., *Nucleic Acids Res.*, 1997, vol. 25, pp. 1339–1346.
98. Baran, N., Lapidot, A., and Manor, H., *Proc. Natl. Acad. Sci. USA*, 1991, vol. 88, pp. 507–511.
99. Gerber, H.P., Seipel, K., Georgiev, O., *et al.*, *Science*, 1994, vol. 263, pp. 808–811.
100. Armour, J.A., Neumann, R., Gobert, S., and Jeffreys, A.J., *Hum. Mol. Genet.*, 1994, vol. 3, pp. 599–565.
101. Peter, M., Michon, J., Vielh, P., *et al.*, *Int. J. Cancer*, 1992, vol. 52, pp. 544–548.
102. Kohwi, Y. and Panchenko, Y., *Genes Dev.*, 1993, vol. 7, pp. 1766–1778.
103. Rooney, S.M. and Moore, P.D., *Proc. Natl. Acad. Sci. USA*, 1995, vol. 92, pp. 2141–2144.
104. Wahls, W.P., Wallace, L.J., and Moore, P.D., *Mol. Cell. Biol.*, 1990, vol. 10, pp. 785–793.
105. Weinreb, A., Collier, D.A., Birshtein, B.K., and Wells, R.D., *J. Biol. Chem.*, 1990, vol. 265, pp. 1352–1359.
106. Wahls, W.P., Wallace, L., and Moore, P.D., *Cell*, 1990, vol. 60, pp. 95–103.

107. Kunkel, T.A., *Nature*, 1993, vol. 365, pp. 207–208.
108. Kang, S., Ohshima, K., Shimizu, M., Amirhaeri, S., and Wells, R.D., *J. Biol. Chem.*, 1995, vol. 270, pp. 27014–27021.
109. Usdin, K. and Woodford, K.J., *Nucleic Acids Res.*, 1995, vol. 23, pp. 4202–4209.
110. Ohshima, K. and Wells, R.D., *J. Biol. Chem.*, 1997, vol. 272, pp. 16798–16806.
111. Shimizu, M., Kubo, K., Matsumoto, U., and Shindo, H., *J. Mol. Biol.*, 1994, vol. 235, pp. 185–197.
112. Freudenreich, C.H., Stavenhagen, J.B., and Zakian, V.A., *Mol. Cell. Biol.*, 1997, vol. 17, pp. 2090–2098.
113. Kang, S., Ohshima, K., Jaworski, A., and Wells, R.D., *J. Mol. Biol.*, 1996, vol. 258, pp. 543–547.
114. Gordenin, D.A., Kunkel, T.A., and Resnick, M.A., *Nat. Genet.*, 1997, vol. 16, pp. 116–118.
115. White, P.J., Borts, R.H., and Hirst, M.C., *Mol. Cell. Biol.*, 1999, vol. 19, pp. 5675–5684.
116. Schweitzer, J.K. and Livingston, D.M., *Hum. Mol. Genet.*, 1998, vol. 7, pp. 69–74.
117. Schweitzer, J.K. and Livingston, D.M., *Genetics*, 1999, vol. 152, pp. 953–963.
118. Kokoska, R.J., Stefanovic, L., Tran, H.T., *et al.*, *Mol. Cell. Biol.*, 1998, vol. 18, pp. 2779–2788.
119. Parenteau, J. and Wellinger, R.J., *Mol. Cell. Biol.*, 1999, vol. 19, pp. 4143–4152.
120. Reiss, A.L., Kazazian, H., Jr., Krebs, C.M., *et al.*, *Hum. Mol. Genet.*, 1994, vol. 3, pp. 393–398.
121. Kunst, C.B. and Warren, S.T., *Cell*, 1994, vol. 77, pp. 853–861.
122. Snow, K., Tester, D.J., Kruckeberg, K.E., Schaid, D.J., and Thibodeau, S.N., *Hum. Mol. Genet.*, 1994, vol. 3, pp. 1543–1551.
123. Jodice, C., Malaspina, P., Persichetti, F., *et al.*, *Am. J. Hum. Genet.*, 1994, vol. 54, pp. 959–965.
124. Trinh, T.Q. and Sinden, R.R., *Nature*, 1991, vol. 352, pp. 544–547.
125. Samadashwily, G.M., Raca, G., and Mirkin, S.M., *Nat. Genet.*, 1997, vol. 17, pp. 298–304.
126. Kornberg, A. and Baker, T., *DNA Replication*, 2nd ed., New York: W. H. Freeman and Co., 1992.
127. Brewer, B.J. and Fangman, W.L., *Cell*, 1987, vol. 51, pp. 463–471.
128. Huberman, J.A., Spotila, L.D., Nawotka, K.A., El-Assouli, S.M., and Davis, L.R., *Cell*, 1987, vol. 51, pp. 473–481.
129. Zhu, L.J., Newlon, C.S., and Huberman, J.A., *Mol. Cell. Biol.*, 1992, vol. 12, pp. 4733–4741.
130. Little, R.D., Platt, T.H., and Schildkraut, C.L., *Mol. Cell. Biol.*, 1993, vol. 13, pp. 6600–6613.
131. MacAllister, T., Khatri, G.S., and Bastia, D., *Proc. Natl. Acad. Sci. USA*, 1990, vol. 87, pp. 2828–2832.
132. Friedman, K.L. and Brewer, B.J., *Meth. Enzymol.*, 1995, vol. 262, pp. 613–627.
133. Richards, R.I., Holman, K., Friend, K., *et al.*, *Nat. Genet.*, 1992, vol. 1, pp. 257–260.
134. Paques, F. and Haber, J.E., *Microb. Molec. Biol. Rev.*, 1999, vol. 63, pp. 349–404.
135. Jakupciak, J.P. and Wells, R.D., *J. Biol. Chem.*, 1999, vol. 274, pp. 23468–23479.
136. Richard, G.-F., Goellner, G.M., McMurray, C.T., and Haber, J.E., *EMBO J.*, 2000, vol. 19, pp. 2381–2390.
137. Michel, B., *TIBS*, 2000, vol. 25, pp. 173–178.
138. Kuzminov, A., *Mol. Microbiol.*, 1995, vol. 16, pp. 373–384.
139. Monckton, D.G., Cayuela, M.L., Gould, F.K., *et al.*, *Hum. Mol. Genet.*, 1999, vol. 8, pp. 2473–2478.
140. Koefoed, P., Hasholt, L., Fenger, K., *et al.*, *Hum. Genet.*, 1998, vol. 103, pp. 564–569.
141. Grewal, R.P., Leeftang, E.P., Zhang, L., and Arnheim, N., *Neurogenetics*, 1998, vol. 1, pp. 249–252.
142. Leeftang, E.P., Tavaré, S., Marjoram, P., *et al.*, *Hum. Mol. Genet.*, 1999, vol. 8, pp. 173–183.
143. Telenius, H., Kremer, B., Goldberg, Y.P., *et al.*, *Nat. Genet.*, 1994, vol. 6, pp. 409–414.
144. Nolin, S.L., Houck, G.E.J., Gargano, A.D., *et al.*, *Am. J. Hum. Genet.*, 1999, vol. 65, pp. 680–688.
145. Reyniers, E., Vits, L., De Boule, K., *et al.*, *Nat. Genet.*, 1993, vol. 4, pp. 143–146.
146. Satoh, N., *Differentiation*, 1982, vol. 22, pp. 156–163.
147. Yano, H., Wang, B.E., Ahmad, I., *et al.*, *Exp. Cell. Res.*, 1999, vol. 251, pp. 388–400.
148. Deissler, H., Behn-Krappa, A., and Doerfler, W., *J. Biol. Chem.*, 1996, vol. 271, pp. 4327–4334.
149. McLaughlin, B.A., Spencer, C., and Eberwine, J., *Am. J. Hum. Genet.*, 1996, vol. 59, pp. 561–569.