Interruptions in the Triplet Repeats of SCA1 and FRAXA Reduce the Propensity and Complexity of Slipped Strand DNA (S-DNA) Formation†

Christopher E. Pearson,*,‡ Evan E. Eichler,||,⊥ Diego Lorenzetti,|| Shannon F. Kramer,‡ Huda Y. Zoghbi,§,|| David L. Nelson,|| and Richard R. Sinden‡

Center for Genome Research, Institute of Biosciences and Technology, Department of Biochemistry and Biophysics, Texas A&M University, Houston, Texas, 77030, and Department of Pediatrics and Howard Hughes Medical Institute and Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, 77030

Received October 14, 1997; Revised Manuscript Received December 17, 1997

ABSTRACT: Models for the disease-associated expansion of trinucleotide repeats involve the participation of alternative DNA structures during replication, repair, or recombination. CAT or AGG interruptions within the $(CAG)_n$ or $(CGG)_n$ repeats of SCA1 or FRAXA, respectively, confer increased genetic stability to the repeats. In this study, we report the formation of slipped strand structures (S-DNA) using genomic sequences containing pure and interrupted SCA1 and FRAXA repeats having lengths above and below the genetic stability thresholds. S-DNA forms within the repeats during annealing of complementary strands containing equal lengths of repeats. Increased lengths of pure repeats led to an increased propensity for S-DNA formation. CAT or AGG interruptions have both quantitative and qualitative effects upon S-DNA formation: they decrease the total amount of slipped structures as well as limit the specific isomers formed. This demonstrates a unifying inhibitory effect of interruptions in both $(CAG)_n$ and $(CGG)_n$ tracts. We also present transmission stability data for SCA1 and FRAXA alleles spanning the thresholds and compare these with the ability to form slipped structures. The effect of both the length and purity of the repeat tract on the propensity of slipped structure formation correlates with their effect on genetic instability and disease, suggesting that S-DNA structures may be models for mutagenic intermediates in instability.

The expansion of triplet repeats is associated with human disease (*1*−*3*). Nine different disease loci have unstable $(CAG)_n$·$(CTG)_n$ repeats as the underlying mutation, while five chromosomal fragile sites are caused by unstable $(CGG)_n$·$(CCG)_n$ repeats. The molecular details of the mechanism of expansion, an important new form of mutation, are unknown. The expansion mutation of the triplet repeats is strongly dependent upon the repeat tract length, such that the expanded alleles are more likely to undergo further expansion than their predecessor alleles (*4*). Some properties of these repeats, possibly the ability to form inter- or intrastrand structures (*5*−*8*), may be important for their genetic instability. To understand the molecular events that may occur during the expansion process, we performed a series of biochemical experiments using biologically and genetically relevant clones. We recently presented the first biochemical evidence for the existence of slipped strand DNA structures, termed S-DNA[1] (*6*−*8*). S-DNA was formed following the denaturation and renaturation (reduplexing) of DNA containing $(CAG)_n$, $(CTG)_n$, or $(CGG)_n$ repeats of the SCA1, DM, or FRAXA loci, respectively. S-DNAs are composed of complementary strands with equal lengths of triplet repeats in each strand. The DNA secondary structure occurs within the repeat tract and is subtended by linear duplex, mixed sequence DNA. S-DNA formation does not require superhelical tension, the structures are remarkably thermostable, and they display minimal interconversion between isomers and/or the linear duplex form under physiological conditions. S-DNA structures likely form by out-of-register base pairing between complementary strands with hairpin formation in slipped-out regions (*6*−*8*).

Sequence interruptions within human di- and trinucleotide repeat tracts confer increased genetic stability to the repeat tracts (*1, 3, 9*). In three of the triplet disease loci, SCA1, SCA2, and FRAXA, the repeat tracts are normally interrupted with CAT, CAA, and AGG, respectively (*1*,*3*). In nonaffected individuals, the $(CAG)_n$ tract of SCA1 contains 1−3 CAT interruptions (*10, 11*), while the $(CGG)_n$ tract of FRAXA contains 1−3 AGG interruptions (*12*−*19*). In each case, the interruptions confer increased genetic stability to the repeat tract, such that upon transmission the interrupted tracts are less likely to expand. Loss of the interruptions, resulting in a longer length of the pure tract, correlates with instability and disease. In each of the disease-associated

† This work was supported by funds from the Institute of Biosciences and Technology and in part from NIH Grant GM52982 and the Howard Hughes Medical Institute. D.L. is a recipient of an Italian Telethon fellowship.

* To whom correspondence should be addressed. Tel.: +1 713 677-7676. Fax: +1 713 677-7689. E-mail: cpearson@ibt.tamu.edu.

‡ Texas A&M University.

§ Department of Pediatrics and Howard Hughes Medical institute, Baylor College of Medicine.

|| Molecular Human Genetics, Baylor College of Medicine.

⊥ Present address: Department of Genetics, Case Western Reserve University, Cleveland Ohio, 44106.

[1] Abbreviations: S-DNA, homoduplex slipped strand DNA, SI-DNA heteroduplex slipped intermediate DNA; SCA1, spino-cerebellar ataxia type 1; DM, myotonic dystrophy; FRAXA, fragile XA mental retardation; PCR, polymerase chain reaction.

S0006-2960(97)02546-4 CCC: $15.00
Published on Web 02/06/1998

trinucleotide repeat tracts, instability and disease onset occurs within a window of between 35 and 40 pure repeats. The stability threshold lengths at which increased instability and disease transmission occur are ≥40 pure (CAG) and ≥34 pure (CGG) for SCA1 and FRAXA, respectively (*11*, *13*). Interruptions within the $(CAG)_n$ tract in the DM locus (*20*), the $(CGG)_n$ tract of the autosomal fragile site FRA16A (*21*), and the $(GAA)_n$ tract in the Friedreich's ataxia loci (*22*, *23*) may also confer increased genetic stability to these repeats. The reasons for the stabilizing effect of the interruptions are not well-understood. It has been suggested that interruptions may provide genetic stability to the repeat tracts by inhibiting strand slippage, although there has been no biophysical evidence to support this. In this paper, we examine the effect of interruptions on the formation of S-DNA using SCA1 and FRAXA repeats, with repeat lengths above and below the genetic stability thresholds.

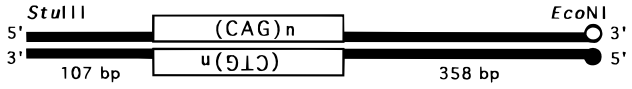
EXPERIMENTAL PROCEDURES

Cloning of SCA1 and FRAXA Trinucleotide Repeat Tracts. We cloned human SCA1 and FRAXA cDNA and genomic repeat tracts, respectively, to obtain highly purified DNA samples containing repeat lengths and interruption patterns that are present within the human population. Because triplet repeats are inherently unstable in *Escherichia coli*, it was necessary to screen many colonies to obtain proper clones. To maximize stability and facilitate cloning, standard precautions were taken to avoid deletions (*6*, *24*). Once obtained, all plasmids were stably propagated under appropriate conditions (*6*, *24*). Preparation and purification of plasmids were as previously described (*6*, *7*, *24*).

The cDNA clone, N4-3 in pBluescript, containing the entire coding region of the SCA1 gene with an interrupted 30 CAG allele (*25*) was used as a backbone for the production of the other SCA1 clones by replacement of the *Sfi*I fragment. *Sfi*I fragments containing CAG tracts of 30, 44, 49, 60, and 74 repeats were generated by PCR from genomic DNA of SCA1 patients with various repeat configurations using primers flanking the repeat (post-2, 5′-CCCGTCACCAGTGCAGTGG-3′; 2-sap, 5′-TGAGGAACCGACTTGCCG-3′). Patient genomic DNA was kindly provided by B. W. Popovich for the production of the longer than threshold length multiply interrupted clone pAG44-24B. pAG74Δ4-1 containing 30 pure repeats, is an SCA1 configuration not observed within the human population and is an *E. coli*-derived deletion product of the pAG74 clone. The repeat configuration of each clone was confirmed by sequence analysis.

PCR amplification of genomic DNAs from individuals having various FRAXA repeat interspersion patterns (Figure 1) utilized primers and conditions previously described (*13*). The pFXA clones were created by insertion of the PCR products into PCRscript (Stratagene). The repeat configuration of each clone was confirmed by sequence analysis.

DNA Treatments and Analysis. Reduplexing reactions were performed as previously described in detail (*6*, *7*). Briefly, linearized and uniquely radiolabeled genomic clones containing triplet repeats were alkaline denatured then reannealed by neutralization. Repeat-containing fragments were liberated by a second restriction digestion and products analyzed by polyacrylamide gel electrophoresis. Gels were dried and exposed to radiographic film, or to a PhosphorImager screen (Molecular Dynamics). Quantitative analysis was performed using the PhosphorImager with ImageQuant software (Molecular Dynamics). The fraction of the repeat-containing fragments migrating as novel anomalous products represents the average of quantitative analysis from 3 to 7 experiments. As previously discussed (*6*, *7*), in addition to the slow-migrating S-DNAs, reduplexing or hetero duplexing of $(CAG)_n$·$(CTG)_n$ or $(CGG)_n$·$(CCG)_n$ repeats results in products that were unable to enter the gel. This DNA may represent incompletely rehybridized strands or may also be composed of higher-order intermolecular complexes. However, unlike the slow migrating S-DNAs, neither the length nor the purity of the repeat tract appeared to dramatically affect the amount of DNA retained in the wells (Figures 2 and 5, C. E. Pearson, data not shown). For this reason, only the DNA that entered the gel was included in quantitative analyses.



SCA1 (CAG)·(CTG) clones

clone	total # repeats	repeat configuration
N4-3	30	12+1+15
pAG74 Δ4-1	30	pure
pAG44-24B	44	12+1+12+1+14
pAG49	49	pure
pAG60	60	pure
pAG74	74	pure

FIGURE 1: SCA1 clones. The upper portion shows a map of the *Eco*NI/*Stu*I SCA1 (CAG)·(CTG) insert. Circles represent the position of the radiolabel, bold lines represent human flanking nonrepetitive sequences. The lower portion of the figure shows the repeat configuration of the various SCA1 cDNA clones. Clones are identical but for their repeat configuration, with CAT interruptions indicated by + signs: $V+W+X+Y+Z = (CAG)_V CAT(CAG)_W CAT(CAG)_X CAT(CAG)_Y CAT(CAG)_Z$.

RESULTS

Effect of CAT Interruptions within the $(CAG)_n$ Tract of SCA1 on S-DNA Formation. To study the effect of repeat length and CAT interruptions on S-DNA formation, we analyzed S-DNA in pure or interrupted SCA1 cDNA clones containing from 30 to 74 CAG repeats (Figure 1), representing normal and expanded diseased alleles, respectively. Reduplexing the SCA1 DNAs containing 30, 49, 60, and 74 pure repeats resulted in slow-migrating S-DNAs (Figure 2A). As discussed previously (*6*, *8*), the reduced mobility of S-DNA is consistent with that expected from the introduction of kinks in the DNA resulting from the junctions associated with slipped-out repeats (*26*). The heterogeneous population of S-DNAs is indicative of a family of structural isomers with bends located at different (or multiple) positions within the DNA molecules (Figure 2A−C, see all lanes containing reduplexed samples). The percentage of S-DNA formed increased with longer repeat lengths (Figure 2A,D). In addition, there was an increase in the number of major S-DNA isomers formed for longer repeat tracts, indicating

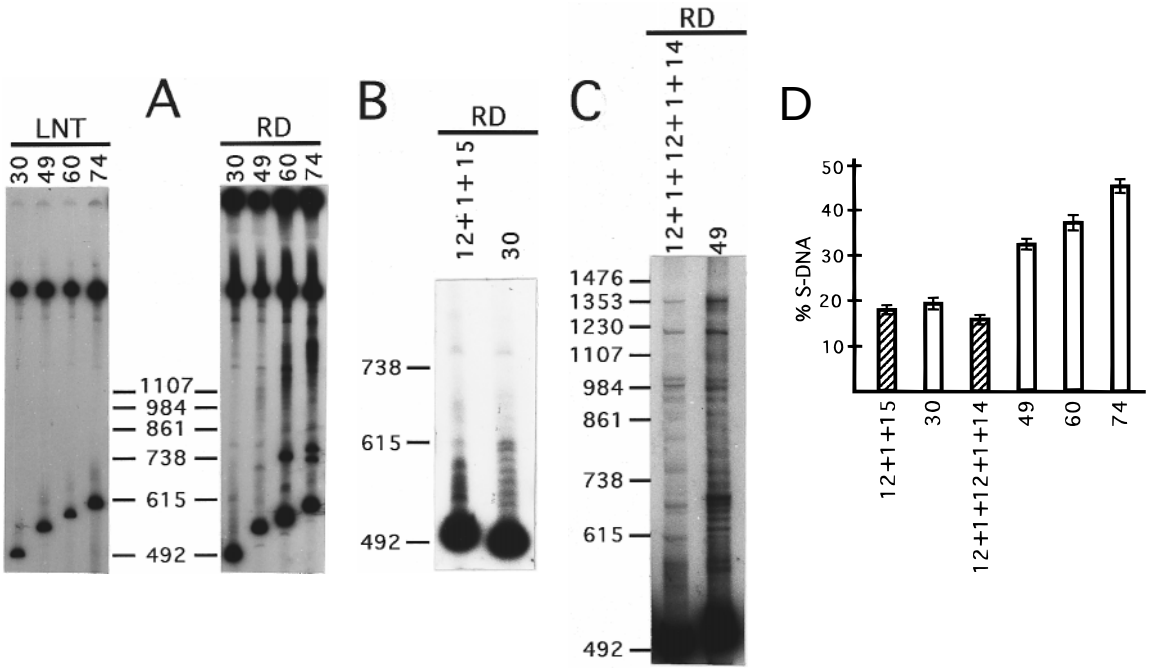


FIGURE 2: S-DNA in (CAG)·(CTG) repeats of SCA1. (A) Reduplex analysis of cDNA clones of the human SCA1 gene (Figure 1). Plasmids were linearized by *Eco*NI digestion, ^{32}P-labeled, reduplexed, and then digested with *Stu*I to liberate the (CAG)·(CTG) insert (Figure 1, top). The pattern of slow-migrating products was identical whether the radiolabel was on the CTG or the CAG strand, indicating that these structures are composed of both strands. The linear nontreated (LNT), the reduplexed (RD) DNAs, and the positions of the 123 bp ladder are shown. (B) Reduplex analysis of the multiply-interrupted and pure 30 repeat alleles. A magnification of the lower portion of the gel is shown. (C) Reduplex analysis of the multiply-interrupted 44 and pure 49 repeat alleles. A magnification of the lower portion of the gel is shown. (D) Densitometric analysis of the reduplexed samples. S-DNA is shown as a percentage of the repeat-containing fragment. Pure and interrupted repeats are represented by hollow and crosshatched bars, respectively. Using the *t* test all comparisons of S-DNA formation by one allele with another, excluding that of the 30 pure with the 12+1+15, were found to be different with *p*-values ranging from 0.0002 to 0.0422, indicating that the differences were statistically significant. Comparison of S-DNA formation by the 30 pure with the 12+1+15 allele was not significantly different, having a *p*-value of 0.1042.

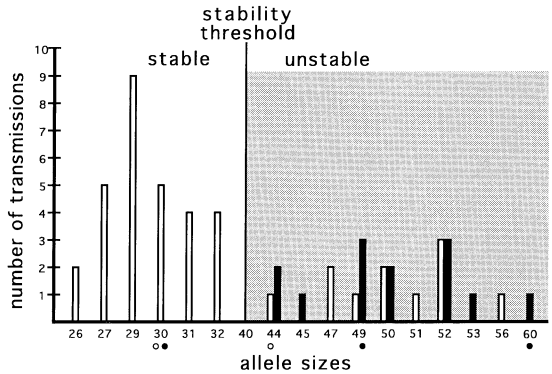
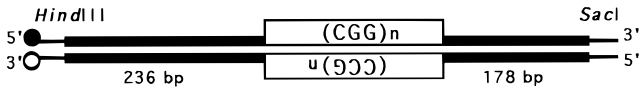


FIGURE 3: Transmission stability of the (CAG)·(CTG) repeat tract of SCA1. The length of the (CAG)·(CTG) repeat tract (in repeat number) is compared against the number of stable (hollow bars) and unstable (filled bars) parent-sibling transmissions observed for each length. All repeat tracts having greater than 39 repeats are pure (except for a single multiply-interrupted $n = 44$ repeat), while all repeat tracts having less than 40 repeats are interrupted. The interrupted alleles used in this study are indicated below the *x*-axis by hollow circles, while the pure alleles used in this study are represented by filled circles. Not all alleles examined structurally are represented by the various transmission class configurations.



FRAXA (CGG)·(CCG) clones

clone	total # repeats	repeat configuration
pFXA17	17	pure
pFXA21	21	pure
pFXA9 +12	22	9+12
pFXA9 +27	37	9+27
pFXA39	39	pure
pFXA9 +9+9+9	39	9+9+9+9

FIGURE 4: FRAXA clones. The upper portion shows a map of the *Hin*dIII/*Sac*I FRAXA (CGG)·(CCG) insert. Circles represent the position of the radiolabel, bold lines represent human flanking nonrepetitive sequences, and thin lines represent plasmid sequences. The lower portion of the figure shows the repeat configuration of the various genomic FRAXA clones. Clones are identical but for their repeat configuration, with AGG interruptions indicated by + signs: $V+W+X+Y+Z = (CGG)_V AGG(CGG)_W AGG(CGG)_X AGG(CGG)_Y AGG(CGG)_Z$.

greater structural complexity with increased tract length. These results are similar to those obtained with DM clones (*6*−*8*). To determine the effect of sequence interruptions, S-DNA formation in the 12+1+15 allele (where + signs represent CAT interruptions), frequent among unaffected individuals, was compared to that in an equal length, uninterrupted repeat (Figure 2B). Both alleles formed a reproducible pattern of slow-migrating products. The pure tract formed a series of S-DNA isomers migrating as distinct bands while the interrupted tract formed a population of S-DNA isomers migrating as closely spaced products with fewer distinct bands (Figure 2B). The longest stably transmitted SCA1 repeat tract in an unaffected individual is 44 repeats in length (*27*), having the configuration 12+1+12+1+14 (Figure 1). This rare allele is longer than

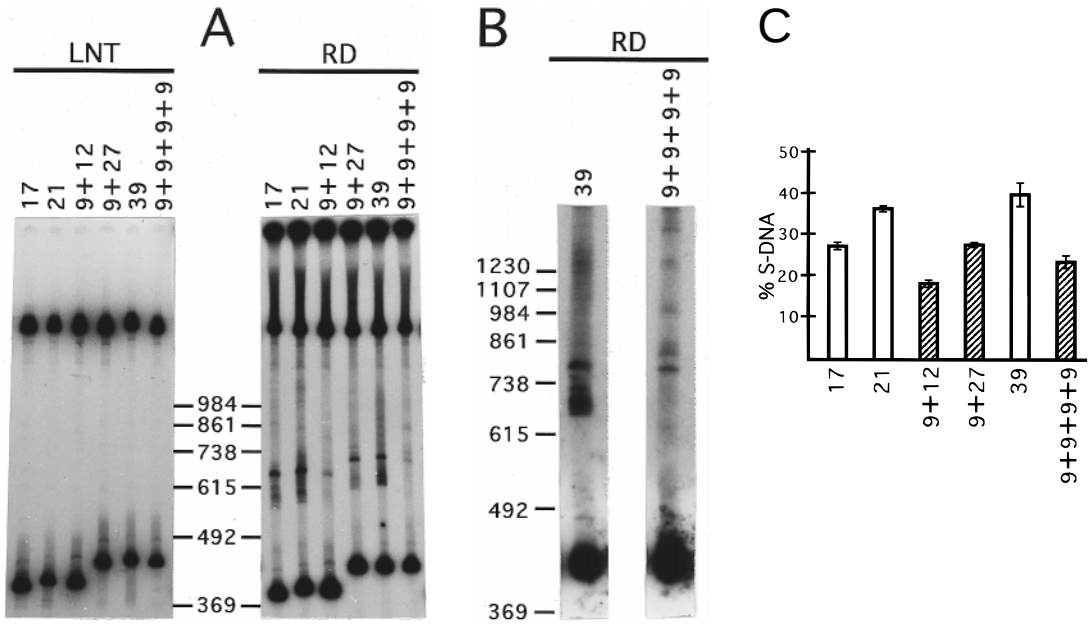


FIGURE 5: S-DNA in (CGG)·(CCG) repeats of FRAXA. (A) Reduplex analysis of genomic clones of the human FRAXA locus (Figure 4). Plasmids were linearized by *Hin*dIII digestion, ^{32}P-labeled, reduplexed, and then digested with *Sac*I to liberate the (CGG)·(CCG) insert (Figure 4, top). The pattern of slow-migrating products was identical whether the radiolabel was on the CGG or the CCG strand, indicating that these structures are composed of both strands. The linear nontreated (LNT), the reduplexed (RD) DNAs, and the positions of the 123 bp ladder are shown. (B) Reduplex analysis of the multiply-interrupted and pure 39 repeat alleles. A magnification of the lower portion of the gel is shown. (C) Densitometric analysis of the reduplexed samples. S-DNA is shown as a percentage of the repeat-containing fragment. Pure and interrupted repeats are represented by hollow and crosshatched bars, respectively. Using the *t* test all comparisons of S-DNA formation by one allele with another, excluding that of the 17 pure with the 9+27, were found to be different with *p*-values ranging from 0.0002 to 0.0286, indicating that the differences were statistically significant. Comparison of S-DNA formation by the 17 pure with the 9+27 allele was not significantly different, having a *p*-value of 0.5334.

the shortest expanded tract of 40 pure repeats found in an SCA1 individual (*10, 11*) and above the genetic stability threshold of 40 repeats. Comparison of the propensity of this multiply interrupted allele to form S-DNA with the similar-length tract of 49 pure CAG repeats is shown in Figure 2C. The interrupted 44 repeat allele formed only 15.8% S-DNA compared to either the similar-length 49 pure tract or the shorter pure 30 repeat allele, which formed 32.5% and 19.3% S-DNA, respectively (Figure 2D). The multiply-interrupted 44 repeat allele also reproducibly formed fewer S-DNA isomers than the 49 pure allele, with most isomers appearing to be a subset of the products formed by the 49 repeat allele (Figure 2C).

CAT Interruptions within the (CAG)$_n$ Tract of SCA1 Provide Genetic Stability to the Repeat Tract. Repeat length is known to affect the genetic stability of trinucleotide repeat tracts (*1, 4, 10, 11*); however, a detailed presentation of the effect of the SCA1 tract length on transmission stability has not been published. Analysis of the (CAG)$_n$ tract of the SCA1 locus in parent−sibling transmissions of normal and diseased individuals demonstrates that the repeat tract is stably transmitted (i.e., no length variations occur) below 40 repeats (Figure 3). As the length of the repeat tract increases above 40 repeats, the probability for genetic length alterations increases. All SCA1 alleles above 39 repeats (excluding a rare case, see below) are pure CAG and found in SCA1 patients (Figure 3, refs *1, 3, 10,* and *11*). There is a single report of an unusually long SCA1 (CAG)$_n$ tract (n = 44 repeats) which was found to be stably transmitted (Figure 3, ref 27). Sequence analysis of this tract revealed that it was multiply-interrupted with CAT units (see above, Figure 1; clone pAG44-24B). In Figure 3, the repeat configuration classes containing some of the alleles used in the present study are indicated. The increased propensity to form S-DNA of longer pure (CAG)$_n$ tracts correlates with their reduced in vivo genetic stability, while the reduced propensity to form S-DNA of the interrupted (CAG)$_n$ tracts correlates with their increased in vivo genetic stability. Together, these data suggest that the loss of the CAT interruptions from the SCA1 (CAG)$_n$ tract may be an important mutagenic event (*1, 10, 11*) and that S-DNA structures may be intermediates during tract length changes.

Effect of AGG Interruptions within the (CGG)$_n$ Tract of FRAXA on S-DNA Formation. To study the effect of AGG interruptions on S-DNA formation, we analyzed S-DNA in pure or interrupted FRAXA genomic clones containing from 17 to 39 CGG repeats (Figure 4). Reduplex analysis of the pure tracts reproducibly showed an increase in the percentage of S-DNA formed as the tract length increased (Figure 5A, C). A reproducible increase in the relative amounts of certain major S-DNA isomers for longer pure repeat tracts was observed, indicating an increase in structural complexity with increased tract length (Figure 5). The presence of AGG interruptions reduced the percentage of S-DNA formed. The pure 39 repeat allele formed 39.6% S-DNA, while the equal-length, multiply-interrupted 9+9+9+9 allele (where + signs represent AGG interruptions) formed only 23.2% S-DNA (Figure 5A−C). The 23.2% S-DNA formed by the multiply-interrupted 9+9+9+9 was less than either the 27% or the 36.1% formed by the shorter pure 17 and 21 repeat alleles, respectively (Figure 5A,C). The presence of even a single AGG interruption had a similar inhibitory effect on S-DNA formation (Figure 5A,C), as seen from a comparison of the 9+12 allele, which formed 18% S-DNA, with either the

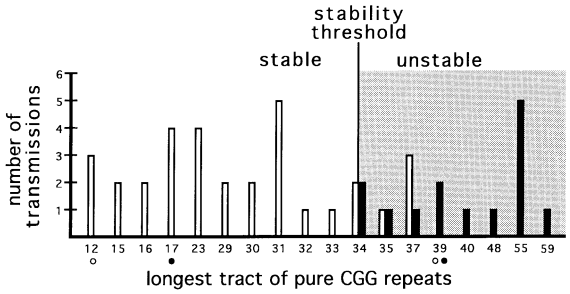


FIGURE 6: Transmission stability of the (CGG)·(CCG) repeat tract of FRAXA. The longest tract of uninterrupted (CGG)·(CCG) repeats (repeat number) is compared against the number of stable (hollow bars) and unstable (filled bars) parent−sibling transmissions observed for each length. The transmissions reported here represent published (*13*) and new observations. The interrupted alleles used in this study are indicated below the *x*-axis by hollow circles, while the pure alleles used in this study are represented by filled circles. Not all alleles examined structurally are represented by the various transmission class configurations.

similar-length pure 21 repeat allele or the shorter pure 17 repeat allele, which formed 36.1% and 27% S-DNA, respectively. Similarly, the 9+27 allele formed only 27.4% S-DNA, while the similar-length pure 39 tract or the shorter pure 21 tract formed 39.6% and 36.1% S-DNA, respectively (Figure 5A and C). Interestingly, the presence of an AGG interruption in the 9+27 allele reduced its propensity to form S-DNA to the level of the 17 pure allele, both of which were less than that of the pure 21 allele (Figure 5C). The structural complexity of S-DNAs formed was also affected by the presence of AGG interruptions. The electrophoretic pattern of the 9+27 S-DNA products showed a reduction in the relative amounts of specific isomers compared with the pattern of the 39 pure S-DNA products. A similar difference is observed between the 9+12 and the 21 pure alleles. A darker exposure of the pure 39 repeat and the 9+9+9+9 alleles showed that fewer S-DNA isomers were formed with the multiply-interrupted allele (Figure 5B).

AGG Interruptions within the $(CGG)_n$ Tract of FRAXA Provide Genetic Stability to the Repeat Tract. Sequence analysis of the $(CGG)_n$ tract of the FRAXA locus in normal, premutant, and diseased individuals revealed that premutant and diseased alleles contain fewer AGG interruptions and have longer pure $(CGG)_n$ tracts (*1, 4, 12*−*19*). The genetic stability of the FRAXA $(CGG)_n$ tract is known to be dependent upon the length of the longest pure tract of CGG units, as we and others have shown (*1, 4, 12*−*19*). An updated analysis of the longest pure $(CGG)_n$ tract of the FRAXA locus in parent−sibling transmissions demonstrates that the repeat tract is stably transmitted (i.e., no length variations occur) below 34 pure repeats (Figure 6). This figure contains new data points encompassing the repeat length range of our clones. As the length of the repeat tract increases above 33 pure repeats, the probability for genetic length alterations increases. Above the length of 59 pure repeats (the premutant range is from 59 to 230 repeats), the probability for genetic length alterations approaches 100% (*1, 12*−*19*). This analysis is consistent with observations by others (*28*). Alleles with pure tracts longer than 33 repeats can undergo mutation, while those with pure tracts shorter than 33 repeats are genetically stable. Thus, two FRAXA repeat tracts of the same length can have very different genetic stabilities. Both the progressive loss of AGG interruptions from normal to premutant to fully expanded repeat tracts and the data on the genetic transmission instability show that AGG interruptions provide genetic stability to the tracts. In Figure 6 the repeat configuration classes containing some of the alleles used in the present study are indicated. As observed for the SCA1 alleles, the increased propensity to form S-DNA of longer pure $(CGG)_n$ tracts correlates with their reduced in vivo genetic stability, while the reduced propensity to form S-DNA of the interrupted $(CGG)_n$ tracts correlates with their increased in vivo genetic instability. Together, these data suggest that the loss of AGG interruptions from the FRAXA $(CGG)_n$ tracts, resulting in genetically unstable longer pure tracts, may be an important mutagenic event (*1, 4, 12*−*19*) and that S-DNA structures may be intermediates during tract length changes.

DISCUSSION

The results shown above demonstrate an effect of both the length and purity of $(CAG)_n$·$(CTG)_n$ and $(CGG)_n$·$(CCG)_n$ repeat tracts on slipped strand structure formation. Both the propensity for S-DNA formation increases and the complexity of S-DNAs formed increases with the length of the pure repeat tract, while sequence interruptions dramatically reduce the propensity for formation and the complexity of S-DNAs. Longer lengths of pure SCA1 $(CAG)_n$·$(CTG)_n$ and FRAXA $(CGG)_n$·$(CCG)_n$ repeats led to an increased propensity for S-DNA formation. These results are similar to those obtained with DM $(CTG)_n$·$(CAG)_n$ clones (*6*−*8*). In a previous study we reported 39% S-DNA formation in a pure $(CTG)_{30}$·$(CAG)_{30}$ DM genomic clone (*7*), which differs from the amount formed by a similar length of repeats in an SCA1 clone. Within each set of clones (DM, SCA1, or FRAXA), all plamids are identical except for the repeat tract length and pattern of interruptions. Significantly, each set contains the nonrepetitive flanking sequences found in their respective chromosomal sites. Thus, direct comparisons of S-DNA formation can only be made between different clones within a set. The different fraction of S-DNA formed for equal $(CTG)_n$·$(CAG)_n$ repeat lengths in DM or SCA1 DNAs demonstrates that nonrepetitive sequences flanking the repeat can influence the percentage and pattern of S-DNA products formed. It is very possible that different chromosomal locations having different sequences flanking (in *cis* to) the triplet repeat tracts will have an affect upon the genetic stability of the repeat tracts. Interestingly, in addition to the various rates of length alteration for each of the trinucleotide repeat loci, the range of the expansion of the DM $(CTG)_n$·$(CAG)_n$ repeat (50−3000 repeats) is greater than that for the SCA1 $(CAG)_n$·$(CTG)_n$ (40−81 repeats) (reviewed in reference *1*). In addition to other factors (such as selection and viability), the flanking sequences may affect the variable rates and magnitudes of genetic change exhibited by the individual unstable loci.

The increase in the structural complexity of S-DNA observed for longer pure repeat tracts is likely due to an increase in the number of modes of out-of-register mispairings. Longer tracts can display an increase in both the number of slip-outs, the number of possible nucleation points for slip-outs, and the number of repeat units contained in a slip-out. This can result in a greater number of slipped isomers formed by longer tracts. The existence of multiple

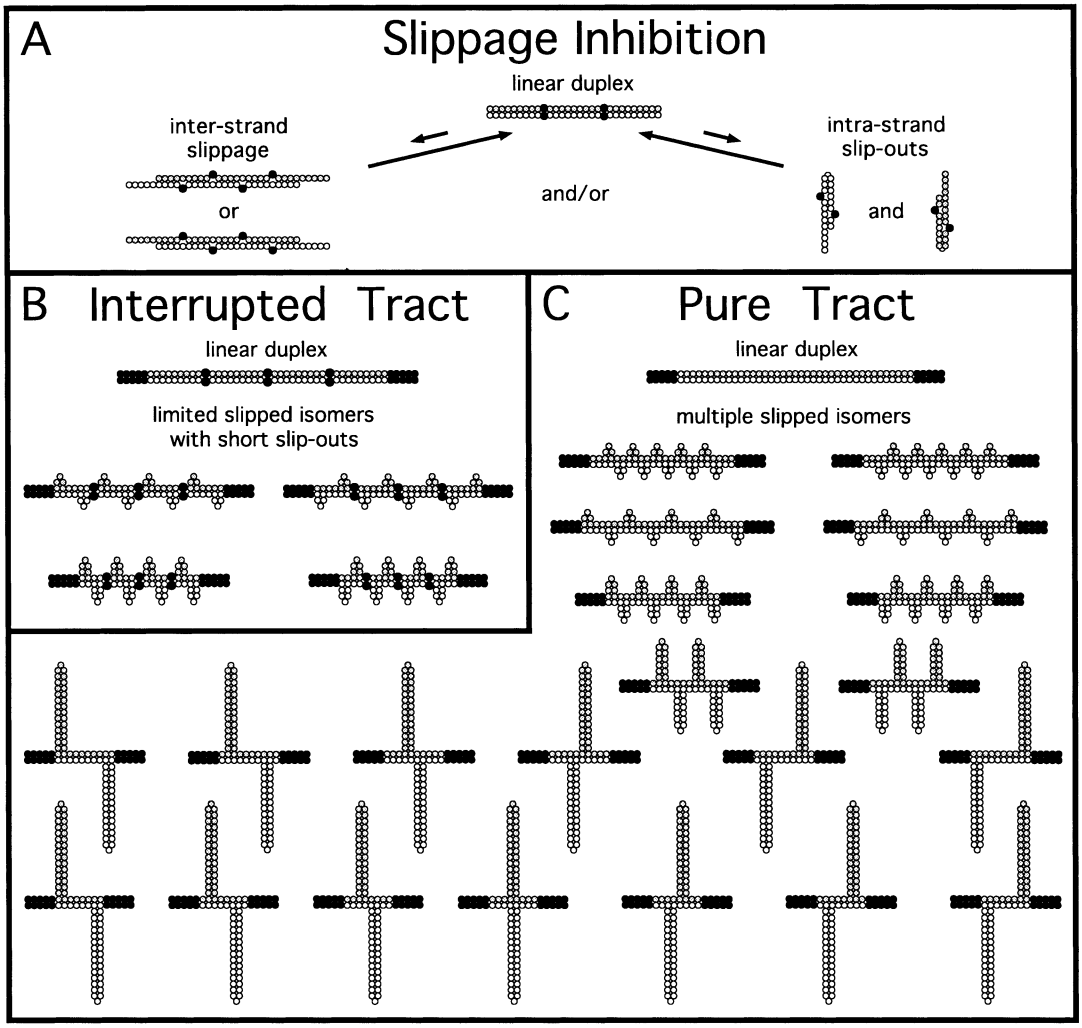


FIGURE 7: Models of possible protective mechanisms of repeat interruptions. (A) For both SCA1 and FRAXA, the open circles represent (CAG)·(CTG) and (CGG)·(CCG) repeats, respectively, while the filled circles represent the (CAT)·(ATG) and (AGG)·(CCT) interruptions. (B) The reduction in S-DNA isomers formed in the interrupted tracts may be due to a limitation on the positions of, and hence, the number of slipped-out repeats. (C) The pure tracts can have slipped-out nucleation points throughout the repeat tract, as well as a greater number of repeats contained in a slip-out. See text for discussion. In both B and C the filled circles on the ends represent nonrepetitive flanking sequences.

structural isomers of gel-purified S-DNAs has been confirmed by electron microscopic analysis (*8*).

Interrupted FRAXA or SCA1 repeat tracts formed reduced amounts of S-DNA relative to similar length pure alleles. The reduced propensity to form S-DNA by interrupted repeats is a striking biophysical effect of a single base pair change (a G·C to a T·A for both SCA1 and FRAXA). Interruptions in both SCA1 or FRAXA repeats caused qualitative changes by reducing the number of specific S-DNA products. In the case of the pure and interrupted $n = 30$ SCA1 alleles, although the *quantitative* difference between the two is not large, the *qualitative* difference (i.e., the electrophoretic patterns) between the two is drastic. Again, this is a striking biophysical effect of a single base pair change. The effect of interruptions on the number of S-DNA isomers formed in the interrupted tracts likely results from a reduction in the number, the size, and location of slip-outs relative to those formed in pure tracts, which can occur throughout the repeat tract (Figure 7).

The expansion of trinucleotide repeats is likely to occur during genome replication, repair−replication, or recombination−replication (*5*−*7*, *29*, reviewed in ref 1) and possibly even as a result of error-prone maintenance of the genome in certain nonproliferative/arrested tissues such as muscle, nerve, and gamete cells (*30*−*32*). The formation of alternative DNA secondary structures, especially hairpins, during DNA replication can promote specific mutational events involving primer−template misalignment (*33*−*36*). Thus, replication slippage mediated by S-DNA or SI-DNA is an attractive model for triplet instability. At the replication fork, interruptions may inhibit both slippage between nascent and template strands and/or hairpin formation within a single strand, reducing the probability of genetic instability. Our results suggest that interruptions in repeat tracts may offer genetic stability by three mechanisms (Figure 7): (1) the interruptions may inhibit *inter*strand slippage between the two strands by maintaining an in-frame register of Watson-Crick base pairing (i.e., anchoring), and/or (2) by inhibiting *intra*strand interactions, and/or (3) by reducing the opportunity for slippage by limiting the position of slip-out nucleation. First, interstrand slippage would be disfavored as it would result in mismatched interruptions, reducing the thermal stability between complementary strands (Figure 7A). Second, S-DNA involves both inter- and intrastrand

base pairings. Theoretical studies predict that hairpins formed with interrupted CAG, CTG, CGG, and CCG repeats would be biophysically destabilized (*37*), which may tend to reduce the formation and biophysical stability of S-DNA (Figure 7A). The third genetic stabilizing mechanism of interruptions is suggested by both the increased structural complexity with increasing length of pure repeats and the reduction in the number of S-DNA isomers formed with interrupted tracts. A reduction in the possible positions of slip-out nucleations would result in fewer opportunities for slippage in the interrupted tracts (Figure 7B,C). Single-stranded oligonucleotides composed of CAG, CTG, CGG, or CCG repeats can form hairpins with as few as 3−5 repeats (*37*−*40*). Interruptions might limit hairpin formation to the short stretches of pure repeats (Figure 7) or to tracts where the interruption would exist at the tip of the hairpin. If interruptions limit hairpin formation to the short stretches of pure repeats, then the length of the repeat contained within a slipped-out region would be shorter, possibly allowing only short length alterations (Figure 7B,C). As noted above, the effects of interruptions on S-DNA formation are both quantitative and qualitative. It is likely that a combination of both these structural effects may contribute to the stability of the tracts. The relative contribution of the qualitative and quantitative effects of interruptions on structure formation may vary depending upon the length of the repeat tract and the spacing of the interruptions. Interestingly, the percentage of S-DNA formed by the FRAXA 9+27 allele was less than that formed in the pure 21 allele, demonstrating that the presence of the AGG interruption "poisons" the ability of the longer 27 pure CGG tract to form slipped structures relative to a shorter pure tract, present in an uninterrupted allele. This poisoning effect of interruptions may be the result of altered base stacking (in single-stranded or double-stranded DNA), a structural effect which can be transmitted along the helical axis of the DNA. Such structural transmission from an interruption may inhibit slippage to flanking pure tracts. This form of structural transmission, also known as "telestability" (*41*) is known to be able to act over distances of 40 bp and may, in some instances, extend over 100 bp. In summary, our results suggest that the first line of defense afforded by the interruptions seems to be via inhibition of slippage.

Another possible stabilizing mechanism of the CAT and AGG interruptions is through a protein-mediated system, possibly mismatch repair (*7*, *42*). The mismatches formed by inter- or intrastrand slippage of DNA containing interruptions (Figure 7A) may be recognized by repair systems resulting in abortive replication and/or perfect strand realignment. However, as recently pointed out by Petes et al. (*43*), the role of mismatch repair in interruption-mediated repeat stabilization may be secondary to that of slippage inhibition as their results indicated that interruptions stabilize dinucleotide repeats independent of mismatch repair.

Small nucleotide changes which result in DNA structural alterations can have significant biological consequences. Holliday junctions containing mismatched bases are structurally destabilized (*44*), which may explain why such junctions are poorly resolved or aborted in several recombination systems (*45*). In *E. coli*, the genetic instability of pure or interrupted Z-DNA forming CpG tracts is directly correlated with either their ability to form Z-DNA (*46*) or with their predicted thermal stabilities in slipped structures. Interruptions within CpG tracts reduce the mutation rate and result in different deletion products, suggesting differences in the molecular intermediates involved in the instability of the pure or interrupted tracts (*46*). The inhibitory effects of interruptions in FRAXA and SCA1 repeats on both the propensity and complexity of S-DNA formation support another possible structure function relationship, associating slipped structure formation in trinucleotide repeat tracts with their genetic/mutagenic potential.

The biological and genetic relevance of both the length and the purity of the FRAXA and SCA1 triplet repeat tracts to stability has been well-established through identification of stable and unstable repeats in patient DNAs. The SCA1 and FRAXA genetic stability transmission data presented here clearly show that around the stability threshold the presence of repeat tract interruptions provides protection from length alterations (Figures 3 and 6). Similarly, during propagation in *E. coli* the FRAXA or SCA1 clones containing interrupted repeat tracts are more stable than clones containing similar length pure repeats, such that the interrupted tracts are less likely to undergo deletions (C. E. Pearson and R. R. Sinden unpublished). Sequence analyses of SCA1 and FRAXA repeat tracts in the stable and unstable ranges in patient DNAs suggest that interruptions may act by inhibiting the formation of slipped mutagenic intermediates. Our in vitro results demonstrate an inhibitory structural phenomenon specific to interruptions of either $(CAG)_n$•$(CTG)_n$ or $(CGG)_n$•$(CCG)_n$ repeats. The effect of repeat length and purity on slipped strand structure formation correlates with their effect upon genetic stability. This correlation suggests a possible molecular explanation for the biology associated with the protective mechanism of interruptions.

ACKNOWLEDGMENT

We thank B. W. Popovich for the genomic DNA to produce the pAG44-24B clone. We thank A. George and P. Hertel for assistance on this project, T. D. Petes and M. Pandolfo for communicating results prior to publication, and J. L. Jennings and P. R. Kramer for enthusiastic support.

REFERENCES

1. Pearson, C. E., and Sinden, R. R. (1998) (Wells, R. D. and Warren, S. T., Eds.) *Genetic Instabilities and Hereditary Neurological Diseases* in press, Academic Press: New York.
2. David, G., Abbas, N., Stevanin, G., Durr, A., Yvert, G., Cancel, G., Weber, C., Imbert, G., Saudou, F., Antoniou, E., Drabkin, H., Gemmill, R., Giunti, P., Benomar, A., Wood, N., Ruberg, M., Agid, Y., Mandel, J. L., and Brice, A. (1997) *Nature Genet. 17*, 65−70.
3. Zoghbi, H. Y. (1996) *Nature Genet. 14*, 237−238.
4. Richards, R. I., and Sutherland, G. R. (1992) *Cell 70*, 709−712.
5. Sinden, R. R., and Wells, R. D. (1992) *Curr. Opin. Biotechnol. 3*, 612−622.
6. Pearson, C. E., and Sinden, R. R. (1996) *Biochemistry 35*, 5041−5053.
7. Pearson, C. E., Ewel, A., Acharya, S., Fishel, R. A., and Sinden, R. R. (1997) *Hum. Mol. Genet. 6*, 1117−1123.
8. Pearson, C. E., Wang, Y.-H., Griffith, J. D., and Sinden, R. R. (1998) *Nucleic Acids Res. 26*, 816−823.
9. Weber, J. L. (1990) *Genomics 7*, 524−530.

10. Chung, M., Ranum, L. P. W., Duvick, L. A., Servadio, A., Zoghbi, H. Y., and Orr, H. T. (1993) *Nature Genet. 5*, 254−258.
11. Chong, S. S., McCall, A. E., Cota, J., Subramony, S. H., Orr, H. T., Hughes, M. R., and Zoghbi, H. Y. (1995) *Nature Genet. 10*, 344−350.
12. Kunst, C. B., and Warren, S. T. (1994) *Cell 77*, 853−861.
13. Eichler, E. E., Holden, J. J. A., Popovich, B. W., Reiss, A. L., Snow, K., Thibodeau, S. N., Richards, C. S., Ward, P. A., and Nelson, D. L. (1994) *Nature Genet. 8*, 88−94.
14. Snow, K., Tester, D. J., Kruckeberg, K. E., Schaid, D. J., and Thibodeau, S. N. (1994) *Hum. Mol. Genet. 9*, 1543−1551.
15. Hirst, M. C., Grewal, P. K., and Davies, K. E. (1994) *Hum. Mol. Genet. 3*, 1553−1560.
16. Zhong, N., Yang, W., Dobkin, C., and Brown, W. T. (1995) *Am. J. Hum. Genet. 57*, 351−361.
17. Eichler, E. E., Hammond, H. A., Macpherson, J. N. Ward, P. A., and Nelson, D. L. (1995) *Hum. Mol. Genet. 4*, 2199−2208.
18. Eichler, E. E., Macpherson, J. N., Murray, A., Jacobs, P. A., Chakravarti, A., and Nelson, D. L. (1996) *Hum. Mol. Genet. 5*, 319−330.
19. Kunst, C. B., Leeflang, E. P., Iber, J. C., Arnheim, N., and Warren, S. T. (1997) *J. Med. Genet. 34*, 627−631.
20. Leeflang, E. P., and Arnheim, N. (1995) *Hum. Mol. Genet. 4*, 135−136.
21. Nancarrow, J. K., Holman, K., Mangelsdorf, M., Hori, T., Denton, M., Sutherland, G. R., and Richards, R. I. (1995) *Hum. Mol. Genet. 4*, 367−372.
22. Cossee, M., Schmitt, M., Campuzano, V., Reutenauer, L., Moutou, C., Mandel, J. L., and Koenig, M. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 7452−7457.
23. Montermini, L., Andermann, E., Labuda, M., Richter, A., Pandolfo, M., Cavalcanti, F., Pianese, L., Iodice, L., Farina, G., Monticelli, A., Turano, M., Filla, A., Demichele, G., and Cocozza, S. (1997) *Hum. Mol. Genet. 6*, 1261−1266.
24. Bowater, R. P., Rosche, W. A., Jaworski, A., Sinden, R. R., and Wells, R. D. (1996) *J. Mol. Biol. 264*, 82−96.
25. Banfi, S., Servadio, A., Chung, M. Y., Kwiatkowski, T. J., Jr., McCall, A. E., Duvick, L. A., Shen, Y., Roth, E. J., Orr, H. T., and Zoghbi, H. Y. (1994) *Nature Genet. 7*, 513−520.
26. Lilley, D. M. J. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 7140−7142.
27. Quan, F., Janas, J., and Popovich, B. W. (1995) *Hum. Mol. Genet. 4*, 2411−2413.
28. Murray, A., Macpherson, J. N., Pound, M. C., Sharrock, A., Youings, S. A., Dennis, N. R., Mckechnie, N., Linehan, P., Morton, N. E., and Jacobs, P. A. (1997) *Hum. Mol. Genet. 6*, 173−184.
29. Dover, G. A. (1993) *Curr. Opin. Genet. Dev. 3*, 902−910.
30. Zatz, M., Passos-Bueno, M. R., Cerqueira, A., Marie, S. K., Vainzof, M., and Pavanello, R. C. (1995) *Hum. Mol. Genet. 4*, 401−406.
31. Telenius, H., Kremer, B., Goldberg, Y. P., Theilmann, J., Andrew, S. E., Zeisler, J., Adam, S., Greenberg, C., Ives, E. J., Clarke, L. A., and Hayden, M. R. (1995) *Nature Genet. 6*, 409−414.
32. Kaytor, M. D., Burright, E. N., Duvick, L.A., Zoghbi, H. Y., and Orr, H. T. (1997) *Hum. Mol. Genet. 6*, 2135−2139.
33. Sinden, R. R., Zheng, G., Brankamp, R. G., and Allen, K. N. (1991) *Genetics 129*, 991−1005.
34. Trinh, T. Q., and Sinden, R. R. (1991) *Nature 352*, 544−547.
35. Trinh, T. Q., and Sinden, R. R. (1993) *Genetics 134*, 409−422.
36. Rosche, W. A., Trinh, T. Q., and Sinden, R. R. (1997) *J. Mol. Biol. 269*, 176−187.
37. Gacy, A. M., Goellner, G., Juranic, N., Macura, S., and McMurry, C. T. (1995) *Cell 81*, 533−540.
38. Chen, X., Mariappan, S. V. S., Catasti, P., Ratliff, R., Moyzis, R. K., Laayoun, A., Smith, S. S., Bradbury, E. M., and Gupta, G. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 5199−5203.
39. Zheng, M. X., Huang, X. N., Smith, G. K., Yang, X. Y., and Gao, X. L. (1996) *J. Mol. Biol. 264*, 323−336.
40. Petruska, J., Arnheim, N., and Goodman, M. F. (1996) *Nucleic Acids Res. 24*, 1992−1998.
41. Wells, R. D., Goodman, T. C., Hillen, W., Horn, G. T., Klein, R. D., Larson, J. E., Muller, U. R., Nuendorf, S. K., Panayotatoes, N., and Stirdivant, S. M. (1980) *Prog. Nucleic Acids Res. 24*, 167−267.
42. Kramer, P. R., Pearson, C. E., and Sinden, R. R. (1996) *Hum. Genet. 98*, 151−157.
43. Petes, T. D., Greenwell, P. W., and Dominska, M. (1997) *Genetics 146*, 491−498.
44. Duckett, D. R., and Lilley, D. M. J. (1991) *J. Mol. Biol. 221*, 147−161.
45. Burgin, A. B., and Nash, H. A. (1995) *Curr. Biol. 5*, 1312−1321.
46. Bichara, M., Schumacher, S., and Fuchs, R. P. (1995) *Genetics 140*, 897−907.

BI972546C