616–623 Nucleic Acids Research, 1999, Vol. 27, No. 2

Nucleotide excision repair affects the stability of long transcribed (CTG•CAG) tracts in an orientation-dependent manner in *Escherichia coli*

Pawel Parniewski, Albino Bacolla, Adam Jaworski[1] and Robert D. Wells*

Institute of Biosciences and Technology, Center for Genome Research, Department of Biochemistry and Biophysics, Texas A&M University, Texas Medical Center, 2121 West Holcombe Boulevard, Houston, TX 77030-3303, USA and [1]Center for Microbiology and Virology, Polish Academy of Sciences, 106 Lodowa Street, 93-232 Lodz, Poland

Received August 24, 1998; Revised and Accepted November 15, 1998

ABSTRACT

The influence of nucleotide excision repair (NER), the principal *in vivo* repair system for DNA damages, was investigated in *Escherichia coli* with *uvrA*, *uvrB* and *uvrAuvrB* mutants with the triplet repeat sequences (TRS) involved in myotonic dystrophy, the fragile X syndrome and Friedreich's ataxia. $(CTG•CAG)_{175}$ was more stable when the (CTG) strand was transcribed than when the (CAG) strand was transcribed in the alternate orientation. A lack of the UvrA protein dramatically increases the instability of this TRS *in vivo* as compared with the stability of the same sequence in *uvrB* mutant, which produces an intact UvrA protein. We propose that transcription transiently dissociates the triplet repeat complementary strands enabling the non-transcribed strand to fold into a hairpin conformation which is then sufficiently stable that replication by-passes the hairpin to give large deletions. If the TRS was not transcribed, fewer deletions were observed. Alternatively, in the uvrA$^-$ mutant, the hairpins existing on the lagging strand will suffer bypass DNA synthesis to generate deleted molecules. Hence, NER, functionally similar in both prokaryotes and eukaryotes, is an important factor in the genetic instabilities of long transcribed TRS implicated in human hereditary neurological diseases.

INTRODUCTION

Neurogenetic diseases including myotonic dystrophy (DM), fragile X syndrome, Huntington's disease, spinobulbar muscular atrophy, spinocerebellar ataxia type 1 and Friedreich's ataxia result from expanded triplet repeat sequences (TRS) [(CTG•CAG), (CGG•CCG) or (GAA•TTC)] within their genes (1,2). The lengths of the TRS influence the age of onset and the increased severity of the neurological diseases through family pedigrees (clinically referred to as anticipation). Long tracts of TRS are unstable and show repeat size polymorphisms in successive generations and in different tissues. The expansion events involved in the etiology of hereditary neurological diseases fall into two categories (Type 1 and Type 2) (2). For relatively small TRS expansions (Type 1 diseases) involving expanded glutamine tracts in the target proteins, small slipped-register expansions and deletions (SSED) may be responsible (3). However, for the massive expansions involved in the Type 2 diseases such as DM and fragile X, a different mechanism involving replication errors (4–10) or recombination (11–16) is likely. In addition to observations in humans, TRS instabilities have been demonstrated in *Escherichia coli* (5,6,17–23), yeast (4,11,24), transgenic mice (25) and cultured cells from patients (26,27).

The molecular mechanism responsible for the genetic instability of TRS is influenced by many factors (1,8). The availability of genetically and biochemically defined systems in *E.coli* enabled detailed investigations of the TRS instabilities. Deletions and expansions occur in *E.coli* (5,6) and TRS instability is influenced by the direction of replication through the TRS (5,6,18,19). Also, a destabilizing effect of active transcription into long (CTG•CAG) tracts was demonstrated (18). Different growth conditions, especially entering a stationary phase by *E.coli* harboring (CTG•CAG) tracts (19), causes increased deletions of the TRS. Functional single-stranded DNA-binding protein (SSB) was an important cellular factor in maintaining the genetic stability of triplet repeats in *E.coli* (22).

Numerous prior investigations have shown the occurrence of minisatelite instabilities caused by inactivation of methyl directed mismatch repair (MMR) or nucleotide excision repair (NER) (28–39). Much less is known, however, about the involvement of the repair systems in TRS instability (3,17,24,40). Long (CTG•CAG) repeats are stabilized in plasmids in *E.coli* with mutations in the MMR genes (*mutS*, *mutL* or *mutH*). Mutations in these genes prevent large (approximate multiples of 40 repeats) TRS deletions. However, in the same genetic background, SSED occur very frequently (3). Similar data also were reported for *pms1* and *msh2* yeast mutants (24).

NER is the major cellular defense system in both prokaryotes and eukaryotes (41). Defects in excision repair cause three human diseases (42): xeroderma pigmentosum, Cockayne's syndrome and trichothiodystrophy. The substrate specificity of the *E.coli* Uvr(A)BC endonuclease is very wide and includes bulky DNA adducts and DNA cross-links that cause significant distortion of the DNA helix as well as less distortive lesions such as methylated bases (43,44). The involvement of NER in the repair of DNA loops has also been reported (37,45).

*To whom correspondence should be addressed. Tel: +1 713 677 7651; Fax: +1 713 677 7689; Email: rwells@ibt.tamu.edu

Herein, we demonstrate that the deficiency of some NER functions dramatically affects the stability of long (CTG•CAG) inserts in plasmids in *E.coli*. This stability is strongly dependent on the orientation of the (CTG•CAG) tract, relative to the ColE1 origin of replication, and on active transcription into the TRS.

MATERIALS AND METHODS

Plasmids

All plasmids used in these experiments containing repeating $(CTG•CAG)_n$ or $(CGG•CCG)_n$ or $(GAA•TTC)_n$ inserts are shown in Table 1.

Table 1. Plasmids used in this study

Plasmids	TRS	Vector	Orientation
pRW3247	$(CAG•CTG)_{175}$	pUC19*Not*I	II
pRW3268	$(CAG•CTG)_{175}$	pUC19	II
pRW3248	$(CTG•CAG)_{175}$	pUC19*Not*I	I
pRW3269	$(CTG•CAG)_{175}$	pUC19	I
pRW4011	$(CAG•CTG)_{50}$	pUC18	II
pRW3032	$(CCG•CGG)_{32}$	pUC19*Not*I	II
pRW4006	$(CGG•CCG)_{32}$	pUC18*Not*I	I
pRW3804	$(GAA•TTC)_{70}$	pUC18*Not*I	I
pRW3803	$(TTC•GAA)_{65}$	pUC19*Not*I	II

pRW3247 and pRW3248 are derivatives of pUC19 *Not*I and were described previously (5,6,17–19). The plasmids contain either $(CAG)_{175}$ (pRW3247, orientation II) or $(CTG)_{175}$ (pRW3248, orientation I) as the leading strand template for replication. This sequence is not homogeneous but contains two G to A interruptions at repeats 28 and 69. Plasmids pRW3268 and pRW3269 contain the same insert cloned in both orientations but the vector lacks the *Pvu*II fragment containing the LacZ′ promoter region such that the (CTG•CAG) insert is not transcribed (5,17,18). pRW4011 is a derivative of pUC18 and contains an uninterrupted $(CAG•CTG)_{50}$ cloned in orientation II. pRW3032 is a pUC19 *Not*I-based plasmid containing the $(CGG•CCG)_{24}$ insert cloned into the *Bam*HI site, where the CCG is in the leading strand template (20). pRW4006 is a pUC18 *Not*I-based plasmid containing the same insert cloned in orientation II, where the CGG is in the leading strand template. pRW3804 is a derivative of pUC18 *Not*I (46) containing the $(GAA•TTC)_{70}$ fragment (47) inserted into the *Bam*HI site of the vector. The plasmid was obtained by the *in vivo* expansion technique (47); the GAA strand is in the leading strand template. pRW3803 is a derivative of pUC19 *Not*I and the $(TTC)_{65}$ is in the leading strand template. These two plasmids were constructed and characterized by K. Ohshima (47).

Bacterial strains

To study the influence of the NER system on the instability of the TRS, the following *E.coli* strains were used: AB1157 [*thr-1, ara-14, leuB6, D(gpt-proA)62, lacY1, tsx-33, qsr′ -, glnV44(AS), galK2, l, arc- hisG4(Oc), rfbD1, mgl-51, rpsL31(sm^R), kdgK51, xylA5, mtl1, argE3(Oc), thi-1*] as a parent (wild-type) of the NER deficient strains; *E.coli* AB1886 is isogenic with AB1157 except for *uvrA6*; *E.coli* AB1885 is isogenic with AB1157 except *uvrB5*; *E.coli* AB2421 is isogenic with AB1157 except *uvrA6uvrB45*. Strains AB1157, AB1886 and AB1885 were obtained from the *E.coli* Genetic Stock Center. *Escherichia coli* AB2421 strain was a gift of Dr E. Tang (University of Texas, M. D. Anderson Cancer Center, Science Park, Research Division, Smithville, TX).

Conditions of bacterial growth

The plasmids containing various TRS were transformed into the appropriate *E.coli* strain and grown for a number of generations, as described (19). Briefly, *E.coli* cells were transformed with the plasmids and an aliquot of this mixture was inoculated into 10 ml of LB containing ampicillin at 100 µg/ml. Incubations of the liquid cultures were continued overnight at 37°C at a shaking rate of 250 r.p.m. The bacteria then were subcultured into fresh liquid media with a dilution factor of 10^7. The cells from each culture were harvested and plasmids were isolated.

Analysis of DNA

The undeleted plasmids were isolated by elution of the full-length supercoiled DNA from the appropriate location on the agarose gel. Transformation of *E.coli* cells with the DNA was performed using electroporation (Electroporator 1000, Stratagene). DNA preparations, agarose and polyacrylamide gel electrophoreses were performed according to standard laboratory protocols (48). Plasmids were prepared by alkaline lysis of 10 ml cultures using the standard alkaline lysis miniprep procedure. Restriction digests were performed following the manufacturer's instructions. All plasmids except for pRW3268 and pRW3269 were digested with *Eco*RI and *Hin*dIII. pRW3268 was cleaved with *Eco*RI and *Nde*I and pRW3269 was cleaved with *Eco*RI and *Afl*III. The restriction analyses were performed by electrophoresis through 7 or 6% polyacrylamide gels in TAE (40 mM Tris-acetate, 1 mM EDTA, pH 8.0) buffer. The gels were stained with ethidium bromide and photographed. The negatives of the gels were scanned (300S, Molecular Dynamics) and the amount of the TRS containing fragment at full-length was quantified. For some experiments the restriction digests were radiolabeled using *E.coli* DNA polymerase I Klenow fragment incubated with [α-^{32}P]dATP, subjected to PAGE and radiolabeled fragments were observed by autoradiography or with a PhosphorImager (400S, Molecular Dynamics). Quantitation of these gels was performed directly on the phosphorimage. The percent of full-length of $(CTG•CAG)_{175}$ was determined relative to all smaller fragments of DNA on the gel.

RESULTS

Effect of orientation of $(CTG•CAG)_{175}$ on its stability in NER proficient and deficient cells: orientation II

The effect of NER on the instability of the $(CTG•CAG)_{175}$ insert cloned in both orientation I and orientation II in plasmids (Table 1) was determined. The orientation of these TRS with respect to the direction of replication within these plasmids was as follows: plasmids containing the CTG sequence as the leading-strand template are designated orientation I; plasmids containing the CAG sequence as the leading-strand template are designated orientation II (5). Several prior studies (5,17–19) revealed that the TRS in orientation I is substantially more stable due to the replication behavior on CTG hairpin loops which are relatively more thermodynamically stable than CAG hairpin loops (49–53). Genetically defined *E.coli* strains were used (see Materials and

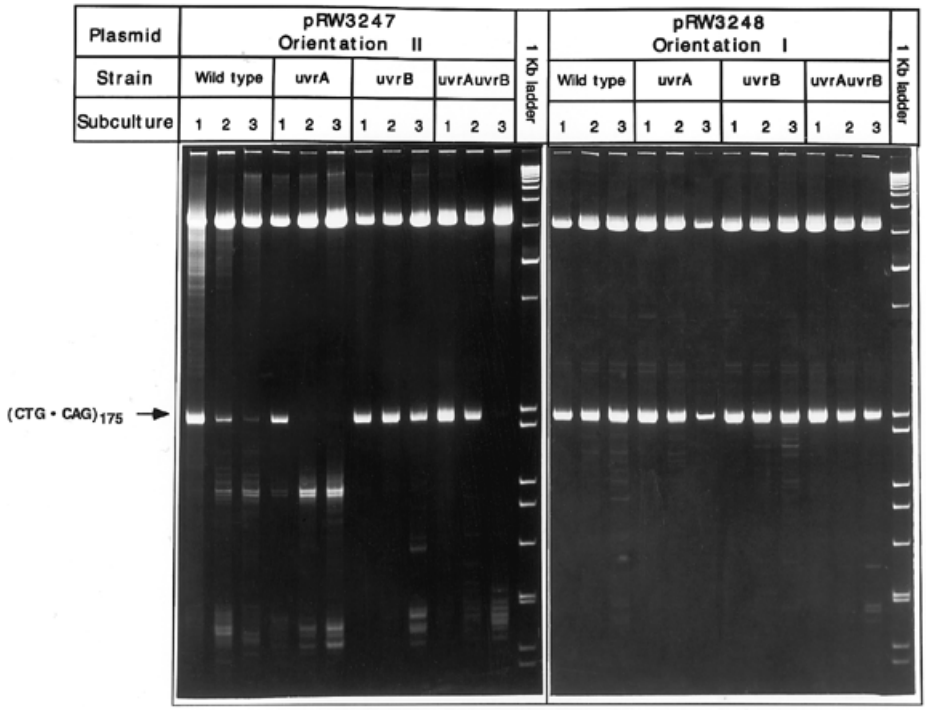

Figure 1. Effect of orientation on instability of transcribed $(CTG•CAG)_{175}$ tracts in plasmids in wild-type *E.coli* and NER mutant strains. The *E.coli* strains were transformed with purified non-deleted monomer of pRW3247 (TRS in orientation II) or pRW3248 (TRS in orientation I) and analyzed as described in the Materials and Methods. The numbers above the samples indicate the number of subcultures. An arrow points to the band corresponding to the full-length, non-deleted $(CTG•CAG)_{175}$. The 1 kb ladder was purchased from Gibco BRL.

Methods). The strains containing mutations in the *uvrA* or *uvrB* or *uvrA* and *uvrB* (double mutant) genes as well as the isogenic wild-type *E.coli* were transformed with pRW3247 (TRS in orientation II) or pRW3248 (TRS in orientation I). The cells were grown overnight and then subcultured into fresh LB medium with a dilution factor of 10^7. After three recultivations of the cells, the DNA from each culture was isolated and plasmids were cleaved with *Eco*RI and *Hin*dIII to release the insert containing the TRS. Restriction digests were analyzed by 7% PAGE (Fig. 1). The quantitation of deletions found for pRW3247 and pRW3248 in the wild-type, *uvrA* and *uvrB* strains is shown in Figure 2A.

A gradual disappearance of the restriction fragment corresponding to the full-length, non-deleted $(CTG•CAG)_{175}$ was observed as analyzed for wild-type cells harboring pRW3247 (TRS in orientation II; Figs 1 and 2, left panel). These data are consistent with the previously reported behavior of the same plasmid as investigated in HB101 and DH5α *E.coli* strains (5,17–19). However, if this plasmid was propagated in cells having a mutation in the *uvrA* gene (*E.coli* 1886), a dramatic increase in the instability of the $(CTG•CAG)_{175}$ insert was observed. The amount of the full-length non-deleted TRS calculated after the first recultivation of the *uvrA* cells harboring pRW3247 was significantly reduced as compared to the amount of non-deleted TRS calculated for the wild-type cells (52 and 76%, respectively). After the second recultivation of the *uvrA* mutant harboring pRW3247, the full-length TRS containing fragment was almost completely deleted (1% of non-deleted insert), whereas the pRW3247 isolated from the wild-type cells still contained a significant amount of the full-length $(CTG•CAG)_{175}$ (24% of non-deleted insert) (Fig. 2A).

The opposite behavior was observed if pRW3247 was grown and subcultured in *uvrB* cells (*E.coli* 1885). The stability of the $(CTG•CAG)_{175}$ insert in this strain was strikingly high and, even after three recultivations of the cells, 45% of the plasmids contained the full-length TRS. After the same number of recultivations of the wild-type and *uvrA* cells harboring pRW3247, the stability of the TRS in these strains was very low (9 and 0% of full-length TRS fragment, respectively). The stability of the $(CTG•CAG)_{175}$ in the *uvrAuvrB* double mutant after the first recultivation of the cells harboring pRW3247 was comparable to that observed in the wild-type strain (77 and 76%, respectively) (Fig. 1A) but the amount of the non-deleted TRS measured after the second growth was higher (48% for the *uvrAuvrB* strain and 24% for the wild-type). However, after three recultivations of the cells, the amount of the full-length $(CTG•CAG)_{175}$ in both the wild-type and double mutant strains was very low (9 and 6%, respectively) (Fig. 1A).

To determine whether the UV irradiation of the DNA (during the isolation of the full-length supercoiled DNA) influenced the genetic instability of the $(CTG•CAG)_{175}$ in plasmids grown in NER proficient and deficient strains, we conducted similar experiments using DNAs that were not exposed to UV (results not shown). No substantial differences were observed between these experiments and the data presented above.

NER in *E.coli* requires both UvrA and UvrB proteins to be functional (41). Since *E.coli* 1886 (*uvrA*), *E.coli* 1885 (*uvrB*) and *E.coli* AB2421 (*uvrAuvrB*) mutants are NER deficient (54), our data suggest that different mechanisms affect the stability of the $(CTG•CAG)_{175}$ in NER proficient and NER deficient cells. As shown in Figure 1, the stability of this TRS after the first and

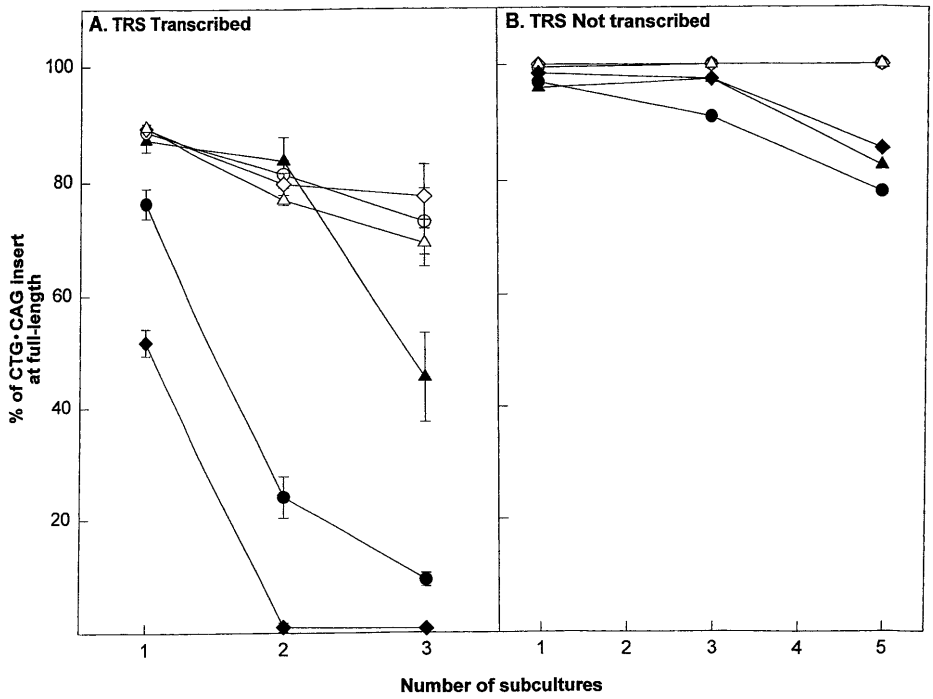

Figure 2. Instability of (**A**) transcribed and (**B**) untranscribed $(CTG•CAG)_{175}$ tracts in plasmids. The plasmids designated were transformed into the wild-type and NER mutants and the cells were grown for the number of recultivations indicated. Restriction digests were analyzed by PAGE as described for Figure 1 and the amount of non-deleted TRS was calculated. (**A**) Quantitation of deletions in pRW3247 and pRW3248 (TRS transcribed). Analyses were performed on digested samples fractionated through 7% PAGE in TAE (as shown in Fig. 1). Each point on the graph represents an average value from three experiments. (**B**) Quantitation of deletions in pRW3269 and pRW3268 (TRS not transcribed). Quantitation was performed on the digested samples, radiolabeled with [α-^{32}P]dATP and separated by 7% PAGE. The reproducibility of the measurement of the amount of the full-length $(CTG•CAG)_{175}$ was estimated to be ±2%. Open symbols, plasmids containing $(CTG•CAG)_{175}$ cloned in orientation I; filled symbols, plasmids containing $(CTG•CAG)_{175}$ cloned in orientation II; circles, wild-type *E.coli*; diamonds, *uvrA* mutant; triangles, *uvrB* mutant.

second recultivations was higher in the double mutant (both UvrA and UvrB proteins not present) as compared to the wild-type strain. However, the difference in the stability of the TRS between *uvrA* and *uvrB* strains, that both are NER deficient, was much more pronounced (Figs 1 and 2). Thus, it seems that the UvrA protein stabilizes the TRS containing insert whereas the UvrB protein, which is a DNA endonuclease, contributes to the instability of the long (CTG•CAG) tract.

Effect of orientation of $(CTG•CAG)_{175}$ on its stability in NER proficient and deficient cells: orientation I

Similar investigations were conducted on the same DM sequence in pRW3248 which contains the $(CTG•CAG)_{175}$ insert cloned in orientation I (Table 1). The NER proficient (wild-type) and deficient (*uvrA* or *uvrB* or *uvrAuvrB*) cells were transformed with pRW3248 and grown for three consecutive subcultures as described. DNA from each culture was isolated and plasmids were cleaved with *Eco*RI and *Hin*dIII to release the insert containing the TRS. Restriction digests were analyzed by 7% PAGE (Fig. 1, right panel). The quantitation of deletion products in pRW3248 is shown in Figure 2A.

Restriction analyses of pRW3248 propagated in NER proficient and deficient strains clearly show that the difference in the stability of the $(CTG•CAG)_{175}$ between the wild-type and NER$^-$ strains is limited to the TRS in orientation II, since no substantial amount of deletions were observed in all *E.coli* strains under investigation harboring pRW3248 (TRS in orientation I) (Fig. 1, right panel and Fig. 2A).

Effect of transcription on the instability of the $(CTG•CAG)_{175}$

We showed previously that active transcription into long (CTG•CAG) tracts markedly increased the frequency of deletions of $(CTG•CAG)_{175}$ from plasmids (18). Instability occurred when *E.coli* cells harboring pRW3247 were grown and subcultured through a number of generations. It is also known that transcription recruits NER to damaged regions of DNA (55–57). Thus, the following experiments were undertaken to evaluate the role of transcription into the TRS as a factor influencing the stability of the $(CTG•CAG)_{175}$ in NER proficient and deficient *E.coli* strains.

Two pairs of plasmids containing the $(CTG•CAG)_{175}$ insert were used: pRW3247 (TRS in orientation II, transcribed) and pRW3248 (TRS in orientation I, transcribed) that contain the $(CTG•CAG)_{175}$ insert cloned in the polylinker of pUC19 *Not*I that contain intact promoter for LacZ′; pRW3268 and pRW3269 contain the same insert (cloned in orientation II and orientation I, respectively) but not located inside a transcription unit (Table 1). The NER proficient (wild-type) and deficient (*uvrA* or *uvrB*) cells were transformed with pRW3268 (TRS in orientation II) or pRW3269 (TRS in orientation I) and grown for five consecutive subcultures as described. DNA from each culture was isolated and plasmids were cleaved with *Eco*RI and *Nde*I (for pRW3268) or with *Eco*RI and *Afl*III (for pRW3269) to release the insert containing the TRS. Restriction digests were radiolabeled (see Materials and Methods), separated by 6% PAGE and analyzed with a PhosphorImager (data not shown). The amount of full-length, non-deleted $(CTG•CAG)_{175}$ was plotted as a function of the

Table 2. Transcription- and orientation-dependent instability of $(CTG•CAG)_{175}$ tract in wild-type and NER mutants

		% of $(CTG•CAG)_{175}$ at full-length			
Transcription		+		–	
Orientation		II	I	II	I
Plasmid		pRW3247	pRW3248	pRW3268	pRW3269
Strain	Wild type	3	98	78	100
	uvrA	0	91	86	100
	uvrB	29	98	83	100

Percentage of the full-length $(CTG•CAG)_{175}$ fragment as calculated after five consecutive recultivations of wild-type *E.coli* or the *uvrA* or *uvrB* mutants harboring pRW3247 or pRW3248 (TRS transcribed) or pRW3268 or pRW3269 (TRS not transcribed). Restriction digests of the designated plasmids were labeled with [α-^{32}P]dATP and analyzed by 6 or 7% PAGE. The numbers represent the percentage of non-deleted $(CTG•CAG)_{175}$ as calculated with a PhosphorImager. The reproducibility of these measurements was ±2%.

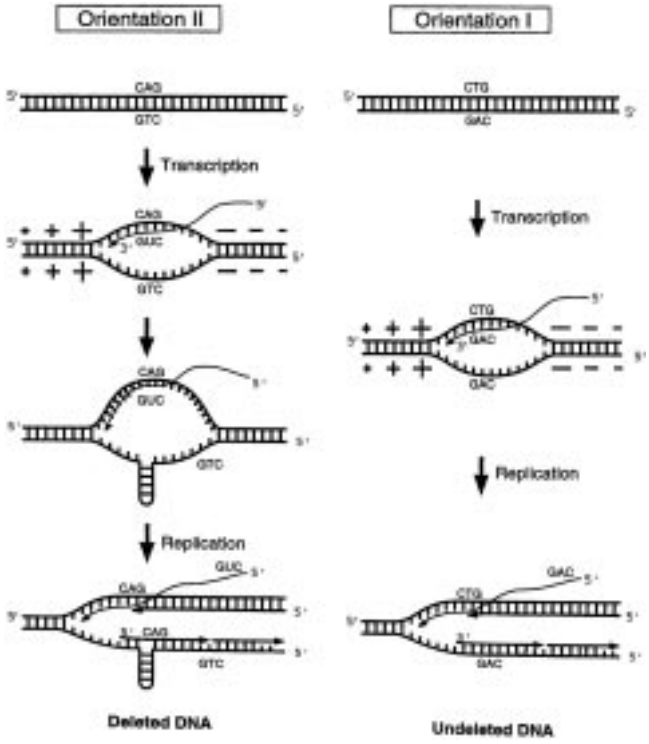

Figure 3. Model for the involvement of transcription and TRS orientation in the instability of the $(CTG•CAG)_n$. Local positive and negative superhelical domains (indicated by pluses and minuses) caused by transcription (64,65) through the $(CTG•CAG)_n$ facilitate an opening of the duplex TRS. The single-stranded (CTG) region formed for orientation II (left side) is not complexed by RNA polymerase, as in the case of orientation I, and therefore this region is able to form the (CTG) hairpin. Since the DNA polymerase complex may bypass this hairpin during lagging-strand synthesis, this would lead to deletions of the $(CTG•CAG)_n$ tract in orientation II but not in orientation I (right side).

number of subcultures (Fig. 2B). No substantial instability of the TRS was observed after five recultivations of the wild-type and NER$^-$ cells harboring pRW3269 (TRS in orientation I). After the same number of generations, the instability of the $(CTG•CAG)_{175}$ insert in the NER proficient and deficient cells harboring pRW3268 (TRS in orientation II) was clearly detectable. These data show similar orientation-dependent instabilities of the two plasmids as reported for DH5α and HB101 *E.coli* strains (5,17–19).

Similar analyses were conducted for the wild-type and NER deficient strains harboring pRW3247 and pRW3248 (TRS transcribed). Figure 2A shows the percentage of full-length $(CTG•CAG)_{175}$ calculated from experiments described earlier and presented in Figure 1. A comparison of the behavior of pRW3247 and pRW3248 (TRS transcribed) (Fig. 2A) and pRW3269 and pRW3268 (TRS not transcribed) (Fig. 2B) clearly shows that NER proteins in *E.coli* affect the stability of the $(CTG•CAG)_{175}$ insert when the (CTG) strand is the lagging-strand template for replication but only if the TRS region is transcribed. The effect of orientation of the (CTG•CAG) tract and transcription going through the TRS after five recultivations of the strains under investigation is summarized in Table 2. These data clearly show the effect of the orientation on the stability of the TRS (pRW3268 versus pRW3269) and the influence of active transcription into the $(CTG•CAG)_{175}$ in orientation II (pRW3247) in NER proficient and deficient *E.coli* strains. The explanation for the involvement of transcription as well as NER proteins in the stability of long (CTG•CAG) tracts in plasmids is shown in Figures 3 and 4 and will be discussed below (Discussion).

Instability of other lengths and types of TRS

Thus far, we have shown that the stability of a long transcribed (CTG•CAG) tract is influenced by mutations in NER genes in *E.coli*. The absence of the *uvrA* gene product greatly enhances the instability of the TRS whereas lack of the functional UvrB protein causes substantial stabilization of $(CTG•CAG)_{175}$ in orientation II.

Additional experiments were performed to determine whether the stability of shorter (CTG•CAG) tracts as well as other repeated sequences, related to human neurodegenerative diseases, are susceptible to the mutations in NER in *E.coli*. To evaluate the effect of length of the (CTG•CAG) tract on its stability in NER proficient and deficient strains, the deletion properties of $(CAG•CTG)_{50}$ in orientation II were investigated. Wild-type as well as *uvrA* or *uvrB E.coli* strains were transformed with the full-length non-deleted monomer of pRW4011 (Table 1) and grown for the indicated number of recultivations. The DNA from each culture was isolated and plasmids were cleaved with *Eco*RI and *Hin*dIII to release the TRS containing fragment. Restriction digests were analyzed by 7% PAGE (data not shown). The quantitation of the deletions (described in Materials and Methods) in pRW4011 for the wild-type, *uvrA* and *uvrB E.coli* strains is shown in Table 3. The extent of instability of $(CAG•CTG)_{50}$ was essentially the same among all strains investigated. Also, studies performed on a plasmid containing $(CTG•CAG)_{50}$ in orientation I (pRW4015) (not shown) showed a higher stability of the TRS and no differences between NER proficient and deficient strains were observed. These data demonstrate that for (CTG•CAG) tracts, besides the effect of the orientation and transcription into the TRS, a certain length of the (CTG•CAG) (>50 units) is required as an additional factor that triggers NER in *E.coli*.

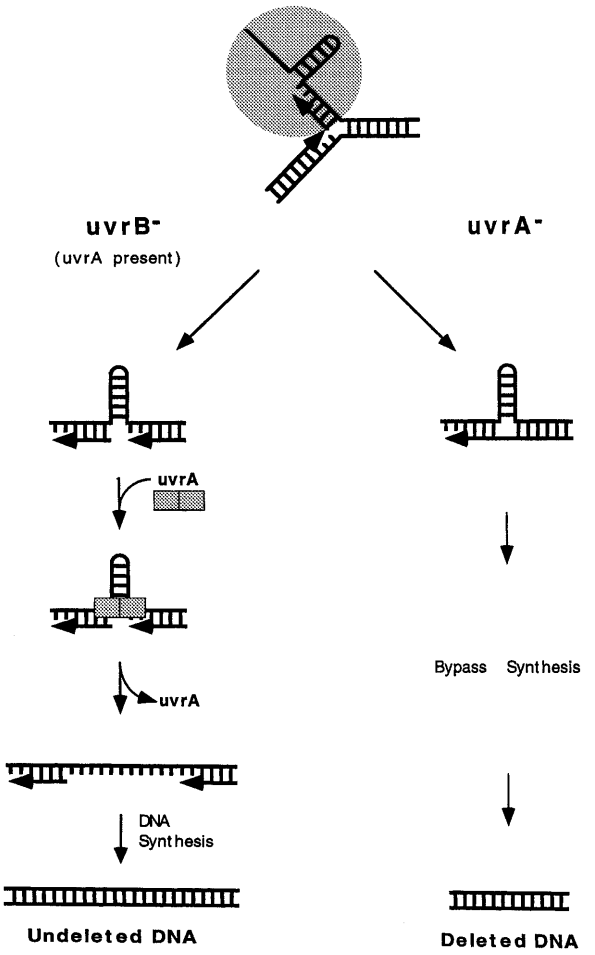

Figure 4. Model for the involvement of the UvrA protein in the stabilization of $(CTG•CAG)_{175}$ *in vivo*. The lower part of the figure shows events occurring on the shaded region of the top replicating molecule. The (CTG) hairpin formed during replication of the lagging strand in orientation II is recognized by the UvrA protein (left side). The interaction of the dimeric UvrA (69) with the (CTG) hairpin causes the removal of the DNA secondary structure and thus obviates bypass synthesis through the (CTG) hairpin. This leads to substantial stabilization of the TRS and gives rise to undeleted DNA. However, for *uvrA*⁻ cells (right side), bypass synthesis through the TRS containing the CTG hairpin (no UvrA protein) leads to deletions.

Similar investigations conducted on the fragile X sequence (CCG•CGG) cloned in both orientation II and orientation I in plasmids pRW3032 and pRW4006, respectively, showed no effect of the genetic background of the *E.coli* strains (Table 3).

The same behavior was observed for pRW3803 and pRW3804 (Table 1) harboring the (GAA•TTC) Friedreich's ataxia sequence cloned in both orientations. The stability of the (GAA•TTC) tracts showed the orientation dependent effect as described (47) but the mutations in *uvrA* and *uvrB* genes had no influence on the amount of deletions in these plasmids (Table 3).

Thus, our results for the (CGG•CCG) and (GAA•TTC) sequences cloned in plasmids show that these TRS may be too short and/or that some different structural properties of (CGG•CCG) and (GAA•TTC) (21,58,59) tracts do not induce the NER in *E.coli*. Other prior investigations have revealed the structural uniqueness of (CTG•CAG) relative to the other TRS (21,60).

Table 3. Instability of other lengths and types of TRS contained on plasmids grown in wild-type *E.coli* and NER mutants

Plasmid	TRS	Wild type			uvrA			UvrB		
		No. of subcultures			No. of subcultures			No. of subcultures		
		1	2	3	1	2	3	1	2	3
pRW4011	$(CAG•CTG)_{50}$	77	70	70	74	73	72	78	72	78
pRW3032	$(CCG•CGG)_{32}$	65	49	35	63	48	38	67	51	42
pRW4006	$(CGG•CCG)_{32}$	87	84	84	86	82	84	88	86	84
pRW3804	$(GAA•TTC)_{-70}$	87	83	79	87	85	82	85	76	71
pRW3803	$(TTC•GAA)_{-65}$	84	79	75	81	77	45[a]	84	72	59[b]

Percentage of the full-length TRS contained on pRW4011 $[(CTG•CAG)_{50}$, orientation II], pRW3032 $[(CCG•CGG)_{32}$, orientation II], pRW4006 $[(CGG•CCG)_{32}$, orientation I], pRW3804 $[(GAA•TTC)_{70}$, orientation I] and pRW3803 $[(TTC•GAA)_{65}$, orientation II] as grown and subcultured in wild-type *E.coli* and the *uvrA* or *uvrB* mutants. Analyses of the restriction digests of plasmids were performed by 7% PAGE and the amount of the non-deleted TRS containing inserts was calculated as described in Materials and Methods. The numbers show the percentage of the full-length TRS and are average values from three experiments with the standard deviations ranging from ±0.2 to ±9.8% except as indicated otherwise.

[a]The average value from six experiments is shown with a standard deviation ±75%.

[b]Standard deviation for this experiment was ±28%.

DISCUSSION

NER (41) has a substantial impact on the genetic stability of long (CTG•CAG) inserts cloned in plasmids in *E.coli*. NER is the major cellular defense system against a variety of DNA damages in both prokaryotes and eukaryotes (61,62). To investigate the potential influence of this system on the stability of the TRS *in vivo*, we employed *E.coli* NER mutants. In the wild-type AB1157 strain, the (CTG•CAG) tract cloned in orientation II, relative to the ColE1 origin of replication, is more unstable compared to orientation I, and its genetic instability is strongly dependent on active transcription through the TRS sequence. These results confirm previous observations obtained with the HB101 strain, and agree with our hypothesis (18) that this instability is caused by a collision of the replication apparatus with the transcription machinery, rather than by host-specific factors.

Figure 3 presents a model that shows how transcription and orientation of a (CTG•CAG) tract may affect its genetic stability. The top strand of the duplex TRS on both sides of the figure is the transcribed strand, as well as the leading strand template for DNA synthesis. The left side of Figure 3 represents orientation II, whereas the right side shows orientation I. Transcription of the (CAG) strand (orientation II) leads to deletions, whereas transcription of the (CTG) strand (orientation I) elicits a much lower frequency of deletions. The model proposes that as the (CAG) strand is being transcribed, the complementary (CTG) strand folds back and forms a hairpin. On the other hand, the non-transcribed (CAG) strand in orientation I is less able to form stable hairpins; these differences in thermodynamic stability are based on studies conducted on single-stranded oligonucleotides (63). The model further envisages that while the TRS is transcribed, it is also replicated. In this case, the (CTG) hairpin in

orientation II will be bypassed by the DNA polymerase complex during lagging-strand synthesis, and this will lead to deletions (instability). Conversely, in orientation I, fewer deletions will be found since fewer secondary structures are formed on the lagging-strand (CAG) template and thus no bypass synthesis occurs.

Therefore, the crucial component of this model is the selective formation of hairpins on the non-transcribed (CTG) strand. Transcription is accompanied by an accumulation of positive supercoils in front of the RNA polymerase complex and negative supercoils behind it (64,65). It is also known that negative supercoiling unwinds the helical duplex DNA and promotes strand separation (66). Measurements of the elastic constants of $(CTG•CAG)_n$ and $(CGG•CCG)_n$ and calculations of their free energy of supercoiling (58,67,68) revealed their flexibility and their writhed structure, compared to random B-DNA. Hence, these TRS were proposed to act as a 'sink' for the accumulation of superhelical density (58,67). In addition to this superhelical tension (free energy of supercoiling), in orientation II the (CTG) strand is not complexed by RNA polymerase, as in the case of orientation I, but rather is transiently single-stranded. We suggest that this 'transient single-strandedness' of the (CTG) strand in orientation II enables the nucleation of a folded-back (CTG) region into a hairpin.

The results obtained for the NER deficient *E.coli* strains show that the absence of certain repair functions dramatically influences the stability of the long transcribed (CTG•CAG) tracts in plasmids. This was manifested by the higher stability of the TRS in the *uvrB* mutant (UvrA protein present) than observed in the *uvrA* strain. Figure 4 presents a model for the involvement of the UvrA protein in the stabilization of the long (CTG•CAG) sequence *in vivo*. The top part of the figure represents a replication fork for the TRS region in a plasmid containing the $(CTG•CAG)_{175}$ in orientation II. The (CTG) hairpin formed during the lagging-strand synthesis (Fig. 3) is shown inside the shaded circle (which is the focus of the remainder of the figure). The left side of Figure 4 shows how the UvrA protein may contribute to the stabilization of the TRS. This part of the model is based on the data obtained from the analyses of the $(CTG•CAG)_{175}$ containing plasmid as grown and analyzed in the *uvrB* mutant. This strain is NER deficient but is able to produce the functional UvrA protein. The model predicts that the UvrA protein specifically recognizes and binds the (CTG) hairpin formed during the lagging-strand synthesis (Fig. 4, left side). This binding of the dimeric UvrA (69) to the (CTG) hairpin causes the destabilization of the hairpin. The resulting molecule contains a single-stranded gap of the size of the previous hairpin. This restores the correct template for complementary strand synthesis. The gap is then filled by DNA polymerase, followed by the ligation at the nicks, and the resulting molecule contains the full-length, non-deleted (CTG•CAG) tract.

On the other hand, in the absence of the UvrA protein (*uvrA E.coli* strain) (Fig. 4, right side), the (CTG) hairpin will not be removed from the DNA template. Thus, bypass DNA synthesis through the TRS may occur, which would cause deletions. The affinity of the UvrA protein to single-stranded DNA (70), specifically to bubbles and loops (45), may be responsible for the recognition and binding to the (CTG) hairpins in their single-stranded loop region. Also, the increased instability of (CTG•CAG) inserts was shown *in vivo* in *E.coli* having a mutation in the *ssb* gene (22). The absence of the SSB *in vivo* led to an increased frequency of large deletions within the triplet repeats. Another possible scenario is that in the absence of the UvrA protein, the (CTG) hairpin is a substrate for the DNA nicking activities (71). Such nicked DNA may be degraded *in vivo* which would also lead to deletions. The greater stability of long (CTG•CAG) tracts in the *uvrB* mutant than observed in wild-type cells, as well as the difference in the stability of the TRS between wild-type cells and the *uvrAuvrB* double mutant, also suggests that the UvrB protein itself may possess some nicking activity toward the CTG hairpin loops.

We propose that the (CTG) hairpins share, at least to some extent, similar properties with a variety of the DNA lesions that are recognized and repaired by the NER system. The substrate specificity of the Uvr(A)BC endonuclease is very wide (41). The important factor involved in the recognition of lesions by UvrA is a distortion of the DNA structure. Cruciform DNA and three-way junction structures, that are basically similar to the hairpins formed by the (CTG) tracts, introduce a significant amount of bending into DNA (72,73). Recently, the triplet repeats were shown to be inherently flexible and writhed (58,67,68,74). Also, (CTG•CAG) tracts can form slipped structures (3,75) and under some circumstances appear as bent/kinked DNA molecules (75). Additionally, the unorthodox DNA conformations of the triplet repeats were shown to be responsible for the DNA polymerase pausing on those sequences *in vitro* (76,77). It is possible that such pausing *in vivo* causes the arrest of replication, the induction of the SOS response and, as a consequence, de-repression of NER protein synthesis. Also, it was shown that repair activity in HeLa extracts was induced by a triple helix, and the authors suggested that other non-duplex DNA structures might trigger repair systems (78).

Our results and the model (Fig. 4) are supported by recent discoveries demonstrating the involvement of the UvrA protein in blocking the DNA repair-dependent *trans*-lesion synthesis pathway operating in *E.coli* (79,80). The binding of the UvrA protein to DNA containing a UV lesion in a single-stranded DNA prevented the bypass synthesis. In the absence of the UvrA protein the *trans*-lesion synthesis was not inhibited and this process is believed to be responsible for SOS mutagenesis in *E.coli* (41). In our case, however, binding of the UvrA protein to the (CTG) hairpin may be sufficient to remove the 'lesion'. Such binding destabilizes the (CTG) hairpin, restores a 'non-mutated' template for replication and therefore prevents deletions.

It is unclear why the (CGG•CCG) and (GAA•TTC) sequences cloned in plasmids did not respond to the mutations in NER in our system. It is possible that these TRS may be too short and/or that some different structural properties of (CGG•CCG) and (GAA•TTC) (21,58,59) do not trigger the repair in *E.coli*. Other prior investigations have revealed the structural uniqueness of (CTG•CAG) relative to the other TRS (21,60).

We conclude that the binding of the UvrA protein to the (CTG) hairpins formed on the lagging-strand template removes its secondary structure and prevents deletions. Preferential formation of hairpins on Okazaki fragments is also believed to result in the expansion of long (CTG•CAG) repeats (5). Therefore, we speculate that these hairpins can be removed by human nucleotide excision proteins thereby stabilizing long (CTG•CAG) tracts and thus prevent expansions.

ACKNOWLEDGEMENTS

We thank A. Pluciennik and R. R. Iyer for helpful discussions. This work was supported by grants from the NIH (GM52982) and the R. A. Welch Foundation.

REFERENCES

1 Wells,R.D. and Warren,S.T. (eds) (1998) *Genetic Instabilities and Hereditary Neurological Diseases*. Academic Press, San Diego, CA.
2 Paulson,H.L. and Fischbeck,K.H. (1996) *Annu. Rev. Neurosci.*, **19**, 79–107.
3 Wells,R.D., Parniewski,P., Pluciennik,A., Bacolla,A., Gellibolian,R. and Jaworski,A. (1998) *J. Biol. Chem.*, **273**, 19532–19541.
4 Freudenreich,C.H., Stavenhagen,J.B. and Zakian,V.A. (1997) *Mol. Cell. Biol.*, **17**, 2090–2098.
5 Kang,S., Jaworski,A., Ohshima,K. and Wells,R.D. (1995) *Nature Genet.*, **10**, 213–218.
6 Kang,S., Ohshima,K., Jaworski,A. and Wells,R.D. (1996) *J. Mol. Biol.*, **258**, 543–547.
7 Richards,R.I. and Sutherland,G.R. (1994) *Nature Genet.*, **6**, 114–116.
8 Wells,R.D. (1996) *J. Biol. Chem.*, **271**, 2875–2878.
9 Maurer,D.J., O'Callaghan,B.L. and Livingston,D.M. (1996) *Mol. Cell. Biol.*, **16**, 6617–6622.
10 Samadashwily,G.M., Gordana,R. and Mirkin,S.M. (1997) *Nature Genet.*, **17**, 298–304.
11 Freudenreich,C.H., Kantrow,S.M. and Zakian,V.A. (1998) *Science*, **279**, 853–856.
12 Zhong,N. Kajanoja,E., Smits,B., Pietrofesa,J., Curley,D., Wang,D., Ju,W., Nolin,S., Dobkin,C., Ryynanen,M. and Brown,W.T. (1996) *Am. J. Med. Genet.*, **64**, 226–233.
13 O'Hoy,K.L., Tsilfidis,C., Mahadevan,M.S., Neville,C.E., Barcelo,J. Hunter,A.G.W. and Korneluk,R.G. (1993) *Science*, **259**, 809–813.
14 Losekoot,M., Hoogendoorn,E. Olmer,R., Jansen,C.C.A.M., Oosterwijk,J.C., van den Ouweland,A.M.W., Halley,D.J.J., Warren,S.T., Willemsen,R., Oostra,B.A. and Bakker,E. (1997) *J. Med. Genet.*, **34**, 924–926.
15 van den Ouweland,A.M.W., Deelen,W.H., Kunst,C.B., Uzielli,M.-L.G., Nelson,D.L., Warren,S.T., Oostra,B.A. and Halley,D.J.J. (1994) *Hum. Mol. Genet.*, **3**, 1823–1927.
16 Brown,W.T., Houck,G.E.,Jr, Diang,X., Zhong,N., Nolin,S., Glicksman,A., Dobkin,C. and Jenkins,E.C. (1996) *Am. J. Med. Genet.*, **64**, 287–292.
17 Jaworski,A., Rosche,W.A., Gellibolian,R., Kang,S., Shimizu,M., Bowater,R.P., Sinden,R.R. and Wells,R.D. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 11019–11023.
18 Bowater,R.P., Jaworski,A., Larson,J.E., Parniewski,P. and Wells,R.D. (1997) *Nucleic Acids Res.*, **25**, 2861–2868.
19 Bowater,R.P., Rosche,W.A., Jaworski,A., Sinden,R.R. and Wells,R.D. (1996) *J. Mol. Biol.*, **264**, 82–96.
20 Shimizu,M., Gellibolian,R., Oostra,B.A. and Wells,R.D. (1996) *J. Mol. Biol.*, **258**, 614–626.
21 Ohshima,K., Kang,S., Larson,J.E. and Wells,R.D. (1996) *J. Biol. Chem.*, **271**, 16773–16783.
22 Rosche,W.A., Jaworski,A., Kang,S.F., Kramer,S., Larson,J.E., Geidroc,D.P., Wells,R.D. and Sinden,R.R. (1996) *J. Bacteriol.*, **178**, 5042–5044.
23 Ohshima,K., Kang,S. and Wells,R.D. (1996) *J. Biol. Chem.*, **271**, 1853–1856.
24 Schweitzer,J.K. and Livingston,D.M. (1997) *Hum. Mol. Genet.*, **6**, 349–355.
25 Monckton,D.G., Coolbaugh,M.I., Ashizawa,K.T., Siciliano,M.J. and Caskey,C.T. (1997) *Nature Genet.*, **15**, 193–196.
26 Ashizawa,T., Monckton,D.G., Vaishnav,S., Patel,B.J., Voskova,A. and Caskey,C.T. (1996) *Genomics*, **36**, 47–53.
27 Steinbach,P., Wohrle,D., Glaser,D. and Vogel,W. (1998) In Wells,R.D. and Warren,S.T. (eds), *Genetic Instabilities and Hereditary Neurological Diseases*. Academic Press, San Diego, CA pp. 509–528.
28 Bronner,C.E., Baker,C.M., Morrison,P.T., Warren,G., Smith,L.G., Lescoe,M.K., Kane,M., Earabino,C., Lipford,J., Lindblom,A., Tanergard,P., Bollag,R.J., Godwin,A.R., Ward,D.C., Nordenskjold,M., Fishel,R., Kolodner,R. and Liskay,R.M. (1994) *Nature*, **369**, 258–261.
29 Papadopoulos,N., Nicolaides,N.C., Wei,Y., Ruben,S.M., Carter,K.C., Rosen,C.A., Haseltine,W.A., Fleischmann,R.D., Fraser,C.M., Adams,M.D., Venter,J.C., Hamilton,S.R., Petersen,G.N., Watson,P., Lynch,H.T., Peltomaki,P., Mecklin,J., de la Chapelle,A., Kinzler,K.W. and Vogelstein,B. (1994) *Science*, **263**, 1625–1629.
30 Umar,A., Boyer,J.C. and Kunkel,T.A. (1994) *Science*, **266**, 814–816.
31 Umar,A., Boyer,J., Thomas,D.C., Nguyen,D.C., Risinger,J.I., Boyd,J., Ionov,Y., Perucho,M. and Kunkel,T.A. (1994) *J. Biol. Chem.*, **269**, 14367–14370.
32 Fishel,R. and Kolodner,R.D. (1995) *Curr. Biol.*, **5**, 382–395.
33 Modrich,P. (1995) *Science*, **266**, 1959–1960.
34 Parsons,R., Li,G.-M., Longley,M.J., Modrich,P., Liu,B., Berk,T., Hamilton,S.R., Kinzler,K.W. and Vogelstein,B. (1995) *Science*, **268**, 738–740.
35 Marsischky,G.T., Filosi,M., Kane,M.F. and Kolodner,R. (1996) *Genes Dev.*, **10**, 407–420.
36 Modrich,P. and Lahue,R. (1996) *Annu. Rev. Biochem.*, **65**, 101–133.
37 Kirkpatrick,D.T. and Petes,T.D. (1997) *Nature*, **387**, 929–931.
38 Sia,E.A., Kokoska,R.J., Dominska,M., Greenwell,P. and Petes,T.D. (1997) *Mol. Cell. Biol.*, **17**, 2851–2858.
39 Strauss,B.S., Sagher,D. and Acharya,S. (1997) *Nucleic Acids Res.*, **25**, 806–813.
40 Kramer,P.R., Pearson,C.E. and Sinden,R.R. (1996) *Hum. Genet.*, **98**, 151–157.
41 Friedberg,E.C., Walker,G.C. and Siede,W. (1995) *DNA Repair and Mutagenesis*. ASM Press, Washington, DC.
42 Cleaver,J.E. and Kraemer,K.H. (1989) In Scriver,C.R., Beaudet,A.L., Sly,W.S. and Valle,D. (eds), *The Metabolic Basis of Inherited Diseases*. McGraw-Hill, New York, NY, Vol. 2, pp. 2949–2971.
43 Huang,J.-C., Hsu,D.S., Kazantsev,A. and Sancar,A. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 12213–12217.
44 Sancar,A. (1996) *Annu. Rev. Biochem.*, **65**, 43–81.
45 Ahn,B. and Grossman,L. (1996) *J. Biol. Chem.*, **271**, 21462–21470.
46 Herrero,M., de Lorenzo,V. and Timmis,K.N. (1990) *J. Bacteriol.*, **172**, 6568–6572.
47 Ohshima,K., Montermini,L., Wells,R.D. and Pandolfo,M. (1998) *J. Biol. Chem.*, **273**, 14588–14595.
48 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
49 Mitas,M., Yu,A., Dill,J., Kamp,T.J., Chambers,E.J. and Haworth,I.S. (1995) *Nucleic Acids Res.*, **23**, 1050–1059.
50 Smith,G.K., Jie,J., Fox,G.E. and Gao,X. (1995) *Nucleic Acids Res.*, **23**, 4303–4311.
51 Yu,A., Dill,J. and Mitas,M. (1995) *Nucleic Acids Res.*, **23**, 4055–4057.
52 Zheng,M., Huang,X., Smith,G.K., Yang,X. and Gao,X. (1996) *J. Mol. Biol.*, **264**, 323–336.
53 Petruska,J., Arnheim,N. and Goodman,M.F. (1996) *Nucleic Acids Res.*, **24**, 1992–1998.
54 Howard-Flanders,P., Boyce,R.P. and Theriot,L. (1966) *Genetics*, **53**, 1119–1136.
55 Selby,C.P. and Sancar,A. (1993) *Science*, **260**, 53–58.
56 Hanawalt,P.C. (1994) *Science*, **266**, 1957–1958.
57 Friedberg,E.C. (1996) *Annu. Rev. Biochem.*, **65**, 15–42.
58 Bacolla,A., Gellibolian,R., Shimizu,M., Amirhaeri,S., Kang,S., Ohshima,K., Larson,J.E., Harvey,S.C., Stollar,B.D. and Wells,R.D. (1997) *J. Biol. Chem.*, **272**, 16783–16792.
59 Wang,Y.-H., Gellibolian,R., Shimizu,M., Wells,R.D. and Griffith,J. (1996) *J. Mol. Biol.*, **263**, 511–516.
60 Wang,Y.-H., Amirhaeri,S., Kang,S., Wells,R.D. and Griffith,J. (1994) *Science*, **265**, 669–671.
61 Sancar,A. (1994) *Science*, **266**, 1954–1956.
62 Sancar,A. (1995) *J. Biol. Chem.*, **270**, 15915–15918.
63 Gao,X., Huang,X., Smith,G.K. and Zheng,M. (1998) In Wells,R.D. and Warren,S.T. (eds), *Genetic Instabilities and Hereditary Neurological Diseases*. Academic Press, San Diego, CA, pp. 623–646.
64 Wang,J.C. (1996) *Annu. Rev. Biochem.*, **65**, 635–692.
65 Rahmouni,A.R. and Wells,R.D. (1989) *Science*, **246**, 358–363.
66 Sinden,R.R. (1994) *DNA Structure and Function*. Academic Press, San Diego, CA, pp. 95–133.
67 Gellibolian,R., Bacolla,A. and Wells,R.D. (1997) *J. Biol. Chem.*, **272**, 16793–16797.
68 Bacolla,A., Bowater,R.P. and Wells,R.D. (1998) In Wells,R.D. and Warren,S.T. (eds), *Genetic Instabilities and Hereditary Neurological Diseases*. Academic Press, San Diego, CA, pp. 467–484.
69 Mazur,S.J. and Grossman,L. (1991) *Biochemistry*, **30**, 4432–4443.
70 Seeberg,E. and Steinum,A.-L. (1982) *Proc. Natl Acad. Sci. USA*, **79**, 988–992.
71 Yao,M. and Kow,Y.W. (1996) *J. Biol. Chem.*, **271**, 30672–30676.
72 Gough,G.W. and Lilley,D.M. (1985) *Nature*, **313**, 154–156.
73 Lilley,D.M.J. and Clegg,R.M. (1993) *Quart. Rev. Biophys.*, **26**, 131–175.
74 Chastain,P.D. and Sinden,R.R. (1998) *J. Mol. Biol.*, **275**, 405–411.
75 Pearson,C.E., Wang,Y.-H., Griffith,J.D. and Sinden,R.R. (1998) *Nucleic Acids Res.*, **26**, 816–823.
76 Kang,S., Ohshima,K., Shimizu,M., Amirhaeri,S. and Wells,R.D. (1995) *J. Biol. Chem.*, **270**, 27014–27021.
77 Ohshima,K. and Wells,R.D. (1997) *J. Biol. Chem.*, **272**, 16798–16806.
78 Wang,G., Seidman,M.M. and Glazer,P.M. (1996) *Science*, **271**, 802–805.
79 Cohen-Fix,O. and Livneh,Z. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 3300–3304.
80 Paz-Elizur,T., Barak,Y. and Livneh,Z. (1997) *J. Biol. Chem.*, **272**, 28906–28911.