Involvement of the Nucleotide Excision Repair Protein UvrA in Instability of CAG-CTG Repeat Sequences in Escherichia coli*

Elena A. Oussatcheva‡, Vera I. Hashem‡, Yue Zou§, Richard R. Sinden‡, and Vladimir N. Potaman†‡

From the ‡Institute of Biosciences and Technology, Texas A&M University System Health Science Center, Houston, Texas 77030, and §Sealy Center for Molecular Science, University of Texas Medical Branch, Galveston, Texas 77555

Several human genetic diseases have been associated with the genetic instability, specifically expansion, of trinucleotide repeat sequences such as (CTG)n, (CGG)n, (CAG)n, (CCG)n, and (GAA)n. Molecular models of repeat instability imply replication slippage and the formation of loops and imperfect hairpins in single strands. Subsequently, these loops or hairpins may be recognized and processed by DNA repair systems. To evaluate the potential role of nucleotide excision repair in repeat instability, we measured the rates of repeat deletion in wild type and excision repair-deficient Escherichia coli strains (using a genetic assay for deletions). The rate of triplet repeat deletion decreased in an E. coli strain deficient in the damage recognition protein UvrA. Moreover, loops containing 23 CTG repeats were less efficiently excised from heteroduplex plasmids after their transformation into the uvrA− strain. As a result, an increased proportion of plasmids containing the full-length repeat were recovered after the replication of heteroduplex plasmids containing unrepaird loops. In biochemical experiments, UvrA bound to heteroduplex substrates containing repeat loops of 1, 2, or 17 CAG repeats with a $K_d$ of about 10–20 nM, which is an affinity about 2 orders of magnitude higher than that of UvrA bound to the control substrates containing (CTG)n/(CAG)n in the linear form. These results suggest that UvrA is involved in triplet repeat instability in cells. Specifically, UvrA may bind to loops formed during replication slippage or in slipped strand DNA and initiate DNA repair events that result in repeat deletion. These results imply a more comprehensive role for UvrA, in addition to the recognition of DNA damage, in maintaining the integrity of the genome.

Repetitive DNA sequences in the human genome are clearly involved in disease-associated mutations or genetic rearrangements (1–4). Several human genetic diseases, including Huntington’s disease, fragile X syndrome, myotonic dystrophy, and Friedreich’s ataxia, have been associated with the genetic instability (length expansion) of trinucleotide repeat sequences such as (CTG)n,(CAG)n, (CGG)n,(CCG)n, and (GAA)n,(TTC)n (see Refs. 5 and 6 for review). Some types of cancer, such as prostate cancer and nonpolyposis colorectal cancer, are also associated with the instability of repeat sequences (7–9). Certain molecular models of repeat instability have implied a role for alternative DNA structures in aberrant DNA replication, repair, and recombination (10). Repeat instability may also involve replication slippage, which would be associated with the formation of loops and imperfect hairpins in single strands (11–13). The data available to date indicate that both short (1–3 repeats) and long (likely in multiples of 40 repeats) repeat tracts may loop out (14–18). The extent of hairpin-stabilizing internal hydrogen bonding in looped-out strands depends on their length and sequence (13, 19, 20). For example, (CTG)n and (CGG)n form more stable hairpins than their corresponding complements, (CAG)n and (CCG)n. Subsequent loop/hairpin recognition and processing by some postreplicative mechanism are likely involved (21). The identities of the systems responsible for loop repair remain unclear.

Whereas the methyl-directed mismatch repair (22–24) is clearly involved in very short repeat instability, one may not expect significant effects on the large loop processing. This is because the methyl-directed mismatch repair system is only able to recognize mismatches/heteroduplexes as large as 3 nucleotides (25, 26). Indeed, mutations in various methyl-directed mismatch repair functions in bacteria, yeast, and humans did not change their ability to remove large nonrepeating loops (27–29). The BscCD nuclease, which has been previously shown to be responsible for cleaving hairpins in vitro (30) and generating double-stranded breaks at inverted repeats in vivo (31), may play a role in the large loop processing. Furthermore, the nucleotide excision repair (NER)1 system (32), which repairs bulky DNA adducts at damaged nucleobases, is likely involved in loop repair because the recognition specificity of NER proteins extends beyond the damaged bases. In particular, strand breaks and flaps and unpaired strands in heteroduplex DNA may serve as binding substrates for UvrA, which is a primary recognition protein in the NER system of Escherichia coli (33–36). We were interested in evaluating the contributions of DNA repair systems that may process relatively large loops arising from the DNA strand slippage. Using a genetic assay for deletions, we found that the rates of triplet repeat deletions decreased in the UvrA-deficient E. coli strain. Heteroduplex plasmids containing trinucleotide repeat loops in one strand were less efficiently repaired in the UvrA-deficient E. coli strain after transformation. Purified UvrA binds to heteroduplex substrates containing repeat loop-outs with an

* This work was supported by National Institutes of Health Grant ES05508 (to R. R. S.) and a grant from the Texas A&M University System Health Science Center Council on Tobacco and Environmental Health Research (to V. N. P.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Institute of Biosciences and Technology, Texas A&M University System Health Science Center, 2121 W. Holcombe Blvd., Houston, TX 77030; Tel.: 713-677-7675; Fax: 713-677-7689; E-mail: vpotaman@ibt.tamu.edu.

‡ The abbreviations used are: NER, nucleotide excision repair; amp, ampicillin; bp, base pair; nt, nucleotide(s); CAT, chloramphenicol acetyl transferase; LB, Luria-Bertani broth.
Involvement of UvrA in Instability of Repeated Sequences

TABLE I

| Wild type | uvrA | uvrB | sscC |
|-----------|------|------|------|
| (CAG)25   | 458  | 413  | 569  |
| (CAG)23   | 3.2  | 1.1  | 1.1  |
| (CAG)19   | 19.2 | 10.1 | 4.1  |
| (CTG)25   | 20.2 | 4.2  | 80.3 |
| (CTG)19   | 0.33 | 0.18 | 0.22 |

Rates (10−4) of reversion to Cm′

- Statistically significant differences in mutation rates for the mutant cells compared to the wild type ones.

Experimental Procedures

Bacterial Strains—The following E. coli strains were used: AB1157 (thr−1, ara−C14, leuB6, Δgpt-proA62, lacY1, tsx−33, mar−, galK2(Oc), λ−, Rac−, hisG416(Oc), rfbD1, rpsL−51, rpoS396(hsAm), rpsL31(strR), kgk51, xylA5, ndt−1, argE3(Oc), thi−1), AB1886 as AB1157 with uvrA6, and AB1885 as AB1157 with uvrB5. Strains AB1157, AB1886, and AB1885 were obtained from the E. coli Genetic Stock Center at Yale University. To study the influence of mutations in the SbcCD nuclease on the trinucleotide repeat sequences, we used strain PF2070 (a gift from P. Foster, Indiana University) as AB1157 with sscC.

Plasmids—Plasmid pUC8NcoI is identical to pUC8, except that it contains a NcoI site as a result of a 1-base pair A-T substitution for the G-C base pair at position 225 of pUC8 sequence introduced by the polymerase chain reaction point mutation technique (37). To assemble (CTG)n (CAG)n repeats from oligonucleotides and clone them in pBR325, short oligonucleotides (CTG)n and (AGC)n were annealed and ligated to create a series of plasmids [(CTG)n (CAG)n]m. With the EcoRI site in the CAT gene that is inactivated by an insert of 25 repeats or more, making cells Cm′ sensitive. Deletion of the repeat from the plasmids was achieved by Cl857 (38). Determinations of rates of triplet repeat deletion were based on reversion to chloramphenicol-resistant (Cm′) phenotype. The CAT gene is inactivated by a (CTG)n insert of 25 repeats or more, making cells chloramphenicol-sensitive (Cm′). Deletion of the repeat tract to n ≤ 24 makes cells chloramphenicol-resistant. Because of the triplet repeat instability in E. coli, the transformants were grown on LB + amp plates for a minimal amount of time to detect colony formation. Colonies were then stored as frozen glycerol stocks for future experiments. For determination of mutation rates, frozen cells were cultured on LB + amp plates to determine the viable cell count and onto LB + amp + Cm plates to determine Cm′ revertants. A minimum of six independent reversion assays from different colonies were performed for each plasmid, and mutation rates were determined according to Drake (39). Median values and confidence intervals are shown. The higher mutation rates for the CAG repeats in the leading strand template were statistically different from those for the CTG repeats in the leading strand (bottom three lines).

Affinity about 2 orders of magnitude higher than that to linear DNA. Altogether, this indicates that UvrA is involved in triplet repeat instability related to the formation of large single-stranded loops.

Determination of Mutation Rates—Plasmid pBR325 contains genes conferring antibiotic resistance to ampicillin (amp) and chloramphenicol. Triplet repeats were cloned in the unique EcoRI site in the CAT gene (38). Determinations of rates of triplet repeat deletion were based on reversion to a chloramphenicol-resistant (Cm′) phenotype. The CAT gene is inactivated by a (CTG)n (CAG)n insert of 25 repeats or more, making cells chloramphenicol-sensitive (Cm′). Deletion of the repeat tract to n ≤ 24 makes cells chloramphenicol-resistant. Because of the triplet repeat instability in E. coli, the transformants were grown on LB + amp plates for a minimal amount of time to detect colony formation. Colonies were then stored as frozen glycerol stocks for future experiments. For determination of mutation rates, frozen cells were distributed onto LB + amp plates and allowed to grow to small colonies. A culture was then grown from a single colony in LB to mid-log phase, after which cell dilutions were plated onto LB + amp plates to determine the viable cell count and onto LB + amp + Cm plates to determine Cm′ revertants. A minimum of six independent reversion assays from different colonies were performed for each plasmid, and mutation rates were determined according to Drake (39). p values were calculated using a Mann-Whitney test. The nature of reversion events was analyzed by polymerase chain reaction and sequence analysis on DNA isolated by the Magic mini prep procedure (Promega).

Analysis of the Stability of (CTG)n and (CAG)n—Hairpin-containing Heteroduplex DNA after Transformation—To form heteroduplexes, plasmids were digested with AIII and NdeI producing 2063-bp and 602-bp AIII-NdeI fragments for pUC8NcoI and similar fragments for pEO(CAG)23H or pEO(CTG)23H. About 5 μg of pUC8NcoI and either pEO(CAG)23H or pEO(CTG)23H cut with AIII-NdeI were mixed in 40 μl of hybridization buffer and hybridized as described previously (40).
Heteroduplex isomers formed by the short fragments of pUC8NcoI and pEO/CAG/23H were isolated from 5% polyacrylamide gel and purified using an UltraClean 15 plasmid purification kit from MoBio (Solana Beach, CA). Heteroduplex species were identified by sequencing using a Magic mini-prep procedure (Promega) and analyzed for the presence of the NcoI site and the (CAG)_{23} insertion. For this, plasmids were digested separately with NcoI and PvuII and analyzed on a 1% agarose gel.

Binding of UvrA Protein to the (CAG)_{n} Loop-containing Heteroduplex DNA—For experiments on UvrA binding to heteroduplexes, each of the plasmids p(CAG)_{6}E, p(CAG)_{7}E, p(CAG)_{8}E, and pEO/CAG/23E was digested with PvuII producing short (301-bp) triplet repeat-containing PvuII-PvuII fragments and longer (2364-bp) fragments of vector DNA. After dephosphorylation, restriction fragments were 5'-end-labeled with [γ-32P]ATP. Upon hybridization of the short PvuII-PvuII fragment of plasmid p(CAG)_{6}E with each of the similar fragments of plasmids p(CAG)_{7}E, p(CAG)_{8}E, and pEO/CAG/23E, the excess repeats formed loop-outs of 3, 6, and 51 nt, respectively. These heteroduplex fragments were separated on a 5% native polyacrylamide gel from the correctly annealed linear species, and, in the case of 51 nt loop-outs, heteroduplex isomers were separated from each other. The 301-bp heteroduplex fragments isolated from the gel were then digested with HindIII to produce mixtures of 205-bp end-labeled heteroduplex fragments with the CAG loop-outs, unlabeled heteroduplex fragments with the CTG loop-outs, and 96-bp labeled linear fragments. After phenol-chloroform extraction and ethanol precipitation, mixtures containing one specific labeled heteroduplex and a labeled linear fragment were used for protein binding experiments. Binding of the UvrA protein to the DNA substrates was determined by gel mobility shift assays. Typically, the substrate (2 nM) was incubated with UvrA with varying concentrations as indicated at 37 °C for 15 min in 20 μl of UvrABC buffer (50 mM Tris-HCl, pH 7.5, 75 mM KCl, 10 mM MgCl₂, and 5 mM dithiothreitol) in the presence of 1 mM ATP. After incubation, 2 μl of 80% (v/v) glycerol was added, and the mixture was loaded immediately onto a 3.5% native polyacrylamide gel in TBE (90 mM Tris borate (pH 8.3), 2.5 mM EDTA) running buffer and electrophoresed at room temperature. After quantification of the radioactivity of the corresponding bands in the gel, the dissociation constants were estimated from the protein concentration at the point where half of the DNA was bound.

RESULTS
Differential Deletion of Large Repeat Fragments in the Wild Type and UvrAB Mutants—Table I shows the effects of mutations in the uvrA, uvrB, and sbcC genes on the deletion rates of (CAG)_{23} in E. coli. Cloning the triplet repeat tracts in the CAT gene results in the insertion of a polyglutamine-coding sequence that inactivates the CAT gene, rendering cells Cm^{R} when an insert is 25 repeats or longer. Partial deletions of the triplet repeats to <25 repeats result in a Cm^{R} phenotype. Therefore, the high mutation rates for plasmids containing 25 repeats correspond to deletions that may be as short as 1–2 repeats, whereas deletions of at least 19 and 55 repeats are required for Cm^{R} reversion for plasmids containing 43 or 79 repeats. For the two possible insert orientations, the mutation rates were 10–50-fold higher when the CAG repeat comprised the leading strand template, and the CTG comprised the lagging strand template. The differences were statistically significant (p < 0.025). This may be explained by a better propensity of the lagging strand CTG repeats to form hairpins that may be bypassed by DNA polymerase, leading to deletion (41). In either repeat orientation, the mutation rates are slightly lower in the uvrA and uvrB cells compared with the wild type strain for repeat lengths of 43 and 79, where deletions of at least 19 and 55 repeats must occur. For (CAG)_{23} in the leading strand, the differences between the wild type and uvrA and uvrB cells were statistically significant (p < 0.025), whereas for the CAG repeats in the leading strand, the difference for (CAG)_{23} was statistically significant (p < 0.05). Thus, these genetic experiments show that inactivation of either UvrA or UvrB makes large deletions of triplet repeats in bacteria less likely. Muta-

**Fig. 2. Preparation of (CTG)_{23} and (CAG)_{23} hairpin-containing plasmids.** A, structural features of the heteroduplex DNA isomers important for construct design and analysis. The correct sequence of the NcoI site is in both cases on the opposite strand to the hairpin-containing strand. B, schematics of the hairpin-containing plasmids for cell transformation. Cloned heteroduplex isomers are shown in gray. An open circle indicates the nucleotide substitutions to form the NcoI site.

**Fig. 3. Analysis of the progeny plasmids after transformation of the heteroduplex-containing plasmids in E. coli.** Progeny plasmids were digested with either PvuII or NcoI, and the products were separated on a 1% agarose gel. A, PvuII digestion. Lane 1, pUC8NcoI; lane 2, pUC8; lane 3, a mixture of plasmids with and without the insert; lane 4, insert-containing plasmid pEO/CAG/23H. B, NcoI digestion. Lanes 1–4, same as those described in A.
Involvement of UvrA in Instability of Repeated Sequences

For each heteroduplex construct, bacterial strains were transformed twice in separate experiments. Plasmids from randomly selected clones were analyzed for the retention of triplet repeat insert sequences, the presence of the marker NcoI site, a mixture of both plasmids, and excision of both the loop and the NcoI site to produce pUC8. Deletions resulting in loss of PvuII sites were not identified. Percentages of clones containing the indicated plasmids are shown.

| Strains         | No. of clones | Inserts | pUC8NcoI | Plasmid mix | pUC8 | Deletions |
|-----------------|---------------|---------|----------|-------------|------|-----------|
| (CAG)_{23} loop in lagging strand |              |         |          |             |      |           |
| Wild type       | 39            | 0       | 64.2     | 25.6        | 5.1  | 5.1       |
| uvrA            | 40            | 2.5     | 70.0     | 25.0        | 0    | 2.5       |
| uvrB            | 40            | 0       | 72.5     | 22.5        | 0    | 5.0       |
| sbcC            | 44            | 0       | 77.2     | 20.5        | 0    | 2.3       |
| (CTG)_{23} loop in lagging strand |              |         |          |             |      |           |
| Wild type       | 33            | 0       | 78.8     | 0           | 3.0  | 18.2      |
| uvrA            | 19            | 0       | 94.7     | 0           | 5.3  | 0         |

Processing of Repeat Hairpin-containing Plasmids in Bacteria—To get further insight into the role that UvrA and UvrB may play in triplet repeat instability, we performed experiments on triplet repeat loop processing in wild type and mutant E. coli strains. To identify which strand of the heteroduplex plasmid was used as the template for replication during repair, heteroduplexes were made from a modified pUC8 plasmid (pUC8NcoI), in which a NcoI site was introduced into the original pUC8 by a single-base mutation, and pUC8-based plasmids carrying triplet repeat inserts but no NcoI site. As a result, one strand of the heteroduplex constructs contained a NcoI site as a marker, whereas the other strand contained a triplet repeat hairpin. Hybridization ofAfIII- and NdeI-digested plasmids pUC8NcoI and pEO(CAG)23H produced heteroduplex DNA fragments that were separated on 5% polyacrylamide gel (Fig. 1). As shown in Fig. 2A, the slower-migrating heteroduplex isomer contained the top strand of pUC8NcoI and the bottom (CTG) strand of pEO(CAG)23H, whereas the faster-migrating heteroduplex isomer contained the top (CAG) strand of pEO(CAG)23H and the bottom strand of pUC8NcoI. Thus, each heteroduplex isomer contained a 1-bp mismatch 40 nt from a 75-nt loop-out (presumably, an imperfect hairpin). The (CTG)- and (CAG)-containing heteroduplex fragments were isolated from the gel and ligated with the 2063-bp AfIII-NdeI fragment of pUC8 so that the (CAG)_{23} loop-out was in the leading strand template for DNA replication, and the (CTG)_{23} loop-out was in the lagging strand (Fig. 2B). Similarly, heteroduplex plasmids with the (CAG)_{23} loop-out in the leading strand template and the (CTG)_{23} loop-out in the lagging strand were made using heteroduplex isomers prepared from pUC8NcoI and pEO(CTG)23H. Because we used plasmids isolated from the E. coli strain HB101, in which dam methylation mechanism is active, both heteroduplex strands are methylated. Therefore, the repair of the mismatch at the NcoI site should proceed randomly with either one or another strand being used as the template for repair (42). Consistent with this assumption, in the progeny plasmids from 20 clones after wild type E. coli transformation with heteroduplex plasmids containing one strand with the sequence of NcoI site and the complementary strand of pUC8 (therefore without the NcoI sequence), 9 plasmids had a pUC8 sequence, 8 had a modified sequence with the NcoI site, and 3 contained a mixture of both.

Plasmids recovered from cells after heteroduplex transformation were analyzed by separate digestions with PvuII and NcoI (Fig. 3). The size of the PvuII-PvuII fragment indicated whether the repeat insert in individual clones had been lost (Fig. 3A, lanes 1 and 2) or retained (lane 4). The latter would occur if the repeat-containing strand were chosen as a template during repair replication. If a heteroduplex plasmid escaped repair, subsequent replication would result in a mixture of the NcoI site- and repeat-containing plasmids in the same cell (lanes 3 in Fig. 3, A and B). For individual clones, the NcoI digestion distinguished the plasmids containing no inserts as either pUC8NcoI (Fig. 3B, lane 1) or pUC8 (lane 2). For a pure plasmid population, susceptibility to NcoI likely means that the NcoI site-containing strand was chosen as a template during repair replication.

Table II shows the effects of mutations in uvrA, uvrB, and sbcC on loop-out processing by E. coli cells when either (CAG)_{23} or (CTG)_{23} loop-outs are in the lagging strand for DNA replication. Upon transformation of the wild-type strain AB1157 with heteroduplex plasmids, no clones with pure insert-containing plasmids were found for either loop-out. Most of the clones contained pUC8NcoI, and minor fractions of clones contained pUC8. This is consistent with loop recognition as a target for excision and with the NcoI-containing strand being chosen as a template for replication during heteroduplex repair. The repair patch seems to be long enough to include both the heteroduplex loop-out and a single-base pair mismatch separated by 40 bp. About one-fourth of the clones formed after cell transformation with the (CAG)_{23} loop-out heteroduplex contained a mixture of plasmids. The plasmid mixture contained both pUC8NcoI and (CAG)_{23}-containing plasmids, consistent with replication of unrepaired heteroduplex construct, so that both strands were replicated, and the two plasmids coexisted in the same transformant colony. No significant differences were detected for processing of heteroduplex constructs in wild type strain AB1157 and strains AB1885 (mutant for UvrB protein, which is required to form the UvrBC-DNA incision complex during NER (43)) and PP2070 (mutant for the SbcCD protein complex, an exonuclease known to prevent replication of long palindromes (21, 30)). Upon heteroduplex transformation into strain AB1886, which is mutant for the damage recognition protein UvrA, 1 of 40 clones contained a pure insert-containing plasmid in the case of the (CAG)_{23} loop-out.

Table III shows the effects of mutations in uvrA, uvrB, and sbcC on loop-out processing by E. coli cells when either (CAG)_{23} or (CTG)_{23} loop-outs are in the leading strand for DNA replication. Similar to the wild-type strain AB1157, no clones with pure insert-containing plasmids were found for strains AB1885 and PP2070 with mutations in the uvrB and sbcC genes, respectively. In all three strains, recovered plasmids predominantly contained the NcoI site. However, in the case of (CTG)_{23} loop-outs, a significant fraction of clones contained a mixture of plasmids. The percentage of cells containing the mixture was somewhat higher for the SbcC-deficient strain compared with the wild type and UvrB-deficient strains. Compared with other strains, in the UvrA-deficient strain, the (CTG)_{23} inserts were retained at a higher rate. Three of 40 clones contained pure...
repeat-containing plasmids, and 17 clones contained a mixture of plasmids. The number of clones containing plasmids with undeleted repeats increased at the expense of clones containing pure pUC8NcoI plasmids. Thus, it appeared that protein UvrA, which recognizes damaged nucleotides and loads UvrB at the damage site, was also able to recognize loop-outs of trinucleotide repeats, and in the absence of such recognition, fewer loop-outs were processed by repair systems, resulting in an increased amount of the insert-containing plasmids.

**UvrA Binding to Heteroduplex Substrates**—In the *E. coli* NER, the functional form of UvrA is a UvrA$_2$ dimer that interacts with UvrB to form a UvrA$_2$B complex and load UvrB to the damage site. In addition, UvrA$_2$ itself recognizes DNA damage (44). To determine whether UvrA$_2$ interacts directly with triplet repeat loop-outs and may thereby influence the deletion of triplet repeats, the binding of purified protein to the heteroduplex loops was tested. One strand of heteroduplex substrates contained 6 CTG repeats, whereas the complementary strands contained either 7, 8, or 23 CAG repeats (Fig. 4A). Thus, the loop sizes were 1, 2, and 17 triplet repeats or 3, 6, and 51 nucleotides. Double-stranded DNA fragments carrying inserts were used as control substrates. The nonspecific UvrA$_2$ binding to control substrates was characterized by dissociation constants in the low micromolar range, consistent with the previously published data (45, 46). In all three cases of heteroduplex substrates, strong protein binding was detected (Fig. 4, B–D). The dissociation constants were about 2 orders of magnitude lower than those for control double-stranded DNA substrates, which indicated specific UvrA$_2$ binding to the heteroduplex DNA (Fig. 4E). However, no UvrB-DNA complex has been detected in the gel mobility shift assay upon incubation of the heteroduplex substrates with the protein complex UvrA$_2$B. Two options are possible: (i) the UvrB-DNA complex has not been detected due to the technical difficulties as reported for some N-2-acetylaminofluorene adducts (47), and (ii) although UvrA$_2$ recognizes both CTG loops and damaged nucleotides, it does not load UvrB onto the loops and influences the repeat loop excision in a different way than the excision of damaged nucleotides. The latter option is consistent with the data on processing of heteroduplex loop-outs in *E. coli* cells, where uvrB mutation had no effect.

**DISCUSSION**

Nucleotide excision repair is a major cellular repair pathway for a variety of DNA damages in both prokaryotes and eukaryotes (32). To investigate the potential contribution of this system to triplet repeat instability, we analyzed repeat stability in *E. coli* strains containing mutations in different genes involved in the nucleotide excision repair. A significant role of the UvrA protein has emerged as a consensus from genetic and biochemical experiments. When UvrA was inactive, the deletion rate of (CAG)$_{79}$/(CTG)$_{79}$ from plasmids propagated in bacteria was significantly lower than that in the wild type cells. UvrA may promote a removal of the replication misalignment loops, thus promoting shortening of the repeat length in progeny plasmids. Consistent with this idea, in the absence of active UvrA, fewer preformed repeat loop-outs were excised after transformation of heteroduplex molecules in *E. coli*. A direct involvement of UvrA-DNA loop interaction in repeat instability was supported by the demonstration of an interac-
Involvement of UvrA in Instability of Repeated Sequences

Several potential models may explain the involvement of UvrA in the instability of triplet repeat heteroduplex loops. First, UvrA may act as a recognition protein of this error of replication (or recombination). UvrA may recruit other repair proteins that may accomplish loop excision. Another possible explanation is that UvrA blocks replication of the loop-containing strand, making it highly inefficient, whereas the other strand is quite normally replicated. This is similar to a suggested replication interference of UvrA bound to the N2-aminofluorene adduct. The replication blockage may result in overrepresentation of progeny plasmids without repeats. Some clones, heteroduplex plasmids did not seem to undergo loop repair and then went into replication, producing mixtures of both repeat-containing and vector plasmids. The ratio of vectors to repeat-containing plasmids is severalfold, so it is possible that a partial block to replication by a UvrA-(CAG) loop complex results in this vector bias. Finally, a block to polymerization may induce a gap that would stall the replication fork and initiate fork restart, thereby inducing recombination events (57–60) that lead to increased rates of deletions.

REFERENCES

1. Cooper, D. N., and Krawczak, M. (1993) Human Gene Mutation, BIOS Scientific Publishers, Oxford, United Kingdom
2. Sinden, R. R., and Wells, R. D. (1992) Curr. Opin. Biotechnol. 3, 612–622
3. Richards, R. I., and Sutherland, G. R. (1997) Trends Biochem. Sci. 22, 432–436
4. Lusaki, J. R. (1996). Trends Genet. 14, 417–422
5. Wells, R. D., and Warren, S. T. (eds) (1998) Genetic Instabilities and Hereditary Neurological Diseases, Academic Press, San Diego, CA
6. Cummins, C. A., and Zoghbi, H. Y. (2000) Hum. Mol. Genet. 9, 909–916
7. Giovannucci, E., Stampfer, M. J., Kriskovan, K., Brown, M., Brufsky, A., Talcott, J., Hennekens, C. H., and Kantoff, P. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 3320–3325
8. Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1997) Nature 386, 623–627
9. Kantoff, P., Giovannucci, E., and Brown, M. (1998) Biochem. Biophys. Acta 1378, C1–C5
10. Pearson C. E., and Sinden, R. R. (1998) in Genetic Instabilities and Hereditary Neurological Diseases (Wells, R. D., and Warren, S. T., eds), pp. 585–621, Academic Press, San Diego, CA
11. Pearson C. E., and Sinden, R. R. (1998) Biochemistry 35, 5041–5053
12. Wang, Z., and Leach, D. R. F. (1998) J. Biol. Chem. 273, 24727–24730
13. Pearson C. E., and Sinden, R. R. (1989) Curr. Opin. Struct. Biol. 5, 231–140
14. Jaworski, A., Rosche, W. A., Gelibblisn, R., Kang, S., Shimizu, M., Bowater, R. R., Sinden, R. R., and Wells, R. D. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 11919–11923
15. Schummer, S., Fuchs, R. P. B., and Bichara, M. (1998) J. Mol. Biol. 279, 1191–1110
16. Wells, R. D., Parniewski, P., Pluciennick, A., Babola, A., Gelibbbisn, R., and Jaworski, A. (1998) J. Biol. Chem. 273, 19532–19541
17. Schmidt, K. H., Abbott, C. M., and Leach, D. R. (2000) Mol. Microbiol. 35, 463–471
18. Parniewski, P., Jaworski, A., Wells, R. D., and Bowater, R. P. (2000) J. Mol. Biol. 299, 865–874
19. Mitra, M. (1997) Nucleic Acids Res. 25, 2245–2253
20. Darlow, J. M., and Leach, D. R. F. (1998) J. Mol. Biol. 275, 3–16
21. Leach, D. R. (1994) Bioessays 16, 883–900
22. Modrich, P., and Lahue, R. (1996) Annu. Rev. Biochem. 65, 101–133
23. Modrich, P. (1997) J. Biol. Chem. 272, 24727–24730
24. Kolodner, R. D., and Marsichky, G. T. (1997) Curr. Opin. Genet. Dev. 9, 89–96
25. Parker, B. O., and Marinus, M. G. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 1730–1734
26. Genschel, J., Littmann, S. J., Drummond, J. T., and Modrich, P. (1998) J. Biol. Chem. 273, 19885–19901
27. Dohet, C., Dizdil, S., Wagner, R., and Radman, M. (1987) Mol. Gen. Genet. 206, 181–184
28. Corretti-Bennett, S. E., Parker, B. O., Mohlan, N. L., and Lahue, R. S. (1999) J. Biol. Chem. 274, 17605–17611
29. Littmann, S. J., Pang, W., and Kodina, P. (1999) J. Biol. Chem. 274, 7474–7481
30. Connelly, J. C., Kirkham, L. A., and Leach, D. R. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 7969–7974
31. Leach, D. R., Okley, E. A., and Finder, D. J. (1997) Mol. Microbiol. 26, 597–606
32. Sancar, A. (1996) Annu. Rev. Biochem. 65, 43–81
33. Ahr, B., and Grossman, L. (1969) J. Biol. Chem. 271, 21462–21470
34. Zou, Y., Walker, R., Bassett, H., Geacintov, N. E., and Van Houten, B. (1997) J. Biol. Chem. 272, 4820–4827
35. Zou, Y., and Van Houten, B. (1999) EMBO J. 18, 4889–4901
36. Moslenaar, G., Babine, M., van Knippenberg, I. C., Visse, R., and Gossen, N. (1998) J. Biol. Chem. 273, 34986–34993
37. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
38. Altman, N. K., and Vaghe, D. (1979) Nature 282, 864–869
39. Drake, J. W. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 7160–7164
40. Oussatsecheva, I., Shiyakhtenko, L. S., Glass, R., Sinden, R. R., Lyubchenko, Y. N., and Potamian, V. N. (1999) J. Mol. Biol. 292, 75–86
30884

Involvement of UvrA in Instability of Repeated Sequences

41. Kang, S., Jaworski, A., Ohshima, K., and Wells, R. D. (1995) Nat. Genet. 10, 213–218
42. Fishel, R. A., and Kolodner, R. (1984) Cold Spring Harbor Symp. Quant. Biol. 49, 603–609
43. Verhoeven, E. E. A., van Kesteren, M., Moolenaar, G. F., Visse, R., and Goosen, N. (2000) J. Biol. Chem. 275, 5120–5123
44. Zou, Y., Liu, T. M., Geacintov, N. E., and Van Houten, B. (1995) Biochemistry 34, 13582–13593
45. Van Houten, B., Gamper, H., Sancar, A., and Hearst, J. E. (1987) J. Biol. Chem. 262, 13180–13187
46. Mazur, S. J., and Grossman, L. (1991) Biochemistry 30, 4432–4443
47. Delagoutte, E., Bertrand-Burggraf, E., Dunand, J., and Fuchs, R. P. P. (1997) J. Mol. Biol. 266, 703–710
48. Parniewski, P., Bacolla, A., Jaworski, A., and Wells, R. D. (1999) Nucleic Acids Res. 27, 616–623
49. Ayares, D., Ganea, D., Chekuri, L., Campbell, C. R., and Kucherlapati, R. (1987) Mol. Cell. Biol. 7, 1656–1662
50. Weiss, U., and Wilson, J. H. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 1619–1623
51. Campbell, C. R., Ayares, D., Watkins, K., Wolski, R., and Kucherlapati, R. (1989) Mutat. Res. 211, 181–188
52. Fishel, R. A., and Kolodner, R. (1989) J. Bacteriol. 171, 3046–3052
53. Carraway, M., and Marinus, M. G. (1993) J. Bacteriol. 175, 3972–3980
54. Husain, I., Griffith, J., and Sancar, A. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 2558–2562
55. Tsao, H., Miao, B., Zhuang, P., Xu, R., Amin, S., and Geacintov, N. E. (1998) Biochemistry 37, 4993–5000
56. Bichara, M., and Fuchs, R. P. (1987) J. Bacteriol. 169, 423–426
57. Cox, M. M. (1998) Genes Cells 3, 65–78
58. Haber, J. E. (1999) Trends Biochem. Sci. 24, 271–275
59. Marians, K. J. (2000) Curr. Opin. Genet. Dev. 10, 151–156
60. Michel, B. (2000) Trends Biochem. Sci. 25, 173–178
Involvement of the Nucleotide Excision Repair Protein UvrA in Instability of CAG·CTG Repeat Sequences in *Escherichia coli*

Elena A. Oussatcheva, Vera I. Hashem, Yue Zou, Richard R. Sinden and Vladimir N. Potaman

*J. Biol. Chem.* 2001, 276:30878-30884.
doi: 10.1074/jbc.M104697200 originally published online June 18, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M104697200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 56 references, 24 of which can be accessed free at [http://www.jbc.org/content/276/33/30878.full.html#ref-list-1](http://www.jbc.org/content/276/33/30878.full.html#ref-list-1)