Nucleotide Sequence of the Wild-Type RAD4 Gene of Saccharomyces cerevisiae and Characterization of Mutant rad4 Alleles

LINDA B. COUTO† AND ERROL C. FRIEDBERG*

Laboratory of Experimental Oncology, Department of Pathology, Stanford University School of Medicine, Stanford, California 94305

Received 31 October 1988/Accepted 6 January 1989

Shuttle plasmids carrying the wild-type RAD4 gene of Saccharomyces cerevisiae cannot be propagated in Escherichia coli (R. Fleer, W. Siede, and E. C. Friedberg, J. Bacteriol. 169:4884-4892, 1987). In order to determine the nucleotide sequence of the cloned gene, we used a plasmid carrying a mutant allele that allows plasmid propagation in E. coli. The wild-type sequence in the region of this mutation was determined from a second plasmid carrying a different mutant rad4 allele. We established the locations and characteristics of a number of spontaneously generated plasmid-borne RAD4 mutations that alleviate the toxicity of the wild-type gene in E. coli and of several mutagen-induced chromosomal mutations that inactivate the excision repair function of RAD4. These mutations are situated in very close proximity to each other, and all are expected to result in the expression of truncated polypeptides missing the carboxy-terminal one-third of the Rad4 polypeptide. This region of the gene may be important both for the toxic effect of the Rad4 protein in E. coli and for its role in DNA repair in S. cerevisiae.

The RAD4 gene is one of at least five genes from the yeast Saccharomyces cerevisiae that are essential for damage-specific incision of DNA during the process of nucleotide excision repair (23, 34). Mutations in any of these five genes render cells abnormally sensitive to physical and chemical agents that generate bulky adducts in DNA and thus result in death of the cells (13). This selectable phenotype facilitated molecular cloning of the RAD1, RAD2, RAD3, and RAD10 genes (9). However, such a cloning strategy could not be employed to isolate the RAD4 gene from a yeast genomic library, because the wild-type gene is toxic to Escherichia coli and plasmids carrying the gene cannot be propagated in this organism (7).

In previous studies, it was determined that plasmids carrying certain mutations that inactivate the RAD4 function in S. cerevisiae can be propagated in E. coli (7, 8). RAD4 is known to be genetically closely linked to the SPT2 gene on chromosome V (35), and a plasmid designated pR169 (originally isolated by Roeder et al. (24)) carrying the spt2-1 allele was shown to also carry an inactivated RAD4 gene (7). This plasmid and plasmids derived from it were used to rescue the wild-type RAD4 allele by gap repair in S. cerevisiae (7). In subsequent studies, the plasmid-borne wild-type RAD4 gene was inactivated experimentally by insertional mutagenesis. This strategy facilitated propagation of the plasmid in E. coli at normal levels. The wild-type allele could then be restored by removing the inserted fragment prior to transformation of S. cerevisiae cells (8).

Plasmids that can be propagated in E. coli, such as pR169, arise spontaneously at a low frequency following transformation with vectors carrying the wild-type RAD4 gene (8). These plasmids presumably carry mutant rad4 alleles that alleviate the lethal effects of RAD4 gene expression in E. coli and hence may define regions of the gene that are important determinants for toxicity of the Rad4 protein in this organism. The rad4-2, rad4-3, and rad4-4 chromosomal mutations were originally generated by mutagenesis of repair-proficient yeast cells (2, 10, 30) and render cells defective in nucleotide excision repair. Information on the nature and location of these mutations is therefore expected to provide insights into regions of RAD4 required for its DNA repair function. With these goals in mind, we have determined the nucleotide sequence of the wild-type RAD4 gene. Additionally, we have mapped and characterized mutations in several plasmid-borne rad4 alleles generated spontaneously in E. coli as well as the mutagen-induced chromosomal alleles rad4-2, rad4-3, and rad4-4.

MATERIALS AND METHODS

Bacterial and yeast strains, plasmids, and culture conditions. The haploid yeast strains carrying rad4-2 or rad4-3 were described previously (7). Strain WS8014-2B (MATa rad4-4 ade2-1 trpl-1 289 ura3-52) was obtained from W. Siede, Department of Pathology, Stanford University. E. coli HB101 was used to propagate plasmids, and E. coli TG1 (a recA− derivative of strain JM101) was used as the recipient of vectors containing cloned RAD4 fragments.

Plasmid pUC119 is a derivative of pUC19 and was obtained from J. Vieira, Rutgers University, Piscataway, N.J. The series of pUC cloning vectors has been described by Vieira and Messing (33). The yeast plasmids pNF411 and pNF417 have 4.75- and 3.4-kilobase (kb) inserts, respectively, that carry the same mutant rad4 allele that is present in plasmid pNF402GR-EC1, which was previously described (7). Plasmid pNF422CEN-GR-IN contains the wild-type RAD4 gene with an inactivating DNA fragment inserted at the BglII site (8). Plasmid pR169 was obtained from Shirleen Roeder, Yale University, New Haven, Conn.

Yeast strains were grown in YPD medium (1% yeast extract, 2% Bacto-Peptone [Difco], 2% glucose). Yeast strains transformed with plasmids carrying the URA3 selectable marker were grown in synthetic medium (1.65% yeast

* Corresponding author.
† Present address: Division of Cell and Molecular Biology, Lawrence Berkeley Laboratory, Berkeley, CA 94720.
nitrogen base, 5% ammonium sulfate, 2% glucose) containing the appropriate amino acids but lacking uracil to maintain selection for plasmid-bearing cells.

*E. coli* HB101 and TG1 cells harboring plasmid pUC119, various derivatives of this plasmid, or plasmid pR169 were grown in L broth containing ampicillin (50 μg/ml). *E. coli* TG1 cells transformed with recombinant pUC119 plasmids were plated on LB plates containing ampicillin (50 μg/ml), isopropylthio-β-D-galactoside (IPTG) (100 μM), and X-Gal (0.005%). White (lac-) (16) colonies were picked, and plasmid DNA was analyzed by digestion with restriction enzymes. *E. coli* strains transformed with plasmid pNF411, pNF417, or pNF422CEN-GR-IN, were grown in L broth containing tetracycline (12.5 μg/ml).

**Cloning and sequencing of DNA fragments.** Plasmid DNA was purified by the alkaline lysis procedure of Maniatis et al. (15) and was prepared for sequencing by the method of Hattori and Sakaki (12). The dideoxynucleotide chain-termination method of Sanger et al. (27) used denatured plasmid DNA templates (12) and either the Klenow fragment of DNA polymerase I plus reagents from Pharmacia or the Sequenase DNA-sequencing kit (U.S. Biochemical Corp.). Two independent clones of each DNA fragment were isolated, and in all cases both strands were sequenced.

In order to sequence the mutant *rad4* allele in plasmid pNF411, a 3.3-kb fragment from the gene was cloned into pUC19. Restriction fragments 150 to 400 base pairs (bp) in size were isolated from agarose gels and subcloned. The strategy for sequencing this allele is shown in Fig. 1A. A *PvuI-EcoRV* fragment from plasmid pNF422CEN-GR-IN, (Fig. 1B) was subcloned and sequenced as described above. Five fragments from plasmid pR169, which include most of the region between the *SphI* and *MstII* sites (Fig. 1), were subcloned and sequenced as described above.

The yeast chromosomal mutations designated *rad4-2*, *rad4-3*, and *rad4-4* were rescued into a centromeric plasmid by the gap repair technique of Orr-Weaver and Szostak (19). Plasmid pNF417-Xba was constructed by deleting a 1-kb *XbaI* fragment, including the putative *RAD4* promoter and the first 500 bp of the *rad4* coding region, from the centromeric plasmid pNF417. Plasmid pNF417-Xba was digested with *XcoI* (*Bsu36I*), *MstII* and *BglII* to generate an 850-bp gap. The gapped plasmid was used to transform *rad4-2*, *rad4-3*, and *rad4-4* strains. Plasmid DNA was isolated from *Ura+* yeast transformants as described previously (29) and subsequently propagated in *E. coli*. Gap-repaired plasmids were preliminarily identified by restriction enzyme analysis. Definitive verification of gap repair was based on the demonstration by DNA sequencing that the plasmids had rescued the wild-type sequence in the *XcmI-PvuI* interval. Appropriate fragments were then sequenced to identify the *rad4-3* and *rad4-4* mutations.

Plasmid pGR4-3 is identical to pNF417-Xba, except that it contains the *rad4-3* chromosomal mutation rather than the mutation generated in *E. coli* on plasmid pNF402GR-EC1. Plasmid pGR4-3 was digested with *Bsr36I* and *BglII*, resulting in an 850-bp gap, which includes the *rad4-3* mutation. This gapped plasmid was used to transform a *rad4-2* strain, and the *rad4-2* mutation was characterized as described above.

The computer program of Queen and Korn (22) was used to analyze the *RAD4* DNA sequence. Nucleotide and amino acid homology searches were carried out through BIONET.

**Enzymes and reagents.** Restriction enzymes and DNA-modifying enzymes (including T4 DNA ligase, T4 DNA polymerase, and the Klenow fragment of *E. coli* DNA polymerase I) were purchased from Bethesda Research Laboratories, New England BioLabs, or U.S. Biochemical Corp. Calf intestinal phosphatase was purchased from Boehringer Mannheim. Reagents for DNA sequencing were purchased from Pharmacia. The Sequenase DNA-sequencing kit was purchased from U.S. Biochemical Corp. [35S]dATP (1,000 Ci/mmol) was obtained from Amersham. IPTG was purchased from Bethesda Research Laboratories, and X-Gal was purchased from International Biotechnologies, Inc.

**RESULTS**

**Sequence of the wild-type *RAD4* gene.** The strategy used to determine the sequence of the wild-type *RAD4* gene is diagrammatically illustrated in Fig. 1. We first determined the sequence of an inactivated mutant *rad4* allele carried on the centromeric plasmid pNF402GR-EC1 (Fig. 1A). This
plasmid was originally isolated following transformation of *E. coli* HB101 with a plasmid (pNF402GR) carrying the wild-type *RAD4* gene (7). Plasmid pNF402GR does not express wild-type *RAD4* activity in yeast cells but can be propagated in *E. coli* at normal levels (7). The inactivating mutation(s) in this plasmid-borne allele was previously mapped by gap repair to a 100-bp *PvuI*-XmnI fragment (7) (Fig. 1A).

Plasmid pNF422CEN-GR-IN_BglII contains the wild type *RAD4* gene with a 415-bp DNA fragment inserted at the *BglII* site (Fig. 1B). This fragment is out of frame with the *RAD4* translational codon and contains multiple stop codons (8). The insertional mutation permits propagation of the plasmid in *E. coli* (8). This plasmid was used to sequence the wild-type *PvuI*-EcoRV fragment (Fig. 1B) that includes the corresponding mutated *PvuI*-XmnI region in plasmid pNF402GR-EC1 (Fig. 1A).

Sequencing 2,796 bp of DNA revealed a 2,262-bp open reading frame (ORF) flanked by 386 bp of 5′ noncoding sequence and 148 bp of 3′ noncoding sequence (Fig. 2). The sequence TCAG, starting at nucleotide position −49, conforms to one of several known consensus sequences [TC(G/A)A] for initiation of transcription in *S. cerevisiae* (11) and corresponds to the previously mapped position of the *RAD4* transcriptional start site (7). The sequence TATAATA, located at nucleotide position −127, is −78 bp upstream of the major *RAD4* transcriptional start site (7) and is a candidate for an *S. cerevisiae* TATA box (consensus sequence TATAAA [28]).

The ATG codon that begins the *RAD4* ORF is ~50 bp downstream of the major transcriptional start site. In eucaryotes, the translational start codon is typically the most 5′ AUG triplet of the message (14). The consensus sequence for eucaryotic translational initiation sites is CCA/GCAA/UUG (where A* = position +1) (14). The purines at positions −3 and +4 are highly conserved, and there is a predominance of C at positions −1, −2, −4, and −5. However, few mRNAs contain all 4 Cs (14). The *RAD4* mRNA sequence (CTAAAAUGA) deduced from the DNA sequence is a good match with the consensus sequence at the translational start site of eucaryotic mRNAs. The sequence context of the next in-frame ATG (at nucleotide position 433 [Fig. 2]) contains pyrimidines at positions −3 and +4. The −3 position of internal AUG codons in eucaryotic mRNAs has a random nucleotide distribution (14). Hence, it is likely that the ATG codon that marks the beginning of the *RAD4* ORF is the translational start codon.

In the 3′ noncoding region, the sequence AATAAA is present beginning at nucleotide position 2309, 44 bp downstream of the TGA stop codon. It has been proposed that this sequence is important for polyadenylation of transcripts in higher eucaryotes (6). Additionally, the tripartate sequence TAG...TATG...TTT is located 27 to 43 bp downstream of the *RAD4* TGA stop codon (Fig. 2). It has been suggested that this sequence also has a role in polyadenylation and, in some yeast genes, in transcriptional termination (37). Consensus yeast mRNA splicing signals were not detected in the 2,796 bp sequences, suggesting that the *RAD4* transcript is not spliced.

Like most yeast genes, the *RAD4* coding region is very rich in AT (60.2%). The predicted amino acid sequence of the *Rad4* polypeptide is shown in Fig. 2. The polypeptide consists of 754 amino acids and is expected to have a molecular mass of 87.1 kilodaltons. The calculated net charge of the protein at pH 7 is +4; it is thus the most basic of the five Rad proteins (Rad1, Rad2, Rad3, Rad4, and Rad10) required for damage-specific incision during nucleotide excision repair (9). However, the carboxy-terminal end of the putative Rad4 polypeptide is distinctly acidic. Of the last 22 amino acids, 11 are Glu or Asp, and of the last 44 amino acids, 19 are Glu or Asp. A hydrophilicity plot indicates the presence of predominantly hydrophilic domains in the *Rad4* polypeptide, particularly at the amino and carboxy termini (Fig. 3).

The *RAD4* gene has no significant codon bias. The calculated codon bias index is 0.41, a value typical of weakly expressed yeast genes (1). A computer search of DNA and protein data bases showed no extensive regions of similarity with other genes or proteins. However, a 25-amino-acid region of the *Rad4* protein is 40% identical and 60% similar to a region of the *S. cerevisiae* DNA photolyase, a product of the *PHR1* gene (26, 36) (Fig. 4). Both of these proteins are involved in the repair of DNA containing pyrimidine dimers, albeit by different mechanisms. Thus, it is possible that the region of amino acid sequence similarity is important for recognition of damage to DNA.

Mapping and characterization of plasmid-borne mutant alleles. We determined the sequences of two *rad4* mutations spontaneously generated during propagation in *E. coli* of plasmids carrying the wild-type gene. One of these mutations is carried on plasmid pNF402GR-EC1 (7). As indicated above, it was previously established that all inactivating mutations in this allele mapped to a 100-bp *PvuI*-XmnI fragment (Fig. 1A). Sequencing of this region revealed the loss of a single G·C base pair at nucleotide position 1169, resulting in a −1 frameshift mutation (Fig. 5 and Table 1). This mutation terminates the ORF at nucleotide position 1213 and is expected to result in the synthesis of a truncated polypeptide 404 amino acids long (Table 1).

A second spontaneously generated mutant allele is contained on plasmid pR169, originally isolated by Roeder et al. (24). These investigators sequenced a 3,351-bp region of the insert in this plasmid and identified the *spt2*-1 ORF as well as one small complete ORF and an incomplete ORF located upstream of and in the same orientation as the *spt2*-1 gene. They reported the sequence of 2,025 bp (24), much of which represents the *spt2*-1 gene and sequences downstream of this gene. The sequence of most of the 1,326 bp located upstream of the *spt2*-1 gene has not been published but was generously provided to us by Michael Smith, University of British Columbia, Vancouver (personal communication).

A comparison between our wild-type *RAD4* sequence and the sequence provided by Michael Smith revealed two differences. We therefore cloned and sequenced 1,477 bp of the *rad4* allele present on plasmid pR169. This sequence included the regions in which differences were noted as well as the portion of the mutant *rad4* ORF not previously sequenced. Our analysis indicates that the two unidentified ORFs identified by Roeder et al. (24) actually represent part of a single ORF from a mutant *rad4* gene. The sequences of this allele and of the wild-type *RAD4* gene differ at a single nucleotide. The mutant allele contains a −1 frameshift mutation resulting from deletion of an A·T base pair from a string of four such base pairs located between nucleotides 1505 and 1508 (Fig. 5 and Table 1). This frameshift terminates the *RAD4* ORF at nucleotide position 1507 and is expected to result in a truncated polypeptide 502 amino acids long (Table 1).

Collectively, these results demonstrate that two independently derived plasmid-borne *RAD4* alleles carry single −1 frameshift mutations located in the middle one-third of the *RAD4* coding region. These mutations are apparently both
**FIG. 2.** Nucleotide sequence of a 2,796-bp region of the *S. cerevisiae* genome that includes the RAD4 ORF (1 to 2262) and 5' and 3' noncoding regions. The predicted amino acid sequence of the Rad4 polypeptide is also shown.
FIG. 3. Hydrophilicity plot of the predicted Rad4 polypeptide. The plot was generated with the DNA Inspector II* program of Textco (West Lebanon, N.H.) by using an averaging length of 20 amino acids.

necessary and sufficient to propagate plasmids carrying the wild-type RAD4 gene in E. coli. As indicated below, the mutation in plasmid pNF402GR-EC1 is identical to the rad4-2 mutation. The mutation present in plasmid pR169 is a new mutant allele, designated rad4-12 (Table 1).

Mapping and characterization of chromosomal rad4 mutant alleles. The yeast chromosomal rad4-2, rad4-3, and rad4-4 mutations were previously mapped by gap repair to a region of the RAD4 ORF spanned by the MstII and BglII sites (Fig. 1A) (7). We rescued these mutations on centromeric plasmids by the gap repair technique of Orr-Weaver and Szostak (19). A priori, we could not be certain that the rescued mutations would necessarily preclude the generation of further mutations during propagation of plasmids in E. coli. In order to obviate this potential problem, we severely crippled the RAD4 gene by deleting 1 kb of DNA that includes 500 bp at the 5' end of the RAD4 coding region (Fig. 6) and 500 bp of upstream sequence that presumably includes part or all of the RAD4 promoter. A plasmid carrying this crippled gene was then gapped in a region that includes the mutation in plasmid pNF402GR-EC1 described above (Fig. 6). Rescue of the chromosomal rad4-3 or rad4-4 allele by gap repair is expected to rescue also the wild-type sequence in this region of the RAD4 gene (Fig. 6). This was determined by screening the restriction patterns of rescued plasmids and was verified by sequencing the appropriate DNA fragments. Gap-repaired plasmids were then subcloned and sequenced to determine the location and nature of the rad4-3 and rad4-4 chromosomal mutations.

The rad4-3 allele was shown to contain an ochre mutation resulting from a change in the coding sequence, CAA, at nucleotide position 1141, to TAA (Table 1). This result directly confirmed previous evidence of an ochre mutation in the rad4-3 strain based on the use of translational suppressors (20). This mutation is located just 28 bp upstream of the mutation present in plasmid pNF402GR-EC1 (Fig. 5). The ochre mutation is expected to result in a truncated polypeptide of 380 amino acids (Table 1).

The rad4-4 allele is missing a single A·T base pair in a string of 5 A·T bp situated between nucleotides 1509 and 1513 (Fig. 5). This deletion resulted in a frame-shift mutation, which terminates the RAD4 ORF at nucleotide position 1534. This sequence is immediately 3' to the T·A stretch that marks the location of the E. coli-derived mutation in plasmid pR169 (Fig. 5). The rad4-4 mutation is expected to result in the translation of a truncated polypeptide of 511 amino acids (Table 1).

Plasmids isolated from five separate rad4-2 transformants were shown by DNA sequencing to be identical to plasmid pNF402GR-EC1. One interpretation of this result is that gap repair failed following transformation of rad4-2 cells, a surprising result in view of the observation that two gapped plasmids isolated from the rad4-3 and rad4-4 strains were both repaired. An alternative explanation is that the rad4-2 mutation and the one carried on plasmid pNF402GR-EC1 are in fact identical. In order to prove this, a plasmid carrying the rad4-3 mutation rather than the mutation in pNF402GR-EC1 was used to rescue the rad4-2 mutation by gap repair (Fig. 6). Gap repair was demonstrated by recovery of the wild-type sequence at the position of the rad4-3 mutation (Fig. 6). Sequencing of appropriate fragments from repaired plasmids identified the rad4-2 mutation, which is indeed identical to that in the plasmid-borne pNF402GR-EC1 allele, i.e., a -1 frameshift mutation resulting from the

FIG. 4. Amino acid sequence similarity in regions of the predicted Rad4 (amino acids 217 to 241) and Phr1 (amino acids 515 to 539) polypeptides. The boxed regions are regions of identical amino acid sequences.

FIG. 5. Partial sequence of the RAD4 gene between nucleotides 1135 and 1518 showing the locations of two plasmid-borne mutations generated spontaneously in E. coli, and of three mutagen-induced chromosomal mutations. ▲, -1 frameshift mutations carried on plasmids pNF402GR-EC1 (nucleotide 1169) and pR169 (between nucleotides 1505 and 1508). The rad4-3 ochre mutation (position 1141) is a G·C to A·T transition. △, rad4-2 (nucleotide 1169) and rad4-4 (between nucleotides 1509 and 1513) -1 frameshift mutations. The brackets indicate regions of sequence redundancy from which a single base pair was deleted.
loss of a single G·C base pair at position 1169 (Fig. 5 and Table 1).

**DISCUSSION**

Unlike the *RAD1, RAD2, RAD3*, and *RAD10* genes, yeast shuttle plasmids carrying the *RAD4* gene of *S. cerevisiae* under the control of its endogenous promoter cannot be propagated in *E. coli* (7). Following transformation of various *E. coli* strains, transformation frequency drops drastically and many of the plasmids recovered are deleted and/or rearranged (8). A minority of the recovered plasmids have a normal restriction pattern but carry inactivated *rad4* alleles, since they do not complement the UV sensitivity of *rad4* mutant strains when present at low copy number (7, 8). Such plasmids can be serially propagated in *E. coli* with normal transformation frequencies (8).

In order to sequence the wild-type *RAD4* gene, we first determined the sequence of an inactivated *rad4* allele carried on a plasmid designated pNF402GR-EC1, which can be propagated in *E. coli*. Previous studies have shown that the inactivating mutation(s) in this allele maps to a 100-bp *XmnI-PvuI* interval in the physically mapped gene, since gap repair of this region alone restores normal *RAD4* activity (7). To determine the wild-type sequence of this 100-bp region, we used a strategy that we refer to as reversible mutational inactivation. This involves disruption of the wild-type *RAD4* gene (isolated directly from yeast cells) by insertion of a DNA fragment at a defined location. Previous studies have shown that shuttle plasmids carrying such insertional mutations can be propagated at normal levels in *E. coli* (8).

Following removal of the inserted fragments, the plasmids regain wild-type *RAD4* functions (8), indicating that no other inactivating mutations are generated during propagation of the plasmid carrying the insertional mutation. In the present study, we cloned and sequenced the 100-bp *XmnI-PvuI* region from a plasmid carrying such an insertional mutation. Comparison of this sequence with that from the inactivated *rad4* allele on pNF402GR-EC1 demonstrated the presence of a single point mutation in the latter (see later discussion).

The observation that plasmids carrying the wild-type *RAD4* gene are lethal to *E. coli* and that this lethality can be eliminated by mutations that also inactivate its function in *S. cerevisiae* suggests that such mutations affect regions of the gene that are relevant both to its toxicity in *E. coli* and to its function in nucleotide excision repair. We identified the location and nature of such spontaneously generated mutations in two independently derived plasmids that were propagated in *E. coli* as part of gene cloning strategies. Both plasmids contain −1 frameshift mutations. One of these mutations resulted from the deletion of a G·C base pair situated in a string of seven otherwise uninterrupted A·T base pairs. The other frameshift mutation resulted from deletion of a single T·A base pair in a string of four such base pairs. Two of the three yeast chromosomal mutations are also frameshift mutations resulting from the deletion of a single base pair in a short redundant sequence. Unlike the *rad4* mutations spontaneously generated in *E. coli*, these mutations were isolated following treatment of yeast cells with ethyl methanesulfonate (*rad4*2) (10, 30) or UV radiation (*rad4*4) (2, 10).

A model involving nucleotide slippage and mispairing during replication of redundant sequences has been proposed by Streisinger et al. to account for the generation of frameshift mutations (32). In support of this model, these investigators demonstrated a hot spot for spontaneous mutations in the lysozyme gene of bacteriophage T4 at a region of six consecutive A·T base pairs, where additions or deletions of a single A·T base pair occur at an abnormally high frequency (18). Other examples of spontaneous mutations and examples of mutagen-induced frameshift mutations that are consistent with the “slipped mispairing” model have been documented in *E. coli* (5, 17), yeast (3, 31), and mammalian cells (4, 25). This model could account for the loss of a single A·T base pair in the redundant sequences TTGT and AAAA, resulting in the *rad4*4-12 (plasmid pR169) and *rad4*4-4 alleles, respectively. Mechanistically, it seems feasible that this model could also explain the loss of a single G·C base pair in the sequence AAAAAAGAAA, accounting for the identical *rad4*2 mutation in plasmid pNF402GR-EC1 and in the *rad4*2 strain.

The *rad4*3 allele contains the only base substitution mutation among the five mutations characterized in this study. The *rad4*3 strain was selected for UV sensitivity following ethyl methanesulfonate mutagenesis of a wild-type strain (10, 30). The demonstration of an ochre mutation in the *rad4*3 allele resulting from a GC- to AT transition is consistent with the observation that in yeast cells ethyl methanesulfonate results predominantly in GC- to AT transition mutations (21).

It is remarkable that one of the two plasmid-borne mutations generated in *E. coli* is identical to one of the three yeast chromosomal mutations, and that the other plasmid-borne mutation is located in very close physical proximity to a second yeast chromosomal mutation. All five mutations studied map to a 372-bp region in the middle one-third of the *RAD4* gene. All five mutations are chain terminating and are expected to encode truncated polypeptides between 380 and 511 amino acids long, missing the carboxy-terminal one-third of the wild-type protein. The *rad4*2, *rad4*3, and *rad4*4 mutant strains were selected for UV sensitivity and are

### Table 1. Summary of *RAD4* Alleles

<table>
<thead>
<tr>
<th>Allele</th>
<th>Nature of mutation*</th>
<th>Site of mutation (nucleotide position)</th>
<th>Predicted size of polypeptide (no. of amino acids)</th>
<th>Source of mutation*</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>RAD4</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>rad4</em>-2</td>
<td>−1 frameshift mutation (loss of G·C bp)</td>
<td>1169</td>
<td>754</td>
<td>Spontaneous in <em>E. coli</em></td>
</tr>
<tr>
<td><em>rad4</em>-12</td>
<td>−1 frameshift mutation (loss of A·T bp)</td>
<td>1505–1508</td>
<td>502</td>
<td>Spontaneous in <em>E. coli</em></td>
</tr>
<tr>
<td><em>rad4</em>-2</td>
<td>−1 frameshift mutation (loss of G·C bp)</td>
<td>1169</td>
<td>404</td>
<td>EMS</td>
</tr>
<tr>
<td><em>rad4</em>-3</td>
<td>ochre mutation (CAA → TAA)</td>
<td>1141</td>
<td>380</td>
<td>EMS</td>
</tr>
<tr>
<td><em>rad4</em>-4</td>
<td>−1 frameshift mutation (loss of A·T bp)</td>
<td>1509–1513</td>
<td>511</td>
<td>UV</td>
</tr>
</tbody>
</table>

*a* bp, Base pair.

*EMS, Ethyl methanesulfonate.*
defective in damage-specific incision of DNA, whereas the mutant alleles carried on the plasmids pNF402GR-EC1 and plasmid pR169 were recovered on the basis of their ability to be propagated in E. coli. These observations suggest that the carboxy-terminal one-third of the Rad4 protein may be necessary both for normal nucleotide excision repair and for lethality in E. coli. Interestingly, this region of the polypeptide is distinctly acidic, with a net charge of -21. Indeed, 11 of the last 22 amino acids (50%) and 19 of the last 44 amino acids (43%) are Glu or Asp. The acidic carboxy terminus of the Rad4 protein may be required for interactions with basic proteins during nucleotide excision repair in S. cerevisiae. In E. coli, interaction of the Rad4 protein with basic proteins may interfere with metabolic events required for the viability of cells.

The specific role of the Rad4 protein in nucleotide excision repair in S. cerevisiae awaits the purification and characterization of this protein from yeast cells in which the cloned gene is overexpressed. Such studies are in progress. The mutant alleles defined in this study provide opportunities for informative structure-function correlations once the biochemical function(s) of the wild-type Rad4 protein is established.

ACKNOWLEDGMENTS

These studies were supported by Public Health Service grant CA-12428 from the National Institutes of Health. L.C. was the recipient of Public Health Service grant NSRA CA-07705 from the National Institutes of Health.

We thank Michael Smith for providing us with unpublished sequence information and Shirleen Roeder for plasmid pR169. We also thank Wolfram Siede and Reinhard Fleer for their contributions to this study and our other laboratory colleagues for helpful discussions and for critical reading of the manuscript.

LITERATURE CITED

14. Kozak, M. 1984. Compilation and analysis of sequences up-

FIG. 6. Diagrammatic representation of the strategy used for rescue of the rad4-3 and rad4-4 mutations (A) and of the rad4-2 mutation (B) by repair of specifically gapped plasmids from homologous chromosomal sequences. The mutant rad4-2, rad4-3, and rad4-4 strains were transformed with a centromeric yeast plasmid containing a gap as shown. The gap included a "diagnostic" rad4 mutant allele (either that present in the plasmid pNF402GR-EC1 [designated here as the EC1 allele] or that present in rad4-3) as well as the mutation to be rescued from the chromosome. Gap repair (and hence rescue of the allele of interest) was established by demonstrating the wild-type sequence at the position of the "diagnostic" mutation.