A Highly Conserved Repeated Chromosomal Sequence in the Radioresistant Bacterium *Deinococcus radiodurans* SARK

EILEEN LENNON,† PABLO D. GUTMAN, HANLONG YAO, AND KENNETH W. MINTON*

Department of Pathology, F. E. Hébert School of Medicine, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, Maryland 20814-4799

Received 10 September 1990/Accepted 5 January 1991

A DNA fragment containing a portion of a DNA damage-inducible gene from *Deinococcus radiodurans* SARK hybridized to numerous fragments of SARK genomic DNA because of a highly conserved repetitive chromosomal element. The element is of variable length, ranging from 150 to 192 bp, depending on the absence or presence of one or two 21-bp sequences located internally. A putative translational start site of the damage-inducible gene is within the reiterated element. The element contains dyad symmetries that suggest modes of transcriptional and/or translational control.

*Deinococcus radiodurans* is the type species of the family *Deinococcaceae*, a small eubacterial family whose members share in common extreme resistance to the lethal and mutagenic effects of ionizing and UV radiation and many genotoxic chemicals. Although repair of DNA damage by these organisms is known to be unusually efficient, the molecular mechanisms responsible remain largely unidentified (6). We recently reported a method that uses transformation of *D. radiodurans* SARK to generate random chromosomal gene fusions with a 5′-truncated lacZ (lac′Z) gene (5). Several of these gene fusions were inducible following exposure of *D. radiodurans* to the DNA-damaging agent mitomycin C, and these gene fusions were cloned (5).

**SARK sequence in pEL21.** A SARK DNA fragment from one cloned mitomycin C-inducible gene fusion (pEL21) hybridized to numerous fragments in Southern blots of SARK genomic DNA, suggesting the presence of a reiterated sequence (Fig. 1). The SARK DNA fragment in pEL21 is 419 bp (Fig. 2). There are two translational terminators, both TGA, in the SARK fragment in frame with lac′Z (Fig. 2). Downstream from these terminators, and in frame with lac′Z, are seven translational start codons: two ATGs, two GTGs, and three TTGs. One ATG (bases 255 to 257 [Fig. 2]) appears most likely to be the translation start site by virtue of its proximity to Shine-Dalgarno (S-D) sequences. The deduced protein-coding sequence for this possible start site encodes 55 amino acids prior to fusion with lac′Z, and both this start site and part of the coding sequence are within the reiterated element (Fig. 2). There is a 31-nucleotide hairpin with almost perfect dyad symmetry, 62 nucleotides upstream from this ATG, which is also within the reiterated element. The ATG itself also falls within a region compatible with formation of an mRNA hairpin with the ATG located on the upstream portion of the stem (Fig. 2).

**Sequence responsible for the multiple hybridization pattern is highly conserved.** When SARK genomic DNA fully digested with *Eco*RI was used to construct a library in the *Eco*RI site of pUC18, about 3% of all *Escherichia coli* colonies hybridized with the *D. radiodurans* insert in pSRE1 (pSRE1 is a subclone of pEL21 [Table 1]; SRE stands for SARK repeated element). Four of the isolates detected by colony hybridization that contained SARK DNA fragments of different sizes, pSRE4, pSRE5, pSRE11, and pSRE29 (Table 1), were selected for sequence analysis of the reiterated element. Sequencing was by the dideoxy method using 35S-dATP. When needed, dITP was used to resolve band compressions. pSRE1 and pSRE11 were sequenced by using forward and reverse primers for pUC18. Once these sequences were determined, two custom oligonucleotides were made to determine the sequences at the ends of the mapped SARK element (Fig. 3). These oligonucleotides were 5′-GGGCACTTGGATTCTTCT-3′ (oligonucleotide 1) and 5′-GGACACCGATTGCTATC-3′ (oligonucleotide 2) and are shown in Fig. 3. Partial sequence analysis of these oligonucleotides revealed that the SARK sequence contains a dyad symmetry and that the target oligonucleotides are located within a variable region that is homologous to the oligonucleotide sequence.

![FIG. 1. Southern blot of *D. radiodurans* DNA probed with insert of pSRE1. *D. radiodurans* SARK genomic DNA was cleaved with *Eco*RI (lane 1), *Hind*III (lane 2), *Stu*I (lane 3), and *Sma*I (lane 4). Each lane contains 1 μg of DNA. The probe was the 360-bp *Bgl*II-*Pvu*II SARK DNA insert from pSRE1 (a subclone of pEL21 [Table 1]). The fragment was radiolabeled with [32P]dCTP by nick translation. Hybridization was at 107 cpm/ml for 18 h in 50% formamide at 37°C. The final washing was at high stringency (65°C in 0.1× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate] for 1 h). *Hind*III-cleavable sites in *D. radiodurans* DNA are rare, and the majority of *Hind*III-cleaved DNA migrates uniformly at the upper limit of resolution (4, 10). The pattern of hybridization to the *Hind*III digest reflects this bias.](http://jb.asm.org/)
FIG. 2. Map of pEL21, DNA sequence of its D. radiodurans portion, and likely RNA secondary structures. The pRF1 fusion vector contains the kanamycin resistance determinant aphA and the 5′-truncated lacZ gene deleted of its promoter and first seven codons (lac′Z, ). A unique BgII site in pRF1 is present adjacent to the eighth codon (5,7). Inserted into the BgII site is the D. radiodurans fragment that contains the mitomycin C-inducible gene translationally fused to lac′Z (5). Arrows indicate direction of transcription. Restriction sites are as follows: B, BgII; C, ClaI; E, EcoRI; H, HindIII; P, PvuII; Ps, PstI; RV, EcoRV; S, Sall; Sa, Sau3AI; BglII fusion; Sm, Smal; Sa, SstI; and St, Stul. The DNA sequence extends from the BgII site at the 5′ end of the D. radiodurans fragment to just beyond the BgII-Sau3AI site at the eighth codon of lacZ. The BgII, Smal, and PvuII sites in the sequence are underlined. At the 3′ end of the sequence, lacZ bases and corresponding amino acids are indicated in italics. Upstream transcriptional terminators that are in frame with lacZ are underlined and labeled End. The two ATG codons that are in frame with lacZ are underlined. S-D sequences are underlined, using the criterion (3) of at least four consecutive nucleotides, including three G’s, from the GGAGG consensus (8). The deduced amino acid sequence begins with the second ATG. The two brackets in the DNA sequence indicate the ends of the repeated sequence detailed in Fig. 3. The two dyad symmetries in the DNA sequence, one near the 5′ end and the other near the 3′ end of the reiterated element, are indicated by opposing arrows above the relevant bases. Hypothetical RNA structures corresponding to the reiterated DNA element are shown with dyad symmetries in stem-loop conformation. Free energies of formation were calculated essentially as previously described (12).

TABLE 1. Bacteria and plasmids

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Description (reference or source)</th>
</tr>
</thead>
</table>
| Strains
| Deinococcus radiodurans SARK* | Wild type (2); UWO 298* |
| Escherichia coli DH5αd | recA1 endA1 hsdR17 (rK- mK+) F- φ80lacZ ΔM15 (lacZYA-argF)U169 (Bethesda Research Laboratories) |
| Plasmids
| pEL21 | 419-bp fragment of SARK chromosomal DNA ligated to a 5′-truncated lacZ gene in the E. coli lacZ translational fusion vector pRF1, Km' (5) |
| pUC18 | 2,686-bp pBR322 derivative, Ap', multiple cloning site in proximal lacZ protein-coding sequence (14) |
| pSRE1 | 360-bp subclone of SARK DNA that contains reiterated sequence from pEL21; pUC18 BamHI-Smal-A:pEL21 BglII-PvuII-D |
| pSRE4 | pUC18 EcoRI::4-kb EcoRI SARK fragment |
| pSRE5 | pUC18 EcoRI::1.2-kb EcoRI SARK fragment |
| pSRE11 | pUC18 EcoRI::2.0-kb EcoRI SARK fragment |
| pSRE29 | pUC18 EcoRI::1.6-kb EcoRI SARK fragment |
| pS19 | Full-length clone of natural D. radiodurans SARK plasmid pEL21 (45 kb) in the EcoRV site of Km' ColEl derivative pS27 (9) |
| pS28 | Full-length clone of natural D. radiodurans SARK plasmid pUE10 (37 kb) in the EcoRV site of pS27 (9) |

* Plasmids for which a reference is not noted were constructed during this study.

D. radiodurans was grown at 32°C in TGY broth or on TGY plates solidified with 1.5% agar (10).

University of Western Ontario culture collection, London, Ontario, Canada.

E. coli was grown in LB broth or LB agar at 37°C. Selection of E. coli strains transformed with pUC18 was on LB agar supplemented with 50 μg of ampicillin per ml. Selection of E. coli containing pEL21 was on agar containing 20 μg of kanamycin per ml.

were employed to obtain the sequence and flanking sequences of the repetitive elements in the other clones. One custom oligonucleotide, 5′-AGAGCGATTTCGCAAAT-3′, is complementary to the sequence of the repeated element from nucleotides 175 to 191 (as numbered in Fig. 3) and was used to sequence the upper strand of the repeated element; the other custom primer, 5′-GAGCATTTCGACAAAG-3′, is the sequence of the repeated element from nucleotides 4 to 20 (as numbered in Fig. 3) and was used to sequence the lower strand. In all cases, both upper and lower strands were sequenced. The sizes of the five repeated elements are 151 nucleotides for SRE1, 171 nucleotides for SRE4 and SRE29, and 192 nucleotides for SRE11 and SRE5 (Fig. 3). The difference in the sizes is accounted for by the presence or absence of one or two 21-nucleotide sequences inserted after nucleotide 90 and the presence of an additional base at nucleotide 157 in SRE1. The two 21-bp inserts are direct repeats. Within the entire reiterated sequence, including the 21-bp inserts, there are only eight nucleotides that show point variations (Fig. 3). The repeated elements have a mean GC content of 56%, which is lower than the 67 to 70% GC content characteristic of D. radiodurans SARK and all other members of the family Deinococcaceae (2). All versions of the repeat contain dyad symmetry (DS-A) at the 5′ end, as shown in the pEL21 (SRE1) gene fusion in Fig. 2; however,
there is a single base change from C to A within DS-A in the other four sequences (Fig. 3).

Hypothetical RNA secondary structure corresponding to the 3' end of the element depends on the number of 21-bp inserts, with the potential formation of either one stem-loop structure or two closely adjacent stem-loop structures (Fig. 3). We speculate that the location of S-D sequences in and adjacent to this region of potential secondary structure may have regulatory significance for translation. Modifications in this portion of the repeated element in other copies compared with SRE1 include elimination of several canonical S-D sequences and an additional potential hairpin structure (SRE4 and SRE29). In SRE5 and SRE11 there is the possible formation of mutually exclusive hairpin structures (DS-B versus DS-C) (Fig. 3). All five of the sequenced elements contain DS-A and DS-D (Fig. 2 and 3). We suggest that DS-A may be a transcriptional terminator because of its nearly perfect dyad symmetry of 26 nucleotides (13 bp; one G-T base pair) in pEL21 (SRE1), a feature common in transcriptional terminators (1).

Possible functions of reiterated element. Since the reiterated element was detected as part of a mitomycin C-inducible gene fusion, it may be a highly conserved consensus sequence for a DNA damage-inducible regulon in D. radio-
durans. If so, our estimate of ≥30 copies would suggest a large number of genes within the regulon, comparable with the number of genes in the SOS regulon of E. coli (13). Using appropriate oligonucleotide primers for this element and polymerase chain reaction conditions sufficient to amplify sequences up to 1.5 kb, we did not find any polymerase chain reaction-amplifiable sequences larger than the sizes of the elements described here (not shown), suggesting that this element is not part of a transposon. Strain SARK contains two cryptic plasmids, pUE11 and pUE10 (9). On the basis of the strong hybridization of pUE11 against SARK chromosomal DNA, we have suggested that pUE11 is an episome (11). Neither pUE11 nor pUE10 hybridizes with the pSRE1 insert (Fig. 4); consequently, homology between pUE11 and the chromosome cannot be ascribed to the reiterated element. Deinococcus spp. are unique in their ability to mend several hundred double-strand breaks per chromosome without any loss of viability (6). In this regard, a novel hypothesis for the occurrence of these conserved repeated sequences is that they function as specific substrate sites for recombinational repair of DNA. Frequent exchange of sequence information by recombination at these sites may serve to retain the repeated elements' high degree of sequence conservation.
The GenBank accession number for the SARK fragment in pEL21 is M59306. The GenBank accession numbers for SRE1, SRE4, SRE5, SRE11, and SRE29 are M59306, M59307, M59308, M59309, and M59310, respectively.

This work was supported by Public Health Service grant GM39933.

REFERENCES