

## Human *ERCC5* cDNA-Cosmid Complementation for Excision Repair and Bipartite Amino Acid Domains Conserved with RAD Proteins of *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*

MARK A. MACINNES,<sup>1\*</sup> JUDITH A. DICKSON,<sup>1</sup> RUDY R. HERNANDEZ,<sup>1</sup> DIANNE LEARMONTH,<sup>2</sup>  
GRACE Y. LIN,<sup>1</sup> JOHN S. MUDGETT,<sup>3</sup> MIN S. PARK,<sup>1</sup> SUSAN SCHAUER,<sup>1</sup>  
RICHARD J. REYNOLDS,<sup>1</sup> GARY F. STRNISTE,<sup>1</sup> AND JOYCE Y. YU<sup>4</sup>

*Life Sciences Division, M888, Los Alamos National Laboratory, Los Alamos, New Mexico 87545<sup>1</sup>;*  
*West Nistaben, Stenness, Orkney, Scotland KW16 3HE, United Kingdom<sup>2</sup>;* *Department of*  
*Molecular Immunology, Merck Research Laboratories, Rahway, New Jersey 07065<sup>3</sup>;*  
*and Department of Biology, University of Texas, Austin, Texas 78731<sup>4</sup>*

Received 20 April 1993/Returned for modification 8 June 1993/Accepted 5 July 1993

Several human genes related to DNA excision repair (ER) have been isolated via ER cross-species complementation (ERCC) of UV-sensitive CHO cells. We have now isolated and characterized cDNAs for the human *ERCC5* gene that complement CHO UV135 cells. The *ERCC5* mRNA size is about 4.6 kb. Our available cDNA clones are partial length, and no single clone was active for UV135 complementation. When cDNAs were mixed pairwise with a cosmid clone containing an overlapping 5'-end segment of the *ERCC5* gene, DNA transfer produced UV-resistant colonies with 60 to 95% correction of UV resistance relative to either a genomic *ERCC5* DNA transformant or the CHO AA8 progenitor cells. cDNA-cosmid transformants regained intermediate levels (20 to 45%) of ER-dependent reactivation of a UV-damaged pSVCATgpt reporter plasmid. Our evidence strongly implicates an in situ recombination mechanism in cDNA-cosmid complementation for ER. The complete deduced amino acid sequence of *ERCC5* was reconstructed from several cDNA clones encoding a predicted protein of 1,186 amino acids. The *ERCC5* protein has extensive sequence similarities, in bipartite domains A and B, to products of *RAD* repair genes of two yeasts, *Saccharomyces cerevisiae* *RAD2* and *Schizosaccharomyces pombe* *rad13*. Sequence, structural, and functional data taken together indicate that *ERCC5* and its relatives are probable functional homologs. A second locus represented by *S. cerevisiae* *YKL510* and *S. pombe* *rad2* genes is structurally distinct from the *ERCC5* locus but retains vestigial A and B domain similarities. Our analyses suggest that *ERCC5* is a nuclear-localized protein with one or more highly conserved helix-loop-helix segments within domains A and B.

DNA repair pathways are ubiquitous enzymatic and DNA replicative processes that renew genomes from diverse types of damage (reviewed in reference 16). Nucleotide excision repair (ER) enzymes have the capacity to recognize a wide variety of helix-distorting and bulky DNA alterations, including pyrimidine dimers, (6-4) photoproducts, and chemically induced adducts. The biochemistry of ER in *Escherichia coli* is fairly well characterized (64). It comprises a highly ordered set of sequential protein complexes and events at damaged sites that produce excision of the alterations (46). Six proteins, including the *uvrABC* excinuclease complex, helicase II (*uvrD*), DNA polymerase I, and DNA ligase are the minimal necessary and sufficient components to excise the lesion containing oligonucleotide in vitro and to replace it by gap-filling DNA synthesis (22, 55). In contrast, the genetic and biochemical evidence in eucaryotes now implicates more than 10 genes and proteins as crucial for ER (6, 17, 31). The increased complexity of eucaryotic ER is also evidenced by the general lack of resemblance of almost all known ER proteins of yeasts and mammals compared with those of *E. coli*.

The importance of ER to human health is evident from at least two human genetic disorders, xeroderma pigmentosum

(XP) and Cockayne's syndrome (CS), and from recent epidemiological studies (reviewed in references 4, 26, 42, and 68). Most XP patients exhibit sun-induced skin degeneration, a high incidence of skin cancer, and frequently ocular and neurologic abnormalities. Cultured fibroblast cells of XP patients are moderately to profoundly deficient in ER (10, 69). There are at least eight XP genetic complementation groups (A through G and variant forms) probably representing different genes (26). Related diseases include at least three CS complementation groups (29) and XP-D/trichothiodystrophy syndrome (30, 56). Cells of CS patients also exhibit partial ER deficiency and increased UV sensitivity (63, 67). CS patients exhibit growth retardation and neurologic and skin degeneration but surprisingly lack increased rates of skin cancer (42). A few XP patients of groups B, D, and G have a more complex presentation that combines XP and CS syndromes (29, 30, 42, 67).

Characterization of XP and CS genes was first made possible by human gene transfer into rodent CHO cell mutants (6). By using this method, human repair genes *ERCC1*, *ERCC2* (XP-D), *ERCC3* (XP-B), and *ERCC6* (CS group B) have been sequenced and found to encode different deduced proteins (14, 62, 63, 66, 67). Conversely, mouse DNA complemented human XP-A cells, resulting in isolation of a functional mouse gene designated *XPAC* (58, 59). The present work follows from the functional cloning of the

\* Corresponding author. Electronic mail address: MacInnes@flovax.lanl.gov.

human *ERCC5* gene in cosmid clones (41). The *ERCC5* gene is approximately 32 kbp in size, and in cosmids it functionally corrects the ER deficiency of CHO UV135 cells and mouse mutant Q31 cells (41). In this study, we report the complete coding sequence of *ERCC5*, its deduced amino acid sequence, and a preliminary functional characterization of the *ERCC5* cDNA. The *ERCC5* mRNA has an apparent length of 4.6 kbp. No full-length, functional cDNA clones have been isolated so far. However, we determined their capacities to functionally complement the UV sensitivity of CHO UV135 cells by a novel cDNA-cosmid gene transfer approach. Our preliminary evidence strongly supports a cDNA-cosmid, homologous recombination mechanism for UV resistance complementation of UV135 cells (41, 43). cDNA-cosmid transformants have partial capacity for ER-mediated reactivation of UV-damaged chloramphenicol acetyltransferase (CAT) reporter gene.

We also report that the *ERCC5* protein has extensive sequence similarities, in bipartite domains A and B, to products of the *RAD* repair genes of two yeasts, *Saccharomyces cerevisiae* *RAD2* and *Schizosaccharomyces pombe* *rad13* (7, 37, 44). Several lines of evidence argue that *ERCC5* and the closely related ER genes of two yeasts are likely representatives of a functionally homologous locus conserved widely among eucaryotes. We show that *ERCC5* also has vestigial similarities in the conserved domains to other yeast genes, including *S. cerevisiae* *YKL510*, an anonymous open reading frame (ORF) (23), and its close relative *S. pombe* *rad2* gene (7). Our amino acid sequence analysis indicates that *ERCC5* is probably a nuclear-localized protein with one or more highly conserved helix-loop-helix (HLH) segments within domains A and B. We could find no sequence evidence in the *ERCC5* protein for nucleoside triphosphate (NTP)- or  $Mg^{2+}$ -binding domains associated with RNA/DNA helicases (32, 66).

## MATERIALS AND METHODS

**Mammalian cell lines and culture conditions.** The UV-sensitive cell line UV135 was derived from parental CHO AA8 cells as previously described (6). UV135 is deficient in the damage site incision step of ER (60). UV-resistant cell line 38.4.4 was derived by two serial genomic DNA transfers into UV135 cells (41). DNA from 38.4.4 was used for cosmid cloning of the *ERCC5* gene. Cells were grown as monolayers in modified Eagle minimal essential medium supplemented with 10% fetal bovine serum, penicillin, and streptomycin (GIBCO) at 37°C in a humidified incubator equilibrated with 5.5%  $CO_2$ .

**UV irradiation and UV survival studies.** Cell monolayers growing in 100-mm-diameter dishes were irradiated with a GE germicidal lamp at a dose rate of  $0.5 J \cdot m^{-2} \cdot s^{-1}$  by methods described previously (36). Determination of colony surviving fractions involved irradiation 3 h after cell plating at densities of 200 or 2,000 cells per 100-mm-diameter dish. These cultures were regrown for 8 to 9 days and then stained with ethanol-crystal violet for enumeration. For selection of UV-resistant transformants, cells were plated at 500,000 per dish and UV irradiated with  $5 J \cdot m^{-2}$  at 20 h postplating and once again 72 h after plating. After the second exposure, the colonies were grown for 6 days and then fixed and stained. UV-resistant colonies from independent experiments were subcloned for further characterization, using cloning rings. Single colonies were regrown to a culture size of  $2 \times 10^6$  to  $3 \times 10^6$  cells. These populations were then reirradiated once with  $5 J \cdot m^{-2}$  to eliminate any remaining UV-sensitive cells

arising from spontaneous segregation or UV shielding. The surviving cells were regrown to populations of  $5 \times 10^7$  cells and then frozen.

**Northern (RNA) analyses.** Probe DNAs for potential hybridization to mRNA on Northern blots were isolated from within the *ERCC5* gene functional boundaries from cosmid clones cH75 and cH44 (41). We found >15 small DNA fragments that by Southern hybridization lacked human low- $C_{ot}$  repetitive DNA sequences ( $C_{ot} < 50$ ). Total cellular RNA and polyadenylated mRNA were isolated from  $2 \times 10^7$  cells by the vanadyl chloride method (2) and by the Fast-Track (Invitrogen, Inc.) isolation procedure, respectively. RNA [10 to 20  $\mu$ g of total RNA, or 1 to 5  $\mu$ g of poly(A)<sup>+</sup> mRNA] was size separated by electrophoresis on 1% agarose gels containing formaldehyde (53). The RNA was transferred from the gel to nitrocellulose membranes (Bethesda Research Laboratories, Inc.) and hybridized to the <sup>32</sup>P-labeled DNA probes as instructed by the manufacturer.

**cDNA library screens.** The human fibroblast pcD2 cDNA library was from H. Okayama (8). Approximately  $3 \times 10^6$  colonies were screened by nylon filter hybridization (Gen-Plaque; NEN) with two or more DNA probes that were mRNA positive in Northern analysis. Filters were hybridized at 65°C as recommended by the manufacturer. Hybridizing clones were repurified to homogeneity through two or three rounds of filter colony isolations and hybridization to yield 85 clones in total.

Several other cDNA libraries were screened for predicted 5'-end segments of the *ERCC5* cDNA by anchored polymerase chain reactions (PCR) (34).  $\lambda$ gt10, and  $\lambda$ gt11 libraries were generous gifts from Davis Chen (Harvard University). The human hydatid mole tissue cDNA library was kindly provided by Joe Gatewood (Los Alamos National Laboratory). The latter library was constructed in the pcDNA<sub>II</sub> vector by Invitrogen. Four other cDNA libraries constructed in an Epstein-Barr virus plasmid vector were the gifts of R. Legerski (48). Each library DNA was prepared by a brief liquid growth culture prior to the isolation of plasmid or phage. Specific oligonucleotide primers (18 to 20 nucleotides) from vector flanking sequences of each library were synthesized on ABI Inc. model 380 and 392 DNA synthesizers, using the manufacturer's phosphoramidite chemistry. Nested primers were used from within available *ERCC5* cDNA sequences and coupled to 5'-end cDNA vector primers by PCR (34). The technique allowed us to extend the available partial cDNA sequences obtained from sequenced pcD library clones. PCR was carried out with an Ericomp Inc. thermocycling instrument by using AmpliTaq DNA polymerase (Perkin-Elmer/Cetus). The longest 5'-end cDNA fragment (620 bp) came from the Invitrogen pcDNA<sub>II</sub> library. The PCR fragment was recloned into pSK II (Stratagene, Inc.). We then sequenced nine clones to obtain a consensus sequence by methods described below. The sequence of one clone (PCR8) matched the consensus sequence perfectly. This fragment extended our cDNA sequence to just beyond the predicted translation start site of the *ERCC5* gene (see Results). Its sequence was also later confirmed by direct comparison with genomic DNA sequences of the appropriate exons in a cosmid clone, cH75 (data not shown).

**Sequencing of cDNA clones and PCR fragments.** Two pcD2 library clones (pcD 59 and pcD F10) were subcloned into the sequencing-mapping plasmid pSK I or pSK II in both orientations. Each subclone was sequenced by using modified T7 DNA polymerase (Sequenase; U.S. Biochemical Corp.) as instructed by the manufacturer and the dideoxy-

strand termination method (54). Sequencing templates were prepared either by single-stranded rescue of pSK DNA as described by Stratagene or with normal double-stranded plasmid templates (61). DNA sequencing of cDNA inserts proceeded by primer walking starting from vector sequence at 5' and 3' ends. The sequencing project was managed by using Genetics Computer Group software (12).

**cDNA-cosmid complementation of UV135 cells.** cDNAs and cosmids were prepared from *E. coli* DH1 and HB101 host cells, respectively, by standard alkaline lysis methods (53). These preparations were further purified by centrifugation through cesium chloride gradients. Cosmids were linearized with by restriction endonucleases (e.g., *PvuI* and *SfiI*) that cut the cosmids once within vector sequence. cDNAs were linearized either with *PvuI* in vector sequence or alternatively with *XhoI* to liberate the entire cDNA insert from the vector with 5'- and 3'-end flanking cuts (8). Transformations into UV135 cells were carried out by standard DNA-CaPO<sub>4</sub> precipitations (53). cDNA-cosmid mixtures (1:1, wt/wt) were precipitated together at a total DNA concentration of 5  $\mu\text{g} \cdot \text{ml}^{-1}$ . One milliliter of precipitate was dispersed into the medium of each dish. After 18 h of incubation, the cells were given a brief glycerol shock. Growth medium was aspirated from each dish, and then an overlay of 3 ml of phosphate-buffered saline-15% (vol/vol) glycerol was applied for 3 min. The plates were rinsed once with 10 ml of growth medium, which was then replaced with fresh medium. After 24 h, cells were removed from each primary dish by trypsin treatment and then replated on three to four dishes (~500,000 per dish) to attach and regrow overnight before UV irradiations as described above.

**DNA repair assay with a CAT reporter gene.** Plasmid pSVCATgpt (21) was prepared as described above and purified by CsCl gradient for CaPO<sub>4</sub>-DNA transformation. Supercoiled plasmid was irradiated with cumulative UV doses ranging from 0 to 750 J  $\cdot \text{m}^{-2}$  and then stored in frozen (-20°C) aliquots until use. Aliquots of either damaged or undamaged plasmids were linearized with restriction endonuclease *EcoRV* prior to transformation. Plasmid DNA (2.5  $\mu\text{g}$  in 250 mM CaCl<sub>2</sub>) was mixed slowly with *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid (HEPES)-buffered saline (1:1 [vol/vol], pH 7.05), and precipitates formed at room temperature (25°C) (53). The precipitation timing and dispersal were critical for achieving high-level transient expression of CAT enzyme activity in CHO cells. We determined that an interval of 15 min of precipitation in Falcon tubes (no. 2059) followed by very vigorous vortexing (10 s) are optimal precipitation conditions for these studies. UV135 cells were plated as monolayers 18 h prior to DNA transformation. After 20 h of precipitate treatment, cell monolayers were exposed to 15% glycerol as described above. After rinsing, growth medium was replaced and cultures were allowed to recover for 20 h. Cultures were then harvested by scraping in TEN buffer (53). CAT enzyme extractions were carried out by multiple freeze-thaw cycles (53). CAT enzyme was assayed by methods described elsewhere (41, 53).

**Nucleotide and protein sequence analysis.** The *ERCC5* cDNA sequence and its protein relatives were analyzed with the Genetics Computer Group suite of DNA-protein analysis programs (12).

**Nucleotide sequence accession number.** The complete nucleotide sequence of *ERCC5* cDNA, a portion of which is shown in Fig. 3, has been submitted to GenBank via Authorin software (Intelligenetics Inc.). The *ERCC5* gene GenBank accession number is L20046.

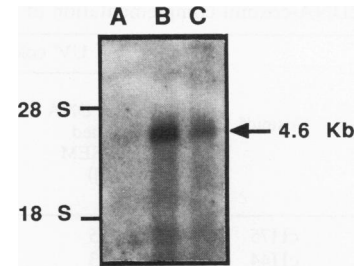


FIG. 1. mRNA expression of the *ERCC5* gene in CHO and HeLa cells. The DNA probe for Northern hybridization was derived from cosmid cH75, and it includes the sequence of translated exon 1 (see Fig. 3 and 5). The Northern blots were prepared as described in Materials and Methods, using 2  $\mu\text{g}$  of poly(A)<sup>+</sup> mRNA from UV135 (lane A), HeLa (lane B), and 38.4.4 (lane C) cells. Autoradiographic exposure of the Northern blot was for 3 days at -70°C.

## RESULTS

***ERCC5* mRNA size.** Nonrepetitive DNA probes from *ERCC5* cosmids cH75 and cH44 (41) were hybridized to Northern blots of RNA from CHO and human cells. A single mRNA species with an apparent molecular size of ~4,600 nucleotides was observed for several probes (Fig. 1). As expected for a human *ERCC5* gene probe, HeLa cell mRNA gave strong hybridization relative to the UV135 lane (Fig. 1, lanes A and B). *ERCC5* gene transformant 38.4.4 exhibited somewhat lower expression level for putative *ERCC5* mRNA (Fig. 1, lane C). The same probes produced weak cross-hybridization to a ~4.6-kb species in CHO UV135 mRNA that was visible only in autoradiographic overexposures of the Northern blots (faintly visible in Fig. 1, lane A).

**Isolation and functionality of *ERCC5* cDNA clones from the pcD2 library.** A human fibroblast cDNA expression library (8) probed with several nonrepetitive cosmid DNA fragments yielded 85 cDNA clones ranging in size from 2.2 to 3.7 kb. It was evident that no single clone was equivalent to the predicted full-length mRNA (estimated insert size, ~4.8 kb). The longest clone inserts with no evident rearrangements were pcD 59 and pcD F10, 3.7 and 3.65 kb, respectively. These two cDNAs and all others that we have tested were not active individually for UV resistance correction of UV135 cells by CaPO<sub>4</sub>-mediated gene transfer (see below). The properties of expressed minigene constructs made from segments of available cDNA and cosmid DNA will be reported elsewhere.

Most partial-length cDNA clones were indeed shown to be functional for correction of UV135 cells by cDNA-cosmid DNA cotransformation. The method was based on our previous demonstration of UV135 complementation by recombination of *ERCC5* gene segments in overlapping clones (41). Two cosmids, cH75 and cH44, together reconstitute the functional *ERCC5* gene by recombination in UV135 cells. Following the same reasoning, these two cosmids were then tested in pairwise combination with cDNA clones of various insert lengths (Table 1). Clones pcD F10 and pcD 33 complemented UV135 cells specifically with cH75 (5' end of *ERCC5*) but not with cH44, which contains the 3' end of the gene (Table 1). The efficiency of UV-resistant colony induction was also strongly dependent on the length of the cDNA insert in these pairwise cotransfers. In the shortest cDNA clone tested (pcD 64), correction of UV135 was not demonstrated by colony counting directly (Table 1). The combination of pcD 64 and cH75 was shown to produce UV-resistant

TABLE 1. cDNA-cosmid complementation of UV135 cells

cDNA			UV <sup>r</sup> colonies	
Clone	Length <sup>b</sup> (kbp)	Cosmid <sup>a</sup>	No./5 μg of DNA precipitated (mean ± SEM [n = 3])	No. of colonies formed in cDNA + cosmid expt/no. formed in control expt <sup>c</sup>
None		cH75	7 ± 5	Control
None		cH44	8 ± 3	Control
F10	3.65	None	5 ± 4	1
F10		cH75	107 ± 25	~18
F10		cH44	8 ± 5	1
33	2.60	None	4 ± 3	1
33		cH75	50 ± 15	~8
33		cH44	7 ± 4	1
64	2.35	None	6 ± 4	1
64		cH75	10 ± 5	<2
64		cH44	6 ± 3	1

<sup>a</sup> Cosmid cH75 has all 5'-end exons of *hERCC5*, including exon 8 (see Fig. 5). Cosmid cH44 has all exons downstream of exon 2.

<sup>b</sup> The position of each cDNA 5' terminus is indicated in Fig. 3. Insert lengths include poly(A) and poly(G/C) tails and vector cloning sites.

<sup>c</sup> Values for all controls, including individual cDNA experiments, were averaged; mean = 6.4.

colonies in UV135 by a sensitive selection-enrichment method used previously (36, 41). cDNA-cosmid gene transfer demonstrated that these cDNAs were authentic, if incomplete, derivatives of *ERCC5* mRNA. Sequence information (5' termini of the cDNA clones are shown in Fig. 3) and cosmid exon mapping would later provide detailed information about the putative mechanism by which cDNA and cosmids together reconstitute a UV resistance gene (see Discussion). Three independent colonies from cDNA-cosmid gene transfers were characterized further for UV resistance and DNA excision repair capacity.

**UV resistance in cDNA-cosmid transformants.** Following colony isolation and regrowth, three independent transformant clones retained resistance to UV light (Fig. 2A). Each transformant displayed an intermediate to normal level of UV resistance for colony formation relative to 38.4.4 cells (37% UV survival dose [ $D_{37}$ ],  $9.5 \text{ J} \cdot \text{m}^{-2}$ ). Colony-forming  $D_{37}$  doses for transformants A and B were  $9 \text{ J} \cdot \text{m}^{-2}$  (~95% survival correction relative to 38.4.4), while transformant C had a  $D_{37}$  of  $6 \text{ J} \cdot \text{m}^{-2}$  (~60% survival correction). In comparison, 38.4.4 had a  $D_{37}$  for UV survival identical to that of CHO AA8, the cell line progenitor to UV135 (data not shown).

**Excision repair in cDNA-cosmid transformants.** The CAT reporter gene has been used extensively to monitor ER in situ for cultured cells (25, 41, 49, 50, 68). UV-irradiated aliquots of plasmid pSVCATgpt were introduced into UV135 and 38.4.4 cells and the cDNA-cosmid transformants as described in Materials and Methods. We found that the UV dose that gave 37% relative CAT activity yield (CAT  $D_{37}$ ) in the ER-proficient cell line 38.4.4 is  $\sim 900 \text{ J} \cdot \text{m}^{-2}$  (Fig. 2B). In UV135 cells, CAT activity declined exponentially with increasing plasmid UV damage, with a CAT  $D_{37}$  of  $110 \text{ J} \cdot \text{m}^{-2}$  (Fig. 2B). CAT  $D_{37}$  values for transformants A, B, and C were 300, 400, and  $180 \text{ J} \cdot \text{m}^{-2}$ , respectively (Fig. 2B). The ~7.5-fold difference in CAT  $D_{37}$  values between UV135 and 38.4.4 cells is comparable to their relative colony survival  $D_{37}$  values (~5-fold difference) (Fig. 2A). Damaged CAT gene molecules are therefore reactivated extensively in

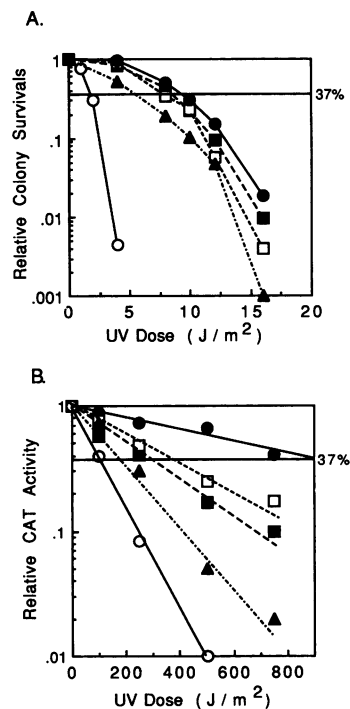


FIG. 2. UV resistance and pSV2CAT gene reactivation in cDNA-cosmid transformants. (A) Dose responses for colony formation after UV irradiation. Cells were grown and then UV irradiated with various single doses as described in Materials and Methods. Cells analyzed were 38.4.4 (closed circles), cDNA-cosmid transformants A (closed squares), B (open squares), and C (closed triangles), and UV135 (open circles). Data represent averages of replicate UV irradiations and colony survival determinations. Errors bars (standard errors of the means) are within the symbol widths. (B) ER-mediated reactivation of a damaged CAT reporter gene in cDNA transformants of UV135. Symbols and cell lines for dose-response curves are the same as for panel A. Plasmid pSVCATgpt was irradiated and  $\text{CaPO}_4$  transformed into various cell lines as described in Materials and Methods. The CAT enzyme assay was performed as described in Materials and Methods and reference 41. Control (nonirradiated) CAT enzyme activity yields varied less than 15% among the cell lines (data not shown). Slopes were determined from replicate experiments at each dose to the plasmid (variations in slope between replicate experiments were <10% for all cell lines).

38.4.4 cells. There is considerably greater heterogeneity among the transformants for CAT ER  $D_{37}$  values than for UV resistance  $D_{37}$  values (Fig. 2). Colony-forming  $D_{37}$  varied less than twofold among the UV resistant cell lines. Although there is some discrepancy between colony survival and CAT  $D_{37}$  assays of ER correction (see Discussion), we conclude that the cDNA-cosmid transformants produced a significant correction of UV135 ER deficiency as measured by either assay.

**Nucleotide and protein-coding sequence of *ERCC5*.** Evidence for ER complementation by the partial cDNAs and the estimated *ERCC5* mRNA size both indicated that longer cDNA inserts should be found. PCR-mediated screens of other cDNA libraries for extension products at the 5' end yielded the PCR8 fragment (Materials and Methods). Since PCR8 was isolated from a cDNA library that was not an expression vector, we decided not to rescreen the library for an entire insert. The locations of 5' termini of PCR8 and of pcD 59, F10, 33, and 64 are shown in Fig. 3. The cDNA



sequence compiled from PCR8 and two pcD clones (59 and F10) gave a long ORF of 3,558 nucleotides encoding a deduced protein of 1,186 amino acids (aa) (shown partly in Fig. 3). The predicted ATG translation start site is preceded by translation terminator codons in all three reading frames (asterisks in Fig. 3). There are no other long ORFs within this sequence. We conclude that the translated ORF of 1,186 aa depicted in Fig. 3 represents the deduced *ERCC5* protein.

The ATG translation start site sequence deviates from one theoretically most optimal for translation efficiency, i.e., GCC(A/G)CCATGG (27). The ATG context in *ERCC5* (GC CGCCTCATGGGGG) has a C in place of a purine at -3 relative to ATG (Fig. 3). The favorable G base in position +4 is present in *ERCC5*. The translation start region contains upstream predicted mRNA sequences exhibiting some potential for stem-loop formation. A short stem-loop may form between 6 bp starting at base -43 (AGAGATGA) with sequence at -24 (TCCGCTCT). The significance of an upstream stem-loop in the *ERCC5* mRNA is not known.

A polyadenylation signal (AATAAA) is present downstream from the translation termination codon, indicated in Fig. 3. We also found evidence for occasional polymorphism of polyadenylation sites in cDNA clones. One other sequenced cDNA clone (not shown) is polyadenylated about 150 nucleotides downstream from the end of the sequence shown in Fig. 3. The two cDNA clones described as functionally complementing with cosmid, as well as many other clone 3' ends, had identical poly(A) tail positions beginning after the last nucleotide in Fig. 3.

**ERCC5 protein similarities to yeast RAD proteins.** The translated protein of the *ERCC5* ORF has extensive regions of amino acid similarity to *S. cerevisiae* RAD2 and *S. pombe* Rad13 deduced proteins (7, 37, 44). Partial alignments of *ERCC5* protein with these RAD proteins and two others (*S. cerevisiae* YKL510 and *S. pombe* Rad2) are shown in Fig. 4). The deduced proteins exhibit greatest sequence similarities in bipartite domains, designated A and B (Fig. 4). These conserved domains are separated by nearly identical-length intervening sequences in *ERCC5*, RAD2, and Rad13. The intervening sequences of these proteins have much lower amino acid sequence homology, but they are similarly enriched in acidic residues (illustrated in Fig. 6). The other two genes, *S. cerevisiae* YKL510 and *S. pombe* rad2, are undoubtedly homologous representatives of a distinct structural and functional locus in their respective species (7, 23). The products of these two genes retain vestigial amino acid similarities with the other three in domains A and B but lack the intervening sequence between the domains (see Fig. 6).

Domain A is >45% identical in pairwise comparisons between *ERCC5*, RAD2, and Rad13, with overall similarities of >70%, allowing for conservative amino acid substitutions. Domain A can be further divided into two segments, A1 and A2, to optimize alignments with YKL510 and Rad2 (Fig. 4A) (23). The A1 segment spans 38 aa beginning at arginine 15 of *ERCC5*. The A1 consensus sequence includes 9 identical and 34 structurally conserved residues in total among the five proteins (Fig. 4A). Secondary protein structure prediction algorithms indicate that the A1 segment may have several folded  $\beta$  sheets (9, 20). The A2 conserved region begins at proline 52 of *ERCC5*. The A2 consensus sequence has 30 structurally conserved amino acids, including 14 identical residues. Two invariant prolines flank a nonvariant core of FVDG. Secondary structure predictions (9, 20) suggest that the A2 segment may form an HLH structure with a flexible loop between the prolines.

Domain B contains the most highly conserved region

between *ERCC5* and RAD2. It corresponds to at least 110 aa starting from glutamine 753 of *ERCC5* (Fig. 4B). The domain B region is >50% identical and >70% conserved in pairwise comparisons between *ERCC5*, RAD2, and Rad13. The other predicted protein sequences, YKL510 and Rad2, both lack the intervening segment of over 600 aa, and consequently their B domains start at leucine 119 in YKL510 and alanine 127 in Rad2 (Fig. 4B). We have divided domain B into three segments, B1, B2, and B3 (Fig. 4B). B1 is 27 aa in length and begins at glutamic acid 771 of *ERCC5* (Fig. 4B). The B1 segment consensus has 22 structurally conserved residues, including 14 identical aa, among the five proteins. Secondary structure predictions (9, 20) and Monte Carlo secondary folding simulations (22a) indicate consistently that the B1 segment may also form an HLH structure. Sequences upstream of B1 exhibit fairly limited homology, but we note a short basic amino acid segment between glutamine 753 and arginine 759 of *ERCC5*. This upstream region bears some resemblance to the basic motifs in the basic-HLH DNA-binding proteins (1). Near segment B2, there is a single potential tyrosine kinase phosphorylation site at *ERCC5* coordinates aa 835 (shown by the asterisk above the sequences in Fig. 4B). This site may be conserved only in RAD2 and Rad13 proteins (Fig. 4B). YKL510 and Rad2 have little B2 segment homology to the other three proteins and lack the potential kinase site. Segment B3 has greater similarities among the five proteins, starting at isoleucine 852 of *ERCC5*. In segment B3, 6 aa are identical and 21 aa are conserved structurally (Fig. 4B). YKL510 and Rad2 differ from *ERCC5* and its close relatives by three cysteine substitutions in the B3 segment. These residues may produce a zinc finger motif that may implicate this region of these two genes as an interface for protein-DNA interactions (15).

## DISCUSSION

We have determined the complete coding sequence of 3,558 bp (1,186 aa) for the human *ERCC5* DNA excision repair gene. All of our cDNA clones contain inserts smaller than the estimated full-length mRNA (~4.6 kb), but they are functional for correction of UV135 cells in pairwise mixtures with an overlapping cosmid 5'-end segment of the *ERCC5* gene (41). Transformants of UV135 cells derived from cDNA-cosmid mixtures were 60 to 95% corrected for colony-forming UV resistance, and they expressed 20 to 45% relative ER reactivation of damaged plasmid compared with a genomic *ERCC5* transformant. Our functionality data and cDNA sequence analysis provide compelling evidence for isolation of the entire *ERCC5* gene protein-coding region.

We approached the question of authenticating the cDNAs in a novel way. Each cDNA was cotransformed by CaPO<sub>4</sub> precipitation with a cosmid clone known to contain the 5' or 3' end of the *ERCC5* gene. cDNA-cosmid cotransformation provided a rapid complementation test permitting selection from among many cDNAs the two longest actively complementing clones for complete sequencing. By this test, a number of cDNAs were also found to be noncomplementing, and they were not studied further. We had found previously that cosmids cH75 (5' end) and cH44 (3' end) recombine in UV135 cells to reconstitute the *ERCC5* gene (41). Our evidence strongly supports this mechanism for cDNA-cosmid correction of UV135.

Relevant structural features of the cDNA and cosmid clones are illustrated in Fig. 5. Exon-intron junction sequences and cosmid restriction mapping will be presented elsewhere. We have shown that cDNA-cosmid (cH75) com-

A. Domain A alignments

			--		A 1		--
A1 Consensus	1	MGV GLW I E...PSA.RPV		ESLEGKR	AIDASIWLYQFLKAVRDQEGNQL		N....
hERCC5	1	---Q---KLL---C-G-Q-SP-A---		IL-V-I---	N-A-G---RH---SIE-....		
ScRAD2	1	---HSF-D-AG...-T-.-P-RL---		D-M-V---	I-----AVK-....		
SpRad13	1	---S---N-L-...-VK-.-P-KL-T-VN---		L-----I-----K---KS...			
YKL510	1	---IK---NA-ISEHV---I-KSDIK-FF-RKV---		MS-----I---Q-D-G---T-EAGE			
SpRad2	1	.....		R.....R.KV---MS-----IQ-S-D-Q---M-EQGE			
Constant AAs		MG		R	A D S	Q L VR	G

			--		A 2 HLH?		--
A2 Consensus		..SHL G F R CKLL FGIKP FVFDG		P LK QTL KR RR			
hERCC5	52	..P--LTL-H-L---F-R-R-I---		DA-L-K---V--RQ-K			
ScRAD2	52	..--IT-F-R-I---Y---R-V---		GV-V--RE-IRQ-KE--			
SpRad13	52	..--VV-F-R-I---F-----V---		GA-S--R--IQ--QA--			
YKL510	57	TT--M-M-Y-TLRMIDN---		CY---KP-D--SHE-T--SS--			
SpRad2	57	TT--M-M-Y-TLRIVDN---		CP---KP-T-S.....			
Constant AAs		H		F R	I P VFDG P LK		R R

B. Domain B Alignments

			-basic-		--		B 1 HLH?		--
B1 Consensus		QK Q R		VT M	E QELL LFGIPYI AP		EAEAQCA L		
hERCC5	752	A--QQ-E-IAAT--GQ-FL-S----		R-----Q-M-----I-DLTDQ					
ScRAD2	756	E-QMKDK-DSDE--MD-IK-V---		SR-----T-M-----E-LQLNL					
SpRad13	741	SK-GSEK-DADE--QV-IK-C----		R---L---V---Q-----SK-LLEKL					
YKL510	119	LE-MK-E-RLVK-SKEHNE-A-K--G-M---		I---T-----E-AKKGK					
SpRad2	127	.....AKRTVK--RQHND-AKR--E-M---		FVN--C-----A-ARSGK					
Constant AAs				V	E LL G P AP EAEAQC L				

			--		B 2		--		Ty.Kin.?
B2 Consensus		V G TDDSD LFG		YRN F	VE Y D		L...		*
hERCC5	803	TS-TI-----IW---ARHV---F-NKNKF--Y-QYV-FHNQ-...							
ScRAD2	806	-D-I-I-----VF---GTKI-K-M-HEKNY--F-DAESILKL-...							
SpRad13	792	-D-IV-----VF---GTRV---M-NQNKF--L-LMD-MKREFNVN							
YKL510	170	-YAAASE-M-TLCYRTPFL-HLTFSEAKK-PIHEI-TELV-R..							
SpRad2	172	-YAAASE-M-TLC-QAPVLL-HLTFSEQRK-PISEYNIEKA-N..							
Constant AAs				D D	E				

			--		B 3		--
B3 Consensus		GLD..		I LA LLGSDYTEGI	GVGPVTA E		EF
hERCC5	845	---.RNKL-N--Y-----PT--C---M-ILN-F					
ScRAD2	848	---.RKNM-E--Q-----N-LK-M---SI-VIA-F					
SpRad13	835	QM-----L-K--H-----M-LSR---L-L-ILH-F					
YKL510	213	---LTIEQFVD-CIM--C--C-S-R-----LKLTKTH					
SpRad2	215	---MSVEQFVD-CI---C--C-P-R---A.....					
Constant AAs		D		L LG DY	G		

FIG. 4. Amino acid alignments of ERCC5 with several related RAD proteins of *S. cerevisiae* and *S. pombe*. Amino acid alignments were derived with slight manual modifications from the FASTA program (33) as implemented in the Genetics Computer Group suite of programs (12). The *S. cerevisiae* RAD2 (ScRAD2) and *S. cerevisiae* YKL510 (ScYKL510) alignment is from reference 23. Partial amino acid sequences for *S. pombe* Rad13 (SpRad13) and *S. pombe* Rad2 (SpRad2) deduced proteins are from reference 7. Amino acid domains A and B are described in the text. A consensus sequence (50% identical residues) is shown above the multiple sequence alignment for each domain segment. Amino acids that are identical with the consensus sequence are represented by dashes; alignment gaps and regions for which no sequence data are available are indicated by dots. Amino acid numbering denotes actual sequence, not alignment gaps. Invariant amino acids are indicated below each multiple sequence alignment. (A) The domain A alignment is partitioned into subregions A1 and A2 as described in the text. An optimized alignment of human ERCC5 (hERCC5) with *S. cerevisiae* YKL510 and *S. pombe* Rad2 proteins is then found (7, 23). (B) The domain B alignment is partitioned into three subregions, B1, B2, and B3. Each segment is separated by short spacers with low sequence similarities. Segment B2 is not well conserved among the five genes and demonstrates in part the structural distinctions between the two loci (see text). An asterisk is located above the sequences of a potential tyrosine kinase site conserved among ERCC5, RAD2, and Rad13 proteins.

plementation of UV135 cells is strongly dependent upon the extent of complementing overlap between clone pairs. cDNA-cosmid overlaps deduced from sequence analysis range from 1,499 bp in pcD F10 (the most active clone) to 197 bp for pcD 64, in which production of UV resistance was marginally detectable. We also find absolute polarity for the cosmid clone (ch75) that cotransfers UV resistance. This is expected from the distribution of exons within each cosmid (Fig. 5B). We presume that cDNA-cosmid homologous recombination is also facilitated by the size and position of exon 8 (~1 kb; Fig. 5). Production of UV-resistant colonies was easily quantified for cDNA-cosmid overlap of >300 bp in exon 8. From these lines of evidence, cDNA-cosmid complementation almost certainly occurs by a conventional mechanism of homologous recombination followed by DNA

integration (43). Cotransfer of cDNA with cosmids or other cloning vectors may have general usefulness in both phenotype and positional cloning strategies, that is, for assessment of the position of homologous recombination with a much larger cloning vehicle such as yeast artificial chromosomes and for phenotype complementation with cosmids or yeast artificial chromosomes.

The cDNA-cosmid transformants have 60 to 95% colony-forming ability after UV irradiation relative to the 38.4.4 cell line, a human genomic ERCC5 DNA transformant. However, the cDNA-cosmid transformants had 20 to 45% relative ER reactivation of UV-damaged CAT plasmid compared with that in 38.4.4 cells. We selected 38.4.4 as the ER-competent control, as it was constructed in the UV135 genetic background. It is therefore nearly isogenic to our

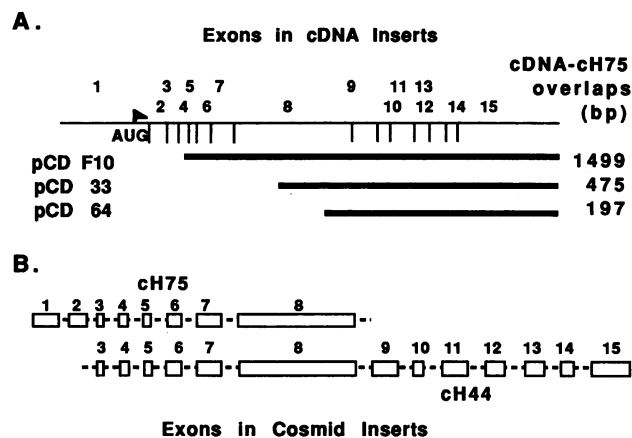


FIG. 5. Exon structure in *ERCC5* cDNA and cosmid clones. The genomic exon-intron map positions and exon junction sequences will be presented elsewhere. The diagram illustrates our current understanding of the corresponding *ERCC5* gene and cDNA exon-intron structures. The first translated exon is designated exon 1. (A) Positions and relative sizes of ~15 translated exons in the combined *ERCC5* cosmid and cDNA sequences. The 5'-end termini of the three cDNAs are also shown in Fig. 3. pCD F10 terminates within translated exon 4. pCD 33 and pCD 64 terminate within exon 8. (B) Exons in each cosmid (shown not to scale) were confirmed by PCR and DNA sequencing of the cosmid clone segments (unpublished results).

cDNA-cosmid transformants. 38.4.4 contains only one functional DNA copy of the *ERCC5* gene (41), as we presume to be the case in the cDNA-cosmid transformants. However, 38.4.4 was created with the normal genomic form of the *ERCC5* gene, whereas the cDNA-cosmid transformants must contain an artificial gene construct. We have no direct evidence about either (i) relative *ERCC5* protein levels in the two types of transformants or (ii) whether there is complete coding sequence integrity of *ERCC5* in the cDNA-cosmid versus genomic *ERCC5* forms. Significant molecular differences, in particular the very high DNA lesion densities in the CAT assay, which exceed by orders of magnitude the DNA lesion levels affecting UV colony survival, also may contribute to the discrepancy in ER corrections observed (5, 11, 24, 25, 35, 39, 40, 49). Although the exact cause(s) of partial CAT plasmid-ER complementation by cDNA-cosmid *ERCC5* genes is still speculative, we conclude that survival and CAT reactivation data taken together provide very good preliminary evidence that cDNA-cosmid transformants contain a functional, hybrid form of the *ERCC5* gene.

The cloned cDNA segments encode a predicted *ERCC5* protein sequence of 1,186 aa (molecular mass, ~133 kDa). Figure 6 is an illustration of structural and sequence similarities of *ERCC5*, *S. cerevisiae* RAD2, and *S. pombe* Rad13 deduced proteins. Overall structural similarities include an acidic central region that is flanked by basic elements near the C terminus. It seems particularly significant that *ERCC5*, *S. cerevisiae* RAD2, and *S. pombe* Rad13 have nearly identical amino acid intervals separating the bipartite, conserved domains A and B. Overall amino acid similarities between *ERCC5*, RAD2, and Rad13, including conservative substitutions, exceed 50% in both domains. It seems highly likely but not yet proved from these comparisons that *ERCC5* and these *RAD* genes are functionally equivalent homologs in their respective species. An evolutionary missing link in this argument is provided by the *S. pombe rad13*

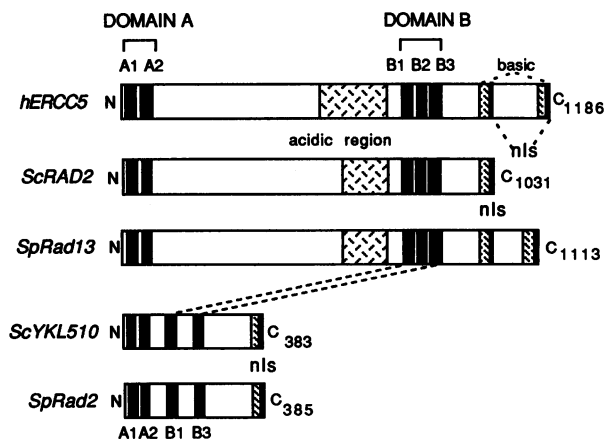


FIG. 6. Structural similarities between *ERCC5* protein and several yeast DNA repair proteins. A diagram of human *ERCC5* protein (*hERCC5*) is compared with published structures for *S. cerevisiae* RAD2, (*ScRAD2*), *S. cerevisiae* YKL510 (*ScYKL510*), *S. pombe* Rad13 (*SpRad13*), and *S. pombe* Rad2 (*SpRad2*) (see text). Two segments in *ERCC5* and its relatives are enriched in basic amino acids (stippled). The central regions of three deduced proteins are acidic (cross-hatched). These regions are enriched in serine, tyrosine, aspartic acid, and glutamic acid. Conserved segments within domains A and B are indicated by vertical solid bars. Putative NLSs are indicated at the C termini of all five deduced proteins.

gene and its cognate mutant cell line (7, 31). The cloned *S. cerevisiae* RAD2 gene functionally complements *S. pombe* mutant *rad13* (38). These two genes are therefore functionally equivalent homologs. From both sequence and structural considerations, the deduced *ERCC5* protein has equivalent similarity to Rad13. There is 55% similarity in domain A and 71% similarity in domain B between *ERCC5* and Rad13. All of the available evidence is consistent with the idea that these three genes are representatives of a functionally homologous locus conserved throughout long evolutionary distances.

The predicted *S. cerevisiae* YKL510 and *S. pombe* Rad2 proteins (7, 23) are strikingly different in overall structure from *ERCC5* and its homologs in lacking the intervening acidic section (Fig. 6). There is also considerable sequence divergence from the other genes within segments B2 and B3. Lehmann and colleagues (7, 31) have concluded that *Rad2* is the likely homolog of *YKL510*. *YKL510* will very likely be assigned an ER function in *S. cerevisiae*. These two genes represent a distinct locus with vestigial sequence similarities to the *ERCC5* locus.

What possible functions can be inferred about A and B domains in *ERCC5* and RAD2? Both *ERCC5* and RAD2 are highly hydrophilic and somewhat acidic proteins (17, 45; this work). These structures suggest that *ERCC5* and RAD2 may interact strongly with other macromolecules. Acidic nuclear proteins often have interactions with chromatin components such as histone (28, 51, 57). Proteins with the basic-HLH motifs have been shown to dock with the major groove of DNA (1). Antibody coprecipitation experiments with RAD2 protein and three other ER proteins gave no indication of RAD2 interaction with RAD1, RAD3, or RAD10 (3). Of course, these observations do not rule out any other interactions of RAD2 and *ERCC5* proteins. *ERCC5* and its homologs have one or more possible basic-HLH DNA-binding motifs. Our analyses also indicate that *ERCC5* protein has at least two nuclear location signals (NLSs) (Fig.



6). Dingwall and colleagues (13, 52) have described a bipartite 15- to 17-aa sequence shown to be necessary and sufficient for nuclear localization of certain proteins. Bipartite NLS sequences are frequently (but not always) present in DNA repair proteins, including RAD2 (18). Matches to this consensus sequence are coincident with the C termini of all of these genes (see Fig. 3 for the ERCC5 C-terminal NLS). We speculate that nuclear localization is a specific function of the C termini in all of these repair genes.

Finally, members of other protein groups, e.g., putative RNA/DNA helicases, have roles in mammalian and yeast ER. The human-*S. cerevisiae* yeast homologs ERCC2-RAD3 and ERCC3-RAD25 and the ERCC6 proteins each have a complete repertoire of seven RNA/DNA helicase motifs (47, 61, 65, 66). We have analyzed the ERCC5 protein sequence and find no evidence for the two most conserved helicase motifs, the nucleotide-binding fold domain I [GxGK(T/S)] (64) or the Mg<sup>2+</sup>-binding domain II (32). We would not expect ERCC5 protein to have the NTP-binding function associated with RNA/DNA helicases by itself.

#### ACKNOWLEDGMENTS

We thank E. Morton Bradbury and Paul Kraemer for critical readings of the manuscript.

This work was supported by the Office of Health and Environmental Research, Department of Energy, under contract W-7405-ENG-36 and by Los Alamos National Laboratory internal research and development funds.

#### ADDENDUM IN PROOF

During review of the manuscript, two groups published papers which indicated that the candidate ERCC5 protein complemented the DNA repair defects of XP-G and UV135 cell extracts in vitro (A. O'Donovan and R. D. Wood, *Nature [London]* **363**:185-188, 1993) and that a RAD2-like cDNA complements lymphoblastoid XP-G cells (D. Scherly, T. Nospikel, J. Corlet, C. UCLA, A. Baroch, and S. G. Clarkson *Nature [London]* **363**:182-185, 1993).

#### REFERENCES

- Anthony-Cahill, S. J., P. A. Benfield, R. Fairman, Z. R. Wassermann, S. L. Brenner, W. F. Stafford, C. Altenbach, W. L. Hubbell, and W. F. deGrado. 1992. Molecular characterization of helix-loop-helix peptides. *Science* **255**:979-983.
- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.). 1990. Current protocols in molecular biology, vol. 1 and 2. John Wiley & Sons, New York.
- Bailly, V., C. H. Sommers, P. Sung, L. Prakash, and S. Prakash. 1992. Specific complex formation between proteins encoded by the yeast DNA repair and recombination genes *RAD1* and *RAD10*. *Proc. Natl. Acad. Sci. USA* **89**:8273-8277.
- Bohr, V. A., M. K. Evans, and A. J. Fornace. 1989. DNA repair and its pathogenetic implications. *Lab. Invest.* **61**:143-161.
- Bohr, V. A., D. S. Okumoto, and P. C. Hanawalt. 1986. Survival of UV-irradiated mammalian cells correlates with efficient DNA repair in an essential gene. *Proc. Natl. Acad. Sci. USA* **83**:3830-3833.
- Busch, D., C. Greiner, K. Lewis, R. Ford, G. Adair, and L. H. Thompson. 1989. Summary of complementation groups of UV-sensitive CHO cell mutants isolated by large-scale screening. *Mutagenesis* **4**:349-354.
- Carr, A. M., K. S. Sheldrick, J. M. Murray, R. Al-Harthy, F. Z. Watts, and A. R. Lehmann. 1993. Evolutionary conservation of excision repair in *Schizosaccharomyces pombe*: evidence for a family of sequences related to the *Saccharomyces cerevisiae* *RAD2* gene. *Nucleic Acids Res.* **21**:1345-1349.
- Chen, C., and H. Okayama. 1987. High-efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell. Biol.* **7**:2745-2752.
- Chou, P. Y., and G. D. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* **47**:45-147.
- Cleaver, J. E. 1968. Defective repair replication of DNA in xeroderma pigmentosum. *Nature (London)* **218**:652-656.
- Cleaver, J. E., F. Cortes, L. H. Lutze, W. F. Morgan, A. N. Player, and D. L. Mitchell. 1987. Unique DNA repair properties of a xeroderma pigmentosum revertant. *Mol. Cell. Biol.* **7**:3353-3357.
- Devereux, J., P. Haeblerli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**:387-395.
- Dingwall, C., and R. A. Laskey. 1991. Nuclear targeting sequences—a consensus? *Trends Biochem. Sci.* **16**:478-481.
- Fletjer, W. L., L. D. McDaniel, D. Johns, E. C. Friedberg, and R. A. Schultz. 1992. Correction of xeroderma pigmentosum complementation group D mutant cell phenotypes by chromosome and gene transfer: involvement of the human *ERCC2* DNA repair gene. *Proc. Natl. Acad. Sci. USA* **89**:261-265.
- Freemont, P. S., I. M. Hanson, and J. Trowsdale. 1990. A novel cysteine-rich sequence motif. *Cell* **64**:483-484.
- Friedberg, E. C. 1985. DNA repair. W. H. Freeman, San Francisco.
- Friedberg, E. C. 1991. Eukaryotic DNA repair: glimpses through the yeast *Saccharomyces cerevisiae*. *BioEssays* **13**:295-302.
- Friedberg, E. C. 1992. Nuclear targeting sequences. *Trends Biochem. Sci.* **17**:347.
- Friedberg, E. C. 1992. Xeroderma pigmentosum, Cockayne's syndrome, helicases, and DNA repair: what's the relationship? *Cell* **71**:887-889.
- Garnier, J., D. J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* **120**:97-120.
- Gorman, C. M., L. F. Moffat, and B. H. Howard. 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* **2**:1044-1051.
- Grossman, L., and A. T. Yeung. 1990. The UvrABC endonuclease of *Escherichia coli*. *Photochem. Photobiol.* **51**:749-755.
- Gupta, G. (Los Alamos National Laboratory). Personal communication.
- Jacquier, A., P. Legrain, and B. Dujon. 1992. Sequence of a 10 · 7 kb segment of yeast chromosome XI identifies the *APN1* and the *BAF1* loci and reveals one tRNA gene and several new open reading frames including homologs to *RAD2* and kinases. *Yeast* **8**:121-132.
- Jones, C. J., J. E. Cleaver, and R. D. Wood. 1992. Repair of damaged DNA by extracts from a xeroderma pigmentosum complementation group A revertant and expression of a protein absent in its parental cell line. *Nucleic Acids Res.* **20**:991-995.
- Klocker, H., R. Schneider, H. J. Burtscher, B. Auer, M. Hirsch-Kauffmann, and M. Schweiger. 1985. Transient expression of a plasmid gene, a tool to study DNA repair in human cells: defect of DNA repair in Cockayne's syndrome; one thymine cyclobutane dimer is sufficient to block transcription. *Eur. J. Cell Biol.* **39**:346-351.
- Kraemer, K. H., and M. M. Lee. 1987. Xeroderma pigmentosum. *Arch. Dermatol.* **123**:241-250.
- Kozak, M. 1991. An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell Biol.* **115**:887-903.
- Lapeyre, B., H. Bourbon, and F. Amainic. 1987. Nucleolin, the major nuclear protein of growing eucaryotic cells: an unusual protein structure revealed by the nucleotide sequence. *Proc. Natl. Acad. Sci. USA* **84**:1472-1476.
- Lehmann, A. R. 1982. Three complementation groups in Cockayne syndrome. *Mutat. Res.* **106**:347-356.
- Lehmann, A. R., C. F. Arlett, B. C. Broughton, S. A. Harcourt, H. Steingrimsdottir, M. Stefanini, A. M. R. Taylor, A. T. Natarajan, S. Green, M. D. King, R. M. MacKie, J. B. P. Stephenson, and J. L. Tolmie. 1988. Trichothiodystrophy, a

- human DNA repair disorder with heterogeneity in the cellular response to ultraviolet light. *Can. Res.* **48**:6090-6096.
31. Lehmann, A. R., A. M. Carr, F. Z. Watts, and J. M. Murray. 1991. DNA repair in the fission yeast, *Schizosaccharomyces pombe*. *Mutat. Res.* **250**:205-210.
  32. Linder, P., P. F. Lasko, M. Ashburner, P. Leroy, P. J. Nielson, K. Nishi, J. Schnier, and P. P. Slonimski. 1989. Birth of the D-E-A-D box. *Nature (London)* **337**:121-122.
  33. Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science* **227**:1435-1441.
  34. Loh, E. Y., J. F. Elliott, S. Cwiria, L. L. Lanier, and M. M. Davis. 1989. Polymerase chain reaction with single-sided specificity: analysis of T cell receptor  $\alpha$  chain. *Science* **243**:217-220.
  35. Lommel, L., and P. Hanawalt. 1993. Increased UV resistance of a xeroderma pigmentosum revertant cell line is correlated with selective repair of the transcribed strand of an expressed gene. *Mol. Cell. Biol.* **13**:970-976.
  36. MacInnes, M. A., J. M. Bingham, L. H. Thompson, and G. F. Strniste. 1984. DNA-mediated cotransfer of excision repair capacity and drug resistance into Chinese hamster ovary mutant cell line UV-135. *Mol. Cell. Biol.* **4**:1152-1158.
  37. Madura, K., and S. Prakash. 1986. Nucleotide sequence, transcript mapping, and regulation of the *RAD2* gene of *Saccharomyces cerevisiae*. *J. Bacteriol.* **166**:914-923.
  38. McCready, S. J., H. Burkill, S. Evans, and B. S. Cox. 1989. The *Saccharomyces cerevisiae RAD2* gene complements a *Schizosaccharomyces pombe* repair mutation. *Curr. Genet.* **15**:27-30.
  39. Mellon, I., G. Spivak, and P. C. Hanawalt. 1987. Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell* **51**:241-249.
  40. Mitchell, D. L., J. E. Vaughan, and R. S. Nairn. 1989. Inhibition of transient gene expression in Chinese hamster ovary cells by cyclobutane dimers and (6-4) photoproducts in transfected ultraviolet-irradiated plasmid DNA. *Plasmid* **21**:21-30.
  41. Mudgett, J. S., and M. A. MacInnes. 1990. Isolation of the functional human excision repair gene *ERCC5* by intercosmid recombination. *Genomics* **8**:623-633.
  42. Nance, M. A., and S. A. Berry. 1992. Cockayne syndrome: review of 140 cases. *Am. J. Med. Genet.* **42**:68-84.
  43. Nickoloff, J. A., and R. J. Reynolds. 1990. Transcription stimulates homologous recombination in mammalian cells. *Mol. Cell. Biol.* **10**:4837-4845.
  44. Nicolet, C. M., J. M. Chenevert, and E. C. Friedberg. 1985. The *RAD2* gene of *Saccharomyces cerevisiae*: nucleotide sequence and transcript mapping. *Gene* **36**:225-234.
  45. Nicolet, C. M., and E. C. Friedberg. 1987. Overexpression of the *RAD2* gene of *S. cerevisiae*: identification and preliminary characterization of *RAD2* protein. *Yeast* **3**:149-160.
  46. Orren, D. K., C. P. Selby, J. E. Hearst, and A. Sancar. 1992. Post-incision steps of nucleotide excision repair in *Escherichia coli*: disassembly of the UvrBC-DNA complex by helicase II and DNA polymerase I. *J. Biol. Chem.* **267**:780-788.
  47. Park, E., S. N. Guzder, M. H. M. Koken, I. Jaspers-Dekker, G. Weeda, J. H. J. Hoeijmakers, S. Prakash, and L. Prakash. 1992. *RAD25 (SSL2)*, the yeast homolog of the human xeroderma pigmentosum group B DNA repair gene, is essential for viability. *Proc. Natl. Acad. Sci. USA* **89**:11416-11420.
  48. Peterson, C., and R. Legerski. 1991. High frequency transformation of human repair-deficient cell lines by an Epstein-Barr virus-based cDNA expression vector. *Gene* **107**:279-284.
  49. Protic-Sabljic, M., and K. H. Kraemer. 1985. One pyrimidine dimer inactivates expression of a transfected gene in xeroderma pigmentosum cells. *Proc. Natl. Acad. Sci. USA* **82**:6622-6626.
  50. Protic-Sabljic, M., and K. H. Kraemer. 1986. Host cell reactivation by human cells of DNA expression vectors damaged by ultraviolet radiation or by acid-heat treatment. *Carcinogenesis* **7**:1765-1770.
  51. Ptashne, M. 1988. How eukaryotic transcriptional activators work. *Nature (London)* **335**:683-689.
  52. Robbins, J., S. M. Dilworth, R. A. Laskey, and C. Dingwall. 1991. Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell* **64**:615-623.
  53. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
  54. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
  55. Selby, C. P., and A. Sancar. 1990. Structure and function of the (A)BC excinuclease of *Escherichia coli*. *Mutat. Res.* **236**:203-211.
  56. Stefanini, M., P. Lagomarsini, C. F. Arlett, S. Marinoni, C. Borroni, F. Crovato, G. Trevisan, G. Cordone, and F. Nuzzo. 1986. Xeroderma pigmentosum (complementation group D) mutation is present in patients affected by trichothiodystrophy with photosensitivity. *Hum. Genet.* **74**:107-112.
  57. Sung, P., S. Prakash, and L. Prakash. 1988. The *RAD6* protein of *Saccharomyces cerevisiae* polyubiquitinates histones, and its acidic domain mediates this activity. *Genes Dev.* **2**:1476-1485.
  58. Tanaka, K., N. Miura, I. Satokata, I. Miyamoto, M. C. Yoshida, Y. Satoh, S. Kondo, A. Yasui, H. Okayama, and Y. Okada. 1990. Analysis of a human DNA excision repair gene involved in group A xeroderma pigmentosum and containing a zinc-finger domain. *Nature (London)* **348**:73-76.
  59. Tanaka, K., I. Satokata, Z. Ogita, T. Uchida, and Y. Okada. 1989. Molecular cloning of a mouse DNA repair gene that complements the defect of group-A xeroderma pigmentosum. *Proc. Natl. Acad. Sci. USA* **86**:5512-5516.
  60. Thompson, L. H., K. W. Brookman, L. E. Dillehay, C. L. Mooney, and A. V. Carrano. 1982. Hypersensitivity to mutation and sister-chromatid exchange induction in CHO cell mutants defective in incising DNA containing UV lesions. *Somat. Cell Genet.* **8**:759-773.
  61. Toneguzzo, F., S. Glynn, E. Levi, S. Jmolsness, and A. Hayday. 1988. Use of a chemically modified T7 DNA polymerase for manual and automated sequencing of supercoiled DNA. *Bio-Techniques* **6**:460-469.
  62. Troelstra, C., A. van Gool, J. de Wit, W. Vermeulen, D. Bootsma, and J. H. J. Hoeijmakers. 1992. ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. *Cell* **71**:939-953.
  63. van Duin, M., G. Vredeveltdt, L. V. Mayne, H. Odijk, W. Vermeulen, B. Klein, G. Weeda, J. H. Hoeijmakers, D. Bootsma, and A. Westerveld. 1989. The cloned human DNA excision repair gene ERCC1 fails to correct xeroderma pigmentosum complementation groups A through I. *Mutat. Res.* **217**:83-92.
  64. van Houten, B. 1989. Nucleotide excision repair in *Escherichia coli*. *Microbiol. Rev.* **54**:18-51.
  65. Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay. 1982. Distantly related sequences in the  $\alpha$ - and  $\beta$ -subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**:945-951.
  66. Weber, C. A., E. P. Salazar, S. A. Stewart, and L. H. Thompson. 1990. ERCC2: cDNA cloning and molecular characterization of a human nucleotide excision repair gene with high homology to yeast *RAD3*. *EMBO J.* **9**:1437-1447.
  67. Weeda, G., R. C. A. van Ham, W. Vermeulen, D. Bootsma, A. J. van der Eb, and J. H. J. Hoeijmakers. 1990. A presumed DNA helicase encoded by *ERCC-3* is involved in the human repair disorders xeroderma pigmentosum and Cockayne's syndrome. *Cell* **62**:777-791.
  68. Wei, Q., G. M. Matanoski, E. R. Farmer, M. A. Hedayati, and L. Grossman. 1993. DNA repair and aging in basal cell carcinoma: a molecular epidemiology study. *Proc. Natl. Acad. Sci. USA* **90**:1614-1618.
  69. Zelle, B., and P. H. M. Lohman. 1979. Repair of UV-endonuclease-susceptible sites in the 7 complementation groups of xeroderma pigmentosum A through G. *Mutat. Res.* **62**:363-368.