MOLECULAR AND CELLULAR BIOLOGY, Dec. 1990, p. 6160–6171
0270-7306/90/126160-12$02.00/0
Copyright © 1990, American Society for Microbiology
Vol. 10, No. 12

Molecular Cloning of the Human *XRCC1* Gene, Which Corrects Defective DNA Strand Break Repair and Sister Chromatid Exchange

LARRY H. THOMPSON,* KERRY W. BROOKMAN, NIGEL J. JONES,† SUSAN A. ALLEN, AND ANTHONY V. CARRANO

Biomedical Sciences Division, L452, Lawrence Livermore National Laboratory, P.O. Box 5507, Livermore, California 94550

Received 20 June 1990/Accepted 30 August 1990

We describe the cloning and function of the human *XRCC1* gene, which is the first mammalian gene isolated that affects cellular sensitivity to ionizing radiation. The CHO mutant EM9 has 10-fold-higher sensitivity to ethyl methanesulfonate, 1.8-fold-higher sensitivity to ionizing radiation, a reduced capacity to rejoin single-strand DNA breaks, and a 10-fold-elevated level of sister chromatid exchange compared with the CHO parental cells. The complementing human gene was cloned from a cosmid library of a tertiary transformant. Two cosmid clones produced transformants that showed ≈100% correction of the repair defect in EM9 cells, as determined by the kinetics of strand break repair, cell survival, and the level of sister chromatid exchange. A nearly full-length clone obtained from the pcD2 human cDNA expression library gave ≈80% correction of EM9, as determined by the level of sister chromatid exchange. Based on an analysis of the nucleotide sequence of the cDNA insert compared with that of the 5′ end of the gene from a cosmid clone, the cDNA clone appeared to be missing ≈100 bp of transcribed sequence, including 26 nucleotides of coding sequence. The cDNA probe detected a single transcript of ≈2.2 kb in HeLa polyadenylated RNA by Northern (RNA) blot hybridization. From the open reading frame and the positions of likely start sites for transcription and translation, the size of the putative XRCC1 protein is 633 amino acids (69.5 kDa). The size of the *XRCC1* gene is 33 kb, as determined by localizing the endpoints on a restriction endonuclease site map of one cosmid clone. The deduced amino acid sequence did not show significant homology with any protein in the protein sequence data bases examined.

The processes that repair DNA damage, as well as those responsible for the rearrangement of chromosomal DNA through recombinational events, are of fundamental importance in mutagenesis and carcinogenesis. In *Escherichia coli*, a large number of genes have been identified that participate in repair and recombination (15, 55, 56). In the simpler eucaryotes *Saccharomyces cerevisiae* and *Drosophila melanogaster*, the genetic analysis of these systems is less advanced, but here too progress has been made in identifying many genes that participate in repair processes (5, 15, 34). Progress has been much slower with mammalian cells because of the larger sizes of genes, longer cell doubling times, and intrinsic limitations on performing genetic analyses.

The use of repair-deficient rodent cell mutants has proved to be the most successful approach to the molecular cloning of human DNA repair genes (45). Mutants of CHO cells that are defective in the nucleotide excision repair pathway (48, 61) have provided a tool for isolating the human genes *ERCC1* (53, 60), *ERCC2* (57, 58), *ERCC3* (59), *ERCC5* (29), and *ERCC6* (18) on the basis of functional complementation of the mutations by transfection of human genomic DNA, followed by cosmid or bacteriophage λ vector cloning. These five *ERCC* (excision repair cross-complementing) genes correct mutations that produce a common phenotype of hypersensitivity to UV radiation, which is analogous to the defect in cells derived from humans with the disease xeroderma pigmentosum.

In the present study we report the cloning of a human gene that corrects the repair defect in the CHO mutant line EM9. This mutant was originally isolated by hypersensitivity to ethyl methanesulfonate and also shows hypersensitivity to ionizing radiation (47). The phenotype of this mutant has two particularly noteworthy features. First, there is a biochemical defect in the rate of repair of DNA single-strand breaks that accompany exposure to either ionizing radiation or alkylation damage, such as that produced by ethyl methanesulfonate. Single-strand breaks appear to have a longer half-life in EM9 cells (47). The net effect of this perturbation is likely the production of greater levels of chromosomal aberrations, leading to lethality (8). The second hallmark of EM9 cells is the very high level of sister chromatid exchange (SCE), approximately 10 times that of the parental CHO cell. The high SCEs appear to be due primarily to incorporated bromodeoxyuridine, which is normally used in measuring SCEs (36). EM9 cells also show another defect in recombination: the efficiency of homologous recombination was reduced to ≈37% of normal in a plasmid transfection assay (19).

A human gene that efficiently corrects the defect in EM9 cells was identified by both DNA transfection (49) and cell fusion to form hybrid cells with human lymphocytes (40). This gene, given the designation *XRCC1* (X-ray repair cross-complementing), was assigned to human chromosome 19 (40) and was later shown to lie on the long arm in the region 19q13.2-13.3 (33, 46).

* Corresponding author.

† Present address: Molecular Biology Research Group, School of Biological Sciences, University College of Swansea, Singleton Park, Swansea SA2 8PP, United Kingdom.

We previously described the production of tertiary transformants of EM9 in which the presence of specific human sequences corrected the cells to a chlorodeoxyuridine (CldUrd)-resistant phenotype (49, 51). The DNA isolated from one secondary transformant was reduced in molecular size to <50 kb by shearing and used to produce five tertiary transformants (51). One of these transformants was used in the present study to isolate the functional *XRCC1* gene by using a cosmid vector. Fragments of *XRCC1* were then used to probe a human cDNA expression library and obtain clones, one of which was partially functional. By analyzing the coding sequence and flanking regions of *XRCC1*, we explain the incomplete correction seen with this cDNA.

MATERIALS AND METHODS

Cells and culture conditions. The parental CHO line clone AA8 was isolated as being heterozygous at the *aprt* locus (50). Mutant EM9 was isolated from AA8 following mutagenesis with ethyl methanesulfonate (47). Doubling times for AA8 and EM9 cells were 12 and 16 h, respectively. Cultures were grown in α-MEM (Hazelton/KC, Lenexa, Kans.) with 10% fetal bovine serum (Hazelton/KC), 100 μg of streptomycin per ml, and 100 U of penicillin per ml. Stocks were grown in suspension culture on a revolving drum and renewed from liquid nitrogen storage every 3 months. Cell survival curves were made by standard procedures for measuring colony-forming ability (>50 cells per colony). Gamma-irradiated and control cells were seeded into 10-cm petri dishes at the appropriate concentrations and incubated until all colonies reached macroscopic size; irradiated cells were incubated longer than controls.

SCE. Bromodeoxyuridine (BrdUrd) was added to 10 μM to 10-ml suspension cultures at a cell concentration of 10^5 per ml. Colcemid was added to 0.1 μg/ml at 23 h, and the cells were harvested 5 h later. Cells were prepared on microscope slides and processed for visualization of SCEs as described previously (8). *XRCC1* transformants of EM9, which had been grown in CldUrd-containing medium prior to freezing, were thawed from liquid nitrogen and grown in normal medium for 5 days before the experiment was begun.

DNA transformation. Clones of cosmid or plasmid DNAs were transfected into cells by the calcium phosphate precipitation procedure (10). Precipitates containing 5 to 10 μg of DNA were added in a volume of 1 ml to 10-cm dishes containing 2×10^6 cells in 10 ml of medium. After ≈19 h, the dishes were rinsed with 10 ml of medium and incubated for 24 h with fresh medium for expression. Cells were then trypsinized and replated at 4×10^5 cells per 100-mm dish for selection of transformants and at 300 cells per dish for determining plating efficiency. Selections with CldUrd were done under conditions that result in 20% replacement of thymine by chlorouracil in the DNA (49). Selection for expression of the *neo* gene, which is present in the pcD2 cDNA vector (9), was done by selection for resistance to Geneticin (G-418 sulfate; GIBCO Laboratories, Grand Island, N.Y.) at a concentration of 1.8 mg/ml. For cosmid clones, selection for expression of the *gpt* gene, which was used previously in the primary through tertiary transfections with genomic DNA, was based on resistance to mycophenolic acid with MAXTA medium (49). Transformant colonies were isolated with Pipetman tips.

Alkaline elution of DNA. Experimental cells were labeled for 18 h with [^{14}C]thymidine (Amersham Corp., Chicago, Ill.; 0.005 μCi/ml, 56 mCi/mmol), suspended in fresh unlabeled medium, and incubated at 37°C for 4 h to allow DNA maturation. Cells were then concentrated to 10^6/ml in ice-cold medium and maintained on ice during and after irradiation with 6 Gy of ^{137}Cs gamma rays (24) while on a rotating turntable. In repair experiments, the temperature of the culture was abruptly raised to 25°C by adding warm medium, and incubation was continued at 25°C for 20 min with agitation. Samples (5 ml) for elution were taken at 0, 5, 10, and 20 min after warming, and each was added to 45 ml of ice-cold phosphate-buffered saline (PBS). Cells were centrifuged and resuspended in 5 ml of ice-cold PBS; 3.0×10^5 cells were loaded onto each polycarbonate filter (2-μm pore size, 25 mm diameter; Nuclepore, Pleasanton, Calif.). Standard cells were labeled with [^{3}H]thymidine (0.045 μCi/ml, 2 Ci/mmol; Amersham), suspended at 0°C in PBS at a concentration of 5×10^5 cells per ml, and kept on ice during and after irradiation with 4 Gy; 2.5×10^5 cells were combined with experimental cells on each filter. Elution of DNA was performed by procedures described previously (24, 27). Data were analyzed with an Apple II computer. Elution profiles were obtained by plotting the amount of ^{14}C-labeled experimental DNA retained on the filter against the amount of the retained ^{3}H-labeled standard DNA (log-transformed data). The initial slopes of elution profiles were taken as a measure of the relative number of strand breaks (14).

Construction and screening of cosmid library. Large (>150 kb) DNA was isolated from a tertiary transformant of EM9 (clone 9TTT3), which was produced as described previously (51). This procedure yielded 17 mg of DNA from 1.9×10^9 cells. The DNA was partially digested with *Mbo*I, and fragments in the range of 20 to 50 kb were isolated from a size-fractioning gradient (1.2 to 5.0 M NaCl). The procedures for ligation to the vector p14B1 (2.3 kb), packaging into λ phage particles, infection of *E. coli* HB101, and library screening were those described for the cloning of the *ERCC2* gene (57).

cDNA library screening. The pcD2 human fibroblast cDNA expression library was generously provided by Hiroto Okayama (9). The cells provided were from the initial amplification of the library and were further amplified ≈10-fold in our laboratory before screening. The procedure for screening this library was detailed before (58), with the following modifications. *Eco*RI restriction fragments spanning 35 kb of the 44-kb insert of cosmid clone pH9T3-7 were isolated for use as probes. To reduce the binding of probe to repetitive sequences, a 10-min pre-reassociation of the nick-translated probe (1 ng/μl; specific activity, 2×10^5 to 4×10^5 cpm/ng) was performed in solution, at a C_0t of 100, with 10 mg of human placental DNA (type XIII; Sigma Chemical Co., St. Louis, Mo.) per ml at 68°C in 5× SSC (1× SSC is 0.15 M NaCl, 0.015 M sodium citrate). Nick translation was done with a kit from Bethesda Research Laboratories (BRL), Gaithersburg, Md. Filters (137 mm) with >10^4 colonies were screened with 20 ng of probe DNA per filter, and filters with 50 to 500 colonies were screened with 10 ng per filter. Putative positive colonies on duplicate filters were purified at low density in secondary and tertiary rounds of screening.

Isolation of cosmid DNA. Small-scale preparation of cosmid DNA for the initial transfection tests was done by a rapid alkaline plasmid extraction procedure (3) modified as described before (57). Large-scale isolation of DNA from cosmids pH9T3-6 and pH9T3-7 was done by a Triton lysis procedure as follows. One or 2 liters of bacterial culture was grown in Luria-Bertani broth supplemented with 100 μg of ampicillin per ml (added three times beginning at an OD_{600} of 0.06). When the OD_{600} reached ≈0.8, cosmid DNA was

amplified by adding chloramphenicol to 200 μg/ml. Cells were suspended in 25% sucrose (BRL)–50 mM Tris hydrochloride (pH 8.0)–2.0 mg of lysozyme (Sigma) per ml and were lysed by addition of EDTA to 0.6 M and Triton X-100 (United States Biochemical Corp., Cleveland, Ohio) to 0.05%. To partially remove protein, 20 g of cesium chloride and 2.5 ml of 10-mg/ml ethidium bromide were added to 20 ml of lysate and centrifuged at ≈20,000 × *g* for 30 to 60 min. The supernatant was collected, and the protein aggregated at the top of the column of liquid was eliminated. After centrifugation, the banded DNA was extracted with isopropyl alcohol equilibrated with TE buffer (10 mM Tris hydrochloride [pH 8.0], 1 mM EDTA) and saturated with NaCl to remove ethidium bromide. The yield was 500 to 700 μg of cosmid DNA per liter of cells. Gradient-purified DNA was digested with 50 μg of proteinase K (Sigma) per ml in 0.5% sodium dodecyl sulfate–20 mM EDTA–100 mM NaCl for 1 h at 65°C. After phenol extraction and ethanol precipitation, recovery of DNA was 50 to 60%.

Isolation of cDNA plasmids. For preparation of cDNA plasmids, bacterial cultures were amplified in 150 μg of chloramphenicol per ml. Lysis and cesium chloride-ethidium bromide gradient purification were performed as described for cosmid DNA. The DNA band from the gradient was processed as described before (C. A. Chen and H. Okayama, Biotechniques **6**:632, 1988). The DNA yield was 1.0 to 1.7 mg per liter of bacterial culture.

Purification of DNA for sequencing. M13 DNA was isolated and purified as described by the Cyclone I Biosystem protocol (International Biotechnologies, Inc., New Haven, Conn.). Phagemid DNA was isolated by the same procedure except that two extractions with chloroform alone preceded and followed five to seven phenol-chloroform-isoamylalcohol (25:24:1) extractions. For plasmid DNA isolation, bacteria were grown in TB medium (K. Tartof and C. Hobbs, Focus [publication of BRL] **9**:12, 1987) supplemented with 100 μg of ampicillin per ml. Lysis was carried out by suspension of cells in 2 mg of lysozyme per ml in 50 mM Tris, pH 7.4, followed by addition of EDTA to 0.1 M and Triton X-100 to 0.08%. The cleared lysate was digested with 20 μg of DNase-free RNase A (Sigma) per ml and extracted five to seven times with PCI. DNA was precipitated with ethanol, suspended in water, and reprecipitated for 4 h on ice with 8.6% polyethylene glycol 8000 (Sigma) and 0.7 M ammonium acetate. The last step removed RNA, which remained in solution after centrifugation at ≈12,000 × *g*. DNA pellets were rinsed with 70% ethanol.

Restriction endonuclease mapping of cosmid DNA. All restriction endonuclease digestions described in this article were performed as recommended by the suppliers (BRL; Boehringer Mannheim Biochemicals, Indianapolis, Ind.; New England BioLabs, Beverly, Mass.; U.S. Biochemical Corp.; Promega Corp., Madison, Wis.). The positions of restriction endonuclease sites within cosmid pH9T3-7 were determined by the Southern cross-restriction mapping technique (37), supplied in kit form by NEN-Dupont (Wilmington, Del.). *Sal*I-linearized cosmid DNA was individually digested with one of a number of restriction endonucleases (*Eco*RI, *Hin*dIII, *Kpn*I, *Sma*I, *Sph*I, *Xba*I, and *Xho*I). Cross-blot hybridizations were prepared as described by the supplier and previously outlined for the *ERCC2* gene (57). Because of the presence of repetitive sequences in the cosmid restriction fragments, the stringency of the reaction was increased compared with the published procedure (37) by raising the temperature of hybridization from 37 to 45°C and that of the washes from 50 to 65°C.

To accurately determine site-to-site distances, cosmid DNA was digested with each possible pair of endonucleases used in the Southern Cross restriction procedure plus *Nae*I and *Pvu*I (presented in Fig. 3). For some digestions, the cosmid DNA was previously linearized by digestion with *Sal*I. Electrophoresis for fragment size determination was performed with the appropriate gels. In some digestions, 15 to 25 ng of DNA was end labeled, gels containing labeled digestion products were dried, and the bands were visualized by autoradiography.

Localization of *XRCC1* on cosmid pH9T3-7. To determine the position and orientation of *XRCC1* on cosmid pH9T3-7, Southern blot hybridization analysis was performed with restriction fragments of the cDNA clone pXR1-30 as probes. *Xho*I (or *Bam*HI) digestion of pXR1-30 releases two fragments from the cDNA insert (9): a 1.9-kb 5′-end fragment and a ≈300-bp 3′-end fragment. Cosmid DNA was digested with individual restriction endonucleases (*Eco*RI, *Hin*dIII, *Kpn*I, *Nae*I, *Pvu*I, *Sma*I, *Sph*I, *Xba*I, and *Xho*I), separated by electrophoresis on two identical 0.75% agarose gels, and transferred to nitrocellulose paper by Southern blotting (41). The 1.9-kb and 0.3-kb restriction fragments were labeled by nick-translation with a BRL kit with [α-^{32}P]dCTP and used as probes. Hybridizations were performed in 5× SSC–5× Denhardt solution (1× Denhardt solution is 0.02% bovine serum albumin, 0.02% polyvinylpyrrolidone, and 0.02% Ficoll)–0.1% sodium dodecyl sulfate–50 μg of denatured herring sperm DNA per ml–10% dextran sulfate by a slight modification of the method of Maniatis et al. (30). Following overnight hybridization, filters were washed sequentially in 2× SSC, 1× SSC, 0.5× SSC, and 0.1× SSC. All hybridizations and washes were done at 68°C.

Subcloning DNA fragments. For sequence analysis, 1.9-kb and 300-bp cDNA restriction endonuclease fragments of pXR1-30 were generated as described above. Cosmid pH9T3-7 was digested with *Hin*dIII, and the 3.0-kb *Hin*dIII fragment containing the 5′ end of *XRCC1* (see Fig. 4) was digested further with *Pst*I for subcloning. The 2.2-kb *Sma*I fragment containing the 3′ end of *XRCC1* was recovered from a *Sma*I digestion of a 9.5-kb *Hin*dIII doublet (9.46 and 9.65 kb). Restriction fragments were separated by electrophoresis through agarose (Ultrapure [BRL] or NuSieve GTG [FMC Bioproducts, Rockland, Maine]) and isolated by standard procedures (1, 30).

Restriction endonuclease-digested vector molecules were dephosphorylated with alkaline phosphatase (Boehringer Mannheim) and ligated to insert DNA by using T4 phage DNA ligase (Stratagene, La Jolla, Calif.) for 4 to 18 h at 18 to 24°C. At various stages of sequence determination, the vectors used were M13mp12 (25) (gift of Pieter de Jong), M13mp19 (BRL), pBS, and pBluescript II KS(+) (both from Stratagene). Recombinant vectors were transferred into *E. coli* NM522 or XL1-Blue (Stratagene). In order to identify M13 clones representing insert DNA in both the sense and antisense orientations, phage pairs were analyzed by the C test as described before (32). Deleted subclones of the 1.9-kb *Bam*HI cDNA fragment were obtained in each orientation with the Cyclone I Biosystem kit (International Biotechnologies). This procedure uses the exonuclease activity of phage T4 DNA polymerase to generate a series of overlapping clones for nucleotide sequence determination (11). Fragments of cosmid pH9T3-7 were subcloned into the pBluescript II KS(+) phagemid and grown in single-stranded form by infection of bacteria with helper phage R408 (Stratagene).

Nucleotide sequencing and sequence comparisons. Se-

quence determination of both strands of the insert in the cDNA plasmid pXR1-30 and of the 5'- and 3'-end regions from cosmid pH9T3-7 was carried out by dideoxynucleotide chain termination protocols and with the reagents of the Sequenase kit (U.S. Biochemical Corp.). Both dGTP and dITP reactions were performed with [α-^{35}S]thio-dATP radiolabel (Amersham). DNA was separated on denaturing wedge gels (0.4 to 1.2 mm) composed of 6% acrylamide (acrylamide-bisacrylamide, 38:2)–47% urea with a model S2 electrophoresis unit (BRL). Reactions were initiated with the following primers: universal (U.S. Biochemical Corp.), T3, T7, KS, or SK (Stratagene), or one of six 17- to 22-nucleotide oligomers (synthesized in our laboratory to permit analysis of sequence distal to original primers).

Sequence analyses were done with a Sun 3/50 computer and the programs and data banks from IntelliGenetics, Inc. (Mountain View, Calif.). The nucleotide sequences of the overlapping deletion subclones obtained from the 1.9-kb cDNA insert of pXR1-30 were compared for homologies and ordered with the program GEL. The 1,899-nucleotide coding region of *XRCC1* (see Fig. 5) was translated by using the program SEQ. The resulting 633-amino-acid sequence was used in the program FASTDB to search for homology with all sequences in the data bank SWISS-PROT (12,305 entries), which includes the Protein Identification Resource data bank. With the program PEP (in modes Align and Search), more detailed homology searches were done on ≈80 sequences of proteins involved in DNA metabolism, including the *RAD* proteins of *S. cerevisiae*.

Northern (RNA) blot hybridization. Total RNA was isolated by using an RNA isolation kit from Stratagene, and enrichment for polyadenylated [poly(A)$^+$] RNA was performed with oligo(dT)-cellulose columns (mRNA Separator; Clontech Laboratories, Palo Alto, Calif.). Denaturing gels were run as described before (1) except that the gel contained 1% agarose and 6% formaldehyde. An RNA size standard (0.24 to 9.5 kb) ladder was obtained from BRL. The RNA from the gel was transferred to a GeneScreen*Plus* filter that was then hybridized and washed according to the protocol from NEN-Dupont. The probe was the 1.9-kb *Bam*HI fragment from plasmid pXR1-30, which was labeled by random priming with [^{32}P]dCTP (kit from U.S. Biochemical Corp.). The hybridization volume of 10 ml contained 5.4 × 10^6 cpm and 1.8 ng of probe per ml of hybridization solution.

Primer extension and S1 nuclease analysis. Primer extension and S1 nuclease analyses of the candidate promoter region were performed as described before (1) with 36- to 39-mer oligonucleotides. S1 nuclease digestion was performed under aqueous hybridization conditions with 19 μg of HeLa poly(A)$^+$ RNA and 270 U of S1 nuclease (Pharmacia LKB Biotechnology Inc., Piscataway, N.J.). Primer extension reaction mixes included 5.0 to 7.5 μg of poly(A)$^+$ RNA, 10^5 cpm of primer, and 0.4 M NaCl in a 30-μl volume; the nucleic acid mixtures were denatured at 75°C for 10 min and hybridized at 65°C for 18 h. Primers were extended with 40 U of Moloney murine leukemia virus reverse transcriptase (Boehringer Mannheim) at 42°C for 90 min. Products of both analyses were subjected to electrophoresis on 6 to 8% acrylamide–urea gels with size standards as described before (58), labeled, however, with [α-^{35}S]thio-dATP.

Nucleotide sequence accession number. The sequence data reported here have been assigned GenBank accession no. M36089.

TABLE 1. Normalization of SCEs and chromosomal aberrations in cosmid transformants of CHO EM9 cells

Cell line[a]	No. of SCEs/cell		No. of aberrations/cell	
	Mean	SEM	Mean	SEM
EM9	105.0	1.97	1.19	0.11
AA8	10.1	0.48	0.15	0.04
H9T3-6-2	10.3	0.62	0.14	0.04
H9T3-6-3	12.6	0.55	0.11	0.03
H9T3-6-4	9.8	0.42	0.12	0.03
H9T3-6-5	11.3	0.50	0.09	0.03
H9T3-7-1	9.2	0.49	0.08	0.03
H9T3-7-2	9.6[b]	1.73	0.03	0.02

[a] Six transformant lines were derived as clonal isolates after transfection with cosmid pH9T3-6 or pH9T3-7.
[b] One cell of 50 had very high SCE levels.

RESULTS

Isolation of cosmid clones and testing for gene function. The DNAs from five tertiary transformants of EM9 were previously shown by Southern blotting and probing with the BLUR8 *Alu* family repetitive sequence to have a single common human-specific *Eco*RI restriction fragment of ≈20 kb (51). Because of the relatively small size of the sheared transfected DNA molecules used in producing these lines (<50 kb), we concluded that this fragment was closely associated with the *XRCC1* gene. A cosmid library was constructed from one of these transformants by using the vector p14B1 (see Materials and Methods) and screened by using nick-translated HeLa DNA as the probe for colony hybridization. Nine clones positive for human sequences were obtained from 8.5 × 10^5 bacterial colonies. (This frequency of ≈1 × 10^{-5} agrees well with the value of 6 × 10^{-6} theoretically expected if one copy of the gene is present in the tertiary transformant.) After the tertiary round of cosmid purification screening, individual colonies were isolated and cosmid DNAs were prepared.

Each cosmid was tested for the integrity of the *XRCC1* gene by DNA transformation of EM9 cells and testing for restoration of repair function as measured by resistance to CldUrd. Two of the nine clones were clearly capable of restoring repair, each giving transformation frequencies of ≈6 × 10^{-5} CldUrd-resistant (CldUrd^r) colonies per μg of DNA. These colonies grew vigorously under selection, suggesting that the *XRCC1* gene functioned well in the hamster cells. The remaining seven cosmids produced no colonies. For the two functional cosmids, colonies growing in CldUrd were tested for expression of the *gpt* gene, which was present as a dominant marker during the tertiary transfection. All colonies died when exposed to MAXTA medium, suggesting that a functional *gpt* gene was not present in the cosmids.

Digestion of the two functional clones (pH9T3-6 and pH9T3-7) with several restriction endonucleases (*Bam*HI, *Eco*RI, *Hin*dIII, *Kpn*I, *Pst*I, *Pvu*II, *Xho*I, and *Sal*I) led to loss of correcting activity of the transfecting DNA. This result indicated that the gene was not contained entirely within any of the restriction fragments produced.

Efficient correction in cosmid transformants of EM9. Individual cosmid transformants were isolated and evaluated in detail to further assess how well the human gene corrected the repair defect. SCE provided a sensitive measure of the defect in EM9 cells, as shown in Table 1. EM9 had 10 times as many SCEs as wild-type AA8 cells. Six clonal isolates from cosmid DNA transfection all had SCE levels that were

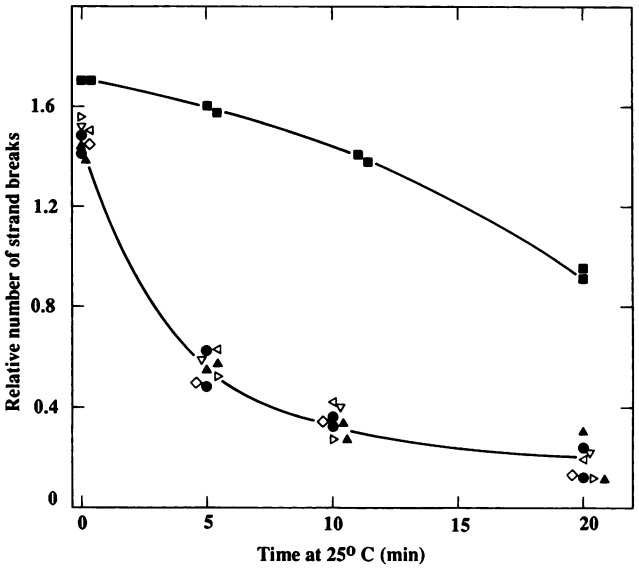

FIG. 1. Efficiency of strand break repair in *XRCC1* transformants of EM9 cells produced by using cosmids pH9T3-6 and pH9T3-7. DNA strand breaks were introduced into cells by irradiation on ice with 6 Gy of ^{137}Cs gamma rays. Cultures were abruptly warmed to 25°C to allow time for repair, and samples were taken at the times indicated. Strand breaks were measured by the method of alkaline elution of DNA from filters. The slopes of the elution curves, when normalized to the elution of internal standard DNA, were used as a measure of the relative number of breaks. Symbols: ●, normal CHO AA8; ■, mutant EM9; ▲, tertiary genomic transformant 9TTT3; cosmid transformants 9T3-6-2 (◁), 9T3-6-3 (▷), 9T3-7-1 (◇), and 9T3-7-2 (▽). Duplicate symbols represent independent experiments.

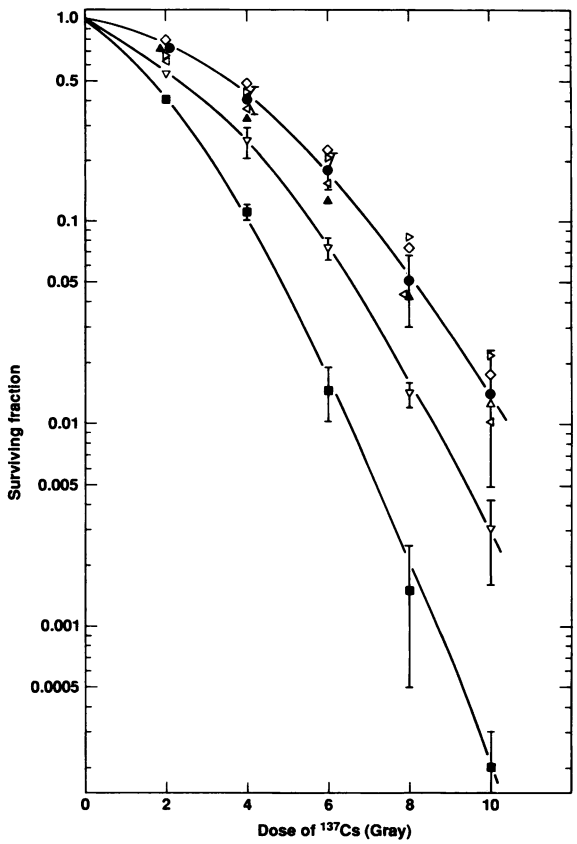

FIG. 2. Survival of colony-forming ability of wild-type CHO AA8, mutant EM9, and *XRCC1* transformants exposed to ^{137}Cs gamma rays. The symbols are the same as defined in the Fig. 1 legend. Error bars show standard errors of the mean for repeat experiments on AA8, EM9, and transformant 9T3-7-2.

very similar to that of AA8 cells. (The 1 cell among 50 that had high SCE levels in the culture of pH9T3-7-2 appears to have lost the expression of the transfected gene.)

The same microscope slides used for SCE analysis were also scored for chromosomal aberrations at the second mitosis after the addition of BrdUrd. EM9 cells showed eight times more aberrations per cell than AA8 cells (Table 1). This increase in aberrations appears to reflect the presence of DNA strand breaks associated with BrdUrd incorporation that are poorly repaired in EM9 cells (14). The aberration levels in the six transformants were very similar to that of the AA8 cells. Thus, by both cytogenetic endpoints, the defects in the cosmid transformants were fully corrected by the expression of *XRCC1*.

A tertiary transformant and four cosmid transformants (two each derived from pH9T3-6 and pH9T3-7) were further analyzed for correction of the repair defect. The repair of DNA strand breaks induced by ionizing radiation, measured by the technique of alkaline elution, was greatly reduced in EM9 cells compared with normal AA8 cells (Fig. 1). All *XRCC1* transformants were fully corrected and gave repair kinetics that matched those of AA8 cells.

The reduced strand break repair in EM9 appears to be responsible for its hypersensitivity to ionizing radiation. Therefore, the effect of the *XRCC1* gene in EM9 cells was also examined in terms of cell survival by colony-forming ability after exposure of transformants to gamma rays (Fig. 2). AA8 cells were ≈1.75 times more resistant than EM9 cells (based on D_{10}, D_{10} being the dose required to reduce the survival to 10%), in agreement with earlier results (47). All but one of the *XRCC1* transformants had cell survival curves that were not distinguishably different from that of AA8 cells (Fig. 2). The exceptional transformant (9T3-7-2) showed an intermediate sensitivity. The reason for this behavior is not apparent but may involve some perturbation in growth characteristics associated with the DNA transformation process (2).

Isolation of cDNA clones and testing for function. The cosmid pH9T3-7 was digested with *Eco*RI (which releases the vector from the insert sequence), and the fragments were separated on a gel. Fragments representing a total of 35 kb of insert were isolated and combined for use as a probe to screen the pcD2 human fibroblast cDNA expression library (9). An analysis of *XRCC1* restriction fragments (from several enzymes) by Southern blotting and hybridizing with total human DNA showed that nearly all restriction fragments contained repetitive sequences (results not shown). Therefore, for the library screening, the cosmid fragments were pre-reassociated with total, sheared human DNA after being labeled by nick translation (see Materials and Methods). From 1.5×10^6 bacterial colonies screened, 6 *XRCC1*-positive colonies were obtained and purified through the tertiary round of screening.

Digestion of the purified pcD2 plasmid clones with either *Bam*HI or *Xho*I releases the human cDNA insert flanked on each side by a small amount of vector sequence (9). Digestion of the six positive clones with *Bam*HI yielded insert sizes of 0.9, 1.0, 1.7, 1.7, 1.7, and 2.3 kb. (An internal *Bam*HI site actually resulted in two fragments, one of ≈300

TABLE 2. Partial correction of SCE frequency in cDNA transformants of CHO EM9 cells

Cell line	No. of SCEs/cell[a]
AA8	9.5
EM9	103.0
9T3-7-1 (cosmid)	9.2
9CD30-1 (cDNA)	31.7
9CD30-2 (cDNA)	28.6
9CD30-3 (cDNA)	28.7

[a] Fifty second-division cells were scored in each culture.

bp in all clones. Each insert released should contain 135 nucleotides of vector sequence [107 5′-ward and 28 3′-ward] plus a G+C tract on the 5′ end of the cDNA and a poly[A-T] tract on the 3′ end. Digestion with *Xho*I released 213 [179 + 34] nucleotides of vector.) Each plasmid clone was transfected into EM9 cells, and selection for CldUrd^r was carried out in both the presence and absence of Geneticin, which selects for expression of the *neo* gene carried on the pcD2 vector. One plasmid clone (pXR1-30) gave CldUrd^r transformants at a frequency of 2.3×10^{-6} per μg with two independent DNA samples, and the other five clones yielded no transformants. The frequency of Geneticin resistance was 4×10^{-5} to 5×10^{-5} per μg and was unaffected by simultaneous CldUrd selection. The colonies arising in CldUrd-containing medium were noticeably smaller and slower-growing than those seen after transfection with cosmid pH9T3-7 DNA. These results suggested that clone pXR1-30 was unable to fully correct the repair defect in EM9. This point was tested further by isolating CldUrd^r pXR1-30 transformants. Three independent clones were analyzed for the level of SCE (Table 2). Each cDNA transformant had an SCE level well above that seen in AA8 or cosmid-transformed cells, and the level was very similar among the three transformants. The degree of correction of the repair defect by this endpoint was ≈80%. Thus, clone pXR1-30 was unable to produce a fully normal phenotype.

Restriction endonuclease site map of cosmid pH9T3-7. Cosmid clone pH9T3-7 was chosen for restriction endonuclease site mapping by the cross-blot hybridization method detailed in Materials and Methods (Fig. 3). Localization of *XRCC1* on the cosmid map was performed by hybridizing each of the two insert fragments resulting from *Bam*HI digestion of pXR1-30 (see Fig. 4B) against a blot containing restriction fragments of the cosmid. From these results, the 3′ end of the gene was localized to the 0.53-kb *Sma*I-*Kpn*I fragment, as shown. The 5′ end of the cDNA was contained within the 5.7-kb *Hin*dIII-*Sph*I fragment but did not show hybridization with the 3.0-kb *Hin*dIII fragment containing the unique *Sal*I site. However, as we show below (see Fig. 5), the 25 5′-terminal nucleotides of the cDNA are represented within this *Hin*dIII fragment but were insufficient for detecting homology.

Digestion of the cosmid at the unique *Sal*I site was known to inactivate the gene with respect to transformation of EM9 cells, indicating that the gene extended 5′ of this site. Because the insert of cDNA clone pXR1-30 contained a unique *Pst*I site beginning at the seventh nucleotide, this site was located on the cosmid. Identification of the appropriate site was made by first isolating the 3.0-kb *Hin*dIII fragment spanning the *Sal*I site and then digesting it with *Pst*I. Of the three resulting fragments (Fig. 4A), the 0.23- and 2.0-kb fragments were subcloned into pBluescript II KS(+), and the resulting single-stranded phagemids were used for nucleotide sequencing. This analysis showed that the *Pst*I site in the cDNA corresponded to the *Pst*I on the 3′ side of the *Sal*I site (Fig. 4A). The *Sal*I site, which appeared to lie within the promoter (see Fig. 5), delineated the minimum extent of the 5′ end of *XRCC1* in Fig. 3. Thus, the gene is ≈33 kb in length, based on the analysis of cosmid clone pH9T3-7.

In order to assess whether rearrangements occurred in the gene during construction and growth of the cosmid clones, the two functional cosmids, clones pH9T3-6 and pH9T3-7, were compared with respect to restriction fragment sizes on ethidium bromide-stained gels after individual digestions with *Eco*RI, *Xho*I, *Hin*dIII, and *Kpn*I. The restriction fragments lying within the gene, as defined by the heavy line in Fig. 3, were in full agreement between the two clones. The region checked for agreement spanned almost all of the gene, extending from the *Hin*dIII site at the 5′ end to the *Xho*I site near the 3′ end. As expected for clones of independent origin, the restriction fragments overlapping the ends of the gene differed in size between the two clones. Because of the similarity of the two functional cosmids, we conclude that rearrangements occurring during the cloning process were unlikely.

Test for rearrangements of *XRCC1* during genomic transfections. Although the *XRCC1* gene in cosmids pH9T3-6 and pH9T3-7 appeared to function normally, there was also a possibility of rearrangement (e.g., loss or gain of sequence) during the lengthy process of constructing the tertiary transformant. To test for gross rearrangement, two samples of human placental DNA were analyzed by Southern hybridization with a collection of cosmid restriction fragments that should detect all segments of the gene in the cosmid upon digestion with either *Hin*dIII or *Kpn*I and *Sal*I as the probe. Restriction fragments spanning the gene, from the unique *Sal*I site to the *Kpn*I site flanking the 3′ terminus, were represented in the digested genomic samples (data not shown; the two samples differed by ≈100 bp in the size of the gene in one region). Therefore, we conclude that the map of *XRCC1* shown in Fig. 3 is probably an accurate representation of the gene that was used in the primary transfection.

Nucleotide sequence of cDNA and genomic 5′ and 3′ ends. Nucleotide sequence determination was performed with the restriction fragments shown in Fig. 4. For the 5′ portion of the cDNA insert (Fig. 4B), the larger *Bam*HI fragment (1.9 kb) was used to generate overlapping deletion subclones in an M13 phage vector (see Materials and Methods). For the remainder of the cDNA, the sequences of *Xho*I and *Bam*HI fragments, each ≈300 bp, were determined directly. For the 5′ and 3′ ends of the gene from the cosmid pH9T3-7, restriction fragments were subcloned into pBluescript II KS(+) and the sequence was determined in either the single- or double-stranded form (Fig. 4A and C).

The nucleotide sequence information from the cDNA and genomic fragments was combined to give the composite sequence shown in Fig. 5. Translation of the open reading frame (ORF) beginning at the methionine codon marked +1 gave a protein of 633 amino acids. Although this ORF could be extended 177 nucleotides 5′ of +1, it contained no other in-frame Met codons. The second in-frame methionine codon was at +280. The 5′ end of the cDNA insert thus lacked 26 nucleotides of presumed translated sequence. The 5′ end of the *XRCC1* genomic sequence contained several possible regulatory elements. The sequence TTTGAAA beginning at position −31 with respect to the putative transcriptional start site, identified as explained below, matched a TATA box consensus sequence (6) in the underlined positions. The first T is usually located between −33

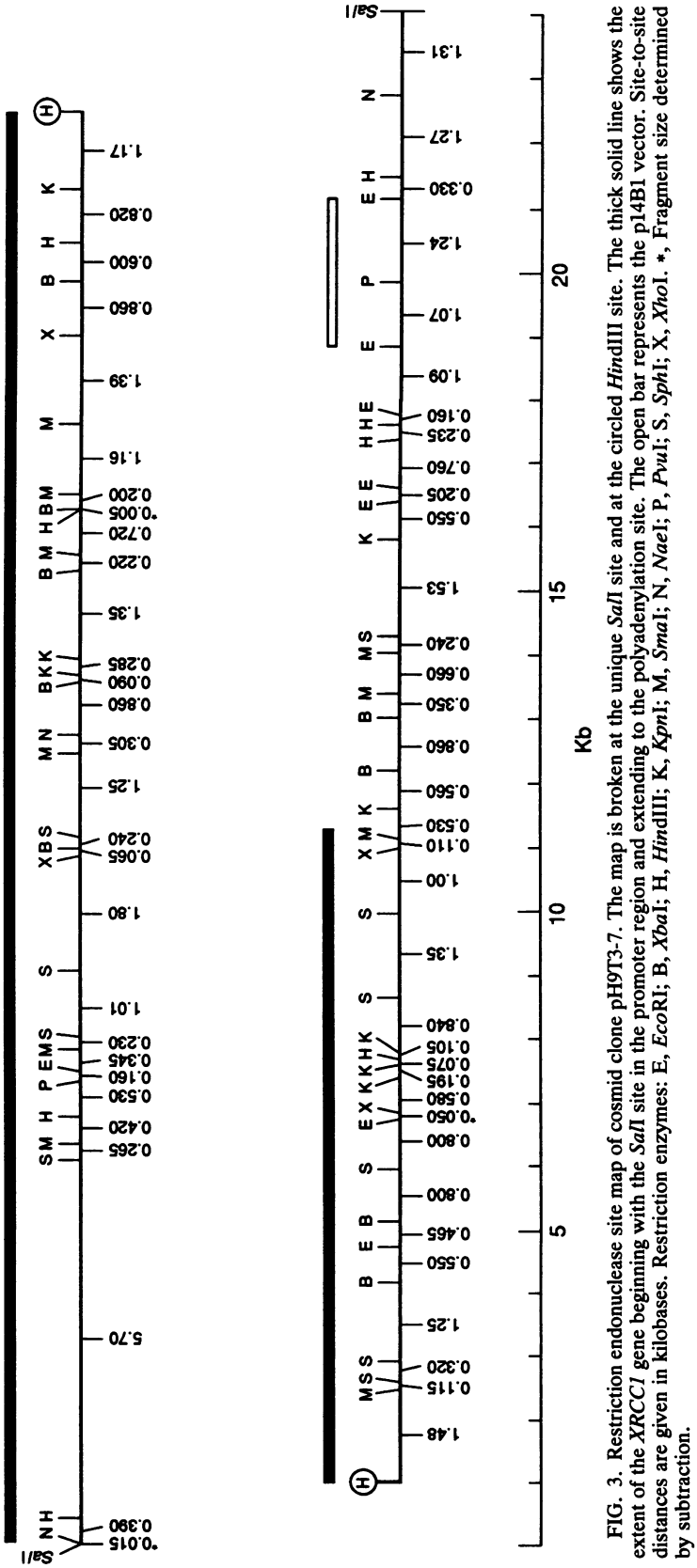

FIG. 3. Restriction endonuclease site map of cosmid clone pH9T3-7. The map is broken at the unique *Sal*I site and at the circled *Hin*dIII site. The thick solid line shows the extent of the *XRCC1* gene beginning with the *Sal*I site in the promoter region and extending to the polyadenylation site. The open bar represents the p14B1 vector. Site-to-site distances are given in kilobases. Restriction enzymes: E, *Eco*RI; B, *Xba*I; H, *Hin*dIII; K, *Kpn*I; M, *Sma*I; N, *Nae*I; P, *Pvu*I; S, *Sph*I; X, *Xho*I. *, Fragment size determined by subtraction.

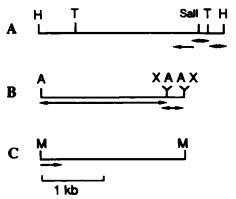

FIG. 4. Restriction enzyme fragments used for sequence determination of genomic and cDNA regions. (A) Cosmid-derived 3.0-kb *Hin*dIII fragment from pH9T3-7 that spans the 5′ end of *XRCC1*. (B) pXR1-30 cDNA insert released by *Bam*HI or *Xho*I digestion. (C) Cosmid-derived 2.3-kb *Sma*I fragment spanning the 3′ end of *XRCC1*. Arrows show directions of sequencing. Restriction site abbreviations: A, *Bam*HI; H, *Hin*dIII; M, *Sma*I; T, *Pst*I; X, *Xho*I.

and −29 from the transcriptional start site (6). There were two elements that resembled a CAAT box, which has the consensus GG(C/T)CAATCT (6): CCAATC beginning at −343 relative to translational start, and the element CCAAT in the reverse orientation, beginning at −287. However, both elements were located quite far upstream from the TATA box (see discussion by Lewin [28]). A GC box (GGGCGG) was also present beginning at position −391. A CpG-rich region included the unique *Sal*I site of cosmid pH9T3-7 as well as other diagnostic restriction endonuclease sites (*Nae*I and *Bss*HII). Other potential control elements were the 34-nucleotide A+T tract (underlined) and nine sequences present as inverted repeats (each 8 or 9 nucleotides long and indicated by the pairs of numbers in italics in Fig. 5). The 3′ end of the *XRCC1* coding region was closely flanked by a 34-nucleotide AC repeat and the prototype polyadenylation consensus sequence AATAAA (4). Sequencing of the 5′ and 3′ ends of genomic *XRCC1* also showed the positions of the first and last intron splice sites, as indicated in Fig. 5.

Transcriptional start sites. Having identified in the cosmid the putative start codon for translation, we performed two kinds of experiments that attempted to define the start point(s) for transcription. The first approach involved primer extension of human poly(A)$^+$ mRNA with reverse transcriptase to produce extension products, which on acrylamide-urea gels should correspond to the 5′ ends of transcripts. The second approach involved S1 nuclease protection, in which end-labeled single-stranded DNAs were annealed with mRNA, digested with S1 nuclease, and analyzed on the same kind of gels. Both kinds of experiments gave many autoradiographic bands but none of relatively high intensity (results not shown). Only one possible start site, at position −105 (marked in Fig. 5), was a common result of the two analyses. This site had features of mRNA start sites recognized in other genes (6); the first base is almost always A and occurs in a consensus sequence (PyNNNPyAPyPyPyPyPy, where Py is a pyrimidine and N is any nucleotide) that closely resembles that of our putative site (TGCGCACTTTA). Thus, the A at position −105 seems likely to be a preferred start site for transcription. This information helped confirm that the start site for translation was correctly identified. The S1 nuclease experiments also suggested about 10 other nearby positions, and primer extension suggested about four other positions.

Detection of *XRCC1* mRNA. To examine the size of the mRNA transcribed from the *XRCC1* gene, poly(A)$^+$ RNA was prepared from both human HeLa cells and a cDNA transformant of EM9 derived by using plasmid pXR1-30 (clone 9CD30-1; Table 2). A Northern blot (Fig. 6) was obtained by using the 1.9-kb *Bam*HI fragment of the cDNA as the probe. Poly(A)$^+$ RNA from HeLa cells gave a single band at ≈2.2 kb (lane 2), and a band in the same position was also weakly visible in total cytoplasmic RNA (lane 3). The upper band in lane 3 corresponds to the position of 28S rRNA and is attributed to background hybridization. Lane 1, containing poly(A)$^+$ RNA from the cDNA transformant, gave a band at ≈2.4 kb that was about fivefold more intense than the band from HeLa cells. This larger size can be attributed to longer leader and/or trailer sequences due to the positions of transcriptional initiation and termination sites of the simian virus 40 (SV40) sequences present in the pcD2 vector (9).

The ≈2.2 kb size of the mRNA in HeLa cells can be compared with the presumed transcribed region shown in Fig. 5 based on the mapping experiments described above. For transcription beginning at position −105, the predicted size of the transcript is 2,083 nucleotides plus a poly(A) tail. If we assume a poly(A) tail of 150 bp, then the expected size of the mRNA agrees with what is shown in Fig. 6 within experimental error.

Additional cDNA library screening. In an attempt to recover missing cDNA sequence, we used a 41-mer oligonucleotide (also used for primer extension experiments) that spanned the translational start site (overlined region in Fig. 5). With this probe, we screened 2.1 × 10^6 clones of the pcD2 library. No positive clones were obtained, although strong signals were seen on control filters containing colonies of cosmid pH9T3-7. In addition, we screened 10^6 clones from a commercial human lymphoblastoid T-cell cDNA library (Clontech), which was constructed by using random primers plus oligo(dT) and had an average insert size of 2.3 kb. Again, no positive clones were seen, suggesting that the 5′ end of *XRCC1* mRNA may be inefficiently copied by reverse transcriptase.

DISCUSSION

This study describes the first isolation and characterization of a human gene involved in repairing DNA strand breaks produced by ionizing radiation and alkylating agents. The recovery of a nearly full-length cDNA clone should provide the necessary sequence for isolating the XRCC1 protein from an overexpression system. The protein can then be studied for its role in repair and SCE formation (44). Also, cloning of the mouse *XRCC1* gene would provide a tool for making mutations by targeted recombination in mouse embryonic stem cells (31), which might then be used to construct transgenic mice that are defective in the mouse *XRCC1* locus.

Technically, the genomic cloning of *XRCC1* was relatively straightforward. First, characterization of genomic transformants and recovery of the gene from the cosmid library were aided by the presence of multiple repetitive elements within the gene. Second, because the entire gene was favorably positioned within two different cosmid inserts, it could be readily identified by its biological activity upon DNA transfection. The 33-kb size for *XRCC1* is larger than those of the first two cloned human repair genes, *ERCC1* and *ERCC2*, which are 15 and 19 kb, respectively (53, 58). The genes *ERCC3* and *ERCC5*, which are very similar in size to *XRCC1*, were partially isolated by using cosmids but were not recovered entirely in individual clones (29, 59).

The longest cDNA clone (pXR1-30) that we were able to isolate was truncated at the 5′ end. Therefore, analysis of genomic nucleotide sequence was required to identify the 5′ end of the coding region of the gene. In comparing the length of the cDNA insert with the size of the *XRCC1* mRNA in

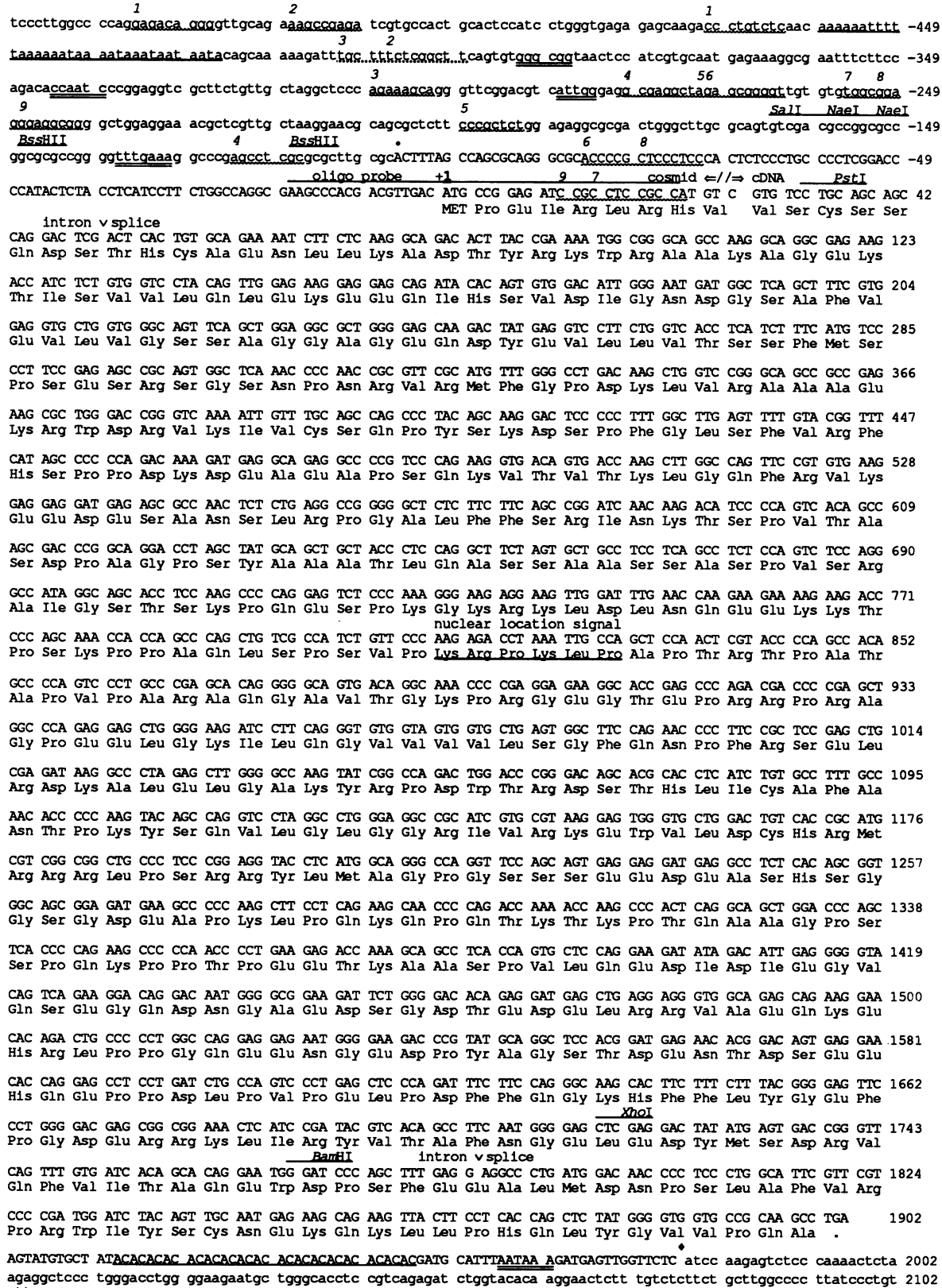

FIG. 5. Composite nucleotide sequence of the pXR1-30 cDNA insert, 5′- and 3′-flanking regions derived from cosmid pH9T3-7, and the amino acid translation of the ORF. The transcribed region is shown in uppercase letters, beginning at the most likely transcriptional initiation site, marked (●) at position −105. Position +1 is the translational initiation site. The cDNA is missing 26 nucleotides of translated sequence.

HeLa cells, we concluded that the insert was missing ≈100 bp of 5′-end sequence. Using both primer extension analysis and S1 nuclease protection experiments, we attempted to determine the transcriptional start site(s). While the results were not definitive, they suggested a preferred site at −105 from the translation initiation site (Fig. 5), which had the appropriate nucleotide sequence for an mRNA start site (6) and was consistent with the calculated length of missing cDNA sequence based on mRNA size. This preferred site is also appropriately positioned with respect to the putative TATA box (6). The ORF identified extended upstream beyond the cDNA 5′ end to position −105. The first in-frame Met codon in the ORF (position +1) was analyzed by the method of Stormo (43) in terms of whether its neighboring nucleotides favor a functional start site. Using Stormo's nucleotide frequency matrix, we calculated a score of 25 on a scale that ranges from −45 to 91. On this scale, the human repair genes *ERCC1* and *ERCC2* score 34 and 54, respectively. Taken together, the above results support the idea that we have correctly identified the start position for XRCC1 protein and that the missing portion of our cDNA insert includes 26 nucleotides of translated sequence. Thus, the translated ORF of 1,899 nucleotides encodes a 69.5-kDa protein of 633 amino acids having a pI value of 5.96. The basic charge on the protein could facilitate binding to DNA. The G+C content of the ORF was 59.2%, similar to that of *ERCC2* (58).

Exhaustive screening of two cDNA libraries with an oligonucleotide probe (Fig. 5) failed to recover clones containing the 5′ sequence encompassing the ATG start codon. As an alternative approach to obtaining the entire protein-coding region for purposes of protein isolation, it should be possible to extend our cDNA insert by adding synthetic oligonucleotides to provide the missing protein-coding sequence.

The 5′- and 3′-flanking regions of *XRCC1* contain several interesting features. The promoter region contains a TATA box, a GC box, and two possible CAAT boxes. The unique *Sal*I site of cosmid pH9T3-7 is located just 5′ of the TATA box. When the cosmid was cut at this site, its transfection activity was abolished, pointing to essential upstream elements, which might include the 34-bp AT tract beginning at −348 and the nine short inverted repeats that are spread over ≈600 bp, as well as the GC and CAAT boxes. At the 3′ end of XRCC1 there is a repeating AC element, which is often associated with the 3′ end of mRNAs (54). A potential nuclear location signal, KRPKLP, which is underlined in Fig. 5, is very similar to a sequence (KRGKLP) identified in another human repair gene, *ERCC2* (58). Both sequences resemble nuclear targeting sequences identified in c-*myc* genes, which have a consensus of KR—KL— (12).

The cDNA clone pXR1-30 gave ≈80% correction of EM9 cells, as measured by the level of SCEs. This incomplete correction is not surprising in view of the breakpoint occurring 3′ of the normal ATG start codon. In Fig. 5, the second in-frame methionine codon occurs at +280 and would produce a highly truncated protein if used to initiate translation.

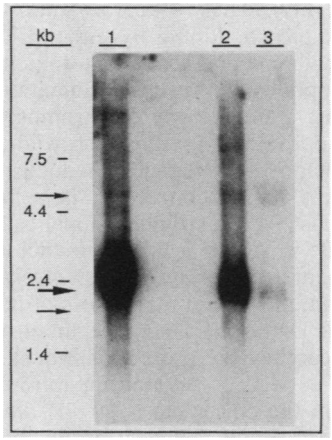

FIG. 6. Northern analysis of the *XRCC1* gene. The probe was the 1.9-kb *Bam*HI fragment from clone pXR1-30 (Fig. 4B). The autoradiogram was exposed for 18 h at −80°C with an intensifying screen. Lane 1, 5 μg of poly(A)$^+$ RNA from the EM9 transformant 9CD30-1; lane 2, 5 μg of poly(A)$^+$ RNA from HeLa cells; lane 3, 10 μg of total RNA from HeLa cells. The large arrow shows the position of the *XRCC1* mRNA band in lane 2, and the small arrows show the positions of the rRNA bands. In lane 3, the upper band is due to background 28S rRNA; the lower band represents *XRCC1* mRNA and lies above the position of 18S rRNA.

Alternatively, the pcD2 vector contains a potential initiation site. In the vector there are two splice acceptor sites downstream of the SV40 early promoter (35). Utilization of the 19S acceptor site results in a transcript containing an ATG triplet that might produce an in-frame fusion protein having nine XRCC1 amino acids replaced by ≈30 abnormal ones in its N-terminus. For this ATG codon to be in-frame with the ORF of pXR1-30, heterogeneity may be required in the length of the poly(G) tract that links the pcD2 vector with the cDNA insert. The length of the poly(G) tract was seen to vary from ≈15 to ≈25 residues in the different M13 phage clones used for sequencing. Thus, overexpression of an abnormal protein from the SV40 promoter appears to be responsible for the partial correction that occurs.

The human *XRCC1* gene in cosmids seems to function with essentially 100% efficiency in the transformed EM9 hamster cells in terms of correcting the repair defect. Two different functional cosmid clones produced transformants that almost always showed full correction by the various endpoints studied: cell survival after gamma irradiation, kinetics of repair of single-strand breaks induced by gamma rays, level of SCEs due to BrdUrd incorporation, and frequency of chromosomal aberrations due to BrdUrd. Although it might be argued that efficient interspecific correction by the heterologous gene could be due to the presence of multiple copies in cosmid transformants, overexpression of *XRCC1* does not seem to be necessary for efficient correction. Several primary genomic transformants described earlier, which should have a single copy of *XRCC1*, had normal

Double underlining marks the putative TATA box beginning at −136, a GC box at −391, possible CAAT boxes at −343 and −287 (reverse orientation), and a polyadenylation consensus sequence in the 3′ end. The polyadenylation site is indicated (◆). Single underlining marks a 34-nucleotide AT tract in the 5′ region and a 34-nucleotide AC repeat preceding the polyadenylation signal. Wavy underlining marks a series of nine inverted repeats in the 5′ end of gene, each 8 or 9 nucleotides in length, which are numbered 1 to 9 in italic numbers above the first base of each sequence. Restriction endonuclease sites within the coding region used for cloning and sequencing are marked, as are the *Sal*I and other sites related to the CpG island in the putative promoter region. The underlined amino acids (KRPKLP) represent a possible nuclear location signal. The overlined region spanning the +1 position represents the oligonucleotide that was used for cDNA library screening. In the last line of the sequence, s indicates either C or G. The positions of the first and last intron splice sites are indicated.

levels of SCE (49). These transformants also had normal levels of resistance to killing by either CldUrd or methyl methanesulfonate in a differential cytotoxicity assay (20) (L. Thompson, unpublished results). Somatic-cell hybrids of EM9 containing a single copy of chromosome 19 also had SCE levels approximating those of wild-type cells (40). Thus, we conclude that the genetic locus represented by *XRCC1* is likely highly conserved, in a functional sense, among mammals. At the structural level of DNA sequence, high homology is supported by the fact that the 1.9-kb *Bam*HI fragment from the pXR1-30 cDNA clone detects restriction fragments in Southern blots of both hamster and mouse genomic DNA (46). Using the amino acid sequence of XRCC1, we searched the protein data banks (see Materials and Methods) as well as several unpublished RAD protein sequences from the eucaryote *S. cerevisiae*. No significant homologies were found. Thus, if an analogous yeast gene exists, it may not yet have been identified. The biochemical defect in EM9 remains unknown. All enzymes assayed so far, including DNA ligases, have appeared to be normal (44).

XRCC1 is the first identified of many genes that affect cell sensitivity to ionizing radiation and simple alkylating agents by playing a role in the repair of strand breaks. The defect in EM9 cells affects single-strand break repair (Fig. 1) of X-ray or chemically induced damage (47) with little or no effect on the capacity for repairing double-strand breaks (52). In Chinese hamster cells, a number of mutations involving strand break repair have been partially characterized, and they can be distinguished by whether the defect involves mostly single- or double-strand break repair upon exposure to ionizing radiation. For example, in CHO cells the mutant irs1SF is ca. twofold deficient in repairing single-strand breaks (16), and BLM-2 (39) is reduced ca. twofold in both single-strand and double-strand break repair. Like EM9, both of these mutants are about twofold more sensitive to killing by ionizing radiation than the parent strain (16, 38). In contrast, the CHO mutants XR-1 (42) and *xrs5* (21) are ca. sevenfold more sensitive to killing by ionizing radiation than the parent strain and have pronounced deficiencies in the repair of double-strand breaks (17, 26). In hamster V79 cells, the mutations *irs1*, *irs2*, and *irs3* (22) are each different (23), and the mutant V-15B belongs to the *xrs5* complementation group (62). *irs1* and *irs2* mutants are not defective in the quantitative efficiency of single- or double-strand break repair as measured in bulk DNA (24), but the *irs1* mutant does show a defect in the fidelity of repair in the vector-mediated double-strand break rejoining assay (13). Among these mutants, six complementation groups were reported (23), but additional ones will undoubtedly be identified. The cloned genes underlying these complementation groups, along with *XRCC1* as reported here, will become tools for determining whether radiation repair genes are inducible by DNA damage or play a role in the radiation resistance of tumor cells, as well as providing a means of purifying the repair proteins.

ACKNOWLEDGMENTS

We thank Sheri Stewart, Collin Collins, and Jason Minkler for technical assistance, Hiroto Okayama for making available the human pcD2 cDNA library, Mark MacInnes for providing information on restriction enzyme sites at the 5′ end of *XRCC1*, Robert Mortimer for making available unpublished sequences of the *S. cerevisiae RAD* genes, and Christine Weber for valuable advice, discussion, and reading of the manuscript.

This work was performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under contract W-7405-ENG-48.

LITERATURE CITED

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl.** 1988. Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
2. **Bardwell, L.** 1989. The mutagenic and carcinogenic effects of gene transfer. Mutagenesis **4**:245–253.
3. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
4. **Birnstiel, M. L., M. Busslinger, and K. Strub.** 1985. Transcription termination and 3′ processing: the end is in site! Cell **41**:349–359.
5. **Boyd, J. B., J. M. Mason, A. H. Yamamoto, R. K. Brodberg, S. A. Banga, and K. Sakaguchi.** 1987. A genetic and molecular analysis of DNA repair in *Drosophila*. J. Cell Sci. Suppl. **6**:39–60.
6. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
7. **Brown, W. M.** 1980. Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc. Natl. Acad. Sci. USA **77**:3605–3609.
8. **Carrano, A. V., J. L. Minkler, L. E. Dillehay, and L. H. Thompson.** 1986. Incorporated bromodeoxyuridine enhances the sister-chromatid exchange and chromosomal aberration frequencies in an EMS-sensitive Chinese hamster cell line. Mutat. Res. **162**:233–239.
9. **Chen, C. A., and H. Okayama.** 1988. High-efficiency transformation of mammalian cells by plasmid DNA. Mol. Cell. Biol. **7**:2745–2752.
10. **Corsaro, C. M., and M. L. Pearson.** 1981. Enhancing the efficiency of DNA-mediated gene transfer in mammalian cells. Somat. Cell Genet. **7**:603–616.
11. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18 S rDNA. Plasmid **13**:31–40.
12. **Dang, C. V., and W. M. F. Lee.** 1988. Identification of the human c-*myc* protein nuclear translocation signal. Mol. Cell. Biol. **8**:4048–4054.
13. **Debenham, P. G., N. J. Jones, and M. B. T. Webb.** 1988. Vector-mediated DNA double-strand break analysis in normal and radiation-sensitive Chinese hamster wild-type cells. Mutat. Res. **199**:1–9.
14. **Dillehay, L. E., L. H. Thompson, and A. V. Carrano.** 1984. DNA-strand breaks associated with halogenated pyrimidine incorporation. Mutat. Res. **131**:129–136.
15. **Friedberg, E. C.** 1988. Deoxyribonucleic acid repair in the yeast *Saccharomyces cerevisiae*. Microbiol. Rev. **52**:70–102.
16. **Fuller, L. F., and R. B. Painter.** 1988. A Chinese hamster ovary cell line hypersensitive to ionizing radiation and deficient in repair replication. Mutat. Res. **193**:109–121.
17. **Giaccia, R., R. Weinstein, J. Hu, and T. D. Stamato.** 1985. Cell cycle-dependent repair of double-strand DNA breaks in a gamma-ray-sensitive Chinese hamster cell. Somat. Cell Mol. Genet. **11**:485–491.
18. **Hoeijmakers, J. H. J., M. van Duin, G. Weeda, A. J. Van der Eb, C. Troelstra, A. P. M. Eker, N. G. J. Jaspers, A. Westerveld, and D. Bootsma.** 1988. Analysis of mammalian excision repair: from mutants to genes and gene products, p. 281–287. *In* E. C. Friedberg and P. C. Hanawalt (ed.), Mechanisms and consequences of DNA damage processing. Alan R. Liss, New York.
19. **Hoy, C. A., J. C. Fuscoe, and L. H. Thompson.** 1987. Recombination and ligation of transfected DNA in CHO mutant EM9, which has high levels of sister chromatid exchange. Mol. Cell. Biol. **7**:2007–2011.
20. **Hoy, C. A., E. P. Salazar, and L. H. Thompson.** 1984. Rapid detection of DNA-damaging agents using repair-deficient CHO cells. Mutat. Res. **130**:321–332.
21. **Jeggo, P. A.** 1983. X-ray-sensitive mutants of the CHO cell line. Isolation and cross-sensitivity to other DNA-damaging agents. Mutat. Res. **112**:313–327.
22. **Jones, N. J., R. Cox, and J. Thacker.** 1987. Isolation and

cross-sensitivity of X-ray-sensitive mutants of V79-4 hamster cells. Mutat. Res. **183**:279–286.
23. **Jones, N. J., R. Cox, and J. Thacker.** 1988. Six complementation groups for ionizing-radiation sensitivity in Chinese hamster cells. Mutat. Res. **193**:139–144.
24. **Jones, N. J., S. A. Stewart, and L. H. Thompson.** 1990. Biochemical and genetic analysis of the Chinese hamster mutants *irs1* and *irs2* and their comparison to cultured ataxia telangiectasia cells. Mutagenesis **5**:15–23.
25. **Karn, J., H. W. D. Matthes, M. J. Gait, and S. Brenner.** 1984. A new selective phage cloning vector, 2001, with sites for *Xba*I, *Bam*HI, *Hin*dIII, *Eco*RI, *Sst*I and *Xho*I. Gene **32**:217–224.
26. **Kemp, L. M., S. G. Sedgwick, and P. Jeggo.** 1984. X-ray-sensitive mutants of CHO cells defective in double-strand break rejoining. Mutat. Res. **132**:189–196.
27. **Kohn, K. W., R. A. G. Ewig, L. C. Erickson, and L. A. Zwelling.** 1981. Measurement of strand breaks and cross-links by alkaline elution, p. 379–401. *In* E. C. Friedberg and P. C. Hanawalt (ed.), DNA repair: a laboratory manual of research procedures, vol. 1, part B. Marcel Dekker, Inc., New York.
28. **Lewin, B.** 1987. Genes III. John Wiley & Sons, Inc., New York.
29. **MacInnes, M. A., and J. S. Mudgett.** 1990. Cloning of the functional human excision repair gene ERCC-5: potential gene regulatory features conserved with other human repair genes, p. 265–274. *In* M. L. Mendelsohn and R. J. Albertini (ed.), Mutation and the environment, part A: progress in clinical and biological research, Vol. 340. Wiley-Liss, New York.
30. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **Mansour, S. L., K. R. Thomas, and M. R. Capecchi.** 1988. Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. Nature (London) **336**:348–352.
32. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
33. **Mohrenweiser, H. W., A. V. Carrano, A. Fertitta, B. Perry, L. H. Thompson, J. D. Tucker, and C. A. Weber.** 1989. Refined mapping of the three DNA repair genes, *ERCC1*, *ERCC2*, and *XRCC1*, on human chromosome 19. Cytogenet. Cell Genet. **52**:11–14.
34. **Myles, G. M., and A. Sancar.** 1989. DNA repair. Chem. Res. Toxiocol. **2**:197–226.
35. **Okayama, H., and P. Berg.** 1983. A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. Mol. Cell. Biol. **3**:280–289.
36. **Pinkel, D., L. H. Thompson, J. W. Gray, and M. Vanderlaan.** 1985. Measurement of sister chromatid exchanges at very low bromodeoxyuridine substitution levels using a monoclonal antibody in Chinese hamster ovary cells. Cancer Res. **45**:5795–5798.
37. **Potter, H., and D. Dressler.** 1986. A "Southern cross" method for the analysis of genome organization and the localization of transcription units. Gene **48**:229–239.
38. **Robson, C. N., A. L. Harris, and I. D. Hickson.** 1985. Isolation and characterization of Chinese hamster ovary cell lines sensitive to mitomycin C and bleomycin. Cancer Res. **45**:5304–5309.
39. **Robson, C. N., A. L. Harris, and I. D. Hickson.** 1989. Defective repair of DNA single- and double-strand breaks in the bleomycin- and X-ray-sensitive Chinese hamster ovary cell mutant BLM-2. Mutat. Res. **217**:93–100.
40. **Siciliano, M. J., A. V. Carrano, and L. H. Thompson.** 1986. Assignment of a human DNA-repair gene associated with sister-chromatid exchange to chromosome 19. Mutat. Res. **174**:303–308.
41. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
42. **Stamato, T. D., R. Weinstein, A. Giaccia, and L. Mackenzie.** 1983. Isolation of a cell cycle-dependent gamma-ray-sensitive Chinese hamster ovary cell. Somat. Cell Genet. **9**:165–173.
43. **Stormo, G. D.** 1987. Identifying coding sequences, p. 231–258. *In* M. J. Bishop and C. J. Rawlings (ed.), Nucleic acid and protein sequence analysis. IRL Press, Oxford.
44. **Thompson, L. H.** 1988. Mammalian cell mutations affecting recombination, p. 597–620. *In* R. Kucherlapati and G. R. Smith (ed.), Genetic recombination. American Society for Microbiology, Washington, D.C.
45. **Thompson, L. H.** 1989. Somatic cell genetics approach to dissecting mammalian DNA repair. Environ. Mol. Mutagenesis **14**:264–281.
46. **Thompson, L. H., L. L. Bachinski, R. L. Stallings, G. Dolf, C. A. Weber, A. Westerveld, and M. J. Siciliano.** 1989. Complementation of repair gene mutations on the hemizygous chromosome 9 in CHO: a third repair gene on human chromosome 19. Genomics **5**:670–679.
47. **Thompson, L. H., K. W. Brookman, L. E. Dillehay, A. V. Carrano, J. A. Mazrimas, C. L. Mooney, and J. L. Minkler.** 1982. A CHO-cell strain having hypersensitivity to mutagens, a defect in strand-break repair, and an extraordinary baseline frequency of sister chromatid exchange. Mutat. Res. **95**:247–440.
48. **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.
49. **Thompson, L. H., K. W. Brookman, J. L. Minkler, J. C. Fuscoe, K. A. Henning, and A. V. Carrano.** 1985. DNA-mediated transfer of a human DNA repair gene that controls sister chromatid exchange. Mol. Cell. Biol. **5**:881–884.
50. **Thompson, L. H., S. Fong, and K. Brookman.** 1980. Validation of conditions for efficient detection of HPRT and APRT mutations in suspension-cultured Chinese hamster ovary cells. Mutat. Res. **74**:21–36.
51. **Thompson, L. H., E. P. Salazar, K. W. Brookman, C. C. Collins, S. A. Stewart, D. B. Busch, and C. A. Weber.** 1987. Recent progress with the DNA repair mutants of Chinese hamster ovary cells. J. Cell Sci. Suppl. **6**:97–110.
52. **vanAnkeren, S. C., D. Murray, and R. E. Meyn.** 1988. Induction and rejoining of γ-ray-induced DNA single- and double-strand breaks in Chinese hamster AA8 cells and in two radiosensitive clones. Radiat. Res. **116**:511–525.
53. **van Duin, M., J. de Wit, H. Odijk, A. Westerveld, A. Yasui, M. H. M. Koken, J. Hoeijmakers, and D. Bootsma.** 1986. Molecular characterization of the human excision repair gene *ERCC-1*: cDNA cloning and amino acid homology with the yeast DNA repair gene *RAD10*. Cell **44**:913–923.
54. **von Heijne, G.** 1987. Sequence analysis in molecular biology. Academic Press, Inc., New York.
55. **Walker, G. C.** 1985. Inducible DNA repair systems. Annu. Rev. Biochem. **54**:425–457.
56. **Walker, G. C., L. Marsh, and L. A. Dobson.** 1985. Genetic analysis of DNA repair: inference and extrapolation. Annu. Rev. Genet. **19**:103–126.
57. **Weber, C. A., E. P. Salazar, S. A. Stewart, and L. H. Thompson.** 1988. Molecular cloning and biological characterization of a human gene, *ERCC2*, that corrects the nucleotide excision repair defect in CHO UV5 cells. Mol. Cell. Biol. **8**:1137–1146.
58. **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.
59. **Weeda, G., R. C. A. van Ham, R. Masurel, A. Westerveld, H. Odijk, J. de Wit, D. Bootsma, A. J. Van der Eb, and J. H. J. Hoeijmakers.** 1990. Molecular cloning and biological characterization of the human excision repair gene *ERCC-3*. Mol. Cell. Biol. **10**:2570–2581.
60. **Westerveld, A., J. H. J. Hoeijmakers, M. van Duin, J. de Wit, H. Odijk, A. Pastink, R. D. Wood, and D. Bootsma.** 1984. Molecular cloning of a human DNA repair gene. Nature (London) **310**:425–429.
61. **Wood, R. D., and H. J. Burki.** 1982. Repair capacity and the cellular age response for killing and mutation induction after UV. Mutat. Res. **95**:505–514.
62. **Zdzienicka, M. Z., Q. Tran, G. P. van der Shans, and J. W. I. M. Simons.** 1988. Characterization of an X-ray-hypersensitive mutant of V79 Chinese hamster cells. Mutat. Res. **194**:239–249.