The EMBO Journal vol.9 no.5 pp.1437–1447, 1990

ERCC2: cDNA cloning and molecular characterization of a human nucleotide excision repair gene with high homology to yeast *RAD3*

Christine A.Weber, Edmund P.Salazar, Sheri A.Stewart and Larry H.Thompson

Biomedical Sciences Division, Lawrence Livermore National Laboratory, PO Box 5507, Livermore, CA 94550, USA

Communicated by D.Bootsma

Human *ERCC2* genomic clones give efficient, stable correction of the nucleotide excision repair defect in UV5 Chinese hamster ovary cells. One clone having a breakpoint just 5′ of classical promoter elements corrects only transiently, implicating further flanking sequences in stable gene expression. The nucleotide sequences of a cDNA clone and genomic flanking regions were determined. The *ERCC2* translated amino acid sequence has 52% identity (73% homology) with the yeast nucleotide excision repair protein RAD3. *RAD3* is essential for cell viability and encodes a protein that is a single-stranded DNA dependent ATPase and an ATP dependent helicase. The similarity of ERCC2 and RAD3 suggests a role for ERCC2 in both cell viability and DNA repair and provides the first insight into the biochemical function of a mammalian nucleotide excision repair gene.

Key words: chromosome 19/DNA repair/gene cloning/nucleotide excision repair

Introduction

DNA repair is a complex biochemical process by which cells attempt to preserve genetic information from the damaging effects of exposure to environmental radiations and chemicals, as well as from spontaneous errors in recombination and replication. The nucleotide excision repair pathway is a major component of this process, removing a wide variety of UV induced photoproducts and bulky chemical adducts. The genetic basis of the nucleotide excision repair process in different organisms has been the focus of much research (reviewed in Rubin, 1988). This pathway has been extensively studied and is best understood in *Escherichia coli*, where five genes encode the proteins that are involved in this repair process (*uvrA*, *uvrB*, *uvrC*, *uvrD* (helicase II), and *polA*; reviewed in Grossman *et al.*, 1988). Eukaryotic systems are more complex and less well characterized. In *Saccharomyces cerevisiae*, 10 genes have been identified as having a role in nucleotide excision repair (known as the *RAD3* epistasis group of which the *RAD3* gene is a member; Friedberg, 1985).

Cells from patients with the classical form of the cancer-prone genetic disorder xeroderma pigmentosum (XP) are defective in performing the incision step of nucleotide excision repair (Cleaver, 1983). Cell fusion studies have identified seven XP complementation groups (Bootsma *et al.*, 1989; Johnson *et al.*, 1989), indicating that at least seven genes may be involved in damage recognition and incision. Because of difficulties in using these human cells to study the genetics of this process, UV sensitive rodent mutants have been isolated in many laboratories (Thompson *et al.*, 1980b; Wood and Burki, 1982; Zdzienicka and Simons, 1987; Busch *et al.*, 1980, 1989; reviewed in Collins and Johnson, 1987; Hickson and Harris, 1988). Among these mutants, eight rodent UV complementation groups have been identified (Thompson *et al.*, 1981; Thompson and Carrano, 1983; Thompson *et al.*, 1987; Zdzienicka *et al.*, 1988a; Thompson *et al.*, 1988). The first five rodent complementation groups are, like XP, defective in the incision step (Thompson *et al.*, 1982), while the sixth group has a defect in the removal of UV induced cyclobutane dimers but not (6–4) pyrimidine–pyrimidone photoproducts (Thompson *et al.*, 1989b).

The cloning of human genes that efficiently and specifically correct the defect in rodent complementation groups 1 and 2 (*ERCC1* and *ERCC2*; Excision Repair Cross Complementing rodent complementation groups 1 and 2, respectively) have been reported (Westerveld *et al.*, 1984; van Duin *et al.*, 1988a; Weber *et al.*, 1989). [The numbering of complementation groups 1 and 2 was recently interchanged to correspond to the numbers of the complementing human gene locus names (Thompson and Bootsma, 1988).] Both *ERCC1* and *ERCC2* have been localized to within 250 kb of each other (Mohrenweiser *et al.*, 1989) on human chromosome 19 bands q13.2–13.3 (Brook *et al.*, 1985; Thompson *et al.*, 1985b; van Duin *et al.*, 1986; Mohrenweiser *et al.*, 1989; Thompson *et al.*, 1989a). Recently, progress has been made by several laboratories in the cloning of human genes correcting other complementation groups (reviewed in Thompson, 1989).

The question of correspondence between the rodent UV complementation groups and the XP complementation groups is important. Cell fusion studies have not revealed overlap but have also yielded very few hybrids, so only a few of the numerous possible crosses have been successfully tested (Thompson *et al.*, 1985a). Complementation analysis with genomic clones of human nucleotide excision repair genes is also limited because XP cells do not take up and integrate large pieces of DNA (Hoeijmakers *et al.*, 1987). However, *ERCC1* has been thoroughly tested and fails to correct XP complementation groups A through I (van Duin *et al.*, 1989). This failure was not unexpected as rodent UV complementation group 1 (as well as group 4) exhibits extreme sensitivity to cross-linking agents such as mitomycin C, a characteristic associated with Fanconi anemia cells rather than XP. Rodent complementation groups 2, 3 and 5 closely parallel XP cells phenotypically.

We report here the cDNA cloning and molecular characterization of the human *ERCC2* gene. We have determined the nucleotide sequence for the largest cDNA insert (from a library in a mammalian expression vector) as well as the 5′- and 3′-flanking regions of genomic clones. Analysis and biological characterization of the transient correction seen in Chinese hamster ovary (CHO) mutant cells

© Oxford University Press

transformed with an atypical genomic cosmid clone having a breakpoint near the 5′-end of the gene are presented. Comparison of the translated amino acid sequence of *ERCC2* with yeast nucleotide excision repair genes revealed extensive homology with RAD3.

Results

cDNA clones: library screening, restriction enzyme site mapping and testing for complementation of UV5 cells

A restriction enzyme site map of *ERCC2* and flanking pSV2gpt sequences in genomic cosmid clones is presented in Figure 1A (adapted from Weber *et al.*, 1988). The pcD2 library, a human fibroblast cDNA library in a mammalian expression vector with a *neo* selectable marker, was screened for *ERCC2* cDNA clones. Probes were the 12 kb *Eco*RI and the 8.3 kb *Eco*RI−*Xho*I fragments from the *ERCC2* region of the cosmid p5T4-1-28 (Figure 1B), which were pre-reassociated with human placental DNA to reduce signal from repetitive sequences. Five positive colonies were obtained from screening 1.1×10^6 colonies with the 12 kb probe (clones pER2-6, 8, 10, 11 and 12), and four positive colonies were obtained from screening 1.9×10^6 colonies with the 8.3 kb probe (clones pER2-14, 16, 17 and 18). *Bam*HI restriction enzyme site maps were generated for each cDNA clone and demonstrated six size classes of insert DNA (pER2-6, 11 = 1.2 kb; 16 = 1.35 kb; 10, 12 = 1.45 kb; 8 = 1.8 kb, additional *Bam*HI site; 17, 18 = 2.05 kb; 14 = 2.6 kb). A detailed restriction enzyme site map of the pER2-14 insert is presented in Figure 2.

Southern blot hybridization analyses indicated that pER2-8 likely contains an unprocessed intron and that both pER2-14 and 17 extend into the 2.1 kb *Xho*I−*Eco*RI fragment from the 5′-end of *ERCC2* (Figure 1B; data not shown). All nine cDNA clones were tested for the ability to confer UV resistance to UV5 cells, and all but pER2-14 were negative. Thus, it appeared that pER2-14 was likely a nearly complete cDNA. Therefore, this clone was examined further for biological activity, and the nucleotide sequence of the insert was determined.

UV survival studies of pER2-14 transformed UV5 cells

The cDNA clone, pER2-14, was found to confer UV resistance to the repair-deficient UV5 CHO cell line in a transient manner. Individual colonies that survived one or two 5 J/m^2 UV exposures were grown to mass culture under conditions selective for the *neo* marker and tested for UV resistance (Figure 3A). The UV survival curves for these cultures were very similar to that for the sensitive mutant, indicating a loss of repair function, but not marker function. Testing individual colonies for UV resistance without further growth indicated the loss of repair function occurred before growth to mass culture. The continued growth in selective medium suggests that gene loss is not the explanation of the transient UV resistance.

Northern analysis confirmed that a pER2-14 transformed UV5 line, 5ER2-14-3, produces RNA of an appropriate size from the pER2-14 *ERCC2* gene at a very high level (unpublished data). Therefore, the loss of repair function does not reflect a loss of gene expression. Transformation of the UV sensitive 5ER2-14-3 cells with an

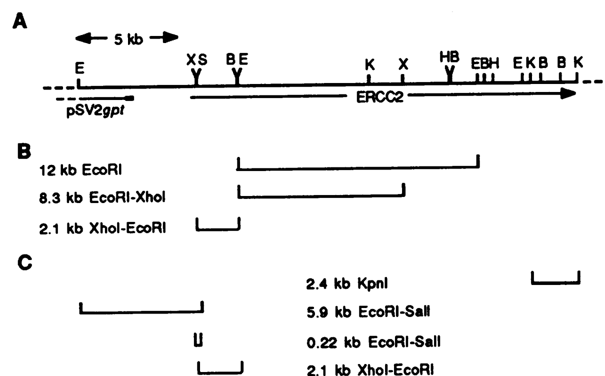

Fig. 1. A restriction enzyme site map of the genomic region of *ERCC2* present in cosmid clones is diagrammed in **A**. Dashed lines indicate continuation of cloned regions (see map in Weber *et al.*, 1988). The thickened portion of the pSV2gpt region indicates the minimum to maximum extent. **B** and **C** show the fragments used as hybridization probes and the fragments subcloned for nucleotide sequence determination, respectively. In C, the 0.22 kb *Eco*RI−*Sal*I fragment is from cosmid p5T4-1-15 (the *Eco*RI site is from the vector). Restriction enzyme abbreviations in A are as follows: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; S, *Sal*I; and X, *Xho*I.

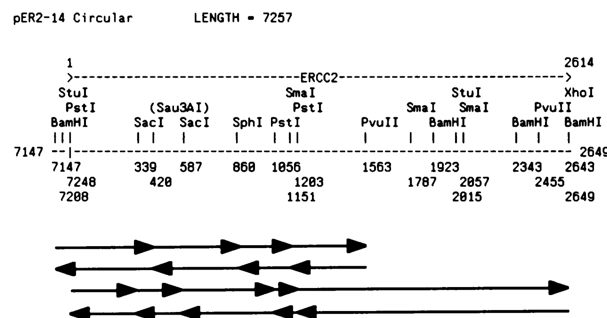

Fig. 2. A restriction enzyme site map of the pER2-14 insert and a schematic of the fragments subcloned for nucleotide sequence determination are shown. The *ERCC2* insert is from position 1 through 2614, including both the 16 bp dG·dC and 40 bp dA·dT tails.

ERCC2-containing cosmid, p5T4-1-24, restored UV resistance (data not shown). This result indicates the UV sensitivity observed in transformations with the cDNA clone is not due to overexpression of the *ERCC2* gene product, which would, if it were the cause of the UV sensitivity, result in a dominant UV sensitive phenotype.

The results thus far suggest that the pER2-14 transformants are retaining the transformed sequences, transcribing the expected RNA, but are not producing a fully functional ERCC2 protein. The level of UV resistance conferred at early stages was tested by irradiating cells 24 h after transformation, along with selection for the *neo* marker (measured as survival within the fraction of cells expressing the selectable marker; Figure 3B). Given that the *neo* gene is linked to the repair gene, it appears that pER2-14 confers an intermediate level of UV resistance at an early stage. This transient, partial UV resistance can be explained in terms of the truncated ERCC2 protein(s) produced (see below).

UV survival studies of p5T4-1-15 transformed UV5 cells

Cosmid p5T4-1-15 was used to delineate the 5′-end of

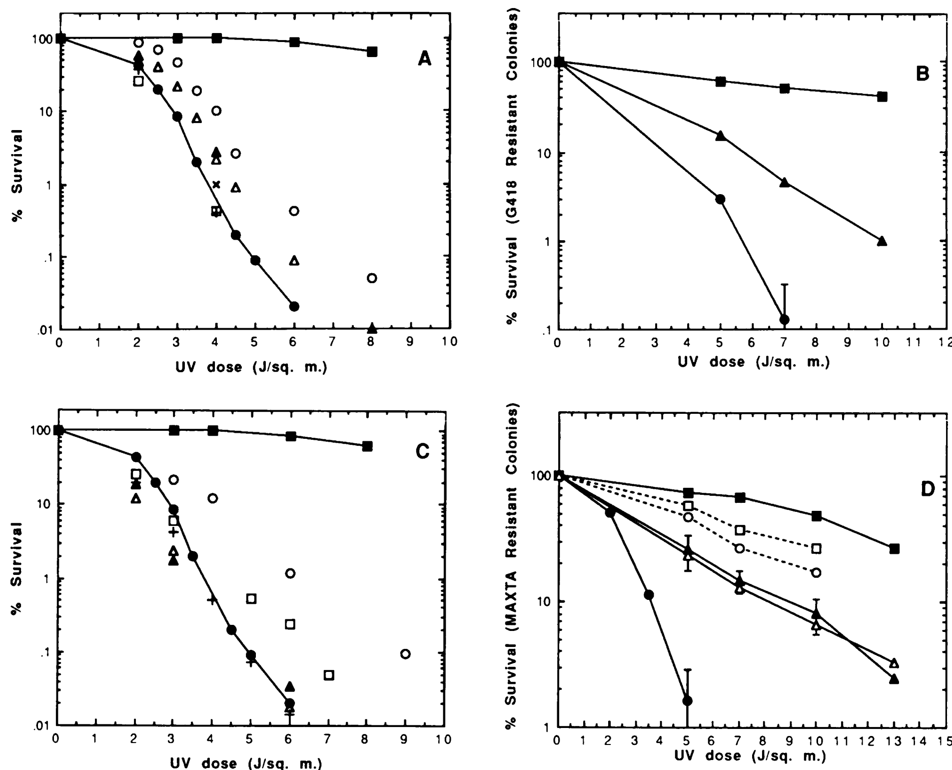

Fig. 3. UV survival curves of colonal isolates of cDNA plasmid pER2-14 transformed (**A**) or genomic cosmid p5T4-1-15 transformed (**C**) UV5 cells (scatter plots) along with wild-type AA8 and mutant UV5 cells (solid lines) are presented. In A, measurements are from one or two experiments, and in C, measurements are from two or three experiments (in most cases, each dose is represented in only one trial). UV survival curves of selectable marker-expressing cells at 24 h after DNA exposure for cDNA transformed (**B**) or cosmid transformed (**D**) UV5 cells along with wild-type AA8 and mutant UV5 cells transformed with selectable markers only are presented. In B, measurements are from one experiment. An error bar representing the Poisson counting error is placed on the point where few colonies were obtained. In D, measurements are from one to three experiments. Error bars represent the standard error of the mean where the point is an average. The standard errors of the mean for the AA8 points are small, and error bars would be within the region of the symbol. Symbols (transforming DNAs in parentheses): (**A**) ■, AA8; ●, UV5; ▲, 5ER2-14-1; □, 5ER2-14-2; ○, 5ER2-14-3; △, 5ER2-14-4; +, 5ER2-14-5; ×, 5ER2-14-6; (**B**) ■. AA8 (*neo*); ●, UV5 (*neo*); ▲, UV5 (pER2-14); (*neo* = pXR1-16, a partial *XRCC1* cDNA clone as negative control); (**C**) ■, AA8; ●, UV5; ▲, 5C15-5; □, 5C15-7; ○, 5C15-13; △, 5C15-15; +, 5C15-16; (**D**) ■, AA8 (pSV2gpt); ●, UV5 (pSV2gpt); ▲, UV5 (p5T4-1-15 + pSV2gpt); □, UV5 (p5T4-1-24); ○, UV5 (p5T4-1-24 + pSV2gpt); △, UV5 (p5T4-1-6 + pSV2gpt).

ERCC2 (Weber *et al.*, 1988) because this cosmid, having the closest 5′-breakpoint, conferred UV resistance to UV5 cells with the same frequency as other genomic cosmid clones. However, in contrast to transformants with genomic cosmid clones extending further 5′ that show stable, wild-type levels of survival (Weber *et al.*, 1988), the UV resistance conferred by p5T4-1-15 is transient. The UV survival curves of the p5T4-1-15 transformant cultures (obtained as above but with *gpt* selection) were very similar to that for the sensitive mutant, indicating a loss of repair function (Figure 3C). These cultures had a plating efficiency in MAXTA medium of 80–100% (except 5C15-5 was 40%), indicating that the UV sensitivity was probably not due to instability of the transformed sequences. Again, loss of UV resistance occurred before growth to mass culture and transformation with p5T4-1-24 demonstrated that the UV sensitivity was not due to over-expression of the *ERCC2* gene product (data not shown).

The results thus far, combined with the absence of nucleotide sequence alterations in the 5′-flanking region (see below), suggest that cosmid p5T4-1-15 lacks some 5′-sequences necessary for stable gene expression. The level of UV resistance conferred at early stages was tested (Figure 3D). Cosmids p5T4-1-15 and p5T4-1-6 (both cotransferred with pSV2gpt) conferred similar levels of UV resistance (p5T4-1-6 confers stable UV resistance to UV5 and, like p5T4-1-15, lacks a linked copy of pSV2gpt; Weber *et al.*, 1988). Thus, it appears that the protein produced by p5T4-1-15 under these conditions is able to function normally.

Nucleotide sequence determination

The nucleotide sequence of the pER2-14 insert and of genomic 3′- and 5′-flanking regions was determined (subcloned fragments are shown in Figures 1C and 2; also see Materials and methods). Comparison of genomic and pER2-14 nucleotide sequences indicated the presence of a partial, unprocessed intron at the 5′-end of the cDNA insert (confirmed by S1 nuclease analysis below; note pER2-8 also likely contains an intron, but not the same one as in pER2-14). The *ERCC2* nucleotide sequence and amino acid translation are diagrammed in Figure 4. The G+C content of the coding region is 59.9%. The transcription initiation site (see below) is designated +1. The first and last intron junctions were identified by analysis of both genomic and cDNA sequences. Using the matrix presented by Stormo (1987) for identifying intron/exon splice sites, these *ERCC2* donor and acceptor sites score well. The *ERCC2* mRNA, from the transcription initiation site to the polyadenylation site observed in pER2-14, is 2397 nucleotides. Poly(A) tail

Fig. 4. The nucleotide sequence of genomic 5'- and 3'-flanking sequences of *ERCC2* and the pER2-14 cDNA insert are shown along with the amino acid translation. Identified introns and non-transcribed sequences are shown in lower case letters. The dotted line after position +2269 indicates the unsequenced portion of the final intron. The 5'-endpoints of cosmid p5T4-1-15 (⇓ ; position −153 at the *Mbo*I site) and of cDNA clones pER2-14 (↑ ; 60 nucleotides into intron 1) and pER2-11 (↓ ; position +1419) are indicated. The polyadenylation sites identified in pER2-14 (first ▲) and in pER2-11 and pER2-8 (second ▲) are marked. The regions corresponding to GC, CAAT and TATA boxes and polyadenylation signals are indicated. Non-initiating ATGs at +3 and +35 are underlined. Other features indicated are a pyrimidine-rich region, an inverted repeat, a putative nuclear location signal region, and six copies of a chromosome 19 specific minisatellite sequence. The transcription initiation site is numbered +1. The 30mer primers used in primer extension experiments are indicated.

addition would give an mRNA size of 2.6−2.7 kb (consistent with Northern analysis of HeLa RNA; unpublished data).

Classical TATA box, reverse CAAT box, GC box, and polyadenylation signals (Breathnach and Chambon, 1981; Kadonaga *et al.*, 1986; Lewin, 1987; Birnstiel *et al.*, 1985) were identified (Figure 4; *ERCC2* sequences are TAATATA, GGATTGACG, CCGCCC, and CATAAA and CTGGTTCC, respectively, where underlined nucleotides match consensus sequences). The reverse CAAT box (−56 to −48) falls in the classic promoter middle region (−70 to −50), and the GC box (−110 to −105) falls in the classic promoter distal region (−110 to −80) (Lewin, 1987). The pER2-8 and 11 inserts terminate 5 nucleotides further 3' than pER2-14, reflecting a small variation in the polyadenylation site.

The *ERCC2* open reading frame indicated that the translation initiation codon is the third one in the mRNA. A matrix for scoring translation initiation sites has been presented (Stormo, 1987) with which most known initiation sites score

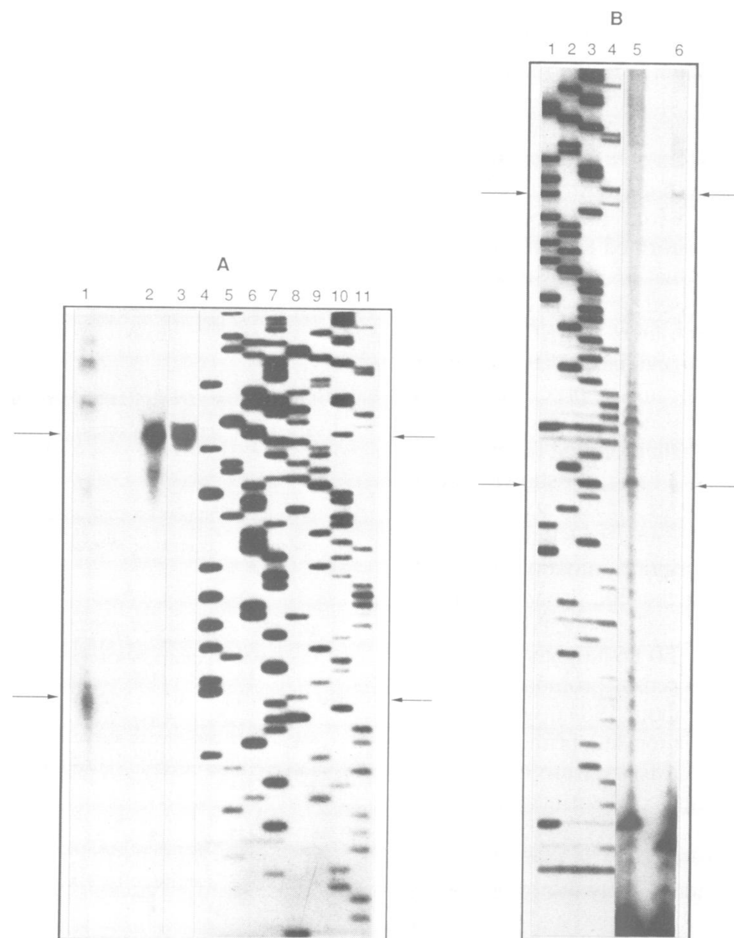

Fig. 5. Panel A shows an autoradiogram of a polyacrylamide−urea gel of the S1 nuclease protection products used to determine the intron 1/exon 2 junction and the transcription initiation site and of the dideoxynucleotide chain termination reactions used as size standards (arrows at 50 and 79 bp). Lane 1: the 202 bp *Sal*I−*Bam*HI region of p5T4-1-15 with 50 μg of HeLa cytoplasmic RNA, lanes 2 and 3: the 353 bp *Sal*I−*Pst*I region of pER2-14 with 50 μg of HeLa or 5ER2-14-3 cytoplasmic RNA, respectively, and lanes 4−7 and 8−11: dideoxynucleotide chain termination reactions (A, T, C and G, respectively) of the 353 bp region (*Sac*I 30mer primer) and of the 202 bp region (*Sal*I 30mer primer); (exposure with intensifying screens at −80°C; lanes 1−2 = 3.7 days and lanes 3−11 = 21 h). **Panel B** shows an autoradiogram of a polyacrylamide−urea gel of the primer extension products used to determine the transcription initiation site and of the dideoxynucleotide chain termination reactions used as size standard (arrows at 53 and 81 bp). Lanes 1−4: dideoxynucleotide chain termination reactions (A, T, C and G, respectively) of the 202 bp *Sal*I−*Bam*HI region of p5T4-1-15 (*Sal*I 30mer primer) and lanes 5 and 6: 5 μg of HeLa poly(A)$^+$ RNA with the *Sal*I and ATG 30mer primers, respectively (room temperature exposure = 4 days).

above +37. In *ERCC2* mRNA, the first ATG is at +3 and, therefore, lacks the important base at −3 from the ATG. The second and third ATGs score +12 and +54, respectively. This analysis is consistent with the current scanning model of translation initiation (Kozak, 1989) and indicates that the third ATG is the translation initiation codon. Thus, the *ERCC2*-encoded protein is 760 amino acids with a mol. wt of 86.9 kd and a net charge of +18 (without potential post-translation modifications).

Other notable features are also marked in Figure 4. There is a 34 base pyrimidine-rich (88%) region in the 5′-flanking sequence (−201 to −168). A 12 base inverted repeat, with the two central bases noncomplementary, is present within the classic promoter distal region at −93 to −82. A highly basic 14 amino acid region of the ERCC2 protein with helix-breaking residues (coded by + 2122 to +2163) was identified as a putative nuclear location signal region. This region is also highly basic in RAD3 (see below for ERCC2 and RAD3 homology), and a portion of this region (KRGKLP) is also present in the human XRCC1 protein (KRPKLP; *XRCC1* is a human repair gene that corrects the strand-break repair defect in the CHO mutant EM9; L.H.Thompson, K.W.Brookman, N.J.Jones, C.C.Collins and A.V.Carrano, in preparation). (For information on known nuclear location signals see: Kalderon *et al.*, 1984; Colledge *et al.*, 1986; Richardson *et al.*, 1986; Bürglin and De Robertis, 1987; Picard and Yamamoto, 1987; Dang and Lee, 1988; Lanford *et al.*, 1988). Three partial and three complete copies of sequences related to a chromosome 19 specific minisatellite family were identified, four in intron 1, one in the 5′-untranslated region, and one in the coding region. This minisatellite sequence of ~37 bp was initially identified in the third intron of the human apolipoprotein C-II gene and localized to ~60 different genomic locations, predominantly between 19q13.3 and 19qter (Das *et al.*, 1987).

cDNA intron confirmation and transcription initiation site determination

Nucleotide sequence analysis indicated that the cDNA clone pER2-14 contains part of intron 1. S1 nuclease analysis using

the 353 bp *Sac*I–*Pst*I region of pER2-14 (Figure 2B, *Sac*I at base 339 in the insert and *Pst*I at base 7248 in the vector; includes the pER2-14 insert through + 162 in Figure 4) was carried out to confirm that this region is indeed normally removed from the *ERCC2* message. A protected fragment of ~79 bp (the predicted size based on nucleotide sequence analysis) was obtained when the *Sac*I–*Pst*I fragment was hybridized with cytoplasmic RNA from HeLa cells or from 5ER2-14-3 cells (a pER2-14-transformed UV5 line) (Figure 5A). These results demonstrate that pER2-14 indeed contains part of the normally processed intron 1. Because the 5ER2-14-3 cells have a very high level of *ERCC2* mRNA from pER2-14 (unpublished data) and the SV40 16S/19S intron donor site in the pcD2 vector allows processing to the *ERCC2* intron 1 acceptor site, these cells provided a positive control. An approximately full length protected fragment was also observed with the 5ER2-14-3 cells (not shown) due to processing to either the SV40 16S or 19S acceptor sites (Chen and Okayama, 1987; Okayama and Berg, 1983).

To localize the *ERCC2* transcription initiation (or CAP) site, S1 nuclease analysis using the 202 bp *Sal*I–*Bam*HI region of p5T4-1-15 (Figure 1C) was conducted with HeLa RNA. A protected fragment of ~50 bp [slightly larger than the 45 bp predicted size from consensus promoter and CAP sequences (Breathnach and Chambon, 1981)] was obtained (Figure 5A). To confirm this initiation site, primer extension experiments with poly(A)$^+$ HeLa RNA were carried out. The *Sal*I and ATG 30mer oligonucleotide primers (indicated in Figure 4) gave extension products of 53 and 81 nucleotides, respectively (Figure 5B). The second band at 58 nucleotides for the *Sal*I primer is likely due to secondary structure as a second 'unextended' primer band is also seen at 35 nucleotides. The transcription initiation site numbered +1 in Figure 4 is based on the primer extension experiment since a background 'ladder' was visible from the starting *Sal*I primer to the extension product, giving an exact size. The *ERCC2* CAP site as determined here is at a G separated from the TATA box by 18 nucleotides. In the Breathnach and Chambon (1981) analysis, the one atypical RNA of 22 studied, the Moloney murine leukemia virus RNA, also has a 'G' CAP site, rather than an 'A' (van Beveren *et al.*, 1980).

Homology to RAD3 and other proteins

Comparison of the translated ERCC2 protein (760 amino acids) with the translations of *RAD* genes from the yeast *S.cerevisiae* revealed a striking 51.7% identity between ERCC2 and RAD3 (778 amino acids). Allowing conservative substitutions, the level of homology increases to 72.5% (Figure 6). The nucleotide binding box and DNA binding box defined as functional domains for RAD3 and the sites of amino acid substitutions in *rad3* mutant alleles are indicated. Seven protein domains identified in one of two related superfamilies of known and putative helicases, including RAD3, are also noted (Gorbalenya *et al.*, 1989). Within these domains, ERCC2 and RAD3 are 73.8% identical and 88.5% homologous. Although the G+C contents of the *ERCC2* and *RAD3* coding regions differ significantly (59.9% and 37.1%, respectively), these nucleotide sequences exhibit 53% identity and require few gaps for alignment (length: *ERCC2* = 2283, *RAD3* = 2337, alignment = 2373; alignment not shown).

Regions of protein homology between ERCC2 and other repair, replication or recombination proteins were identified

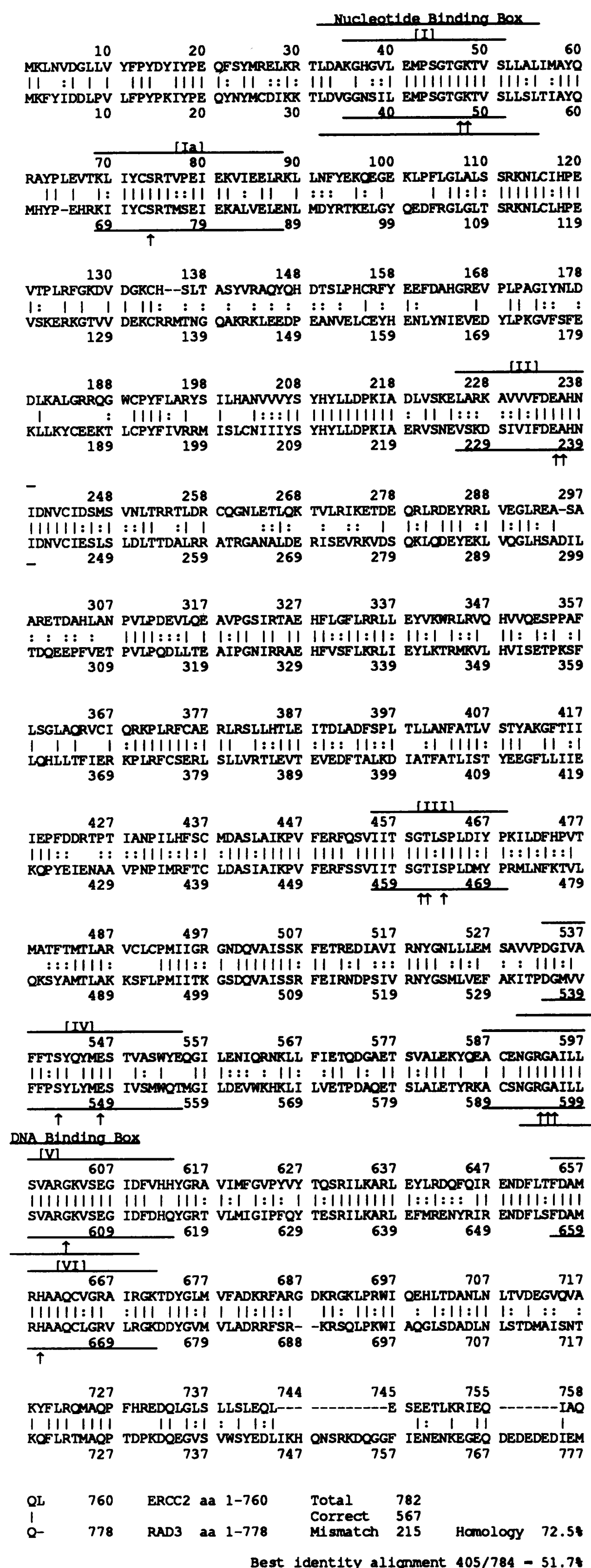

Fig. 6. The best alignment between ERCC2 and RAD3 (databases accession numbers K03293, X02368, and P06839) is shown, allowing conservative substitutions based on the functional group and polarity of the residues as a match (A=G=P=S=T, D=E=N=Q, F=W=Y, H=K=R, and I=L=M=V). Homology is indicated as | for identical amino acids and : for conservative substitutions. Conserved helicase domains (identified by roman numerals) and functional domains of RAD3 are over/underlined (after Gorbalenya *et al.*, 1989). Amino acid positions of characterized missense mutations in *rad3* alleles are indicated (↑).

A

```
                                       _   ____
RAD3   (30-56)         KTLDVGGNSILEMPSGTGKTVSLLSLT
                       :|||   | ::||||||||||||||:|
ERCC2  (30-56)         RTLDAKGHGVLEMPSGTGKTVSLLALI
                         |   |  ::| ::|||||  :: ||
RECB   (11-37)         LRLPLQGERLIEASAGTGKTFTIAALY
                                :| :::|||||  |:| |
RECD   (159-185)       AVALTRRISVISGGPGTGKTTTVAKLL
                       ||| :|    :: :|:|:|||   : :
UVRD   (17-43)         AVAAPRSNLLVLAGAGSGKTRVLVHRI
                       ::  : :    :|: :|||||   :   |
UVRB   (27-53)         GLEDGLAHQTLLGVTGSGKTFTIANVI
                                : ||:||||:   | : :
UVRA   (628-654)       LTLPVGLFTCITGVSGSGKSTLINDTL
                       | :|      :||:|||||||:|  |||
UVRA   (19-45)         LVIPRDKLIVVTGLSGSGKSSLAFDTL
                       |       :|:||: || |||:    :
T7 GENE 4 A/B          LGARGGEVIMVTSGSGMGKSTFVRQQA
(300-326/237-263)       |    | :: : :: : ||:|    |
RECA   (54-80)         GGLPMGRIVEIYGPESSGKTTLTLQVI
                       :|| |: :: : :   | |||| :::::
DNAB   (219-245)       AGLQPSDLIIVAARPSMGKTTFAMNLV
                                 _     ______
```

B

```
                          *          ** *  *  **        *            *         *
                          * *  *     ** *** ** *** ** **          *     *      *     **
UVRB   (200-252)   IDIFPAESDDIALRVELFDEEVERLSLFDPLTGQIVSTIPRFTIYPKTHYVTP
                   |:  |         ||  |    |:  |: | |||        :     | |:|    :
ERCC2  (367-418)   IQRKPLRFCAERLRSLLHTLEITDLADFSPLTLLANFATLVST-YAKGFTIII
                    : | |        || ||: |   :|  || :    |:   |    | -   |
RAD1   (389-440)   YESKQLVKDIRFLRHLLKMLVTSDAVDFFGEIQLSLDANKPSV-SRKYSESPW
                    ::    : :          : |: |:  ||           |:   :| -   |      :
SIR3   (529-579)   FNADINLEESIRESLQKRELLKSQVEDFTRIFLPIYDSLMSSQ-N-KLFYITN
                    :        : |        : |  ::: ||: : |        : : | -   | | |
ERCC2  (367-418)   IQRKPLRFCAERLRSLLHTLEITDLADFSPLTLLANFATLVST-YAKGFTIII
                      | |        |   |  ::||
DNAB   (370-390)   ALAKELNVPVVALSQLNRSLE

                    D K L        LR LLR LE SDL DF PL          A     S       K    IT
                    E             S EVH  L TQV    GI          S                   VP
                    N               Q K  V  E     A           T                   L
                    Q

RAD3   (369-420)   IERKPLRFCSERLSLLVRTLEVTEVEDFTALKDIATFATLIST-YEEGFLLII
```

C

```
RAD3   (369-420)   IERKPLRFCSERLSLLVRTLEVTEVEDFTALKDIATFATLISTYEEGFLLII
                   |:||||||||:|||  |::||||:|:: ||::|  :| ||||:|||  || :||
ERCC2  (367-418)   IQRKPLRFCAERLRSLLHTLEITDLADFSPLTLLANFATLVSTYAKGFTIII
                   |  | |:  ::||: |:                |: || :   |   :    :   : ::
UVRA   (455-506)   IAEKILKEIGDRLKFLVNVGLNYLTLSRSAETLSGGEAQRIRLASQIGAGLV

RAD3   (625-641)   IPFQYTESRI-LKARLEF
                   :|: ||:|||-||||||:
ERCC2  (623-639)   VPYVYTQSRI-LKARLEY
                   :  : :| :: |||||:|
UVSY   (46-63)     MLRIEAQKKVALKARLDY

RAD3   (164-182)   NIEVEDYLPKGVFSFEKLL
                        |  || |::  : |
ERCC2  (163-181)   AHGREVPLPAGIYNLDDLK
                   : :|::| : || |||:|:
T-AG   (480-498)   GESRDLPSGQGINNLDNLR

RAD3   (560-610)   LDEVWKHKLILVETPDAQETSLALETYRKACSNGRGAILLSVARGKVSEGI
                   |::: : ||: :|| |: |||:||| |   || ||||||||||||||||||||
ERCC2  (558-608)   LENIQRNKLLFIETQDGAETSVALEKYQEACENGRGAILLSVARGKVSEGI
                   |  :  |  | :| :|:   ::    : |   |||    | || :| : ::
RHO    (136-186)   LTPLHANSRLRMERGNGSTEDLTARVLDLASPIGRGQRGLIVAPPKAGKTM
```

```
RAD3       (339-448)  IEYLKTRMKVLHVISETPK--SFLQHLKQLTFIERKPLRFCSERLSLLVRTLEVTE
                      :||:| |::| ||: |:| --:||  | |   |:||||||||:|||  |::|||:|:
ERCC2      (337-446)  LEYVKWRLRVQHVVQESPP--AFLSGLAQRVCIQRKPLRFCAERLRSLLHTLEITD
                      |||  : | | | |   : |  || :  ::         :  | |  : ::| : : ::
28K HYP.   (37-148)   LEYDYYSLCVRHPVPFTRPKVAFYTNYPEAWVSYYQAKNFLA--IDPVLNPENFSQ
UVRC REG.
                 (27) IEPYEIENAAVPNPIMRFT-CLDASIAIKP
                      |||::    ::::|||::|:-|:|||:||||
                 (27) IEPFDDRTPTIANPILHFS-CMDASLAIKP
                      : :  : : | | : | || |     ::| :
                 (28) VHSVFNAAQTGALGFLSFSRCSRREIPILS

RAD3       (285-298)  EYEKLVQGLHSADI
                      || :||:||: |
ERCC2      (284-297)  EYRRLVEGLREASA
                      |  :||||| || |
RECC       (150-163)  ETGHLVEGLGEAQA

RAD3       (443-490)  SIAIKPVFERFSSVIITSGTISPLDMYPRMLNFKTVLQKSYAMTLAKK
                      |:|||||||||| ||||||||:||||:||::|:|::|    :::||||:
ERCC2      (441-488)  SLAIKPVFERFQSVIITSGTLSPLDIYPKILDFHPVTMATFTMTLARV
                      ::    :    | |    || |      :    :::|||    | ||   |    |
VASA       (350-397)  GIVYGGTSFRHQNECITRGCHVVIATPGRLLDFVDRTFITFEDTRFVV
```

Fig. 7. **Panel A** shows a comparsion of the ATP binding region of RAD3 and ERCC2 to other known and putative nucleotide binding proteins. The over/underlined region indicates the consensus ATP binding site. **Panel B** shows the homology between a 52 amino acid region of ERCC2 and four other proteins. (ERCC2 is presented twice to compare directly ERCC2 with each of the other proteins). The amino acids listed below a position reflect a consensus for this domain. (RAD3 is included in determining the consensus.) A single or double star above a position indicates consensus based on 3 or ≥4 of the proteins, respectively. (RAD3 is not counted.) **Panel C** illustrates alignments of regions of homology between ERCC2 and other proteins that may have functional significance. Homology is indicated at | for identical amino acids and : for conservative substitutions (see Figure 6). 28K HYP. UVRC REG: *E.coli*, UVRC regulatory region 28K hypothetical protein, P07026 (database accession number). DNAB: *E.coli*, replication, helicase, DNA dependent ATPase, P03005. ERCC2: human, excision repair. RAD1: *S.cerevisiae*, excision repair, P06777. RAD3: *S.cerevisiae*, excision repair, helicase, P06839. RECA: *E.coli*, SOS response, repair, replication, ATPase, P03017. RECB, RECC, RECD:*E.coli*, exonuclease V subunits, P08394, P07648, P04993, RHO: *E.coli*, transcription termination factor, ATPase, helicase, P03002. SIR3: *S.cerevisiae*, transcription regulator for mating type genes, P06701. T7 GENE 4 A/B: bacteriophage T7, primase chains A and B, A04314. T-AG: SV40, large T-antigen, V01380. UVRA, UVRB:*E.coli* excision repair, P07671, P07025. UVRD: *E.coli*, excision repair, helicase, P03018, (position 224 conflict D versus N). UVSY: bacteriophage T4, recombination/repair protein, P04537. VASA: *Drososphila melanogaster*, required in female germ line, oocyte formation, similarity to helicases, P09052.

by computer analysis (Figure 7). These regions, except the SV40 T-antigen, are highly conserved between ERCC2 and RAD3. Regions <20 amino acids in length were presented only if there was a significant amount of identity. The region in Figure 7A contains the RAD3 nucleotide binding box, and many of these proteins are known to bind ATP. A 52 amino acid region of ERCC2 shows homologies to UVRB, RAD1, SIR3 and DNAB, as well as to RAD3 (Figure 7B). The similarity among these proteins in this region suggests it as a putative functional domain. This region of ERCC2 also has homology with UVRA (shown separately in Figure 7C because only part of the region fits the domain consensus). The homology to RHO includes the DNA binding box (domain V), and the homology to the UVRC regulatory region 28K hypothetical protein is striking in its length (a 54 and a 30 amino acid region, separated by a 27–28 amino acid region without sequence similarity) (Figure 7C).

Discussion

Genomic clones of the human *ERCC2* DNA repair gene have previously been isolated and characterized (Weber *et al.*, 1988). This human gene specifically corrected the repair deficiency of CHO complementation group 2 represented by the UV5 mutant cell line. This correction was quantitatively efficient as demonstrated by the return to approximately wild-type levels of UV survival, UV induced mutation at the *aprt* locus, and UV induced strand incision. The genomic region required for function was determined to be ~19 kb. The cloning and characterization, both biological and molecular, of the cDNA for the human *ERCC2* gene are presented here. Although no full length cDNA clone was obtained, complete nucleotide sequence was determined using genomic DNA fragments from the 5′-region.

Nucleotide excision repair in eukaryotes

The single-celled yeast *S.cerevisiae* has been a model system for more complex eukaryotes. The conservation of nucleotide excision repair proteins found here for ERCC2 and RAD3 and previously for ERCC1 and RAD10 (van Duin *et al.*, 1986) indicates that repair processes in these evolutionarily divergent species may have similarity (see Friedberg, 1987

for a discussion of the biochemistry and genetics of repair in *E.coli*, yeast and mammalian cells).

Both cyclobutane dimers and (6−4) pyrimidine−pyrimidone photoproducts are preferentially repaired in actively transcribed regions in mammalian cells (Bohr *et al.*, 1985; Madhani *et al.*, 1986; Mellon *et al.*, 1986; Thomas *et al.*, 1989). Phenotypic heterogeneity in the levels of DNA incision and repair replication, without a difference in UV sensitivity, has been observed among Chinese hamster cell mutants in complementation group 2 (Zdzienicka *et al.*, 1988b). UV5 is completely incision deficient, whereas V-H1 has intermediate incision levels, correlating with the level of (6−4) pyrimidine−pyrimidone photoproduct removal (Mitchell *et al.*, 1989). Since the levels of UV sensitivity are similar, it appears that this removal is not aiding survival. These findings suggest that ERCC2 may have a role in preferential repair.

Promoter sequences and the transient correction of UV5 with p5T4-1-15

Analysis of the *ERCC2* 5′-flanking region revealed classic consensus promoter sequences for a single GC box, a CAAT box (reverse orientation), and a TATA box (see Lewin, 1987 for a discussion of promoter elements). In contrast, classic promoter elements were not found within the 170 bp of *ERCC1* 5′-flanking region that was determined to contain the *ERCC1* promoter (van Duin *et al.*, 1987). Comparison of the mouse *ERCC1* homolog flanking sequence with the human *ERCC1* promoter revealed two regions of conservation, one including a 32 base pyrimidine-rich (81%) region from −97 to −66 (van Duin *et al.*, 1987; 1988b). *ERCC2* has a similar pyrimidine-rich region in the 5′-flanking sequence (34 bases, 88%, −201 to −168; see Figure 4). Cosmid p5T4-1-15, which confers only transient UV resistance to UV5 cells and ends at −153, appears to lack some 5′ sequences necessary for stable *ERCC2* expression. This cosmid contains the GC, CAAT and TATA boxes but lacks the pyrimidine-rich stretch, suggesting a role for this sequence in *ERCC2* expression.

Consensus polyadenylation signals were identified in *ERCC2* cDNA and genomic clones, and no indication of alternative polyadenylation signals was found. In contrast, *ERCC1* was found to have two sets of polyadenylation signals (van Duin *et al.*, 1987).

Transient correction of UV5 with pER2-14

A combination of nucleotide sequence analysis and S1 nuclease analysis confirmed that the cDNA clone insert of pER2-14 is atypical, beginning within an unprocessed copy of intron 1. The message produced from the SV40 promoter when pER2-14 is expressed in UV5 cells should be processed in several ways: from the SV40 16S/19S splice donor site to the SV40 19S, 16S or *ERCC2* intron 1 splice acceptor sites. Using the Stormo (1987) matrix and the Kozak (1989) model, the translation products of the pER2-14 mRNAs are predicted to be predominantly an 18 amino acid peptide and a truncated ERCC2 protein synthesized from reinitiation at codon 57 of the normal *ERCC2* message. This truncated protein would lack the nucleotide binding site and, thus, be unlikely to function in repair. The initial intermediate, transient correction seen is presumably due to the extremely high levels of message produced in the pre-integration stage of transformation. These high levels of message may result in a sufficient amount of putatively functional ERCC2 protein synthesized from the first weak initiation site at codon 24 of the normal message. Work is in progress to construct full length cDNA expression clones under control of the SV40 or *ERCC2* genomic promoters to assess the effects of over-expression of *ERCC2* in CHO cells and to test *ERCC2* for the ability to correct the various XP complementation groups.

Homologies between human and yeast nucleotide excision repair genes

Nucleotide excision repair has been extensively studied in the yeast *S.cerevisiae* (reviewed in Freidberg, 1988). Five of the 10 genes in the *RAD3* epistasis group are completely required for the incision step, which is also the defective step associated with XP. As the human nucleotide excision repair genes are being cloned and characterized, homologies between the proteins from these evolutionarily distant species are being revealed. The 52% identity (and 73% homology) over the entire protein found here for ERCC2 and RAD3 is quite striking. Additionally, ERCC1 (297 amino acids) was found to have 35% identity with RAD10 (210 amino acids) over 108 amino acids (van Duin *et al.*, 1986), and RAD10 is able to partially correct CHO mutants in complementation group 1 (Lambert *et al.*, 1988).

RAD3 is the best characterized of the yeast nucleotide excision repair genes and proteins (reviewed in Friedberg *et al.*, 1986 and Friedberg, 1988). The *RAD3* gene has been cloned and the nucleotide sequence determined (Higgins *et al.*, 1983; Reynolds *et al.*, 1985; Naumovski and Friedberg, 1982; Naumovski *et al.*, 1985). *RAD3* is essential for viability in the absence of DNA damage (Higgins *et al.*, 1983; Naumovski and Friedberg, 1983) as well as for the incision step of nucleotide excision repair. The *rem1* mutants have recently been determined to be alleles of *RAD3* (Montelone *et al.*, 1988). These mutants are phenotypically distinct from classic *rad3* mutants in lacking UV sensitivity and demonstrating high levels of spontaneous mitotic gene conversion, crossing over and mutation. Studies of *rem1* mutants suggest a role for RAD3 in DNA replication.

Rad3 mutant alleles, including two *rem1* alleles, with characterized missense mutations at each of 15 amino acid positions have changes in amino acids conserved (identical) between ERCC2 and RAD3 (Naumovski *et al.*, 1985; Naumovski and Friedberg, 1986; 1987; Sung *et al.*, 1988; B.Montelone, personal communication). Five of these codons were selected for site-directed mutagenesis based on their location within putative functional domains. However, the other 10 codon changes were due to random mutation. All 15 of these amino acid positions are within identified domains of the putative helicase superfamily, and all seven domains are included.

ERCC2 is 18 amino acids shorter than RAD3, and the best alignment between them adjusts for this almost entirely in the carboxy-terminal region, where the composition of the proteins is also dissimilar. This region of RAD3 is highly acidic and was therefore hypothesized to have some important role. However, deletions removing the carboxy-terminal 25 amino acids of RAD3 have no effect on RAD3 functions (Naumovski *et al.*, 1985; Reynolds *et al.*, 1985). The ERCC2 structure is consistent with the highly acidic carboxy terminus of RAD3 not having an important functional role.

Helicase domains

RAD3 protein has been purified and demonstrated to have a single-stranded DNA dependent ATPase activity and a helicase activity dependent on ATP hydrolysis (Sung *et al.*, 1987a; 1987b; 1988; Harosh *et al.*, 1989). The protein regions proposed to be involved in ATP and DNA binding and other domains common to this helicase superfamily are extremely conserved between RAD3 and ERCC2 (Figure 6). Therefore, the human ERCC2 protein is putatively an ATP dependent helicase with similar properties. Helicase domains I and II are involved in binding ATP. The functions of the other domains are as yet unknown, although domain V includes the RAD3 DNA binding box. A 52 amino acid region of ERCC2 was found to have homology to regions of five proteins in addition to RAD3, suggesting that this domain may also be functionally important (Figure 7B and C). At this time, the biological significance, if any, of the identified partial protein homologies, other than the ATP binding domain, is unknown.

An essential function for ERCC2?

Evidence for an essential role for *ERCC2* in mammalian cell viability, as observed for *RAD3* in yeast, was found in analyzing the results of mutant hunts in CHO cells (Busch *et al.*, 1989). *ERCC1* and *ERCC2* are closely linked on human chromosome 19 and the hemizygous CHO chromosome 9 (Thompson *et al.*, 1989a), making both of these repair genes single-copy in CHO cells. Given that the *RAD3* essential function was found to be extremely difficult to affect by point mutation, comparison of the yield of mutants in complementation groups 1 and 2 from agents causing point mutations and from agents causing frame-shifts addresses the question of an essential function for *ERCC2*. Point mutations result primarily in amino acid substitutions, whereas frame-shifts result primarily in null mutations, a lethal event in an essential gene. Assuming target size equals coding region, the relative yield of mutations in the *ERCC2* locus versus the *ERCC1* locus with point mutagens was as expected, but the yield with frame-shift agents was 40-fold lower than expected. Although this is not definitive proof, it is certainly suggestive of an essential role for *ERCC2* in mammalian cell viability.

Conclusion

The results presented here provide the first indication of the specific biochemical function of a mammalian nucleotide excision repair gene. The human *ERCC2* gene encodes a 760 amino acid protein with a high degree of homology to the yeast RAD3 protein. *ERCC2* has not only classical promoter elements but also a pyrimidine-rich region in the 5′-flanking sequence that may, by analogy to *ERCC1*, be relevant to gene expression. The similarity to RAD3 and other initial data suggest that ERCC2 is an ATP dependent helicase that has a nucleotide excision repair function, a role in replication and an essential function for cell viability.

Materials and methods

cDNA library screening

The pcD2 library is a human fibroblast cDNA library in a mammalian expression vector with a *neo* selectable marker (Chen and Okayama, 1987). RNA containing the cDNA insert is transcribed from an SV40 early promoter, contains the alternatively processed SV40 16S/19S intron (with the middle region deleted), and is polyadenylated from SV40 signals at the 3′-end (Chen and Okayama, 1987; Okayama and Berg, 1983). The alternative processing makes available the ATG translation initiation site of the SV40 capsid protein VP2 for translation of partial cDNAs. *E.coli* DH1 cells from the initial amplification of this library were kindly provided by H.Okayama (National Institute of Mental Health, Bethesda, MD) and were amplified once in our laboratory. Bacterial cells were plated at high density (~50 000 colonies per 137 mm filter), and filters processed as described in Weber *et al.* (1988). The hybridization probes were restriction enzyme digest fragments from *ERCC2*-containing cosmid clones (Figure 1B; Weber *et al.*, 1988) and were nick-translated with the Bethesda Research Laboratories, Inc. (BRL; Gaithersburg, MD) kit with [α-^{32}P]dCTP (Amersham Corp., Arlington Heights, IL) to a specific activity of $1.5-5 \times 10^8$ c.p.m./μg. Signal from repetitive sequences in the probes was reduced by pre-reassociation in solution with excess human placental DNA (after Sealey *et al.*, 1985). In the reassociation reactions, the 12 kb probe was at 2.5×10^{-4} mg/ml ($t_{1/2}$ in 5× SSC = 36 min; 1× SSC is 0.15 M NaCl, 0.015 M sodium citrate), and the 8.3 kb probe was at 1.7×10^{-4} mg/ml ($t_{1/2}$ in 5× SSC = 37 min). Human placental DNA was at 10 mg/ml in 5× SSC, and reassociation was conducted at 68°C for 35 min (i.e. to an equivalent of C_0t = 465 in 0.12 M phosphate buffer). Prehybridization and hybridization were as described (Sealey *et al.*, 1985), except prehybridization was for 4–6 h and buffers were not degassed (10–15 ng probe DNA per 137 mm filter; 6 ng probe DNA per 88 mm filter). After one room temperature rinse (in the 4× wash buffer), washes were at 55°C in 5× Denhardt solution, 0.1% SDS, 0.1% sodium pyrophosphate with various concentrations of SSC as follows: 4× twice for 20 min, 3× and 1× three times each for 15 min, and 0.1× three times for 20 min. The final rinse was at room temperature in 0.1× SSC. Secondary and tertiary rounds of screening to purify positive colonies were conducted similarly but at lower density (50–300 colonies per 82 mm filter).

DNA preparation

Small-scale preparation of plasmid DNAs was a modification of a rapid alkaline plasmid extraction procedure (Birnboim and Doly, 1979) as described for cosmid DNAs (Weber *et al.*, 1988). Large-scale preparation of plasmid DNAs was from bacterial cells grown to an optical density at 600 nm of 0.6, with chloramphenicol for plasmid amplification and SDS for cell lysis as described (Maniatis *et al.*, 1982).

Single-stranded DNA of M13 clones was prepared as described for the Cyclone I Biosystem [International Biotechnologies, Inc. (IBI), New Haven, CT] except vortexing during extractions was for 20 s and only two extractions were done. For deletion subcloning, the procedure was scaled up to 10 ml with centrifugation in a Sorvall SM24 rotor at 16 000 r.p.m. and resuspension in 200 μl of TE. Single-stranded DNA of pIBI clones was prepared as recommended for the Clonar I biosystem (IBI) with the modifications described above for M13.

Restriction endonuclease digestion and fragment isolation

Restriction enzyme digestions were carried out as recommended by the manufacturer (BRL; New England Biolabs, Inc., Beverly, MA). Electrophoresis of digested DNA was carried out in 0.7–1.0% agarose or 4% acrylamide gels. Gel regions were excised, and the DNA was recovered by electroelution in a 'unidirectional electroeluter' (model UEA; IBI).

Southern blot hybridization

Southern blot hybridization was conducted as described in Weber *et al.* (1988) with hybridizations and washes at 68°C, hybridizations at 5× SSC, and final washes at 0.1× SSC.

Cell strains and culture conditions

AA8 is a repair-competent CHO line (Thompson *et al.*, 1980a) and is the parental line from which the UV sensitive line UV5, a member of complementation group 2, was derived (Thompson *et al.*, 1980b, 1981; Thompson and Bootsma, 1988). Stock cultures of AA8 and UV5 were maintained as previously described (Weber *et al.*, 1988). Transformant lines containing the *E.coli gpt* or *neo* genes under SV40 regulatory sequences were selected and cultured in MAXTA meidum (see recipe in Weber *et al.*, 1988) or in normal medium with 1.6 mg G418 (Gibco Laboratories, Grand Island, NY) per ml, respectively.

Initial testing of cDNAs for complementation of UV5

Calcium phosphate precipitates of cDNAs (~1–2 μg DNA/ml; DNA was from small-scale plasmid preparations) were prepared as described previously (Corsaro and Pearson, 1981). One milliliter of precipitate was added to each dish containing ~2×10^6 UV5 cells. Cells were exposed to DNA for 4 h with glycerol shock treatment (Thompson *et al.*, 1987) or 16–20 h without shock, rinsed with 10 ml of medium (three times for glycerol shock treated dishes), and replenished with 20 ml of medium for 24 or 48 h to allow time for integration and expression of transformed sequences. Each

dish was trypsinized and the cells were respread at 2×10^5 cells per dish in G418 medium and $3-7 \times 10^5$ in normal medium. The normal medium dishes were irradiated with a 5 J/m^2 dose after 4 h. After 2 days, dishes were either irradiated with a second 5 J/m^2 dose or changed to G418 medium. After an additional 8 days, surviving colonies in the dishes that received two UV doses were tested for growth in G418 medium, and those in the single UV dose/G418 medium dishes were tested for UV resistance.

UV survival studies

Isolated colonies. Isolated colonies were grown to mass culture with selection for the transferred marker gene. These cells were cultured in α-MEM medium for 3 days before being plated at 5×10^5 per 100 mm dish. Cells were exposed to UV radiation 18 h after plating, trypsinized and replated at appropriate densities to determine survival.

At 24 h after transformation. UV5 cells were exposed to calcium phosphate precipitates containing 5 μg of each DNA (large-scale plasmid preparations) and treated with glycerol shock (see above). After 24 h, the cells were trypsinized and plated in normal medium and in G418 or MAXTA medium to determine plating efficiency and selectable marker frequency and at 5×10^5 to 2×10^6 cells per dish in normal medium for UV exposures. After 2 h, the latter dishes were exposed to appropriate doses of UV radiation, and after 24 h, the medium was changed to selective medium. Dishes were examined for colonies 12−14 days after plating.

Subcloning DNA fragments for nucleotide sequence determination

DNA fragments isolated as described above were ligated into appropriately restriction enzyme digested cloning vectors [M13mp18 or 19 (BRL), M13um20 or 21, or pIBI24 or 25 (IBI)] and transformed into *E.coli* NM522 (Gough and Murray, 1983). The insert-containing *Xho*I fragments of several cDNA clones were subcloned into M13 vectors. The pER2-14 insert was unstable as an M13 clone in *E.coli* strains JM101 (Yanich-Perron *et al.*, 1985), MV1190 (a *rec*A$^-$ derivative of JM101), and NM522. Therefore, a large fragment from the 3′ region of the pER2-14 insert was cloned for use in deletion subcloning, and the remainder was cloned in smaller pieces that would accommodate nucleotide sequence determination directly (Figure 2). Deleted subclones were obtained using the Cyclone I Biosystem (IBI) based on the method of Dale *et al.* (1985). Complete *Hin*dIII (New England Biolabs, Inc.) digestion was obtained by using the IBI 'digestion 18mer' under the conditions recommended by Dale *et al.* (1985).

Nucleotide sequence determination

Clones were identified using the dideoxynucleotide chain termination reaction of the Klenow based Sequel-SS/KF Biosystem (IBI). Complete nucleotide sequence of both strands was determined using the dideoxynucleotide chain termination reaction of the Sequenase kit (modified T7 DNA polymerase; United States Biochemical Corp., Cleveland, OH) with both dGTP and dITP reactions, labeling with [α-^{35}S]thio-dATP (Amersham), and analyzing the products in 6% acrylamide (38:2)-urea gels. For large genomic fragments from the 3′- and 5′-flanking regions of *ERCC2*, the nucleotide sequence for a portion of one strand was determined using the universal primer and for the second strand using a primer complementary to the most distant sequence determined on the first strand.

RNA isolation and poly(A)$^+$ purification

For primer extension experiments, RNA was prepared and poly(A)$^+$ purified using the RNA Isolation Kit (Stratagene, La Jolla, CA) and the oligo(dT)-cellulose spun column mRNA Purification Kit (Pharmacia LKB Biotechnology, Inc., Piscataway, NJ). For S1 nuclease analysis, cytoplasmic RNA was prepared by a variation of the method described (Ausubel *et al.*, 1989). Cells ($\leq 2 \times 10^7$) were resuspended in 150 μl of buffer (10 mM Tris−HCl pH 7.8, 150 mM NaCl, 10 mM vanadyl ribonucleoside complex) and transferred to a microfuge tube. Nonidet P40 was added to 1%, and the mixture incubated on ice for 1 min. After centrifugation, 200 μl of lysis buffer (50 nM Tris pH 7.8, 50 mM Na_2 EDTA, 875 mM NaCl) and 50 μl of 10% SDS were added. No proteinase K was used. The remainder of the procedure was as described, except precipitation was without added salt and with 2.5 vol of ethanol.

Primer extension and S1 nuclease analysis

Primer extension was carried out as described (Ausubel *et al.*, 1989) using 30mer oligonucleotide primers with 5 μg poly(A)$^+$ RNA. The end-labeled primer was resuspended in water. The RNA, 1×10^5 c.p.m. of primer, and 9 μl 3× aqueous hybridization buffer containing 0.75 M NaCl were brought to a final volume of 30 μl, and the primer/RNA mix was denatured at 80°C for 10 min prior to hybridization at 60°C for 13 h. Extension reactions used 40 U of Moloney murine leukemia virus reverse transcriptase (Stratagene). S1 nuclease analysis was carried out as described (Ausubel *et al.*, 1989), using the aqueous hybridization procedure, 50 μg of cytoplasmic RNA, and 200 U of S1 nuclease per reaction. Products of both primer extension and S1 nuclease reactions were analyzed on 8% acrylamide (38:2)−urea gels. Size standards were dideoxynucleotide chain termination reaction products generated from M13 clones of corresponding genomic or cDNA regions with the Sequenase kit using the same primers as were used in the reactions being analyzed and [α-^{32}P]dCTP (Amersham).

DNA and protein sequence analysis

All nucleotide and protein sequence analysis and comparison was done using the IntelliGenetics Suite of programs (IntelliGenetics, Inc., Mountain View, CA).

Acknowledgements

We wish to thank Hiroto Okayama for providing the pcD2 cDNA library, Kerry Brookman and Nigel Jones of helpful discussions, Anita Avery for her valuable assistance, Malgorzata Zdzienicka for providing Northern blot hybridization information, and Harvey Mohrenweiser and Errol Friedberg for helpful discussions and reading the manuscript.

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

References

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1989) *Current Protocols in Molecular Biology*. John Wiley and Sons, New York.

Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513−1523.

Birnstiel,M.L., Busslinger,M. and Strub,K. (1985) *Cell*, **41**, 349−359.

Bohr,V.A., Smith,C.A., Okumoto,D.S. and Hanawalt,P.C. (1985) *Cell*, **40**, 359−369.

Bootsma,D., Keijzer,W., Jung,E.G. and Bohnert,E. (1989) *Mutat. Res.*, **218**, 149−151.

Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349−383.

Brook,J.D., Shaw,D.J., Meredith,A.L., Worwood,M., Cowell,J., Scott,J., Knott,T.J., Litt,M., Bufton,L. and Harper,P.S. (1985) *Cytogenet. Cell Genet.*, **40**, 590−591.

Bürglin,T.R. and de Robertis,E.M. (1987) *EMBO J.*, **6**, 2617−2625.

Busch,D.B., Cleaver,J.E. and Glaser,D.A. (1980) *Somatic Cell Genet.*, **6**, 407−418.

Busch,D., Greiner,C., Lewis,K., Ford,R., Adair,G. and Thompson,L. (1989) *Mutagenesis*, **4**, 349−354.

Chen,C. and Okayama,H. (1987) *Mol. Cell. Biol.*, **7**, 2745−2752.

Cleaver,J.E. (1983) In Stanbury,J.B., Wyngaarden,J.B., Fredrickson,D.S., Goldstein,J.L. and Brown,M.S. (eds), *The Metabolic Basis of Inherited Disease*. McGraw-Hill Book Co., New York, pp. 1227−1248.

Colledge,W.H., Richardson,W.D., Edge,M.D. and Smith,A.E. (1986) *Mol. Cell. Biol.*, **6**, 4136−4139.

Collins,A. and Johnson,R.T. (1987) *J. Cell Sci. Suppl.*, **6**, 61−82.

Corsaro,C.M. and Pearson,M.L. (1981) *Somatic Cell Genet.*, **7**, 603−616.

Dale,R.M.K., McClure,B.A. and Houchins,J.P. (1985) *Plasmid*, **13**, 31−40.

Dang,C.V. and Lee,W.M.F. (1988) *Mol. Cell. Biol.*, **8**, 4048−4054.

Das,H.K., Jackson,C.L., Miller,D.A., Leff,T. and Breslow,J.L. (1987) *J. Biol. Chem.*, **262**, 4787−4793.

Friedberg,E.C. (1985) *DNA Repair*. W.H. Freeman and Co., New York.

Freidberg,E.C. (1987) *J. Cell Sci. Suppl.*, **6**, 1−23.

Friedberg,E.C. (1988) *Microbiol Rev.*,**52**, 70−102.

Friedberg,E.C., Fleer,R., Naumovski,L., Nicolet,C.M., Robinson,G.W., Weiss,W.A. and Yang,E. (1986) In Shankel,D.M., Hartman,P.E., Kada,T. and Hollaender,A. (eds), *Antimutagenesis and Anticarcinogenesis Mechanisms*. Plenum Publishing Corp., New York, pp. 231−242.

Gorbalenya,A.E., Koonin,E.V., Donchenko,A.P. and Blinov,V.M. (1989) *Nucleic Acids Res.*,**17**, 4713−4730.

Gough,J.A. and Murray,N.E. (1983) *J. Mol. Biol.*, **166**, 1−19.

Grossman,L., Caron,P.R., Mazur,S.J. and Oh,E.Y. (1988) *FASEB J.*, **2**, 2696−2701.

Harosh,I., Naumovski,L. and Friedberg,E.C. (1989) *J. Biol. Chem.*, **264**, 20532−20539.

Hickson,I.D. and Harris,A.L. (1988) *Trends Genet.*, **4**, 101−106.

Higgins,D.R., Prakash,S., Reynolds,P., Polakowska,R., Weber,S. and

Prakash,L. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 5680–5684.
Hoeijmakers,J.H.J., Odijk,H. and Westerveld,A. (1987) *Exp. Cell Res.*, **169**, 111–119.
Johnson,R.T., Elliott,G.C., Squires,S. and Joysey,V.C. (1989) *Hum. Genet.*, **81**, 203–210.
Kadonaga,J.T., Jones,K.A. and Tjian,R. (1986) *Trends Biochem. Sci.*, **11**, 20–23.
Kalderon,D, Richardson,W.D., Markham,A.F. and Smith,A.E. (1984) *Nature*, **311**, 33–38.
Kozak,M. (1989) *J. Cell Biol.*, **108**, 229–241.
Lambert,C., Couto,L.B., Weiss,W.A., Schultz,R.A., Thompson,L.H. and Friedberg,E.C. (1988) *EMBO J.*, **7**, 3245–3253.
Lanford,R.E., White,R.G., Dunham,R.G. and Kanda,P. (1988) *Mol. Cell Biol.*, **8**, 2722–2729.
Lewin,B. (1987) *Genes III*. John Wiley and Sons, New York.
Madhani,H.D., Bohr,V.A. and Hanawalt,P.C. (1986) *Cell*, **45**, 417–423.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Mellon,I., Bohr,V.A., Smith,C.A. and Hanawalt,P.C. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8878–8882.
Mitchell,D.L., Zdzienicka,M.Z., van Zeeland,A.A. and Nairn,R.S. (1989) *Mutat. Res.*, **226**, 43–47.
Mohrenweiser,H.W., Carrano,A.V., Fertitta,A., Perry,B., Thompson,L.H., Tucker,J.D. and Weber,C.A (1989) *Cytogenet. Cell Genet.*, **52**, 11–14.
Montelone,B.A., Hoekstra,M.F. and Malone,R.E. (1988) *Genetics*, **119**, 289–301.
Naumovski,L. and Friedberg,E.C. (1982) *J. Bacteriol*, **152**, 323–331.
Naumovski,L. and Friedberg,E.C. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4818–4821.
Naumovski,L. and Friedberg,E.C. (1986) *Mol. Cell. Biol.*, **6**, 1218–1227.
Naumovksi,L. and Friedberg,E.C. (1987) *Mol. Gen. Genet.*, **209**, 458–466.
Naumovski,L., Chu,G., Berg,P. and Freidberg,E.C. (1985) *Mol. Cell Biol.*, **5**, 17–26.
Okayama,H. and Berg,P. (1983) *Mol. Cell Biol.*, **3**, 280–289.
Picard,D. and Yamamoto,K.R. (1987) *EMBO J.*, **6**, 3333–3340.
Reynolds,P., Higgins,D.R., Prakash,L. and Prakash,S. (1985) *Nucleic Acids Res.*, **13**, 2357–2372.
Richardson,W.D., Roberts,B.L. and Smith,A.E. (1986) *Cell*, **44**, 77–85.
Rubin,J.S. (1988) *Int. J. Radiat. Biol.*, **54**, 309–365.
Sealy,P.G., Whittaker,P.A. and Southern,E.M. (1985) *Nucleic Acids Res.*, **13**, 1905–1922.
Stormo,G.D. (1987) In Bishop,M.J. and Rawlings,C.J. (eds), *Nucleic Acid and Protein Sequence Analysis: A Practical Approach*. IRL Press Ltd, Oxford, UK, pp. 231–258.
Sung,P., Prakash,L., Matson,S.W. and Prakash,S. (1987a) *Proc. Natl Acad. Sci. USA*, **84**,8951–8955.
Sung,P., Prakash,L., Weber,S. and Prakash,S. (1987b) *Proc. Natl. Acad. Sci. USA*, **84**, 6045–6049.
Sung,P., Higgins,D., Prakash,L. and Prakash,S. (1988)*EMBO J.*, **7**, 3263–3269.
Thomas,D.C., Okumoto,D.S, Sancar,A. and Bohr,V.A. (1989) *J. Biol. Chem.*, **264**, 18005–18010.
Thompson,L.H. (1989) *Environ. Mol. Mutagenesis*, **14**, 264–281.
Thompson,L.H. and Bootsma,D. (1988) In Friedberg,E.C. and Hanawalt,P.C. (eds), *Mechanisms and Consequences of DNA Damage Processing*. Alan R. Liss, Inc., New York pp. 279.
Thompson,L.H. and Carrano,A.V. (1983) In Friedberg,E.C. and Bridges,B.A. (eds), *Cellular Responses to DNA Damage*. Alan R. Liss, Inc., New York, pp. 125–143.
Thompson,L.H., Fong,S. and Brookman,K. (1980a) *Mutat. Res.*, **74**, 21–36.
Thompson.L.H., Rubin,J.S., Cleaver,J.E., Whitmore,G.F. and Brookman,K. (1980b) *Somatic Cell Genet.*, **6**, 391–405.
Thompson,L.H., Busch,D.B., Brookman,K., Mooney,C.L. and Glaser,D.A. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 3734–3737.
Thompson,L.H., Brookman,K.W., Dillehay,L.E., Mooney,C.L. and Carrano,A.V. (1982) *Somatic Cell Genet.*, **8**, 759–773.
Thompson,L.H., Mooney,C.L. and Brookman,K.W. (1985a) *Mutat. Res.*, **150**, 423–429.
Thompson,L.H., Mooney,C.L., Burkhart-Schultz,K., Carrano,A.V. and Siciliano,M.J. (1985b) *Somatic Cell Mol. Genet.*, **11**, 87–92.
Thompson,L.H., Salazar,E.P., Brookman,K.W., Collins,C.C., Stewart,S.A., Busch,D.B. and Weber,C.A. (1987) *J. Cell Sci. Suppl.*, **6**, 97–110.
Thompson,L.H., Shiomi,T., Salazar,E.P. and Stewart,S.A. (1988) *Somatic Cell Mol. Genet.*, **14**, 605–612.
Thompson,L.H., Bachinski,L.L., Stallings,R.L., Dolf,G., Weber,C.A., Westerveld,A. and Siciliano,M.J. (1989a) *Genomics*, **5**, 670–679.
Thompson,L.H., Mitchell,D.L., Regan,J.D., Bouffler,S.D., Stewart,S.A., Carrier,W.L., Nairn,R.S. and Johnson,R.T. (1989b) *Mutagenesis*, **4**, 140–146.
van Beveren,C., Goddard,J.G., Berns,A. and Verma,I.M. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 3307–3311.
van Duin,M., de Wit,J., Odijk,H., Westerveld,A., Yasui,A., Koken,M.H.M., Hoeijmakers,J.H.J. and Bootsma,D. (1986) *Cell*, **44**, 913–923.
van Duin,M., Koken,M.H.M., van den Tol,J., ten Dijke,P., Odijk,H., Westerveld,A., Bootsma,D. and Hoeijmakers,J.H.J. (1987) *Nucleic Acids Res.* **15**, 9195–9213.
van Duin,M., Janssen,J.H., de Wit,J., Hoeijmakers,J.H.J., Thompson,L.H., Bootsma,D. and Westerveld,A. (1988a) *Mutat. Res.*, **193**, 123–130.
van Duin,M., van den Tol,J., Warmerdam,P., Odijk,H., Meijer,D., Westerveld,A., Bootsma,D. and Hoeijmakers,J.H.J. (1989b) *Nucleic Acids Res.*, **16**, 5305–5322.
van Duin,M., Vredeveldt,G., Mayne,L.V., Odijk,H., Vermeulen,W., Klein,B., Weeda,G., Hoeijmakers,J.H.J., Bootsma,D. and Westerveld,A. (1989) *Mutat. Res.*, **217**, 83–92.
Weber,C.A., Salarzar,E.P., Stewart,S.A. and Thompson,L.H. (1988) *Mol. Cell Biol.*, **8**, 1137–1146.
Westerveld,A., Hoeijmakers,J.H.J., van Duin,M., de Wit,J., Odijk,H., Pastink,A., Wood,R.D. and Bootsma,D. (1984) *Nature*, **310**, 425–428.
Wood,R.D. and Burki,H.J. (1982) *Mutat. Res.*, **95**, 505–514.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.
Zdzienicka,M.Z. and Simons,J.W.I.M. (1987) *Mutat. Res.*, **178**, 235–244.
Zdzienicka,M.Z., van der Schans,G.P. and Simons,J.W.I.M. (1988a) *Mutat. Res.* **194**, 165–170.
Zdzienicka,M.Z., van der Schans,G.P., Westerveld,A., van Zeeland,A.A. and Simons,J.W.I.M. (1988b) *Mutat. Res.*, **193**, 31–41.

Received on January 26, 1990

Note added in proof

The nucleotide sequence data reported in this paper will appear in the EMBL, GenBank and DDBJ nucleotide sequence databases under the accession numbers X52470