Identification of DNA Repair Genes in the Human Genome

J.H.J. HOEIJMAKERS, M. VAN DUIN, A. WESTERVELD, A. YASUI, AND D. BOOTSMA
Department of Cell Biology and Genetics, Erasmus University, 3000 DR Rotterdam, The Netherlands

The ubiquitous presence of DNA-damaging agents and the instability of certain chemical bonds in DNA have made it necessary for living organisms to develop DNA repair systems. These repair processes help prevent lesions from interfering with essential DNA functions or from converting into permanent mutations that cause cellular malfunctioning and cell death and in higher organisms contribute to malignancy and possibly aging. In view of the wide spectrum of possible lesions, a network of repair systems has evolved to cope with different types of DNA damage (for an extensive review on DNA damage and repair, see Friedberg 1985). In recent years there have been considerable advances in understanding the mechanism and genetic control of two important repair systems in *Escherichia coli*: the excision-repair pathway operating at UV-induced DNA lesions and bulky DNA adducts and the adaptive response directed toward alkylation lesions (recently reviewed by Walker 1985). Genetic studies have identified a number of genes involved in these processes. Application of recombinant DNA techniques has enabled the cloning of the genes and the subsequent purification of the corresponding proteins. The excision pathway, which constitutes part of the SOS response in *E. coli*, is mediated by the concerted action of the *uvrA*, *-B*, *-C*, and *-D* gene products. First, *uvrA* binds to the DNA. In conjunction with *uvrB* a stable complex is formed at the site of the lesion. Incision of the damaged strand (which occurs on both sides of the damage) is catalyzed in the presence of *uvrC*. The unwinding activity of the *uvrD* gene product stimulates the release of the 12- to 13-nucleotide fragment containing the lesion, after which DNA synthesis and ligation complete the excision-repair process.

Current understanding of the mechanisms of repair pathways in eukaryotes is rather limited compared with the knowledge of these pathways in *E. coli*. In the lower eukaryote *Saccharomyces cerevisiae,* more than 30 *RAD* (radiation sensitive) loci have been identified (for a review, see Haynes and Kunz 1981). These fall into three epistatic groups considered to reflect three distinct cellular responses to DNA injury. The *RAD3* epistasis group (consisting of more than 10 loci) is deficient in excision-repair of UV-induced pyrimidine dimers and cross-links. The *RAD6* group (at least 14 members) is disturbed in postreplication repair and double-strand (ds) break repair, and the *RAD52* group (composed of >10 mutants) in a process called recombinational repair. With the recent cloning of a series of genes functioning in these repair pathways (e.g., Yasui and Chevallier 1983; Adzuma et al. 1984; Naumovski et al. 1985; Nicolet et al. 1985; Reynolds et al. 1985a,b), prospects for rapid advances in our understanding of the mechanisms of DNA repair in yeast look favorable.

As far as higher eukaryotes are concerned, valuable tools for the study of repair and the isolation of genes are available in the form of naturally occurring human and laboratory-induced rodent mutants. The human mutants are cell lines derived from patients suffering from hereditary disorders such as xeroderma pigmentosum (XP), Fanconi's anemia, ataxia telangiectasia (AT), and Bloom's syndrome. These rare diseases are characterized by hypersensitivity to specific categories of damaging agents and, in general, display strongly increased incidence of neoplasia (for a review on the clinical and biochemical aspects of these heritable disorders, see Kraemer 1983).

The most extensively investigated repair syndrome is XP. Individuals with this autosomal recessive disorder clinically present extreme skin sensitivity to sun (UV) exposure. Furthermore, XP is associated with predisposition to skin cancer and frequently with neurological abnormalities. The primary defect in most XP patients resides in the excision-repair of UV-induced lesions (Cleaver 1968); a minority of XP cases (XP variants) is suggested to be deficient in a process termed "postreplication repair" (Lehmann et al. 1977). Genetic studies involving cell hybridization have disclosed the existence of nine complementation groups within the excision-deficient class of XP patients (de Weerd-Kastelein et al. 1972; Fischer et al. 1985) and at least five within excision-deficient mutants generated from Chinese hamster ovary (CHO) cells (Thompson et al. 1981, 1982; Thompson and Carrano 1983). All these mutants are unable to perform efficiently the first step postulated in the excision pathway, i.e., the incision of the damaged DNA strand. Since the relationship between the affected human and rodent loci is unknown, these data suggest the involvement of at least 9 and perhaps more than 13 genes and proteins in nucleotide excision. Considerable genetic complexity is also found for AT, in which the response to ionizing radiation is disturbed (Jaspers and Bootsma 1982; Murnane and Painter 1982). Up to now progress with respect to the isolation of the components involved in these processes is limited, due to the complexity of the systems and the experimental limitations of the organisms. Recently, we have cloned the first human gene implicated in the excision of lesions induced by UV and cross-linking agents. In this paper we summarize a molecular and

cell biological characterization of this gene. Detailed information on some of the results presented here has appeared elsewhere (Westerveld et al. 1984; van Duin et al. 1986).

MATERIALS AND METHODS

The isolation of* ERCC-1 *cDNA clones. A human cDNA expression library was kindly provided by Dr. H. Okayama (Okayama and Berg 1983). This library was constructed from poly(A)$^+$ RNA of SV40-transformed human fibroblasts and allows for the expression of the cDNA inserts directed by the SV40 early promoter. Using colony filter hybridization (Maniatis et al. 1982), we have isolated several cDNA clones using a nick-translated, ^{32}P-labeled, 1-kb genomic *Pvu*II fragment as a probe.

Cell culture and DNA transfection. To determine the biological function of the different *ERCC-1* cDNAs, CHO 43-3B cells were transfected with each of the cDNA clones in coprecipitate with the dominant marker pSV3gptH. Culture, transfection, and selection conditions were essentially as previously described (Westerveld et al. 1984). One day prior to transfection, 5×10^5 cells were seeded in 10-cm petri dishes. After 10–14 days of selection, the cells were fixed and clones were counted.

Northern blotting and hybridization. The total RNA of HeLa cells was isolated by the LiCl procedure as described elsewhere (Auffray and Rougeon 1980). Poly(A)$^+$ RNA was obtained by two passages over oligo(dT)-cellulose and electrophoresed in 1% agarose gels containing formaldehyde. Blotting was as described by Maniatis et al. (1982) on nitrocellulose filters (Schleicher & Schuel, ph79). Filters were hybridized at 42°C in the presence of 50% formamide (1 M NaCl) to ^{32}P-labeled, nick-translated *ERCC-1* probes.

DNA sequencing. The complete *ERCC-1* cDNA sequence was determined in two directions following the base-specific chemical cleavage technique of Maxam and Gilbert (1980). To facilitate the sequence strategy, *ERCC-1* cDNA fragments were subcloned in pUC vectors as described by Maniatis et al. (1982).

RESULTS AND DISCUSSION

Molecular Cloning of *ERCC-1*

The strategy followed for the isolation of the human repair gene *ERCC-1* involved genomic DNA transfection of partially digested human DNA fragments to the excision-deficient mutant 43-3B. This mutant, isolated by Wood and Burki (1982), is very sensitive to UV irradiation as well as the cross-linking agent mitomycin C (MM-C) and falls into complementation group 2 described by Thompson et al. (1981). Prior to transfection, a dominant marker gene (*Ecogpt,* Mulligan and Berg 1981) was ligated in vitro to the 40- to 60-kb human restriction fragments to facilitate selection protocols and to tag the transferred gene. After selection for the uptake of the dominant marker gene and subsequently for UV or MM-C resistance, primary transformants were obtained with UV and MM-C survival close to the parental CHO cells. After a second round of transfection and selection using (undigested) DNA of a primary transformant, resistant secondary clones were isolated that contained only a small amount of human sequences and a few copies of the dominant marker gene physically attached to the human DNA. The linkage of *Ecogpt* to the human fragments was used to identify cosmids carrying human sequences in a cosmid library constructed from the DNA of a secondary transformant. One of the cosmids isolated in this way (cos 43-34) was able to confer UV and MM-C resistance to the 43-3B cells with a very high efficiency and apparently carried the human gene (designated *ERCC-1*) responsible for the correction (Westerveld et al. 1984).

Genetic Correction of 43-3B Cells by *ERCC-1*

After transfection *ERCC-1* induces the repair-proficient phenotype in the mutant cells as judged by all repair parameters investigated. These include: pyrimidine dimer removal (measured by the T4 endonuclease assay), UV-induced unscheduled DNA synthesis, UV and MM-C survival, mutability, and induced chromosomal aberrations (Westerveld et al. 1984; M. Zdzienicka; F. Daroudi; both pers. comm.; J. Hoeijmakers et al., unpubl.). Furthermore, correction is complementation-group-specific, since *ERCC-1* does not compensate for excision defects in CHO mutants from other complementation groups than group 2.

43-3B is not corrected in all repair end points to the wild-type level; e.g., colony-forming ability after UV irradiation or in the presence of MM-C is somewhat below that of the parental CHO line (Westerveld et al. 1984). This level is not due to a gene dose effect, since the difference between transformants and wild-type cells is not diminished when amplified *ERCC-1* gene copies are integrated into the genome of the mutant cell. A possible explanation might be that the human gene product is unable to replace fully the Chinese hamster equivalent in the excision process. Further experiments are required to decide between this and other explanations.

Molecular Characterization of *ERCC-1*

Localization of the* ERCC-1 *gene. By restriction enzyme digestion, subcloning of portions of the cos 43-34 in λ vectors, and Southern blot analysis of independent genomic 43-3B transformants, the location of *ERCC-1* could be narrowed down to a 15- to 17-kb region on cos 43-34. A provisional physical map of this region is presented in Figure 1. The isolation of unique probes from this area was hampered by the abundance

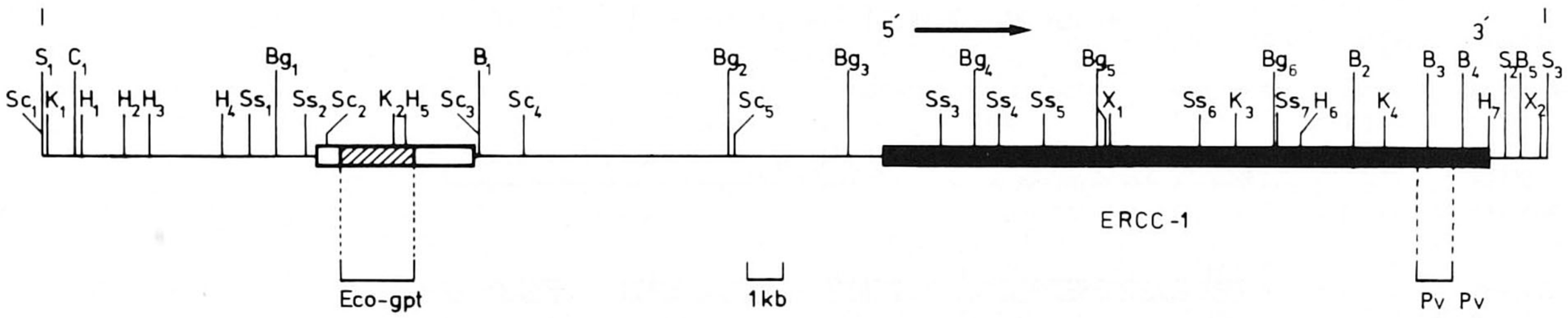

Figure 1. Physical map of the human DNA insert of cosmid 43-34. Cosmid 43-34 contains, in addition to the human excision-repair gene *ERCC-1*, the dominant marker gene *Ecogpt*, which renders transfectants resistant to mycophenolic acid and which was ligated to the human DNA prior to transfection (see text). The position of the genes and the location of the unique 1-kb *Pvu*II probe are indicated. Symbol designation for restriction enzyme cleavage sites: (B) *Bam*HI; (Bg) *Bgl*II; (C) *Cla*I; (H) *Hin*dIII; (K) *Kpn*I; (Pv) *Pvu*II; (S) *Sal*I; (Sc) *Sca*I; (Ss) *Sst*I; (X) *Xho*I. Sites S1 and S3 are from the cosmid vector. Not all *Pvu*II sites are indicated.

of repetitive elements. Nevertheless, the 1-kb *Pvu*II fragment situated at one end of the identified segment (Fig. 1) was found to be largely free of repeats. This fragment was used as probe for Southern and northern hybridizations as well as for screening cDNA libraries. Hybridization with DNA from a well-characterized panel of human/rodent hybrids demonstrated that the presence or absence of the human *ERCC-1* gene correlated well with the presence or absence of chromosome 19, indicating that this chromosome harbored *ERCC-1* (van Duin et al. 1986). Subsequent subchromosomal localization carried out by Brook et al. (1985), using various chromosome 19 translocation hybrids, assigned *ERCC-1* to band q13.2–13.3.

Expression of* ERCC-1 *and isolation of cDNA clones. To obtain information on the size of *ERCC-1* transcript(s), poly(A)$^+$ RNA of HeLa cells was analyzed by northern hybridization. Main hybridization was observed with an RNA species migrating as a relatively broad band at the position corresponding with a size of 1.0–1.1 kb. Weak hybridization was also found with a ~3-kb transcript (Fig. 2). Similar hybridization patterns were obtained with RNA isolated from human fibroblasts, keratinocytes, and blood cells. We conclude from these results that *ERCC-1* is constitutively expressed in a variety of human cell types, and that the 1.0- to 1.1-kb transcript is probably the mature mRNA. Additional evidence (J. Hoeijmakers et al., unpubl.) indicates that the ~3-kb minor transcript is most likely the result of the use of an alternative polyadenylation signal located downstream of the termination site used for the main *ERCC-1* messenger. The ~3-kb RNA species, therefore, possesses a correspondingly longer 3′ untranslated region. Finally, the constitutive level of *ERCC-1* transcripts in HeLa cells is not significantly altered by UV irradiation or MM-C treatment (J. Hoeijmakers et al., unpubl.), suggesting that this repair gene does not belong to a family of inducible genes, as in the case of the SOS response in *E. coli*.

Using the unique *Pvu*II probe from the *ERCC-1* gene region, three cDNA clones (pcD3A, pcD3B7, pcD3C) varying in size from 800 to 1000 bp were isolated from a human expression cDNA library constructed by Okayama and Berg (1983). The physical maps of the inserts are given in Figure 3. Extensive characterization, including sequence analysis, revealed that all three clones lacked different portions of the *ERCC-1* transcript. However, by combining different segments of each clone, a complete cDNA version of the *ERCC-1* transcript could be generated (pcDE, Fig. 3). Comparison of the cDNA sequence with the genomic DNA demonstrated that the gene consists of 10 exons (see Fig. 4), one of which (exon VIII, 72 bp) corresponded exactly with a 72-bp region missing in one of the cDNA

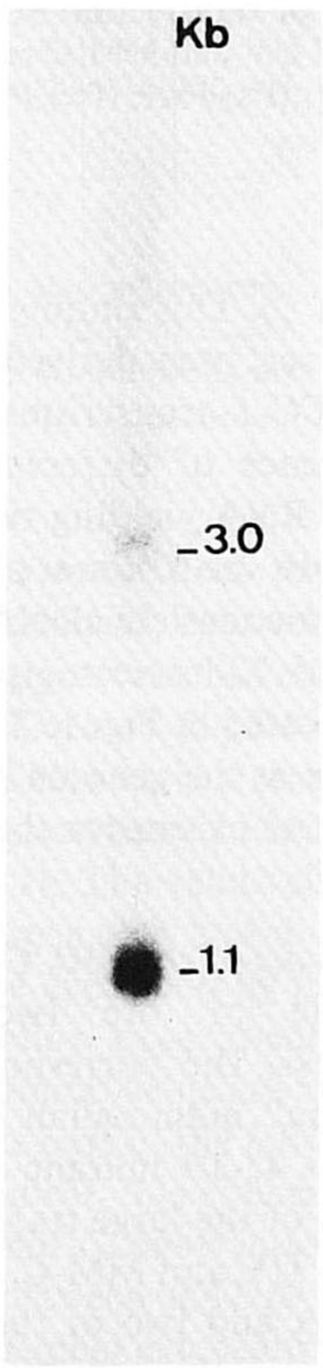

Figure 2. Northern blot analysis of poly(A)$^+$ RNA. Poly(A)$^+$RNA (20 μg) of HeLa cells was size-fractionated by agarose gel electrophoresis in the presence of formaldehyde. After transfer to nitrocellulose, *ERCC-1* mRNA was visualized by hybridization with a ^{32}P-labeled probe from the *ERCC-1* gene.

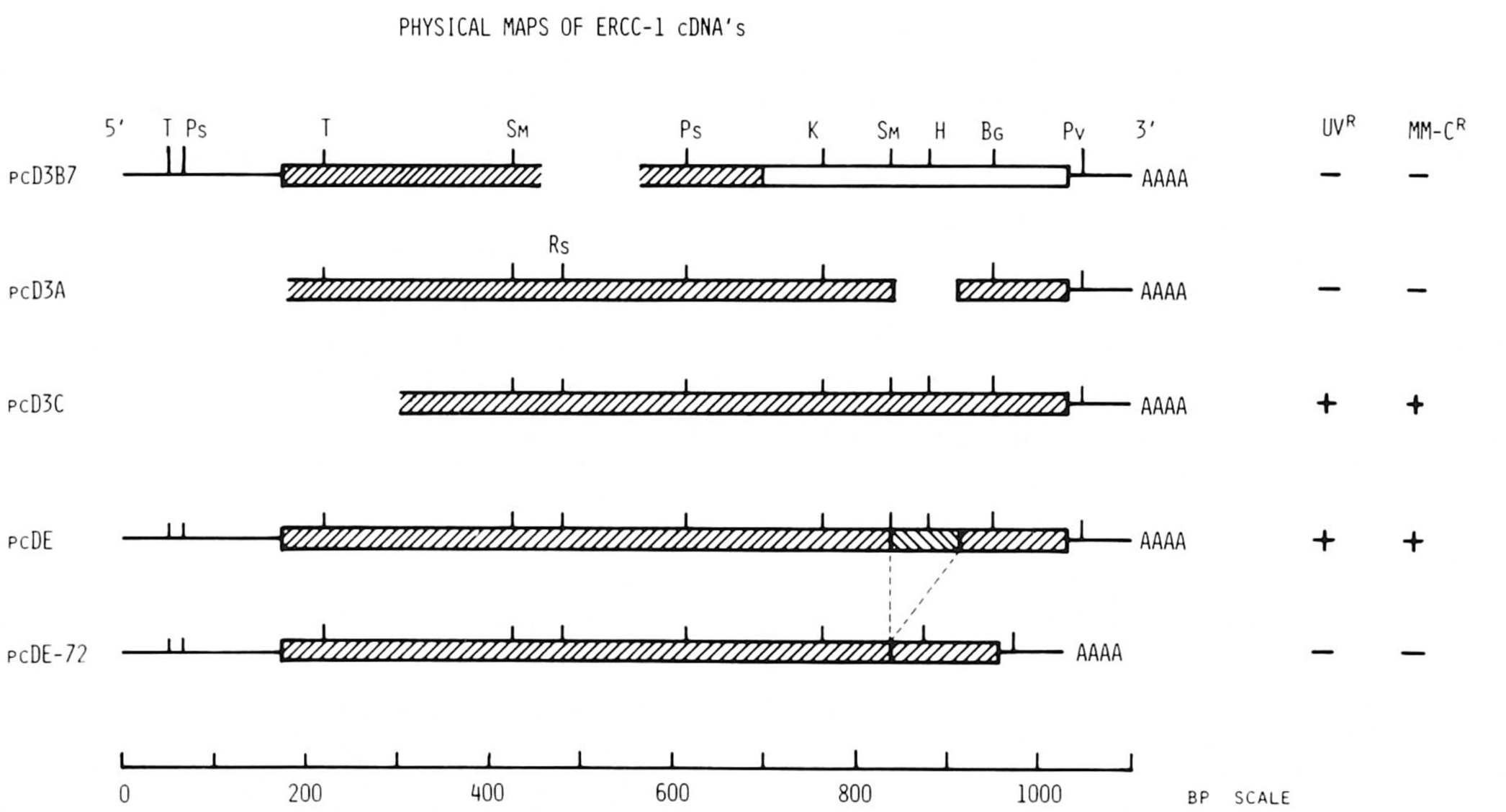

Figure 3. Physical maps of *ERCC-1* cDNAs. The three *ERCC-1* clones isolated from the human expression cDNA library using a genomic *ERCC-1* probe (Fig. 1) are depicted in the upper three rows. Clone pcD3B7, which extends most to the 5′ end has an internal deletion of 104 bp causing a frameshift in the reading frame (shaded area) and premature termination of the coding sequence. Clone pcD3A lacks an internal region of 72 bp, which appears to correspond with exon VIII of the *ERCC-1* gene. The incomplete clone pcD3C stops 161 nucleotides before the start of the open reading frame and therefore lacks 54 amino-terminal codons (a potential ATG start codon is present in the vector). See Fig. 4 for detailed information on the regions covered by these cDNAs. cDNA clone pcDE was constructed by substituting the internal *Sma*I fragment of pcD3B7 with the corresponding fragment of pcD3C. Clone pcDE-72 was obtained by ligation of the *Sma*I–*Bgl*II fragment of pcD3A to the equivalent sites in pcDE. The results of transfection of the cDNA clones to 43-3B cells are summarized in the right part of the figure; (UV^R, $MM\text{-}C^R$) induction of UV and MM-C resistance in 43-3B. Symbol designation for restriction enzyme cleavage sites: (Bg) *Bgl*II; (H) *Hin*dII; (K) *Kpn*I; (Ps) *Pst*I; (Pv) *Pvu*II; (Rs) *Rsa*I; (Sm) *Sma*I; (T) *Taq*I.

clones (pcD3A, Fig. 3). This finding made it very likely that this cDNA insert was derived from an alternatively spliced *ERCC-1* transcript. S1 analysis confirmed the occurrence of differential processing of *ERCC-1* precursor RNA yielding two mature mRNAs of 1.1 and 1.0 kb (see van Duin et al. 1986 for further details). Therefore a complete cDNA clone lacking the alternatively spliced 72-bp exon (pcDE-72) was also constructed as indicated in Figure 2. S1 analysis of the transcriptional start at the genome indicated that both cDNAs were complete except for the first 9–11 nucleotides.

Biological functions of the* ERCC-1 *transcripts. To examine the role of the two transcripts in the excision process, the corresponding cDNAs inserted in "Okayama" mammalian expression vectors were transfected to 43-3B mutant cells. As shown in Figure 3 the cDNA of the large transcript (pcDE) conferred resistance to UV and MM-C, in contrast to that of the small mRNA and two of the three incomplete cDNAs. This rules out the possibility that one of the full-length clones functions in the removal of UV-induced photolesions, and the other in the repair of cross-links. Apparently, the presence of the differentially spliced exon is essential for correction of the mutation in 43-3B cells. The role of the small transcript, if any, is unknown because it is not required for complementing the repair deficiency in the mutant cell.

An interesting observation is provided by the finding that the incomplete clone of pcD3C also compensates for the 43-3B defect (Fig. 3). This clone lacks the first 54 amino acids of the putative *ERCC-1* gene product. Apparently, these are not essential for correction.

Sequence analysis of the* ERCC-1 *cDNAs. The nucleotide and deduced amino acid sequence was determined for both cDNAs and is presented in Figure 4. The sequence contains open reading frames (ORFs) for largely identical polypeptides of 297 amino acids (in the case of pcDE) and 273 amino acids (for pcDE-72). The ORF is preceded by an untranslated region of 142 bp and followed by a 3′ noncoding sequence of 65 nucleotides, which contains the common polyadenylation signal (AATAAA) at the expected distance (~20 nucleotides) from the poly(A) tail. The alternatively spliced exon (Fig. 4, in italics) encodes an internal protein part of 24 amino acids, which is rather rich in threonine residues (~30%). The calculated molecular mass for the two predicted gene products is 32,562 and 29,993 daltons. The hydrophobicity value of −26 and the absence of long hydrophobic stretches suggest that the polypeptides should be water soluble and probably do not represent membrane-associated proteins. Experi-

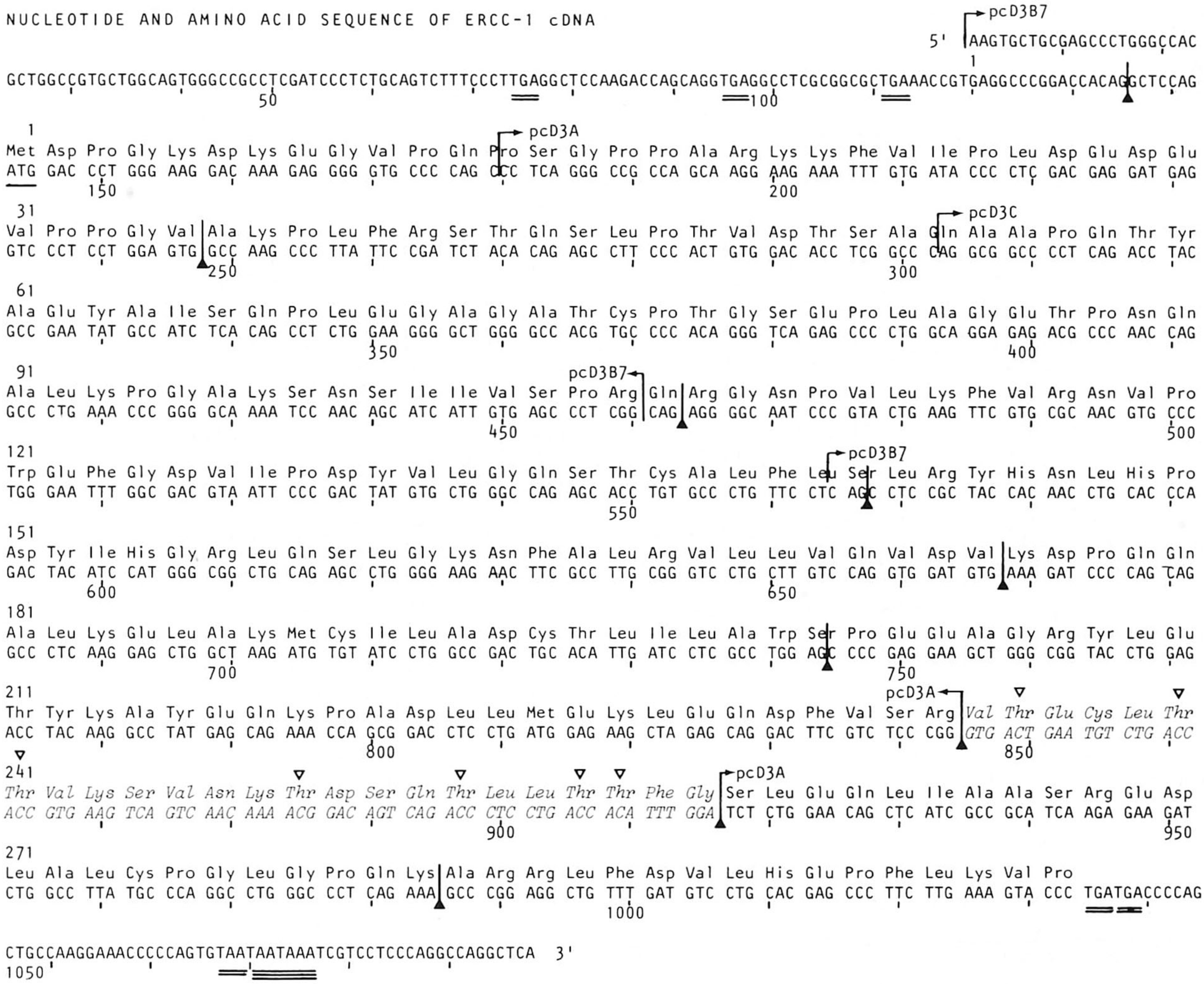

Figure 4. Nucleotide sequence of the *ERCC-1* cDNA clone pcDE and deduced amino acid sequences. The position of exon borders is indicated by ▼ . The alternatively spliced 72-bp exon absent in pcDE-72 is printed in italics. Arrows indicate the regions covered by the various cDNA clones isolated from the Okayama cDNA library. Stop codons and the polyadenylation signal AATAAA are underlined. Sequencing was done by the Maxam and Gilbert procedure.

ments to isolate *ERCC-1* gene product(s) using *E. coli* expression systems and gene amplification in CHO cells are in progress.

Homology of* ERCC-1 *gene products with other proteins. Using DIAGON software (Staden 1982), the *ERCC-1* nucleotide and predicted amino acid sequence was compared with a number of prokaryotic and yeast repair genes and proteins. Significant amino acid homology was detected with the yeast excision-repair protein *RAD10*. Figure 5A presents a possible alignment of both proteins, and quantitative data are compiled in Figure 5B. Since *RAD10* is predicted to be only 210 amino acids (Reynolds et al. 1985b), the *ERCC-1* gene product possesses an extra carboxy-terminal region of 83 amino acids, which harbors the protein segment encoded by the alternatively spliced exon. For the remainder the proteins show 26.5% identity. If substitutions of physicochemically closely related amino acids are allowed (see legend to Fig. 5 for explanation), the overall homology raises to 39%. Particularly, the carboxy-terminal 125 amino acids display a high degree of similarity (34% identical, 51% strongly homologous residues). The finding of extensive homology leads us to believe that *ERCC-1* and *RAD10* probably are evolutionarily related. This idea is strengthened by the fact that the mutant phenotypes of *RAD10* and 43-3B are very similar (Table 1). The relatively low level of homology found between the amino-terminal parts of both proteins can be explained by the indication from the pcD3C transfection experiment (Fig. 3) that the amino-terminal 54 amino acids are not required for 43-3B correction. If this part of the *ERCC-1* protein is not essential for its function, it is conceivable that it is less subject to evolutionary conservation. The main and striking difference between *RAD10* and *ERCC-1* concerns the carboxy-terminal extension of the human gene product absent in the predicted yeast homolog. In this part resides the 24-amino-acid stretch specified by the alternative exon that is shown to be essential for genetic complementation of the 43-3B mutation. It might be that *ERCC-1* has gained additional functions not

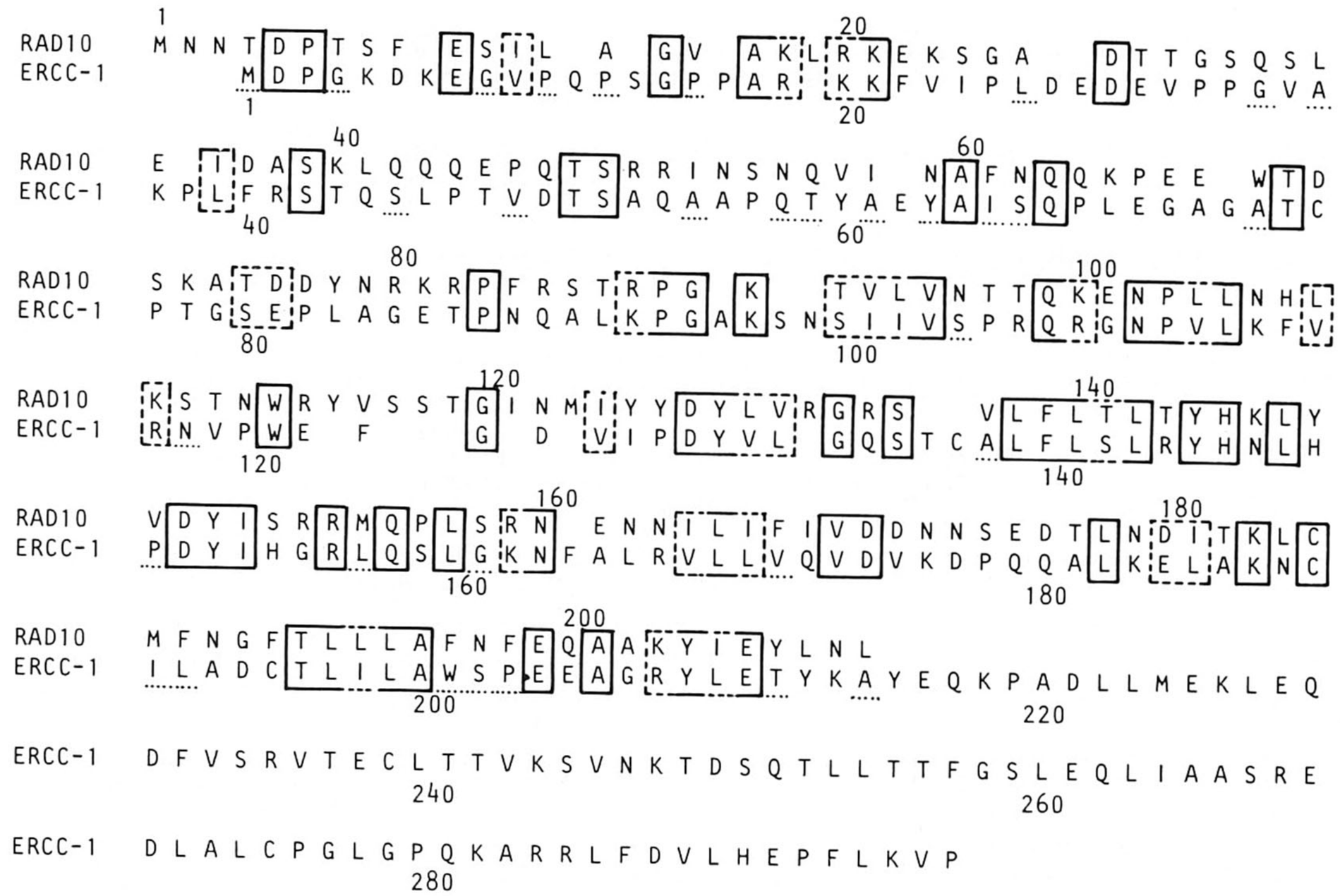

Figure 5. Homologies of the predicted protein sequence of *ERCC-1* to that of *RAD10* and *uvrA*. (*A*) Alignment of amino acid sequence of *ERCC-1* and *RAD10*. (*B*) Schematic representation showing *ERCC-1/RAD10* and *ERCC-1/uvrA* homologies. The table on the right gives percentage of identity and close homology of the region indicated by the arrows. The numbering refers to the amino acid sequence of the respective proteins. (*C*) Alignment of amino acid sequences of part of *ERCC-1* and *uvrA*. Sequence identities are indicated by solid-line boxes; physico-chemically closely related amino acids (K,R; D,E; I,L,V) by dashed-line boxes; and weakly related amino acids (see Schwartz and Dayhoff 1978 for group classification) by dotted underlining. The standard one-letter amino acid abbreviations are used. The *RAD10* and *uvrA* sequences are from Reynolds et al. (1985b) and Husain et al. (1986), respectively.

Table 1. Phenotypic Comparison of Excision-repair Mutants CHO 43-3B and *S. cerevisiae RAD10*

	CHO-43-3B[a]	*RAD10*[a]	References[c] CHO 43-3B	References[c] *RAD10*
UV sensitivity[b]	6–7×	~10×	1	3
UV-induced mutagenesis	38× (0.5–2 J/M²)	15–50× (5–25 J/M²)	1	4
4-NQO sensitivity[b]	5×	~4×	1	5
4-NQO-induced mutagenesis	1.5–3×	~2×	1	5
MMS sensitivity[b]	1.5–2×	1.3×	1	6
X-ray sensitivity	–	–	1	3
Excision deficient (incision⁻)	+	+	2	7
MM-C sensitivity	>100×	unknown	1	

[a]Compared with wild type.

[b]Sensitivities compared at D_{10}.

[c]References: 1. Zdzienicka and Simons (1986); 2. Wood and Burki (1982); 3. Cox and Parry (1968); 4. Lawrence and Christensen (1976); 5. Prakash (1976); 6. Zimmermann (1968); 7. Prakash (1977).

specified by the yeast protein. However, other explanations have not been ruled out, and further research, including mutual complementation experiments using the yeast and human genes and mutants, is required to resolve this issue. Such studies have been initiated.

In addition to the extensive homology between *ERCC-1* and *RAD10*, computer analysis revealed also a regional homology between *ERCC-1* and *uvrA*, which is shown in Figures 5B and C. This concerns a stretch of 42 amino acids that exhibits 31% identity and 52% strong homology. As depicted in Figure 5B, the homologous part in *ERCC-1* overlaps just with the point where the *RAD10* homology stops. If the *uvrA/ERCC-1* homology represents a functional domain shared by the two proteins, it is apparently not present in a complete form in *RAD10*. As nothing is known about the function of this particular region in *uvrA*, the significance of this observation remains to be determined. Comparison of *ERCC-1* with published sequences of other repair proteins (*uvrB*, *alkA*, *phr* of *E. coli* and yeast, bacteriophage T4 *denV*, and yeast *RAD1, -3, -6,* and *-52*) did not reveal significant homologies.

To obtain additional information on possible functional properties of the *ERCC-1* gene product(s), a search was made for amino acid homology with known functional protein domains. Since nucleotide excision is anticipated to take place in the nucleus, the *ERCC-1* amino acid sequence was screened for the presence of sequence motifs resembling nuclear location signals (NLS). The most precisely defined NLS is harbored by the SV40 large T antigens and consists of a series of predominantly basic amino acids, of which Lys-128 by mutation analysis is shown to be essential (Kalderon et al. 1984a) (Fig. 6A). When inserted into other proteins, the sequence is sufficient to redirect the hybrid polypeptide to the nucleus (Kalderon et al. 1984b). A related, but not identical, sequence found at a similar position in polyoma large T exhibits the same properties (Richardson et al. 1986). Figure 6A compares a region in the amino terminus of *ERCC-1* with the two viral NLS. The high level of similarity at important positions strongly suggests that the *ERCC-1* gene product is a nuclear protein. The putative NLS region is located in the apparently nonessential amino-terminal part of the protein. The finding that its absence does not inactivate *ERCC-1* can be explained in two ways. First, the predicted size of *ERCC-1* gene product(s) is such that it can enter the nucleus also by passive diffusion (Paine et al. 1975). Although this process is probably less efficient, it might be sufficient to allow transformants to survive our UV and MM-C selection protocol. Second, the absence of a NLS might be less serious in rapidly dividing cells, such as CHO cells, in which the nuclear membrane is frequently absent due to the high rate of mitotic events.

In addition to a NLS, we have examined whether DNA binding properties could be inferred from the *ERCC-1* amino acid sequence. A well-characterized helix-turn-helix motif has been associated with DNA binding potential for a number of prokaryotic DNA binding proteins (Pabo and Sauer 1984) and recently for homeobox proteins (e.g., Desplan et al. 1985) and yeast mating-type regulatory proteins (e.g., Porter and Smith 1986). A compilation of some of these protein domains is presented in Figure 6B. Specific amino acid positions 5, 8–10, and 15 are considered to be of structural importance, whereas residues 11–13, 16, 17, and 20 (Wharton and Ptashne 1985) and possibly 14 and 19 (Laughon and Scott 1984) are thought to be involved in determining the DNA sequence specificity and hence might vary between different proteins. Comparison of *ERCC-1* with these sequences reveals that identical or related amino acids are present at the important points. Therefore this region, which appears to be the most strongly conserved part between *ERCC-1* and *RAD10*, might comprise a DNA-binding domain. However, direct proof for this hypothesis awaits experiments with the isolated *ERCC-1* protein.

Finally, a region of *ERCC-1* was identified that shared homology with the consensus sequence for ADP-ribosylation found in a family of proteins called G proteins, or guanine nucleotide–binding proteins. Transducin and the *ras* proteins are prominent members of this family. These polypeptides interact with specific receptors and transduce signals by influencing the activity of enzymes that determine the intracellular concentration of "second messengers" such as cAMP

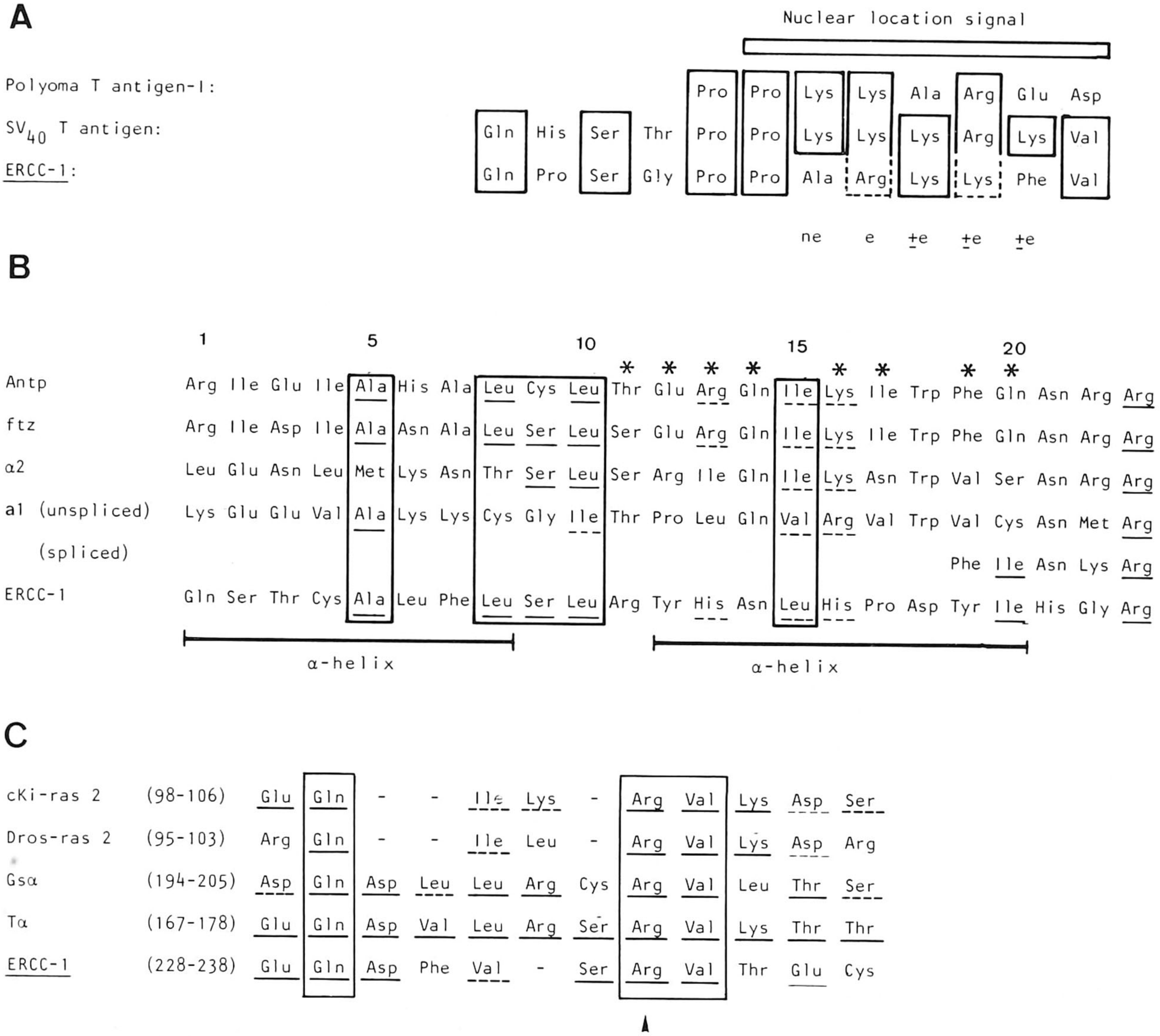

Figure 6. Homologies in amino acid sequences between the predicted *ERCC-1* gene product(s) and various functional protein domains. (*A*) Homology of part of the *ERCC-1* amino acid sequence with NLS of SV40 and polyoma T antigens. Extensive characterization of the NLS of SV40 T antigen is described by Kalderon et al. (1984a,b), and that of the polyoma T antigen by Richardson et al. (1986). Identity is indicated by solid-line boxes, and close homology by dashed-line boxes. The positions of essential (e), important (±e), nonessential (ne) amino acids are deduced from the mutation analysis of the SV40 NLS. The *ERCC-1* region shown corresponds with residues 12–23 of the *ERCC-1* amino acid sequence in Fig. 4. (*B*) Alignment of various DNA-binding protein domains with part of the *ERCC-1* amino acid sequence. The residues at positions 5, 8–10, and 15, which are considered to play an important role in determining the orientation of the two α-helices with respect to each other, are boxed; those amino acids thought to be involved in determining sequence specificity for DNA binding are indicated by an asterisk. The sequence from the *Drosophila* fushi tarazu (*ftz*) and Antennapedia (*Antp*) homeo box proteins and yeast mating-type regulatory proteins a_1 and α2 are taken from the compilation of Shepherd et al. (1984). (*C*) Homology of the *ERCC-1* amino acid sequence with the site for ADP-ribosylation of the α-subunits of transducin (Tα) and the GTP-binding stimulatory protein of adenylate cyclase (Gsα) and homologous regions of two *ras* proteins. Residues that have been retained in all proteins are boxed. Identical and closely related amino acids are indicated by solid and dashed underlining, respectively. The arrowhead points to the arginine residue that is the actual site of ADP-ribosylation by cholera toxin (Van Dop et al. 1984). The *ras* proteins appear not to be good substrates for this reaction (Beckner et al. 1985). The amino acid sequences are from Tanabe et al. (1985) and Robishaw et al. (1986). The residue numbers of the protein part shown are between the brackets.

and perhaps Ca^{++} (Lochrie et al. 1985). ADP ribosylation (e.g., by various bacterial toxins) freezes G proteins in a GTP- or GDP-bound form, thereby interfering with signal transduction. The cholera toxin site for mono-ADP-ribosylation of the α-subunit of transducin has been localized to Arg-174 (Van Dop et al. 1984) (Fig. 6C). Among the residues conserved around this site, the valine immediately following the arginine and a glutamine preceding it at some distance are most strongly conserved (Robishaw et al. 1986). Recently, it has been found that the same acceptor as used by toxins may be modified by the enzymes of host cell origin

(see Ueda and Hayaishi 1985 for a recent review on ADP-ribosylation). Poly-ADP-ribosylation of specific nuclear proteins is closely related to DNA repair events; however, its exact role is not clear. Although it is uncertain whether mono-ADP-ribosyl proteins provide the initiation site for poly-ADP-ribosylation, the homology displayed by *ERCC-1* with the ADP-ribosylation sites of G proteins (Fig. 6C) is suggestive of at least mono-ADP-ribosylation. It is worth noting that the location of the putative ribosylation site is such that alternative splicing of the *ERCC-1* precursor RNA would disrupt the Arg–Val sequence (see Fig. 4). It is likely that this destroys the putative ADP-ribosyl acceptor function of this domain. Evidence for in vivo mono- and/or poly-ADP-ribosylation of one of the *ERCC-1* gene products and analysis of its function can only be obtained by direct experiments.

CONCLUDING REMARKS

We have described molecular and cell biological characterizations of the first human excision repair gene *ERCC-1*. It is clear that more work needs to be done concerning this gene, but also that many more genes implicated in the nucleotide excision process should be isolated. SV40-transformed XP fibroblasts seem to be ideal starting material for the cloning of additional (complementing) repair genes. However, so far they have not successfully been employed for this purpose, notwithstanding many attempts in different laboratories, including ours (Lehmann 1985). We think that this is due at least in part to the fact that the human cells used are very restricted in the amount of DNA that becomes integrated in their genome after genomic DNA transfections. On the average a 20-fold to 100-fold lower amount of exogenous DNA is found in such human transformants, compared for example, with CHO cells (Hoeijmakers et al. 1986). Therefore, it is not surprising that up to now the most promising results have been obtained using laboratory-induced rodent mutants. Transformants corrected by the uptake of human genes have been isolated for a number of CHO repair mutants (Rubin et al. 1983; MacInnes et al. 1984; Thompson et al. 1985), indicating that the isolation of more repair genes is underway. Another potential route to human repair genes is opened by the finding of extensive homology between the *ERCC-1* and *RAD10* gene products. This might indicate that the excision-repair system in toto is strongly conserved during eukaryotic evolution. If so, this may be utilized for the isolation of mammalian genes by virtue of homology with yeast excision-repair genes. The degree of divergence at the nucleotide level between *ERCC-1* and *RAD10* renders it unlikely that the evolutionary distance between both ends of the eukaryotic spectrum can be overcome in one step. Therefore, evolutionary intermediate "stations on the road" from yeast to mammals may have to be taken. At the same time, this study will yield valuable information on evolution of the genes investigated, and the evolutionary conservation and importance of specific parts or functional domains of the gene products. As in the case of *E. coli*, progress in the understanding of mammalian repair depends very much on the success with which these and other approaches will yield genes.

ACKNOWLEDGMENTS

We are very grateful Mrs. H. Odijk, J. van den Tol, and Mr. J. de Wit for excellent technical assistance, Mr. M. Koken and P. ten Dijke for help in some of the experiments, Dr. R. van Gorcom (Medical Biological Laboratory, TNO, Rijswijk, The Netherlands) for indispensible help with the computer, Dr. B. van Houten (University of North Carolina) for giving permission to use the *uvrA* sequence before publication, Drs. M. Zdzienicka and J. Simons for providing the characterization of the 43-3B mutant prior to publication, and Dr. H. Okayama (NIH, Bethesda, Maryland) for the generous gift of the cDNA library. Furthermore, we thank Mrs. R. Boucke for skillful typing of the manuscript and Mr. T. van Os for photography. This work was supported by FUNGO (Foundation of Medical Scientific Research in the Netherlands) and EURATOM contract number BIO-E-404-NL.

REFERENCES

Adzuma, K., T. Ogawa, and H. Ogawa. 1984. Primary structure of the RAD52 gene in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4:** 2735.

Auffray, C. and F. Rougeon. 1980. Purification of mouse immunoglobulin heavy chain messenger RNAs from total myeloma tumor RNA. *Eur. J. Biochem.* **107:** 303.

Beckner, S.K., S. Hattari, and T.Y. Shih. 1985. The *ras* oncogene product p21 is not a regulatory component of adenylate cyclase. *Nature* **317:** 71.

Brook, J.D., D.J. Shaw, A.L. Meredith, M. Warwood, J. Cowell, J. Scott, T.J. Knott, M. Litt, L. Bufton, and P.S. Harper. 1985. A somatic cell hybrid panel for chromosome 19 localization of human genes and RFLPs and orientation of the linkage group. *Cytogenet. Cell Genet.* **40:** 590.

Cleaver, J.E. 1968. Defective repair replication of DNA in xeroderma pigmentosum. *Nature* **218:** 652.

Cox, B.S. and J.M. Parry. 1968. The isolation, genetics and survival characteristics of ultraviolet light-sensitive mutants in yeast. *Mutat. Res.* **6:** 37.

Desplan, C., J. Theis, and P.H. O'Farell. 1985. The *Drosophila* developmental gene, *engrailed* encodes a sequence-specific DNA binding activity. *Nature* **318:** 630.

de Weerd-Kastelein, E.A., W. Keijzer, and D. Bootsma. 1972. Genetic heterogeneity of xeroderma pigmentosum demonstrated by somatic cell hybridization. *Nat. New Biol.* **238:** 80.

Fischer, E., W. Keijzer, H.W. Thielmann, O. Popanda, E. Bohnert, L. Edler, E.G. Jung, and D. Bootsma. 1985. A ninth complementation group in xeroderma pigmentosum, XP-I. *Mut. Res.* **145:** 217.

Friedberg, E.C. 1985. *DNA repair*. W.H. Freeman, San Francisco.

Haynes, R.H. and B.A. Kunz. 1981. DNA repair and mutagenesis in yeast. In *The molecular biology of the yeast* Saccharomyces: *Life cycle and inheritance*. (ed. J. Strathern et al.), p. 371. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Hoeijmakers, J.H.J., H. Odijk, and A. Westerveld. 1986. Differences between rodent and human cell lines in the

amount of integrated DNA after transfection. *Exp. Cell Res.* (in press).

Husain, I., B. van Houten, D.C. Thomas, and A. Sancar. 1986. Sequence of *Escherichia coli uvrA* gene and protein reveal two potential ATP binding sites. *J. Biol. Chem.* **261:** 4895.

Jaspers, N.G.J. and D. Bootsma. 1982. Genetic heterogeneity in ataxia-telangiectasia studied by cell fusion. *Proc. Natl. Acad. Sci.* **79:** 2641.

Kalderon, D., W.D. Richardson, A.F. Markham, and A.E. Smith. 1984a. Sequence requirements for nuclear localization of SV40 large T-antigen. *Nature* **311:** 33.

Kalderon D., B.L. Roberts, W.D. Richardson, and A.E. Smith. 1984b. A short amino acid sequence able to specify nuclear location. *Cell* **39:** 499.

Kraemer, K.H. 1983. Heritable diseases with increased sensitivity to cellular injury. *Update: Dermatology in general medicine* (ed. T.B. Fitzpatrick et al.), p. 113. McGraw-Hill, New York.

Laughon, A. and M.P. Scott. 1984. Sequence of a *Drosophila* segmentation gene: Protein structure homology with DNA-binding proteins. *Nature* **310:** 25.

Lawrence, C.W. and R. Christensen. 1976. UV mutagenesis in radiation sensitive strains of yeast. *Genetics* **82:** 207.

Lehmann, A.R. 1985. Use of recombinant DNA techniques in cloning DNA repair genes and in the study of mutagenesis in human cells. *Mut. Res.* **150:** 61.

Lehmann, A., S. Kirk-Bell, C. Arlett, S.A. Harcourt, E.A. de Weerd-Kastelein, W. Keijzer, and P. Hall-Smith. 1977. Repair of ultraviolet light damage in a variety of human fibroblast cell strains. *Cancer Res.* **37:** 904.

Lochrie, M.A., J.B. Hurley, and M.I. Simon. 1985. Sequence of the alpha subunit of photoreceptor G protein: Homologies between transducin, *ras*, and elongation factors. *Science* **228:** 96.

MacInnes, M.A., J.D. Bingham, L.H. Thompson, and G.F. Strniste. 1984. DNA-mediated cotransfer of excision repair capacity and drug resistance into Chinese hamster ovary cell line UV-135. *Mol. Cell. Biol.* **4:** 1152.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Maxam, A.M. and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65:** 499.

Mulligan, R.C. and P. Berg. 1981. Selection for animal cells that express the *Escherichia coli* gene coding for xanthine-guanine phosphoribosyl transferase. *Proc. Natl. Acad Sci.* **78:** 2072.

Murnane, J.P. and R.P. Painter. 1982. Complementation of the defects in DNA synthesis in irradiated and unirradiated ataxia-telangiectasia cells. *Proc. Natl. Acad. Sci.* **79:** 1960.

Naumovski, L., G. Chu, P. Berg, and E.C. Friedberg. 1985. *RAD3* gene of *Saccharomyces cerevisiae:* Nucleotide sequence of wild type and mutant alleles, transcript mapping, and aspects of gene regulation. *Mol. Cell Biol.* **5:** 17.

Nicolet, C.M., J.M. Chenevert, and E.C. Friedberg. 1985. The *RAD2* gene of *Saccharomyces cerevisiae*: Nucleotide sequence and transcript mapping. *Gene* **36:** 225.

Okayama, H. and P. Berg. 1983. A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. *Mol. Cell. Biol.* **3:** 280.

Pabo, C.O. and R.T. Sauer. 1984. Protein–DNA recognition. *Annu. Rev. Biochem.* **53:** 293.

Paine, P.L., L.C. Moore, and S.B. Horowitz. 1975. Nuclear envelope permeability. *Nature* **254:** 109.

Porter, S.D. and M. Smith. 1986. Homoeo-domain homology in yeast MATα2 is essential for repressor activity. *Nature* **320:** 766.

Prakash, L. 1976. Effect of genes controlling radiation sensitivity on chemically induced mutations in *Saccharomyces cerevisiae*. *Genetics* **83:** 285.

———. 1977. Defective thymidine dimer excision in radiation sensitive mutants *RAD10* and *RAD16* of *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **152:** 128.

Reynolds, P., S. Weber, and L. Prakash. 1985a. *RAD6* gene of *Saccharomyces cerevisiae* encodes a protein containing a tract of 13 consecutive aspartates. *Proc. Natl. Acad. Sci.* **82:** 168.

Reynolds, P., L. Prakash, D. Dumais, G. Perozzi, and S. Prakash. 1985b. Nucleotide sequence of the *RAD10* gene of *Saccharomyces cerevisiae*. *EMBO J.* **4:** 3549.

Richardson, W.D., B.L. Roberts, and A.E. Smith. 1986. Nuclear location signals in polyoma virus large T. *Cell* **44:** 77.

Robishaw, J.D., D.W. Russell, B.A. Harris, M.D. Smigel, and A.G. Gilman. 1986. Deduced primary structure of the α-subunit of the GTP-binding stimulatory protein of adenylate cyclase. *Proc. Natl. Acad. Sci.* **83:** 1251.

Rubin, J.S., A.L. Joyner, A. Bernstein, and G.F. Whitmore. 1983. Molecular identification of a human DNA repair gene following DNA-mediated gene transfer. *Nature* **306:** 206.

Schwartz, R.M. and M.D. Dayhoff. 1978. Matrices for detecting distant relationships. In *Atlas of protein sequence and structure* (ed. M.D. Dayhoff), vol. 5, suppl. 3, p. 353. National Biomedical Research Foundation, Washington, D.C.

Shepherd, J.C.W., W. McGinnes, A.E. Carrasco, E.M. De Robertis, and W.J. Gehring. 1984. Fly and frog homoeo domains show homologies with yeast mating type regulatory proteins. *Nature* **310:** 70.

Staden, R. 1982. An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Res.* **10:** 2951.

Tanabe, T., T. Nukada, Y. Nishikawa, K. Sugimoto, H. Suzuki, H. Takahashi, M. Noda, T. Haga, A. Ichiyama, K. Kangawa, N. Minamino, H. Matsuo, and S. Numa. 1985. Primary structure of the α-subunit of transducin and its relationship to *ras* proteins. *Nature* **315:** 242.

Thompson, L.H. and A.V. Carrano. 1983. Analysis of mammalian cell mutagenesis and DNA repair using in vitro selected CHO cell mutants. *UCLA Symp. Mol. Cell. Biol. New Ser.* **11:** 125.

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.

Thompson, L.H., D.B. Busch, K. Brookman, C.L. Mooney, and P.A. Glaser. 1981. Genetic diversity of UV-sensitive DNA-repair mutants of Chinese hamster ovary cells. *Proc. Natl. Acad. Sci.* **78:** 3734.

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.

Ueda, K. and O. Hayaishi. 1985. ADP-ribosylation. *Annu. Rev. Biochem.* **54:** 73.

Van Dop, C., M. Tsubokawa, H.R. Bourne, and J. Ramachandran. 1984. Amino acid sequence of retinal transducin at the site ADP-ribosylated by cholera toxin. *J. Biol. Chem.* **259:** 696.

van Duin, M., J. de Wit, H. Odijk, A. Westerveld, A. Yasui, M.H.M. Koken, J.H.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.

Walker, G.C. 1985. Inducible DNA repair systems. *Annu. Rev. Biochem.* **54:** 425.

Westerveld, A., J.H.J. Hoeijmakers, M. van Duin, J. de Wit, H. Odijk, A. Pastink, and D. Bootsma. 1984. Molecular cloning of a human DNA repair gene. *Nature* **310:** 425.

Wharton, R.P. and M. Ptashne. 1985. Changing the binding

specificity of a repressor by redesigning an α-helix. *Nature* **316:** 601.

Wood, R.D. and H.J. Burki. 1982. Repair capability and the cellular age response for killing and mutation induction after UV. *Mutat. Res.* **95:** 505.

Yasui, A. and M.-R. Chevallier. 1983. Cloning of photoreactivation repair gene and excision repair gene of the yeast *Saccharomyces cerevisiae*. *Curr. Genet.* **7:** 191.

Zdzienicka, M.Z. and J.W.I.M. Simons. 1986. Analysis of repair processes by the determination of the induction of cell killing and mutations in two repair deficient Chinese hamster ovary cell lines. *Mutat. Res.* **166:** 59.

Zimmermann, F.K. 1968. Enzyme studies on the products of mitotic gene conversion in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **101:** 171.