Molecular Characterization of the Human Excision Repair Gene ERCC-1: cDNA Cloning and Amino Acid Homology with the Yeast DNA Repair Gene RAD10

Marcel van Duin, Jan de Wit, Hanny Odijk, Andries Westerveld, Akira Yasui, Marcel H. M. Koken, Jan H. J. Hoetmakers, and Dirk Bootsma
Department of Cell Biology and Genetics
Erasmus University Rotterdam
P. O. Box 1738
3000 DR Rotterdam, The Netherlands

Summary
The human excision repair gene ERCC-1 was cloned after DNA mediated gene transfer to the CHO mutant 43-3B, which is sensitive to ultraviolet light and mitomycin-C. We describe the cloning and sequence analysis of the ERCC-1 cDNA and partial characterization of the gene. ERCC-1 has a size of 15 kb and is located on human chromosome 19. The ERCC-1 precursor RNA is subject to alternative splicing of an internal 72 bp coding exon. Only the cDNA of the larger 1.1 kb transcript, encoding a protein of 297 amino acids, was able to confer resistance to ultraviolet light and mitomycin-C on 43-3B cells. Significant amino acid sequence homology was found between the ERCC-1 gene product and the yeast excision repair protein RAD10. The most homologous region displayed structural homology with DNA binding domains of various polypeptides.

Introduction
Nucleotide excision repair, which removes DNA lesions like pyrimidine dimers induced by ultraviolet (UV) light and bulky chemical adducts, is the major DNA repair pathway in mammalian cells (Friedberg, 1985). In the hereditary disease xeroderma pigmentosum (XP) defective DNA excision repair is believed to underlie the extreme sensitivity of patients to sunlight and their predisposition to develop skin tumors on exposed parts of the skin (for review see Kraemer, 1983). XP displays a considerable genetic heterogeneity; cell fusion experiments have demonstrated the presence of at least nine complementation groups (de Weerd-Kastelein et al., 1972; Fischer et al., 1985). The genes or gene products that are mutated in this cancer prone disorder are unknown. As an approach to the elucidation of these mutations and the understanding of mammalian DNA repair, a number of Chinese hamster ovary (CHO) cell lines that are sensitive to UV light have been isolated (Wood and Burki, 1982; Thompson et al., 1981; Thompson and Carrano, 1983). Genetic complementation revealed that these mutants constitute at least five different complementation groups (Thompson et al., 1981; Thompson and Carrano, 1983), which are, like XP, all defective in the incision step of the excision repair pathway (Thompson et al., 1982).

With the aid of DNA mediated gene transfer, we recently cloned a human excision repair gene designated ERCC-1 (Westerveld et al., 1984). This gene was cloned by virtue of its ability to correct the excision repair defect in CHO mutant 43-3B, which belongs to complementation group two in the classification of CHO mutants sensitive to UV light (Wood and Burki, 1982) and is also sensitive to mitomycin-C (MM-C). To isolate this gene, human genomic DNA was partially digested with PstI and ligated to the dominant marker pSV3gpH. In a primary transfection of this DNA to 43-3B cells, transformants resistant to mycophenolic acid, which selects for the presence of pSV3gpH, and to UV light or MM-C were isolated. Using genomic DNA of these primary transformants in a secondary transfection, linked transfer of pSV3gpH and the correcting human gene to 43-3B cells could be achieved. This made it possible to isolate ERCC-1 from a cosmid library of a secondary transformant using pSV3gpH probes (Westerveld et al., 1994). The extensive use of XP cells in this approach did not result in the generation of repair proficient transformants (for review see Lehmann, 1985). In contrast, a number of other successful transfections of CHO mutants using human genomic DNA have been reported (Rubin et al., 1983; MacInnes et al., 1984; Thompson et al., 1985a).

We report the cDNA cloning and partial genomic characterization of ERCC-1. Significant homology was found at the amino acid level between ERCC-1 and the Saccharomyces cerevisiae excision repair gene RAD10, suggesting the evolutionary conservation of DNA excision repair. Part of the homologous region has structural homology with DNA binding domains of other polypeptides.

Results
Localization of ERCC-1 on Cos43-34
The cloning strategy of ERCC-1 involved the screening of a cosmid library of a repair proficient secondary 43-3B transformant (Westerveld et al., 1984) of the seven overlapping cosmids isolated from this library, one (cos43-34) contained a functional ERCC-1 gene. The overlap of cos43-34 with the six other cosmids concerned the left-hand region of the insert, indicating that relevant ERCC-1 sequences are on the right-hand part (Westerveld et al., 1984). To narrow down the location of ERCC-1, cos43-34 DNA was partially digested with Sau3A and size fractionated fragments of 15–20 kb were cloned in λEMBL-3 replacement vector. A suitable set of λ-recombinants covering the putative ERCC-1 region was selected for transfection to 43-3B cells in order to screen for a functional ERCC-1 gene. The results of these experiments are shown in Figure 1. Two recombinant clones extending the farthest to the left did not give repair proficient 43-3B transformants. All other phages were positive in this assay. These results limit the position of ERCC-1 on cos43-34 to the 15–17 kb region depicted in Figure 1. The higher number of MM-C resistant transformants generated by...
Cloning of ERCC-1 cDNA

The isolation of unique probes from the genomic ERCC-1 region was hampered by the abundance of repetitive elements. However, a 1.05 kb PvuII fragment situated close to the right end of ERCC-1 (Figure 1) was found to be free of repeats and was used as a probe for screening the human expression cDNA library generously provided by Dr. H. Okayama (Okayama and Berg, 1983). This resulted in a number of hybridizing clones, three of which (pcDSA, pcD3C, and pcD3B7), varying in size from 800 to 1000 bp, will be described in more detail. Northern blot analysis of poly(A)+ RNA of HeLa cells revealed that the cDNA clones hybridized mainly to an mRNA of 1.0-1.1 kb. Faint hybridization with a 3.0 kb RNA species was also observed (Figure 2). Identical results were obtained in a Northern blot analysis of the human chronic myelogenous leukemia cell line K562 and an SV40 transformed human fibroblast line. The simplest interpretation of these results is that the 1.0-1.1 kb mRNA is the mature ERCC-1 transcript. The 3.0 kb band may represent a precursor RNA species.

The aligned physical maps of the three cDNA clones are shown in Figure 3. Sequence analysis of these clones (shown below) revealed that the largest clone, pcD3B7, lacked 104 bp of an internal cDNA region and that clone pcD3A lacked a stretch of 72 bp. However, by substitution of the internal SmaI fragment of pcD3C in the corresponding sites of pcD3B7 a complete ERCC-1 cDNA could be constructed. This clone, designated pcDC, combined all sequences present in the three cDNA clones.

Surprisingly, the 72 bp region that is absent in pcD3A appeared to correspond exactly to a single exon at the genome level. Sequence analysis of genomic ERCC-1 DNA at this position revealed the sequence 5'-caccctttccag-GTGAC...TTGGAgtaaggaatggct-3', which showed that the 72 bp region (capital) is flanked by expected splice donor and acceptor sequences (underlined). Since the chance that an artificial deletion coincides precisely with a single exon is extremely low, this finding rendered it very likely that clone pcD3A was derived from an alternatively spliced ERCC-1 mRNA lacking this 72 bp exon. To obtain additional evidence for differential processing of the ERCC-1 transcript, S1 nuclease analysis was performed. A BamHI-PvuII fragment from cDNA clones pcD3A and pcD3B7 labeled with 32P-ATP at the 3' PvuII site was hybridized to human poly(A)+ RNA and subsequently treated with nuclease S1. The results of these experi-
Figure 3. Cloning of the ERCC-1 cDNA

(A) The screening of the human cDNA expression library with a genomic ERCC-1 probe yielded three overlapping clones: pcD3A, pcD3C, and pcD3B7. These clones were sequenced according to the indicated strategy (arrows). The dotted parts in the aligned physical maps represent deletions. A complete ERCC-1 cDNA (pcDE) was constructed by cloning the internal SmaI fragment of pcD3C in the corresponding sites of pcD3B7. Clone pcDE-72 was obtained by ligation of the SmaI-BglII fragment of pcD3A to the corresponding sites of pcDE.

(B) All stop codons and ATGs in the three reading frames of the ERCC-1 sequence of pcDE (shown in Figure 5) are indicated. Dots: ATG triplets. Vertical bars: stop codons.

ments are shown in Figure 4. After incubation of poly(A)+ RNA from HeLa cells with the pcD3A probe two protected bands of 129 bp and 856 bp were found. These bands can be explained by hybridization of two mRNAs, one completely homologous to the probe and the other differing at a distance of 129 bp from the labeled PvuII site. This position corresponds exactly with the 3` border of the 72 bp deletion found in cDNA clone pcD3A. S1 analysis of poly(A)+ RNA from HeLa and K562 cells with a 3`-labeled probe from pcD3B7 (which includes the 72 bp exon) also yielded a protected band of 129 bp, indicating the presence of ERCC-1 transcripts without the 72 bases. These data indicate that the ERCC-1 precursor RNA is subject to alternative splicing. To obtain complete cDNA clones from both transcripts, in addition to clone pcDE, a cDNA clone (pcDE-72) lacking the 72 bp fragment was constructed by replacing the internal SmaI-BglII fragment of pcDE with the corresponding fragment of pcD3A (see Figure 3).

The construction of the cDNA library by the method developed by Okayama and Berg (1983) enables the expression of full length cDNAs in mammalian cells due to the presence of a strong SV40 promoter. This promoter functions optimally in primate cells, but it also displays considerable activity in CHO cells (Simonsen and Levinson, 1983; Scott McVor et al., 1985; Gorman et al., 1983). In order to investigate the integrity of the cloned ERCC-1 cDNAs, these cDNAs were transfected to the UV light and MM-C sensitive 43-3B cells. The results of these experiments are summarized in Table 1. In contrast to clone pcD3B7, the reconstructed cDNA clone pcDE conferred resistance to UV light and MM-C after transfection to 43-3B cells. This suggests that the 104 bp deletion in pcD3B7 has inactivated the ERCC-1 gene and is most likely a cloning artifact. Reconstructed clone pcDE-72 and clone pcD3A did not compensate for the repair defect in 43-3B cells, indicating that the 72 bp region which is absent in the 3` part of these clones is essential for ERCC-1 functioning in these mutant cells. Surprisingly, clone pcD3C was positive in three independent experiments of this transfection assay. However, sequence analysis (discussed below) revealed that this clone lacks 302 bp of the 5` part including the translational start and 54 N-terminal encoded amino acids. Inspection of the sequence of the pcD3C insert did not show any potential start codons followed by a long open reading frame (see Figure 5B). Apparently, in 43-3B cells transfected with pcD3C the repair defect is corrected by a truncated ERCC-1 protein that is translated from an ATG present in the 5` region of the cDNA expression vector (Okayama and Berg, 1983).

Sequence of ERCC-1 cDNA
Following the strategy depicted in Figure 3A the nucleotide sequence of the ERCC-1 cDNA clones was determined by the chemical modification method developed by Maxam and Gilbert (1980). The nucleotide sequence and deduced amino acid sequence of the 1097 bp insert of
Figure 4. S1 Nuclease Analysis of ERCC-7 mRNA

The probes used were BamHI-PvuII fragments, labeled at the PvuII sites of pcD3A and pcD3B7, which span from the BamHI site (B) in the SV40 part of the vector to the 3' PvuII site (asterisk) in the cDNA. After hybridization with RNA and nuclease S1 treatment the samples were separated on a 6% polyacrylamide gel next to a sequence ladder starting from the PvuII site in pcD3B7. The pcD3A probe was incubated with yeast tRNA (lane 1) and poly(A)+ RNA (lane 2) from HeLa cells. The pcD3B7 probe was incubated with yeast tRNA (lane 3) and poly(A)+ RNA from HeLa cells (lane 4) and K562 cells (lane 5). The arrowheads indicate S1 protected bands. The sequence on the right with numbering according to Figure 5 shows the position at which the two ERCC-1 mRNAs deviate. The diagram shows the protection of two bands of 129 and 856 bp after hybridization of human poly(A)+ RNA with the pcD3A probe.

Clone pcDE is shown in Figure 5. The position of all possible start and stop codons is depicted in Figure 3B. The first ATG of the sequence is followed by the longest open reading frame, of 891 bases. A computer search for protein coding regions based on codon preference (Staden and McLachlan, 1982) gave a strong bias in favor of this reading frame. Moreover, since the other reading frames encode much smaller polypeptides (maximum of 60 amino acids) we conclude that the reading frame translated in Figure 5 specifies the ERCC-1 protein. This corresponds well with the finding that in 95% of all reported cases the 5' proximal ATG serves as the start codon for translation (Kozak, 1984). The open reading frame is preceded by an untranslated region of 142 bp containing three in-frame termination codons. However, it is worth noting that the purine residue found at the -3 position of most eukaryotic ATG start codons (Kozak, 1984) is not present in front of the ERCC-1 translational start codon. The genomic DNA sequence at this point confirmed the cDNA sequence (not shown), ruling out the possibility that this deviation is due to cDNA cloning artifacts. However, the G residue frequently found at position +4 (Kozak, 1984) is present in ERCC-1. In conclusion, the ERCC-1 cDNA clone pcDE encodes a protein of 297 amino acids with a calculated molecular weight of 32,562. The cDNA clone lacking the alternatively spliced 72 bp region might encode a protein of 29,993 daltons. It is remarkable that the region of 24 amino acids lacking in this putative protein is exceptionally rich in threonine residues (about 30%).

Inspection of the 3' region revealed the sequence of the common polyadenylation signal AATAAA (Wickens and Stephenson, 1984) varying in the different cDNA clones from 19 to 21 bases upstream of the poly(A) tail. However, the pentanucleotide sequence CAYTG, which is found adjacent to the polyadenylation site in many eukaryotic mRNAs (Berget, 1984), is not found in the 3' part of the ERCC-1 sequence.

Table 1. Complementation of the Sensitivity of 43-38 Cells to UV Light and MM-C by Transfection with ERCC-1 cDNA

<table>
<thead>
<tr>
<th>cDNA Clone</th>
<th>Correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>pcD3A</td>
<td></td>
</tr>
<tr>
<td>pcD3C</td>
<td>+</td>
</tr>
<tr>
<td>pcD3B7</td>
<td></td>
</tr>
<tr>
<td>pcDE</td>
<td>+</td>
</tr>
<tr>
<td>pcDE-72</td>
<td></td>
</tr>
</tbody>
</table>

ERCC-1 and Yeast RAD10 Protein Show Significant Homology

To determine whether ERCC-1 is partially homologous to other prokaryotic and yeast DNA repair genes a computer analysis was performed using the DIAGON software developed by Staden (1982). No significant homologies were found with published sequences of the genes encoding the E. coli uvrC, Phr, Ack A, bacteriophage T4 den V, and yeast RAD1, RAD3, RAD6, and RAD52 proteins. However, at the protein level significant similarity was found between ERCC-1 and the recently cloned yeast RAD10 DNA repair gene (Weiss and Friedberg, 1985; Prakash et al., 1985; Reynolds et al., 1985). The RAD10 gene encodes a protein of 210 amino acids (Reynolds et al., 1985) which is approximately 90 amino acids smaller than ERCC-1. A comparison of the C-terminal half of RAD10 with the middle part of ERCC-1 is shown in Figures 6A and 6B. Over a region of approximately 110 amino acids 35% homology exists. At the center of this region a stretch of 25 amino acids shows 56% homology. If the amino acids are classified into four groups (Schwartz and Dayhoff, 1978), the group homology in the 110 amino acid region is 63%.

Since ERCC-1 is involved in DNA repair, we have investigated whether DNA binding properties of the protein could be deduced from the amino acid sequence. Based
on amino acid homology with a number of well characterized prokaryotic DNA binding proteins, DNA binding properties have been proposed for the homeo-box proteins of various eukaryotes and yeast mating type regulatory (MAT) proteins (Laughon and Scott, 1984; Shepherd et al., 1984). In case of the yeast MATa2 gene product, DNA binding capacity has been demonstrated (Johnson and Herskowitz, 1985). A comparison of ERCC-7 and RADlO comprises a DNA binding domain (Pabo and Sauer, 1984). Moreover, similarities between the compared protein domains were also found in the flanking positions (e.g. positions 22-24). From these observations it is tempting to speculate that the most conserved region of human ERCC-7 sequences. A representative Southern blot analysis is shown in Figure 7. The overall results of this screening are presented in Table 2. The highest correlation is found with chromosome 19. No hybrids were found in the important category in which chromosome 19 is present and ERCC-7 is absent (+I- column, Table 2) in which ERCC-7 by 19 could be detected cytologically, but in which ERCC-7 by

<table>
<thead>
<tr>
<th>Amino Acid Sequence of Human ERCC-7 cDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acids are numbered on the left and nucleotides below each line. The sequence of 1097 bp is derived from the ERCC-7 cDNA clones as shown in Figure 3 and represents the insert of pcDE. The alternatively spliced exon and polyadenylation signal are underlined. The (asterisks mark the translational stop codons.</td>
</tr>
</tbody>
</table>

Figure 3. Composite Nucleotide and Encoded Amino Acid Sequence of Human ERCC-1 cDNA

Amino acids are numbered on the left and nucleotides below each line. The sequence of 1097 bp is derived from the ERCC-1 cDNA clones as shown in Figure 3 and represents the insert of pcDE. The alternatively spliced exon and polyadenylation signal are underlined. The (asterisks mark the translational stop codons. |
A

| RAD 10 | Thr Val Leu Val Asn Thr Thr Glu Val Asp Pro Leu Leu Asn His Leu Val Ser Thr Asn Arg Tyr Tyr Val Ser Ser Thr Thr Glu Met |
| ERCC-1 | Ser Ile Ile Ser Pro Arg Glu Asp Pro Val Lys Phe Val Arg Asp Val Pro Thr |

| RAD 10 | Asn Met Ile Tyr Tyr Asp Tyr Leu Val Arg Gly Arg Ser Val Leu Phe Leu Leu Thr Tyr His Lys Leu Tyr Val Asp Tyr Ile |
| FER 1 | Asp Val Pro |

| RAD 10 | Ser Arg Asn Glu Asp Leu Glu Ser Ser Asn Leu Leu Ala Leu Arg Val Leu Leu Val Leu Val Arg Glu Val Asp Pro Gin Glu Ala Leu Lys |
| ERCC-1 | His Gly Arg Leu Gin Ser Ser Gin Ser Leu Gly Phe Asp Phe Ala Ala Leu Arg Val Leu Leu Val Leu Val Arg Glu Val Asp Pro Gin Glu Ala Leu Lys |

| RAD 10 | Asp Ile Ile Thr Leu Cys Met Phe Asp Gly Phe Thr Leu Leu Leu Leu Ala Phe Asn Phe Glu Gin Gin Ala Ala Tyr Ile Thr Ile Tyr Leu Leu |
| ERCC-1 | Glu Leu Ala Met Met Leu Leu Ala Asp Cys Thr Leu Leu Leu Leu Ala Trp Ser Pro Gin Gin Glu Gin Gin Gly Arg Arg Thr Thr Tyr Lys |

B

Figure 6. Comparison of ERCC-1 with Other Proteins

(A) ERCC-1 amino acids 100-213 aligned with the yeast RAD10 C-terminal part starting from amino acid 92 to 209. The RAD10 sequence is from Reynolds et al. (1985). The earlier published RAD10 nucleotide and deduced amino acid sequence of Weiss and Friedberg (1985) deviates in the 3' region (from codon 170) from that of Reynolds et al. (1985). Only with the latter sequence, a substantial homology is found in this region. There is now agreement that the sequence determined by Reynolds et al. (1985) is correct (Friedberg, personal communication). Closed boxes and open boxes indicate homologous and related basic or acidic amino acid residues respectively.

(B) Schematic diagram showing ERCC-1 and RAD10 proteins. The table on the right gives the percentage of homology of the regions indicated by arrows between the dashed lines. The numbering corresponds to the amino acid sequence of both proteins.

(C) Amino acid comparison of ERCC-1 and RAD10 and UMA binding domains in Drosophila fushi tarazu (ftz) and Antennapedia (Antp) homeo box proteins and yeast mating type regulatory proteins α1 and α2 as compiled by Shepherd et al. (1984). The positions of the two α-helices of the DNA binding domain are shown at the bottom. The amino acids at positions 5, 8-10, and 15, which are important for the formation of the α-helical structures, are boxed. Amino acids homologous to ERCC-1 are underlined by a solid bar. In addition, thin lines refer to amino acids homologous to RAD10. Dashed lines indicate related basic and acidic amino acid residues.

Southern blot hybridization was shown to be present. When these exceptional hybrids were analyzed by enzyme electrophoresis, it appeared that the chromosome 19 glucose phosphate isomerase (EC 5.3.1.9) was present, indicating that these hybrids had retained cytologically undetectable fragments of chromosome 19. From these data we conclude that ERCC-1 is in human chromosome 19. Additional screening of a number of hybrids with translocations in chromosome 19 (Worwood et al., 1985) revealed that ERCC-1 is most likely located on 19q13.2-q13.3 (Brook et al., 1985).

Discussion

We have presented the molecular characterization of the human excision repair gene ERCC-1. Transfection of mu-
Human Excision Repair Gene ERCC-7

Figure 7. Southern Blot Analysis of PstI Digested DNA (15 μg) from Seven Human × Chinese Hamster Somatic Cell Hybrids and Control HeLa and CHO Cells with a 32P-Labeled ERCC-1 cDNA Probe. The arrow indicates hybridization with the Chinese hamster ERCC-1 gene. Hybrids 2, 3, 4, and 6 retained the human ERCC-1 sequences, whereas the others did not.

Table 2. Relationship between the Human ERCC-1 Gene and Human Chromosomes in 45 Human-Rodent Somatic Cell Hybrid Clones

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>+/+</th>
<th>+/-</th>
<th>-/+</th>
<th>-/-</th>
<th>% Concordance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
<td>5</td>
<td>18</td>
<td>8</td>
<td>49</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>2</td>
<td>27</td>
<td>11</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>4</td>
<td>15</td>
<td>9</td>
<td>58</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>3</td>
<td>22</td>
<td>10</td>
<td>44</td>
</tr>
<tr>
<td>5</td>
<td>17</td>
<td>4</td>
<td>15</td>
<td>9</td>
<td>58</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>5</td>
<td>13</td>
<td>8</td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td>3</td>
<td>17</td>
<td>10</td>
<td>56</td>
</tr>
<tr>
<td>8</td>
<td>18</td>
<td>5</td>
<td>14</td>
<td>8</td>
<td>58</td>
</tr>
<tr>
<td>9</td>
<td>13</td>
<td>3</td>
<td>19</td>
<td>10</td>
<td>51</td>
</tr>
<tr>
<td>10</td>
<td>14</td>
<td>3</td>
<td>18</td>
<td>10</td>
<td>51</td>
</tr>
<tr>
<td>11</td>
<td>20</td>
<td>5</td>
<td>12</td>
<td>8</td>
<td>62</td>
</tr>
<tr>
<td>12</td>
<td>21</td>
<td>2</td>
<td>11</td>
<td>11</td>
<td>71</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>3</td>
<td>19</td>
<td>10</td>
<td>51</td>
</tr>
<tr>
<td>14</td>
<td>16</td>
<td>4</td>
<td>16</td>
<td>9</td>
<td>55</td>
</tr>
<tr>
<td>15</td>
<td>13</td>
<td>2</td>
<td>19</td>
<td>11</td>
<td>53</td>
</tr>
<tr>
<td>16</td>
<td>18</td>
<td>3</td>
<td>14</td>
<td>10</td>
<td>62</td>
</tr>
<tr>
<td>17</td>
<td>13</td>
<td>4</td>
<td>19</td>
<td>9</td>
<td>49</td>
</tr>
<tr>
<td>18</td>
<td>13</td>
<td>3</td>
<td>19</td>
<td>10</td>
<td>51</td>
</tr>
<tr>
<td>19</td>
<td>29</td>
<td>0</td>
<td>3</td>
<td>13</td>
<td>93</td>
</tr>
<tr>
<td>20</td>
<td>15</td>
<td>3</td>
<td>17</td>
<td>10</td>
<td>56</td>
</tr>
<tr>
<td>21</td>
<td>25</td>
<td>4</td>
<td>7</td>
<td>9</td>
<td>76</td>
</tr>
<tr>
<td>22a</td>
<td>19</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>66</td>
</tr>
<tr>
<td>X</td>
<td>23</td>
<td>5</td>
<td>9</td>
<td>8</td>
<td>69</td>
</tr>
</tbody>
</table>

a +/+ and -/- indicate the presence and absence of the human chromosome. +/+ and -/− refer to the presence and absence of human ERCC-1 sequences as detected by Southern hybridization.

b The arrow indicates the chromosome with the highest concordance. Four clones containing chromosome 22 translocations were excluded.
gene family. This rules out the possibility that cDNA clone pcD3A is derived from a transcribed related gene.

Alternative splicing has been found in a number of other gene systems. As a result of this mode of RNA processing single genes are able to meet different functional demands of the cell. In most cases the reported differences between alternatively spliced mRNAs concern the 5' and 3' ends of the transcript. In these cases the use of alternative promoters and polyadenylation sites provokes the generation of different mRNAs. The inclusion or exclusion of a separate internal coding exon, as occurs during the processing of ERCC-1 mRNA, has been reported in a few other cases such as the Drosophila myosin (Rezok and Davidson, 1983), the rodent αA-crystallin (King and Platigorsky, 1983), and the bovine preprotachikinin (Nawar et al., 1984) genes. However, in the first two cases it is not known whether the alternative splicing serves any function. Only for the preprotachikinin gene a tissue specific splicing of a single exon, yielding two functional mRNAs, has been reported (Nawar et al., 1984).

In the case of ERCC-1, our transfection experiments indicate that only the cDNA derived from the larger transcript is able to complement fully the excision defect in 43-38 cells. This rules out the possibility that one of the ERCC-1 gene products is involved in the repair of lesions such as those caused by UV light and the other in the removal of damages caused by cross-linking agents such as MM-C. The significance of the smaller transcript of ERCC-1 is still uncertain. First, the possibilities that it has no function and that it is the result of an artifact in the splicing system have not been excluded. Second, it is possible that the smaller gene product is involved in the removal of lesions other than those induced by UV light or MM-C. Finally, the mutation of 43-38 cells might inactivate only the larger gene product. Consequently, one cannot expect complementation with the smaller ERCC-7 cDNA, although this might be essential for the repair process as well. Further studies are required to discriminate between these possibilities.

Comparison of the ERCC-1 encoded amino acid sequence with the sequences of isolated DNA repair genes from prokaryotes and yeast revealed striking homology with the S. cerevisiae excision repair gene RAD10. The sequence of this gene has recently been determined by Reynolds et al. (1985). The putative RAD10 and ERCC-1 proteins consist of 210 and 297 amino acids respectively. The C-terminal half of RAD10 displays significant homology with the central 110 amino acids of ERCC-1. Furthermore, from amino acid sequence homology we have tentatively identified a DNA binding domain in the most homologous part of this region. A general feature of DNA binding protein domains is the presence of an α-helix-turn-α-helix motif, which is involved in the DNA-protein interaction (for review see Pabo and Sauer, 1984). With respect to the amino acids that are important for the configuration of these α-helices, structural homology was found between a number of prokaryotic DNA binding proteins, eukaryotic homeo-box domains, and yeast MAT proteins (Laughon and Scott, 1984; Shepherd et al., 1984). A comparison of ERCC-1 and RAD10 with the eukaryotic proteins revealed that the most conserved region of ERCC-1 and RAD10 shows a structural homology with the putative DNA binding domains of these polypeptides. Concerning ERCC-1, the α-helical propensities of the individual amino acids of the putative DNA binding region (calculated according to Finkelstein and Pitéan, 1979) are not incompatible with the assumed α-helical structures that comprise the DNA binding domain. However, the presence of the proline at position 17 in the middle of the C-terminal α3 helix of ERCC-1 (Figure 6) is worth noting. In general, prolines are considered strong helix breakers, introducing kinks in the secondary structure of a protein (Chou and Fasman, 1974). Although proline residues are found in the α3 helix of some of the prokaryotic DNA binding proteins they are located at the N-terminal part (and rarely at the C terminus) of the helix but not in the center (Pabo and Sauer, 1984). This helix is believed to be in direct contact with the major groove of the DNA; therefore it is tempting to speculate that the deformation of the putative α3 helix in ERCC-1 caused by proline might be related to the structural deformation in the DNA helix caused by DNA lesions induced by UV light or MM-C. However, it is evident that X-ray diffraction and two-dimensional NMR studies on the purified protein-DNA complexes are required to test this hypothesis. In conclusion, we consider the structural homologies between the highly conserved region of ERCC-1 and RAD10 and the DNA binding domains of various eukaryotic DNA binding proteins strong enough to suggest that the role of ERCC-1 and RAD10 in the removal of DNA damage is mediated through a DNA-protein interaction.

Since the homology covers only a part of the RAD10 and ERCC-1 proteins one can question whether the homology only reflects a common DNA binding property or whether both genes are evolutionary related and serve a common DNA repair function. Several observations favor the last option. First, the homology in the DNA binding domains of unrelated peptides predominantly concerns amino acids at fixed positions (Pabo and Sauer, 1984) and spans a region of 20–25 amino acids, which is much smaller than the homologous region of ERCC-1 and RAD10. Second, the positive transfection experiments with the truncated ERCC-1 cDNA clone pcD3C showed that the absence of 54 N-terminal amino acids does not inactivate the ERCC-1 gene product in 43-3B cells. If this region of the protein is not essential for its function it is conceivable that it is less subject to evolutionary conservation. Third, the mutant phenotypes of the 43-3B cell line and the yeast rad10 strain have a number of characteristics in common. Similar to 43-3B, the rad10 strain is sensitive to UV light, associated with a defective incision step of the excision repair pathway. In addition, 43-3B and rad10 both are sensitive to 4-nitroquinoline-1-oxide (4NQO) and methylmethanesulfonate and both have enhanced mutagenesis induced by UV light and 4NQO (Prakash, 1976; Zdzienicka and Simons, 1986; Haynes and Ku, 1981). To confirm the evolutionary relationship between ERCC-1 and RAD10 it will be of interest to establish whether rad10 is also sensitive to cross-linking agents like MM-C, and whether the repair defects in rad10 and 43-3B can be corrected by introduction of the human ERCC-1
and the yeast RAD10 genes respectively. If such studies support the idea that ERCC-1 and RAD10 are derived from the same ancestor gene, it is evident that evolutionary changes have resulted in a large difference in size between both proteins. ERCC-1 has approximately 90 more C-terminal amino acids than RAD10. The alternatively spliced coding exon from ERCC-1 is located in this region and was essential for ERCC-1 activity in 43-36 cells. If ERCC-1 and RAD10 provide a similar function in DNA repair this would mean either that the functional domains of the two proteins are organized in a different fashion or that additional yeast proteins serve the function of the C-terminal part of the ERCC-1 gene product.

With respect to the first possibility it is interesting that computer analysis also revealed homology between the C-terminal "extra" part of ERCC-1 and the region of RAD10 that already exhibited extensive homology with the middle portion of ERCC-1. This concerns a stretch of 18 amino acids (residues 222-239 in ERCC-1 and 140-157 in RAD10) that has a homology of 38% (Figure 8B). The finding that within ERCC-1 there are two regions which share homology with the same part of RAD10 suggests the existence of internal homology within the ERCC-1 protein. Indeed, the indicated 18 amino acids have partial homology with a more central region of ERCC-1 (residues 143-160, Figure 8B). In addition, a very homologous region of 10 amino acids occurs twice in ERCC-1 at positions 88-97 and 178-187 (Figure 8A). Taken together, these data provide evidence suggestive of a duplication of part of the ERCC-1 gene in the course of evolution, which might explain the difference in size between ERCC-1 and RAD10.

Genetic analysis of a large number of yeast and higher eukaryotic mutant cells deficient in excision repair led to the identification of different complementation groups. The interspecies relationship of these complementation groups is unknown. Our finding that ERCC-1 and RAD10 are in part homologous is the first indication for evolutionary conservation of DNA repair systems. This emphasizes the importance of yeast and rodent DNA repair deficient mutants in studies which are aimed at the identification of the primary defect in human hereditary diseases associated with defects in DNA repair.

### Experimental Procedures

#### General Procedures

Purification of nucleic acids, restriction enzyme digestion, gel electrophoresis, transfer of DNA and RNA to nitrocellulose, nick translation, and filter hybridization were performed according to established procedures as described by Maniatis et al. (1982).

#### DNA Transfection

In order to screen for a functional ERCC-1 gene, DNA from λEMBL3 recombinants containing fragments of cos43-34 or ERCC-1 cDNA clones were cotransfected with the dominant marker pSV3gptH to CHO 43-38 cells. Usually 1-3 μg of DNA was cotransfected with 2 μg of pSV3gptH DNA. Transfection and selection procedures were essentially as previously described (Westerveld et al., 1984). One day before transfection ±5 x 10⁶ cells were seeded in 100 mm petri dishes. After 10-14 days of selection the cells were fixed and clones were counted.

#### Molecular Cloning in λEMBL3

Cos43-34 was partially digested with Sau3A and separated in a low melting agarose gel. Fragments of 15-20 kb were isolated and cloned in the BamHI site of a λEMBL3 replacement vector (Frischauf et al., 1983).

#### Isolation of cDNA Clones

A cDNA library made from a human SV40 transformed fibroblast was generously provided by Dr. H. Okayama (Okayama and Berg, 1983). In this cDNA expression library the cloned cDNA insert is under the direction of the SV40 early region promoter sequences, allowing a direct use of the cDNA clones in transfection experiments with mammalian cells. The library was screened with colony filter hybridization with a 3²P-labeled nick-translated genomic 105 kb PvuII probe.

#### Subcloning and DNA sequencing

For detailed mapping of restriction sites, parts of cos43-34 were subcloned in pUC-vectors by standard procedures (Maniatis et al., 1982). The nucleotide sequence of the ERCC-1 cDNA clones was determined following the chemical modification procedure developed by Maxam and Gilbert (1980).

#### S1 Mapping Procedure

The nuclelease S1 protection experiments were carried out according to the procedure described by Grosveld et al. (1981). For the preparation of the probes, the 105 PvuII sites of cDNA clones pCD3A and pCD3B were labeled with polynucleotide kinase and γ-3²P-ATP.

#### Chromosomal Localization

Southern blot analysis was done on DNA from a panel of 45 human × rodent (mouse or Chinese hamster) somatic cell hybrids. Three different Chinese hamster cell lines, designated A3, E36, and CHO,
and two mouse cell lines. P109 and Wehi-3B, have been used for the construction of the hybrids according to previously described procedures (de Wit et al., 1979, Geurts van Kessel et al., 1981). Prior to DNA isolation the chromosomal content of the hybrids was determined. At least 16 metaphases per hybrid were analyzed.

Acknowledgments

We are very grateful to Dr. H. Okayama for providing the cDNA library, Drs. L. Prakash and S. Prakash for giving permission to use the RAD10 sequence before publication, Dr. A. H. M. Geurts van Kessel for the hybrid panel used for the chromosomal localization, Drs. D. Shaw and T. Hulsebos for the regional chromosomal mapping, P. ten Djike and R. Kanaar for the excellent help in the experiments, T. Kartasova for contribution in the cDNA sequencing, and Dr. G. C. Grosveld for helpful suggestions. Furthermore, we thank Mrs. R. Boucher for typing the manuscript and Mr. T. van Os for photography. This work was supported by Eukarot contract no. J66-141 NL and FUNGO, Foundation of Medical Scientific Research in the Netherlands.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 7, 1985; revised December 16, 1985.

References


Note Added in Proof

We have recently identified a region in the deduced ERCC-1 protein (amino acids 12–23) that shows a striking homology with the nuclear location signal of the SV40 large T antigen (Kalderon et al., Cell 39, 499–509, 1984; Kalderon et al., Nature 317, 33–38, 1984).