Volume 15 Number 22 1987 Nucleic Acids Research

Genomic characterization of the human DNA excision repair gene *ERCC-1*

Marcel van Duin, Marcel H.M.Koken, Jacolien van den Tol, Peter ten Dijke, Hanny Odijk, Andries Westerveld, Dirk Bootsma and Jan H.J.Hoeijmakers

Department of Cell Biology and Genetics, Erasmus University, PO Box 1738, 3000 DR Rotterdam, The Netherlands

Received September 11, 1987; Accepted October 22, 1987

ABSTRACT

In this report the genomic characterization of the human excision repair gene ERCC-1 is presented. The gene consists of 10 exons spread over approximately 15 kb. By means of a transfection assay the ERCC-1 promoter was confined to a region of ± 170 bp upstream of the transcriptional start site. Classical promoter elements like CAAT, TATA and GC-boxes are absent from this region. Furthermore, ERCC-1 transcription is not UV-inducible. A possible explanation is provided for the previously reported alternative splicing of exon VIII. Analysis of ERCC-1 cDNA clones revealed the occurrence of differential polyadenylation which gives ERCC-1 transcripts of 3.4 and 3.8 kb in addition to the major 1.1 kb mRNA. Apparent evolutionary conservation of differential polyadenylation of ERCC-1 transcripts suggests a possible role for this mode of RNA processing in the ERCC-1 repair function.

INTRODUCTION

In order to cope with DNA-lesions induced by mutagenic and carcinogenic agents, living organisms have developed a variety of DNA repair systems. One of the most important and best studied repair processes is the excision repair pathway (see 1 for a review). In Escherichia coli this system constitutes part of the damage inducible SOS-response. Lesions causing a relatively strong deformation of the normal helix structure (such as bulky adducts or ultraviolet light (UV) induced pyrimidine dimers) are recognized and removed by the concerted action of at least 4 excision repair proteins: uvr A,B,C and D (2). In yeast, mutant analysis and gene cloning have demonstrated the existance of more than 10 genetic loci implicated in excision repair (1). In higher eukaryotes cell hybridization experiments have revealed the presence of at least 9 complementation groups in cells from excision deficient xeroderma pigmentosum patients (3,4) and of at

© IRL Press Limited, Oxford, England.

least 5 groups in laboratory induced CHO repair mutant cells (5,6). This extensive genetic heterogeneity suggests a considerable biochemical complexity underlying the excision repair pathway. However, the mechanism of this system in eukaryotes is poorly understood.

We have recently cloned the first human repair gene, designated ERCC-1, that corrects the excision defect in CHO mutants of complementation group 2 (7). The major ERCC-1 mRNA is 1.1 kb and specifies a protein of 32.5 kD. Evidence was obtained for the occurrence of alternative splicing yielding in addition a transcript of 1.0 kb, lacking a 72 bp coding exon (8). However, only the cDNA corresponding to the 1.1 kb mRNA, inserted in a mammalian expression vector was able to correct the repair defect in CHO mutants of group 2 (8). At the amino acid level a significant homology was found with the yeast repair protein RAD10 and parts of the E.coli uvrA and C gene products (8,9,10) indicating strong evolutionary conservation of DNA repair systems.

In this report we present the genomic organization of the human ERCC-1 gene. We have identified the ERCC-1 promoter region and provide evidence that the ERCC-1 mRNA is subject to differential polyadenylation during processing.

MATERIALS AND METHODS

Cell culture and transfections

UV-sensitive Chinese hamster ovary (CHO) cell line 43-3B (11) and HeLa cells were grown in DMEM/F10$^+$ (1:1) medium supplemented with 10% fetal calf serum and antibiotics (penicillin, 100 U/ml and streptomycin 0.1 mg/ml). ERCC-1 DNA constructs were cotransfected with 0.5 - 3.0 µg pSV3gptH to 43-3B cells as described previously (8). After 10-14 days of selection with mitomycin-C (MM-C) and mycophenolic acid resistant clones were fixed and counted.

Identification of intron-exon borders

Using general procedures (12) all BglII fragments of cosmid43-34 that hybridized to ^{32}-labeled ERCC-1 cDNA probes were subcloned in pUC-vectors. Hybridizing parts of these clones were

sequenced by the chemical degradation method of Maxam and Gilbert (13).

Isolation of cDNAs

A human cDNA expression library made of poly(A)$^+$ RNA from a SV40 transformed fibroblast cell line was kindly provided by Dr. H. Okayama and screened with a 1.05 kb PvuII fragment of the 3′ region of ERCC-1 (8).

Plasmid construction of the ERCC-1 promoter

The genomic 3.7 kb BglII fragment Bg3-Bg4 of cosmid43-34 (7) was subcloned in pUC9 yielding p56-4. The ERCC-1 cDNA clone pcDE has been described previously (8). The plasmids pROM-1 and pUCPROMH-1 containing the putative ERCC-1 promoter region linked to the ERCC-1 cDNA and pERCC-SVP were constructed as follows: pcDE was cleaved with HindIII, treated with Klenow enzyme, and further digested with StuI which removes the SV40 promoter region (HindIII-StuI) and an ERCC-1 StuI fragment. The resulting blunt ended vector part of pcDE was ligated to this StuI fragment yielding the promoterless construct pERCC-SVP. The pROM-1 construct was obtained by the consecutive cloning of a 1.3 kb genomic PvuII-StuI fragment of 56-4 harboring the 5′ half of exon 1 at its 3′ end and the ERCC-1 StuI fragment in the HindIII (Klenow)-StuI digested pcDE vector. A 1.3 kb Klenow treated HinfI-BamHI fragment of pROM-1 carrying the complete ERCC-1 cDNA and 170 bp of 5′ genomic DNA was ligated to HindII linearized pUC18 resulting in pUCPROMH-1. Plasmid pROM-2 is similar to pROM-1 but has the ERCC-1 StuI fragment in the anti-sense orientation.

Nuclease S1 analysis

S1-mapping was performed as described by Grosveld et al. (14). A uniformly ^{32}P-labeled single stranded probe was synthesized from M13-templates using the method of Burke (15) and annealed with poly(A)$^+$RNA (10 µg) of K562 cells. After treatment with nuclease S1 the protected fragments were loaded on 6% polyacrylamide sequence gels next to a sequence ladder as molecular weight marker.

Northern blotting

Total RNA was prepared from cell cultures and isolated nuclei as described by Auffray and Rougeon (16). The mRNA

fraction was purified by two passages over oligo dT cellulose. For Northern blot analysis poly(A)$^+$ RNA (± 20 µg) was size fractionated by electrophoresis through 1% agarose/formaldehyde gels, transferred to nitrocellulose, and hybridized to ^{32}P-labeled nick-translated probes.

UV-induction experiments

HeLa cells were grown to near confluency on 15 cm Petri dishes and irradiated with 254 nm UV (0.6 J/m^2). Subsequently the cells were rinsed with PBS and incubated with medium. After several time intervals HeLa cells of 6 Petri dishes were harvested by polishman and poly(A)$^+$ RNA was isolated.

RESULTS

Gene Structure

Cosmid43-34, harboring the human ERCC-1 gene, was isolated from a cosmid library of a repair proficient CHO 43-3B transformant (7). The size of the ERCC-1 gene and its location on cosmid 43-34 (cos43-34) was determined in 3 ways: i. By transfection of restriction enzyme digested cos43-34 DNA to 43-3B cells to test for intactness of the gene (7). ii. By subcloning of Sau3A fragments of cos43-34 in EMBL-3 and screening for recombinant phages carrying a functional gene (8). iii. By Southern blot analysis of independent genomic 43-3B transformants for the presence or absence of specific probes from the ERCC-1 flanking regions (unpublished results). From these results it appeared that ERCC-1 covered a region of 15 kb. A detailed physical map of the deduced ERCC-1 region on cos43-34 is presented in Figure 1A. To determine the structural organization of the gene all BglII fragments in this region were subcloned in pUC-vectors and hybridized with ^{32}P-labeled probes of ERCC-1 cDNA clone pcDE (8). Subsequently, appropriate hybridizing fragments were further subcloned to sequence all exons and intron-exon junctions. The sequence strategy, deduced transcriptional orientation and the genomic organization of ERCC-1 are depicted in Figure 1B. The human ERCC-1 gene appears to consist of 10 exons ranging in size between 60 and 216 bp, spread over a region of 15 kb. The nucleotide sequence of the exons was completely in accordance with the previously reported ERCC-1 cDNA sequence (8). As shown

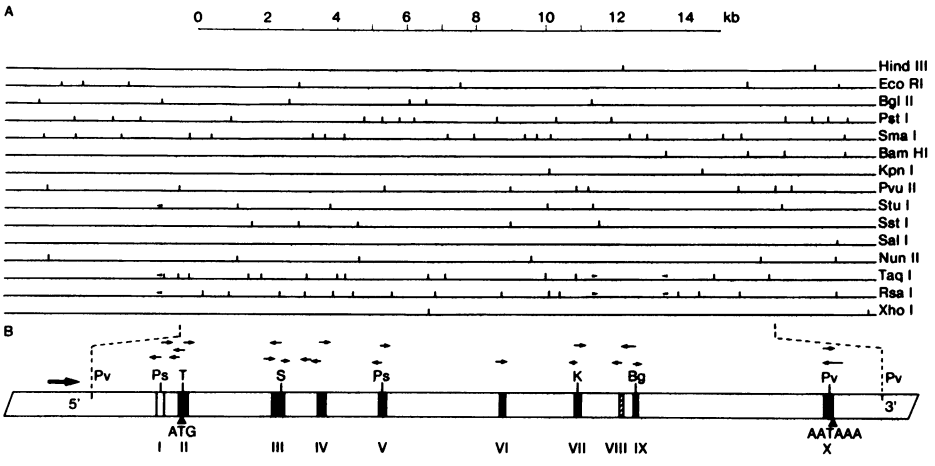


Figure 1. Genomic organization of ERCC-1.
Detailed physical map (A) and intron-exon structure (B) of the human excision repair gene ERCC-1. Arrow heads (in A) define regions for which not all restriction enzyme cutting sites have been mapped. Transcriptional orientation is indicated by the large arrow. Exons (depicted as boxes) are indicated by Roman numbers. The sequence strategy to determine intron-exon junctions is indicated by small arrows. Filled boxes represent coding sequences beginning with the indicated ATG. The differentially spliced exon VIII is shown as a shaded box. Abbreviations: Ps : PstI; T : TaqI; S : SmaI; K : KpnI; Bg : BglII; Pv : PvuII.

in Figure 2, the sequences around the intron-exon borders are consistent with the consensus donor and acceptor splicing signals (17) The much less conserved sequence PyNPyTPuAPy which is involved in branchpoint formation during the splicing process (18,19) could tentatively be identified in most introns at the appropriate distance (20 to 40 nucleotides) proximal to the splice acceptor sites (Figure 2, see also Discussion).

The ERCC-1 promoter region

The nucleotide sequence upstream from the first ERCC-1 exon is shown in Figure 3. A perfect alignment of 142 bp with the 5′ terminus of the previously reported cDNA sequence (8) was found. The precise start site of the ERCC-1 mRNA was determined by nuclease S1 analysis as demonstrated in Figure 4. A 2.0 kb genomic fragment spanning from the PstI site in exon I (position 60 in cDNA clone pcDE, (8)) to an upstream BglII site was

INTRON SPLICE ACCEPTOR	EXON				INTRON SPLICE DONOR	
	I	(153)	ACCACAG 34		GT GCGGGAGGCG	(0.3)
TGAGAACGTAGACCCCAGTGG CTCACTCGCTCCTGGCACCTTCCCTTTCAG	GCTCCAG 35	II	(112)	GLY VAL GGA GTG 107	GT AGGACAA	(1.9)
CTCAGATGCCTCGCTCACCCCAACCA GCCCCTGCAAACTCCCTTTTCTCCCCACAG	ALA LYS GCC AAG 108	III	(216)	ARG GLN CGG CAG 142	GT GAGGAGGGAG	(0.7)
AT TGATGGCTTCTGCCCTTCGTCCCTCCCCAG	ARG GLY AGG GGC 143	IV	(104)	LEU SE CTC AG 175	GT GAGCTCTGC	(1.2)
CTTTGGGGGTCTTGAAGAGACTGAGAC CTCTCAACTCCCGCTCCCAGCCCCCACCAG	R LEU C CTC 176	V	(100)	ASP VAL GAT GTG 201	GT AAGCAGGGGG	(2.3)
GCTAATTTTTGTAATTCATGGCTTCT AGGGTTTCTAATTCTGATTTTCTCCTCCAG	LYS ASP AAA GAT 202	VI	(77)	TRP SE TGG AG 234	GT GAGATGAGGG	(1.8)
AATCTGATACCTCACC TCCCGGCTGCCCTGTATCCTGTTATCCCAG	R PRO C CCC 235	VII	(100)	SER ARG TCC CGG 258	GT GAGGCCACC	(0.9)
GCTCAGTCTTGCTGGCCTTTCTTCT CCGACACACTCCTGCCTCCACCCTTTCCAG	VAL THR GTG ACT 259	VIII	(72)	PHE GLY TTT GGA 281	GT AAGGAATGGCT	(0.2)
GGACAGGGGAGGCTTTTGTGCTCAA CTGCCTCGACCCCTCGCTTTCACCTTTCAG	SER LEU TCT CTG 282	IX	(60)	GLN LYS CAG AAA	GT AAGAGCTCT	(4.1)
CCAGAAAACAGCTCCTTTAATGCT GGTTCCTCATTTTTTTCTCCCTCCATCCAG	ALA ARG GCC CGG	X	(112)	3' POLY A TAIL		

Figure 2. Structural details of the ERCC-1 intron-exon organization. The nucleotide sequence of each intron-exon junction is shown. Vertical lines represent intron-exon borders. Splice acceptor and donor sequences are in accordance with reported consensus sequence $(C)_nNCAG/G$ and AAG/GTA respectively (17). The corresponding amino acids at the beginning and end of each exon are shown in the three letter code and numbered as reported previously (8). Exon numbering is as used in Fig. 1. The size of each intron and exon (respectively kb and bp) is given in parentheses. The length of exon I is based on the results presented in Fig. 3 and 4. Heptamers which display the best match with the splicing branchpoint consensus sequence PyNPyTPuAPy are underlined.

subcloned in M13-mp8. Homogeneously ^{32}P-labeled single stranded DNA of this fragment was annealed to poly(A)$^+$ RNA of K562 cells and treated with nuclease S1. Figure 4 shows a protected band of 73-75 bp. We conclude from this result that the transcription most likely initiates at the A residues indicated in Figure 3 since these are generally preferred as 'capping'-site (17). From these results it appears that cDNA clone pcDE is almost complete (lacking only 9-11 bp at the 5' end, Figure 4) and consequently, that exon 1 has a size of 151-153 bp. The ERCC-1 5' flanking sequence is lacking the TATA and CAAT promoter signals. However at position -100 a sequence with a low level of homology with the

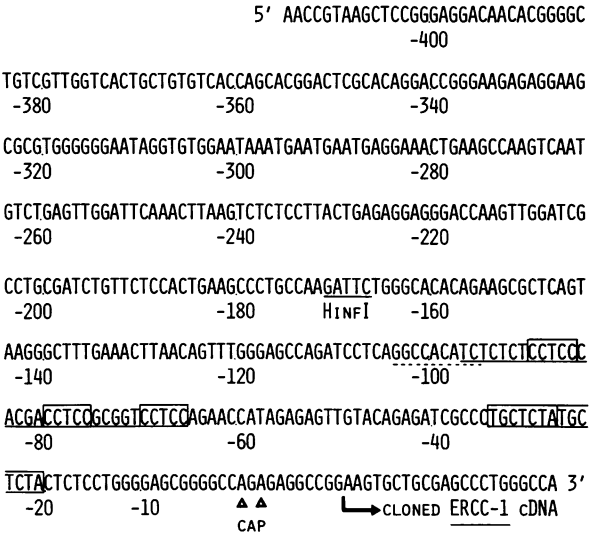


Figure 3. Genomic sequence upstream of ERCC-1 exon I. The 5' border of the cloned ERCC-1 cDNA (8) is indicated. Transcriptional start site(s) (as determined in Fig. 4) are shown by triangles. Tandem sequence TGCTCTA and the 3 times repeated sequence CCTCC are boxed. An 'abbreviated' CAAT box at -100 is indicated by a dashed line. The pyrimidine rich stretch of nucleotides and the HinfI site used for promoter construct pUCPROMHI (Fig. 4) are underlined.

CAAT-box consensus sequence GGTCAATCT can be found (see Figure 3). At position -30 the sequence TCTATGC very remotely resembles the TATA$^{T}_{A}$A$^{T}_{A}$ consensus. Several other sequence motifs are present in the ERCC-1 promoter region. At position -20 to -30 a tandem repeat of TGCTCTA is located. Further downstream the pentanucleotide CCTCC occurs three times with intervals of 5 nucleotides (see Figure 3). This region (-65 to -100) is very rich in pyrimidine residues (78%). The most upstream CCTCC motif is part of a stretch of 13 pyrimidines.

To verify whether this region can drive ERCC-1 transcription two genomic fragments (a 1.3 kb PvuII-StuI and 180 bp HinfI-StuI fragment) were constructed in front of the ERCC-1 cDNA yielding pROM-1 and pUCPROMHI respectively (see Figure 4 and Materials and Methods). These DNA constructs were transfected to 43-3B cells in coprecipitate with the dominant vector pSV3gptH. Transformants

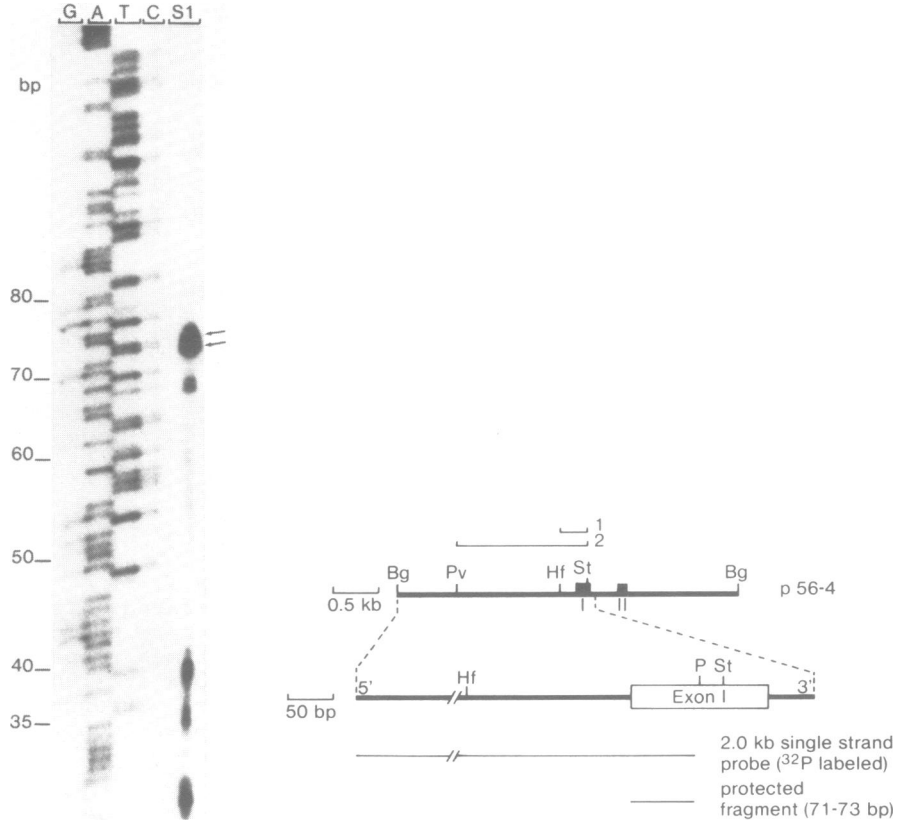


Figure 4. Analysis of ERCC-1 transcriptional start site. A two kb BglII-Pst fragment of genomic ERCC-1 subclone p56-4 (harboring exon I and II) was subcloned in M13-mp8. Single stranded ^{32}P-labeled DNA of this fragment was annealed to K562 poly(A)$^+$RNA and treated with nuclease S1 (4000U, 1 hr, 16^{o}C). Protected fragments (shown by arrows) were visualized by polyacrylamide gelelectrophoresis and autoradiography. A Maxam-Gilbert sequence ladder of an exon I fragment labeled at the StuI site was used as a molecular weight marker. After incubation of the probe with yeast tRNA no protected fragments were observed (not shown). DNA fragments 1 and 2 are genomic PvuII-StuI and HinfI-StuI fragments that were used for constructing pROM-1 and pUCPROMH-1 respectively (see Table 1).

were selected with mycophenolic acid for expression of the E.coli gpt gene and with MM-C for ERCC-1 expression.

The results summarized in Table 1 show that both ERCC-1 plasmids can confer MM-C resistance to 43-3B cells with an efficiency comparable to the genomic ERCC-1 gene on cosmid 43-34. This

Table 1.
Transfected DNAs are described in Materials and Methods.

Transfected DNA		Number of mpa and MM-C resistant clones/µg transfected DNA[a)]
cos43-34	(genomic ERCC-1)	25
pcDE	(cDNA + SV40 early promoter)	5
pROM-1	(cDNA + 1.3 kb promoter fragment)	22
pUCPROMH-1	(cDNA + 170 bp promoter fragment)	32
pROM-2	(inactivated ERCC-1 cDNA)	0
pERCC-SVP	(cDNA without promoter)	2

a) average of three dishes
Identification of the ERCC-1 promoter region. ERCC-1 promoter activity was determined by scoring the number of mitomycin-C and mycophenolic acid (mpa) resistant clones after cotransfection of ERCC-1 DNA and pSV3gptH to 43-3B cells.

indicates that the two genomic regions tested display promoter activity and confines the ERCC-1 promoter to a stretch of 170 bp upstream of the transcriptional start site. Surprisingly, with pcDE, harboring ERCC-1 cDNA under the direction of the strong SV40 early promoter, always a lower transfection frequency is observed. No repair-proficient transformants are obtained with a non-functional ERCC-1 construct (pROM-2). However, a very low but consistent transfection frequency is found with pERCC-SVP which is lacking promoter sequences. This may be explained by insertion of the SV40 early promoter derived from cotransfected pSV3gptH or by integration of pERCC-SVP DNA in the vicinity of endogenous promoters in the genome of 43-3B cells.

Northern blot analysis of different human cell lines revealed a low ERCC-1 transcription level (8, data not shown). To investigate whether the ERCC-1 promoter is DNA-damage inducible, Northern blot analysis was performed on poly(A)$^+$ RNA from Hela cells at various periods of time after UV irradiation. The results are presented in Figure 5. Hybridization with the probe for the glyceraldeyde-3-phosphate dehydrogenase gene (GAPDH; 20) included as control, indicates that approximately equal amounts of poly(A)$^+$ RNA are loaded in each lane. It is evident that UV irradiation with a UV dose of 1 J/m^2 did not result in detectable changes of ERCC-1 transcription within a

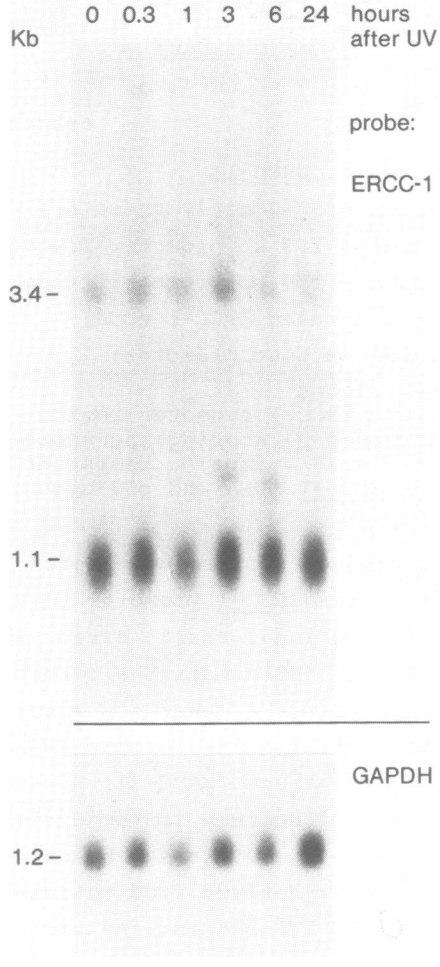
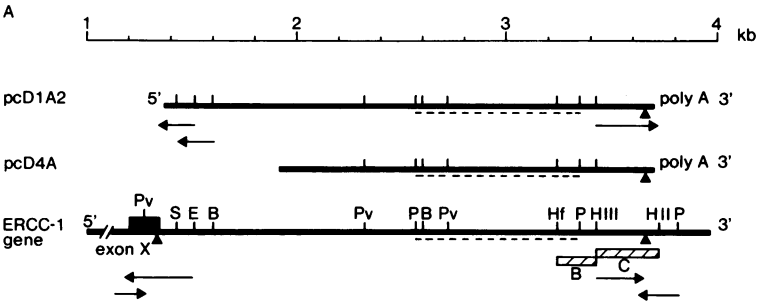


Figure 5. Effect of UV-irradiation on ERCC-1 transcription. Poly(A)$^+$RNA (20 µg) of exponentially growing UV-irradiated (1 J/m^2) HeLa cells was size fractionated on 1% agarose gel and after blotting to nitrocellulose hybridized to a nick-translated ^{32}P-labeled ERCC-1 cDNA probe. To verify that equal amounts of RNA were loaded on the gel the filter was rehybridized with a GAPDH probe (20).

period of 24 hours. Nor did a UV dose of 10 J/m^2 or treatment of HeLa cells with MM-C (10^{-8}M) affect the level of ERCC-1 transcription (data not shown).

The 3′ region

The ERCC-1 cDNA was isolated from a human expression cDNA



B

CAGCTGCCAA GGAAACCCCC AGTGTAATAA TAAATCGTCC TCCCAGGCCA GGCTCCTGCT
PvuII

GCTGCGCTGG TGCAGTCTCT GGGGAGGGAT TCTGGGGGTG TCACCTTCTG GTGGCCCAGG
pcD1A2

TGGGCACCTT CAGCTTTCTT TAGTTCCTCA GTTTCCCGGG GGCAGACTAC ACAGGCTGCT

GCTGCTGCGC TTCTTCCGCT TCTTGTCCCG GCCTGTGGGA GCCTCCTCCC CAGACTCTGA

ATTCAGTGGC CGGCCCTGGC ATCTCCTCTT GGGGCACTGT CTCTGGCATC CGGCTTTCCT
EcoRI

GACTCTGCTT CTTCCTCTTC TTGGTGGATC CCGGAGTTGC C--- ±1750 BP--- AGAA
BamHI

GTCCAAATGT CCTAAGAATG CAGCCCCCAT TACTTAAATA TAACATAACA TAAATGAGCA

AGCTTAGGTT GCAGGATTAA TGGTCGTGGA TAACACCAAT AGCCCCTACC TTTAGCGAGC
HindIII

TTATCTGCAC ACTCCAAGTT TAACTATAGT TCCTTATAGT TTCTTATAAG TAGAAATACT

AACAAAGGGC CGTGGGTTTC TCCCCCTGCT TTCTGAGGAC ACTCTACTCT GTAAAGGAGT

AGTTTCCAAT AAACTTGTTT CTTTCACTGT GCTCTGTGAC TCACCTCTAA TTCTTTCCTG
HinfI

TGACAAATTC AAGAACCCGC TCTTGGGGTC TGGACCGGGA CCCTCTTTTC CGTCAACAAG
HindII

ACCACTCACC ATTC AAAAC AAAGTCTACA GCCAAG-3'

Figure 6. Detailed analysis of the 3' ERCC-1 flanking region. **A.** Alignment of cDNA clones pcD4A and pcD1A2 with the genomic region flanking the 3' site of ERCC-1 exon X. Triangles (▲) indicate AATAAA polyadenylation signals. The dashed line specifies fragments containing repetitive sequences as found by Southern hybridization using human repetitive DNA as a probe (not shown). Arrows indicate the sequence strategy and the shaded boxes B and C are probes used for the Northern blot analysis (see Fig. 7). Not all restriction enzyme sites are shown. Abbreviations: Pv : PvuII; S : SmaI; E : EcoRI; B : BamHI; P : PstI; Hf : HinfI; HIII : HindIII; HII : HindII. **B.** Composition of genomic and cDNA sequence data. Appropriate restriction enzyme sites shown in Fig.6A are indicated. The polyadenylation signals AATAAA are indicated by double underlining. The boxed sequence TTC might represent a polymorphism as it was not found in the genomic sequence. Triangles (▲) indicate polyadenylation sites as found by sequence analysis.

library (21) using a 3′ 1.05 kb genomic PvuII fragment as a probe. In addition to the earlier described ERCC-1 cDNA clones (8) several other clones (pcD1A2 and pcD4A) were isolated with this probe. As illustrated in Figure 6A the physical maps of these clones are colinear with the genomic map of the 3′ ERCC-1 flanking region. The longest clone pcD1A2 has an insert size of approximately 2.2 kb. Furthermore, Southern blot hybridization with human repeats as a probe revealed the presence of repetitive sequences in both cDNA clones as well as in the corresponding genomic fragments (not shown). The nucleotide sequence of the 5′ and 3′ termini of pcD1A2 and the corresponding genomic regions of cosmid 43-34 (7) were determined by the chemical cleavage method of Maxam and Gilbert (13). A composition of both sequences is presented in Figure 6B. The 3′ end of pcD1A2 harbors a poly(A) signal followed by a poly(A) tail indicating a transcriptional orientation which is similar to that of ERCC-1. In the 5′ direction clone pcD1A2 extends to 40 bp downstream of the polyadenylation site of the 1.1 kb ERCC-1 mRNA (see Figure 6B). This renders it conceivable that pcD1A2 is a partial cDNA clone from a longer ERCC-1 mRNA with an alternative polyadenylation site. To investigate this hypothesis poly(A)$^+$ RNA of HeLa nuclei and human primary fibroblasts was analysed by Northern blotting using a coding ERCC-1 cDNA probe and two genomic probes of the pcD1A2 region (see Figure 6A). The autoradiogram of this experiment is shown in Figure 7. ERCC-1 probe 3A3 which encodes the N-terminal half of the ERCC-1 protein recognizes the mature 1.0 - 1.1 kb ERCC-1 mRNA, a 3.4 kb transcript in fibroblasts and a 3.4 and 3.8 kb mRNAs in HeLa nuclei. These 3.4 and 3.8 kb RNA species are also recognized by two probes (B and C) flanking the HindIII site in the 3′ pcD1A2 region. The 3.8 kb transcript seems to be absent in primary fibroblasts. However, although with a much lower hybridization signal than the 3.4 band, the 3.8 kb could be detected in poly(A)$^+$ RNAs of a number of other cell lines (e.g. Figure 5, data not shown). We conclude from these results that pcD4A and pcD1A2 are partial cDNA clones derived from the 3.4 kb ERCC-1 transcript. The 3.8 kb molecule most likely terminates at a yet unidentified polyadenylation site further downstream.

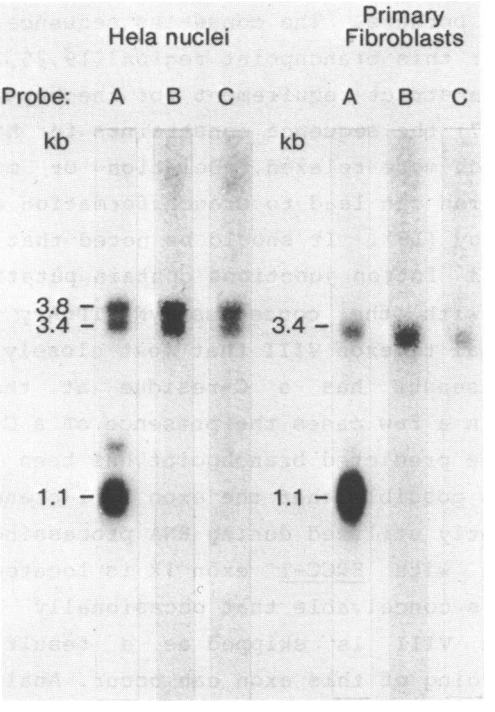


Figure 7. Northern blot analysis of ERCC-1 transcripts. Poly(A)$^+$RNA (± 20 µg) of HeLa nuclei and primary fibroblasts was size fractionated by agarose gelelectrophoresis (1% agarose) and after blotting to nitrocellulose filters hybridized to three ^{32}P-labeled DNA probes. A: 3A3, a 5′ PstI-fragment of ERCC-1 cDNA clone pcDE (8) B and C are genomic DNA fragments (HinfI-HindIII and HindIII-HindII respectively) from the 3′ terminal region of the ERCC-1 gene shown in Fig. 6A.

DISCUSSION

The excision repair gene ERCC-1 has been assigned to human chromosome 19 (8,22,23) and regionally mapped to 19q13.2-13.3 (24). Here we report the structural organization of the ERCC-1 gene. The gene is about 15 kb in length and harbors 10 exons. We have recently shown that the coding exon VIII of 72 bp is subject to alternative splicing. Only the cDNA of the larger ERCC-1 transcript (which includes exon VIII) inserted in a mammalian expression vector can correct the excision repair mutation in 43-3B cells (8). During pre-mRNA processing excision of introns requires branchpoint formation at a distance of -40 to -20 bp

from the 5′ exon borders. The consensus sequence PyNPyTPuAPy has been proposed for this branchpoint region (19,25,26). In contrast to yeast where a strict requirement for the heptamer TACTAAC has been observed (27) the sequence constraints for higher eukaryotes seem to be much more relaxed. Deletion or mutation of the branchpoint region can lead to branch formation at other criptic positions close by (19). It should be noted that except for exon VIII all ERCC-1 3′ intron junctions contain putative branchpoint motifs matching with the consensus PyNPyTPuAPy (Figure 2). The sequence proximal to exon VIII that most closely resembles the branchpoint consensus has a C-residue at the conserved T-position. Only in a few cases the presence of a C-residue at this position of the predicted branchpoint has been observed (18). Therefore, it is possible that the exon VIII branchpoint sequence is less efficiently utilized during RNA processing. Since the 3′ intron junction with ERCC-1 exon IX is located only 0.3 kb downstream it is conceivable that occasionally the branchpoint signal of exon VIII is skipped as a result of which the alternative splicing of this exon can occur. Analysis of a number of mouse ERCC-1 cDNA clones (manuscript in preparation) have not given any evidence for differential RNA procesing in mouse brain and testes tissue. It is therefore not excluded that the observed alternative splicing of ERCC-1 mRNA is the result of an inaccurate splicing event in human cells rather than a means to provide additional biological functions to the ERCC-1 protein.

Using a functional assay we have shown that the ERCC-1 promoter is located in a stretch of 170 bp proximal to the transcriptional start site. In this region there is no clear cut TATA box or AT-rich region which is typically positioned between -20 and -30 from the CAP-site (17). An 'abbreviated' version of a CAAT-box might be located at -100. The apparent absence of these promoter motifs is frequently found in housekeeping genes that are driven by promoters containing the transcription factor Sp1 binding box GGGCGG (28). The ERCC-1 promoter is also devoid of such GC-boxes. The absence of classical transcription elements like CAAT, TATA and GC-boxes has recently been reported for the weakly expressed c-mos proto-oncogene (29) and is exceptional for eukaryotic genes. Therefore, it is possible that the ERCC-1

promoter represents a specific class of promoters. Several other sequence motifs, of which the significance is still unknown, have been notified in the ERCC-1 promoter region (Figure 3). Three CCTCC repeats are located in a pyrimidine rich region from -65 to -100. The promoter of the EGF-receptor gene harbors pyrimidine rich stretches of nucleotides with 4 times repeated TCC elements (30). These regions might have a regulatory function in EGF-receptor expression since nuclear proteins were found to bind to the TCC-containing regions (31). Possibly, ERCC-1 expression is mediated through a similar type of interaction between regulatory proteins and the CT-rich promoter elements.

We have not quantitatively determined the transcription activity of the ERCC-1 gene. However, qualitative observations from Northern blot analysis of poly(A)$^+$ RNA of different cell lines indicate low levels of ERCC-1 mRNA. Low constitutive expression of repair proteins has also been observed in E.coli and yeast (1,32). In bacteria an efficient removal of DNA lesions is provided by an excision repair system that is part of the DNA-damage inducible SOS-response. In yeast, the agents that elicit this response induce the excision repair gene RAD2 whereas transcription of RAD1, RAD3 and RAD10 is not affected by DNA-damage (33). From our experiments it can be inferred that ERCC-1 expression is not influenced by UV or MM-C induced DNA damage.

Several observations suggest very low levels of intracellular ERCC-1 protein. Firstly, as mentioned before, ERCC-1 transcripts belong to a class of low abundant mRNAs in a variety of human cell lines (8, results not shown). Secondly, the AUG start codon of the ERCC-1 protein lies in a very unfavourable sequence context (TCCAGATGG) for translation initiation (34) because of the presence of a pyrimidine at position -3. More than 95% of eukaryotic mRNAs examined have an A (75%) or a G-residue (20%) at that position and the consensus sequence CCACCATGG that emerged from a compilation of the mRNA initiation sites used by higher eukaryotes completely matched the most optimal translational start context as found by mutation analysis (35,36). In those studies the presence of a T at -3, as found in ERCC-1 mRNA, caused a 10-20 fold drop in the efficiency of translation initiation. Hence, the presence of a pyrimidine at

position -3 of the ERCC-1 mRNA suggests a low constitutive level of ERCC-1 translation.

The transfection efficiency of ERCC-1 constructs seems to be promoter dependent (Table 1), since the strong SV40 promoter yields less transformants than the weak ERCC-1 promoter. Further experiments are required to study the effect of the ERCC-1 expression level, however the preliminary data presented in Table 1 suggest that overexpression of ERCC-1 can be deleterious to the cell and that mimimum amounts of ERCC-1 protein are required for repair of DNA damages.

ERCC-1 mRNA molecules of 1.1, 3.4 and 3.8 kb have been identified by Northern blot analysis using coding ERCC-1 cDNA probes. The 1.1 kb transcript which is the major mRNA species encodes the protein that corrects the 43-3B mutation and harbors all exons shown in Figure 1. Although no full length cDNA clones have been isolated of the longer transcripts, there is strong evidence that differential polyadenylation gives rise to the 3.4 and 3.8 kb mRNAs: i. Both transcripts are recognized by 5′ ERCC-1 cDNA probes. ii. Genomic probes located approximately 2 kb 3′ of the 1.1 kb mRNA polyadenylation site hybridize with the 3.4 and 3.8 kb molecules. iii. cDNA clones of the 3.4 kb mRNA have the ERCC-1 transcriptional orientation and have a poly(A) tail preceeded by the polyadenylation signal AATAAA. These cDNA clones contain human repetitive sequences and are completely colinear with the genomic DNA as shown by restriction enzyme and partial sequence analysis. iv. Recently, functional mouse ERCC-1 cDNA clones have been isolated with different 3′ untranslated regions due to alternative polyadenylation (van Duin, in preparation).

In addition to the hexanucleotide AATAAA several other loosely conserved sequence elements have been suggested to play a role in mRNA 3′ end formation (see 37 for a review). The pentanucleotide CAYTG is found 3′ of the AATAAA before or after the poly(A) site in many genes and is suggested to mediate poly(A) formation through interaction with RNA of the U4 small ribonucleoprotein (38). Furthermore, T-rich and GT-rich sequences motifs are located downstream of poly(A) sites in many transcripts (37). Recent studies of Gil and Proudfoot (39) and McDevitt et al. (40) have demonstrated that these motifs function

synergistically and are important for efficient and accurate 3′ end formation. Inspection of the 3′ ERCC-1 sequence revealed the presence of CAYTG box and a T-rich element near the second AATAAA suggesting that this can be a better target for 3′ end processing than the more proximal AATAAA.

Multiple polyadenylation sites occurring in 3′ untranslated regions have been reported for a number of transcripts (41,42,43). It is unknown whether the selection of poly(A) sites within one transcriptional unit is a regulatory event. In this respect it is interesting to note that 3′ untranslated sequences of several transiently expressed genes have been implicated in mRNA stability and that recently AU-rich sequences with AUUUA motifs have been identified defining mRNA instability (44,45). Deletion of such sequences from the c-fos proto-oncogene abolishes the rapid turnover of c-fos transcripts (46) and confers a transforming potential to the gene (47). Such elements have not been found in the partially sequenced 3′ ERCC-1 flanking region. However, the apparent evolutionary conservation of differential polyadenylation of ERCC-1 transcripts could imply that mRNA 3′ end processing plays a regulatory role in ERCC-1 expression.

ACKNOWLEDGEMENTS

We thank Anneke Graus and Petra Warmerdam for technical assistance with the promoter studies. Dr. H. Okayama is kindly acknowledged for providing the cDNA library and Dr. F.J. Benham for giving the GAPDH probe. We are indebted to Rita Boucke, Tar van Os and Mirco Kuit for preparing the manuscript.
This work was financially supported by EURATOM (contract no. B16-141-NL) and MEDIGON, Foundation of Medical Scientific Research in the Netherlands (contract no. 900-501-091).

REFERENCES

1. Friedberg, E.C. (1985) DNA repair, Freeman & Company, San Francisco
2. Walker, G.C. (1985) Annu. Rev. Biochem. 54, 425-457.
3. De Weerd-Kastelein, E.A., Keijzer, W. and Bootsma, D. (1972) Nature 238, 80-83.
4. Fischer, E., Keijzer, W., Thielman, H.W., Popanda, O., Bohnert, E., Edler, E.G., Jung, E.G. and Bootsma, D. (1985) Mutation Res. 145, 217-225.

5. Thompson, L.H., Busch, D.B., Brookman, K.W., Mooney, C.L. and Glaser, P.A. (1981) Proc.Natl.Acad.Sci.USA 78, 3734-3737.
6. Thompson, L.H. and Carrano, A.V. (1983) UCLA Symposium on Molecular and Cell Biology New Series (Friedberg, E.C. and Bridges, B.R., eds.) vol. 11, 125-143.
7. 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, 425-429.
8. 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.
9. Doolittle, R.F., Johnson, M.S., Husain, I., van Houten, B., Thomas, D.C., Sancar, A. (1986) Nature 323, 451-453.
10. Hoeijmakers, J.H.J., van Duin, M., Westerveld, A., Yasui, A. and Bootsma, D. (1986) Cold Spring Harbor Symposia on Quantitative Biology vol. LI, 91-101
11. Wood, R.D. and Burki, H.J. (1982) Mutation Res. 95, 505-514.
12. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning. Cold Spring Harbor, New York.
13. Maxam, A.M. and Gilbert, W. (1980) Meth.Enzymol. 65, 499-560.
14. Grosveld, G.C., Koster, A and Flavell, R.A. (1981) Cell 23, 573-584.
15. Burke, J. (1984) Gene 30, 63-68.
16. Auffray C. and Rougeon, F. (1980) Eur.J.Biochem. 107, 303-314.
17. Breatnach, R. and Chambon, P. (1981) Annu.Rev.Biochem. 50, 349-384.
18. Keller, E.B. and Noon, W.A. (1984) Proc.Natl.Acad.Sci.USA 81, 7417-7420.
19. Reed, R. and Maniatis, T. (1985) Cell 41, 95-105.
20. Benham, F.J., Hodgkinson, S., Davies, K.E. (1984) EMBO J. 3, 2635-2640
21. Okayama, H. and Berg, P. (1983) Mol.Cell.Biol. 3, 280-289
22. Rubin, J.S., Prideaux, V.R., Huntington, F.W., Dulhanty, A.M., Whitmore, G.F. and Bernstein, A. (1985) Mol.Cell.Biol. 5, 398-405.
23. Thompson, L.H., Mooney, C.L., Burkhart-Schultz, K., Carrano, A.V. and Siciliano, M.J. (1985) Mol.Cell.Biol. 5, 881-884.
24. 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) Cytogen.Cell Genet. 40, 590-591.
25. Keller, W. (1984) Cell 39, 423-425.
26. Aebi, M. and Weissmann, C. (1987) Trends in Genet. 3, 102-107.
27. Langford, C.J., Klin, F.J., Donath, C. and Gallwitz, C. (1984) Cell 36, 645-653.
28. Kim, S.H., Moores, J.C., David, D., Respess, J.G., Jolly, J. and Friedman, T. (1986) Nucleic Acids Res. 14, 3108-3118.
29. Propst, F.,Rosenberg, M.P., Lyer, A., Kaul, K. and Van de Woude, G.F. (1987) Mol. Cell. Biol. 7,1629-1637.
30. Ishii, S., Xu, Y.-H., Stratton, R.H., Roe, B.A. Merlino, G.T. and Pastan, I. (1985) Proc.Natl.Acad.Sci.USA 82, 4920-4924.
31. Ishii, S., Imamato, F., Yamanashi, Y., Toyoshima, K. and Yamamoto, T. (1987) Proc.Natl.Acad.Sci.USA 84, 4373-4378.

32. Friedberg, E.C. (1987) J.Cell Sci.Suppl. 6, 1-23.
33. Robinson, G.W., Nicolet, C.M., Kalainov, D. and Friedberg, E.C. (1986) Proc.Natl.Acad.Sci.USA 83, 1842-18466.
34. Kozak, M. (1986) Cell 47, 481-483.
35 Kozak, M. (1984) Nucl.Acids Res. 12, 857-872.
36. Kozak, M. (1986) Cell 44, 283-392.
37. Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) Cell 41, 349-359.
38. Berget, S.M. (1984) Nature 309, 179-181.
39. Gil, A. and Proudfoot, N.J. (1987) Cell 49, 399-406.
40. McDevitt, M.A., Hart, R.P., Wong, W.W. and Nevins, J.R. (1986) EMBO J. 5, 2907-2913.
41. LeMeur, M.A., Galliot, B. and Gerlinger, P. (1984) EMBO J. 3, 2779-2786.
42. Hagenbüchle, O., Wellauer, P.K., Cribbs, D.L. and Schibler, U. (1984) Cell 38, 737-744.
43. Chen, M.-J., Shimada, T., Moulton, A.D., Cline, A., Humphries, R.K., Maizel, J. and Nienhuis, A.W. (1984) J.Biol.Chem. 259, 3933-3943.
44. Shaw, G. and Kamen, R. (1986) Cell 46, 659-667.
45. Brawerman, G. (1987) Cell 48, 5-6.
46. Treisman, R. (1985) Cell 42, 889-902.
47. Meylink, F., Curran, T., Miller, A.D. and Verma, I.M. (1985) Proc.Natl.Acad.Sci. USA 82, 4987-4991.