Volume 16 Number 12 1988 Nucleic Acids Research

Evolution and mutagenesis of the mammalian excision repair gene *ERCC*-1

M.van Duin, J.van den Tol, P.Warmerdam, H.Odijk, D.Meijer, A.Westerveld, D.Bootsma and J.H.J.Hoeijmakers

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

Received April 7, 1988; Revised and Accepted May 19, 1988 Accession nos X07413, X07414

ABSTRACT

The human DNA excision repair protein ERCC-1 exhibits homology to the yeast RAD10 repair protein and its longer C-terminus displays similarity to parts of the E.coli repair proteins uvrA and uvrC. To study the evolution of this 'mosaic' ERCC-1 gene we have isolated the mouse homologue. Mouse ERCC-1 harbors the same pattern of homology with RAD10 and has a comparable C-terminal extension as its human equivalent. Mutation studies show that the strongly conserved C-terminus is essential in contrast to the less conserved N-terminus which is even dispensible. The mouse ERCC-1 amino acid sequence is compatible with a previously postulated nuclear location signal and DNA-binding domain. The ERCC-1 promoter harbors a region which is highly conserved in mouse and man. Since the ERCC-1 promoter is devoid of all classical promoter elements this region may be responsible for the low constitutive level of expression in all mouse tissues and stages of embryogenesis examined .

INTRODUCTION

Cell hybridization experiments have identified 6 complementation groups within DNA excision repair deficient UV-sensitive Chinese hamster ovary (CHO) cell lines (1-3). In many respects these mutants resemble cells from the genetic disorder xeroderma pigmentosum (XP) in which defects in at least nine genes (4,5) underly the extreme sensitivity of XP patients to sun exposure (UV light) and a predisposition to skin cancer. Cell fusion studies - although incomplete - have thusfar not revealed any overlap between these two classes of mutants (6). Hence, it is possible that 15 or more genes play a role in the excision of UV induced DNA damage in mammalian cells. Also in yeast and prokaryotes many loci have been found to be implicated in excision repair (see 7 for a review). In E.coli the uvrABC and D

© IRL Press Limited, Oxford, England.

gene products play a key role in the excision repair machinery and it is likely that comparable multiprotein complexes are operative in higher organisms.

By applying genomic DNA transfer we have isolated the human ERCC-1 gene which corrects the repair defect of UV and mitomycin-C (MMC) sensitive CHO 43-3B cells belonging to complementation group 1*) (8). Recently, also the ERCC-2 gene complementing group 2*) mutants has been isolated (9). The ERCC-1 gene only corrects CHO mutants of complementation group 1 (10) and restores all impaired repair characteristics to wild type levels (11) suggesting that ERCC-1 is the human homologue of the mutated gene in CHO group 1 mutants.

Based on similarity of the predicted ERCC-1 amino acid sequence with functional domains of other proteins, a putative nuclear location signal (NLS), DNA binding domain and ADP-monoribosylation site have been assigned to the ERCC-1 protein (12). Furthermore, a significant homology with yeast RAD10 and parts of the E.coli uvrA and uvrC repair proteins has been found (12-14). This suggests that DNA repair systems are well conserved during evolution. In this respect it is worth noting that the RAD6 protein which is involved in cellular response to DNA damage in yeast and recently found to be a ubiquitin conjugating enzyme, is also very strongly conserved up to mammals (15,16). The extent of homology between ERCC-1 and RAD10 makes it tempting to speculate that both proteins are functionally equivalent although ERCC-1 has a C-terminal extension of 83 amino acids which is absent in RAD10 (12). It is intriguing that this extra ERCC-1 part displays similarity with bacterial excision repair proteins. It is possible that evolution has endowed ERCC-1 with functional domains of distinct repair proteins in prokaryotes or that in the course of evolution the tail of RAD10 was lost. To investigate these possibilities and to further establish the significance of the postulated functional domains in the ERCC-1 protein, it is of interest to characterize the ERCC-1 gene of other organisms. Here we report the characterization of the mouse ERCC-1 gene and present mutation studies on the human ERCC-1 cDNA.

MATERIALS AND METHODS

Cell culture and transfection.

CHO 43-3B cells (17) were routinely grown in DMEM/F10 (1:1) medium with 5% fetal calf serum and antibiotics. To test for a functional ERCC-1 gene, DNA constructs (5-10 µg) were cotransfected with pSV3gptH (2-5 µg) to 5.10^5 43-3B cells in three 9 cm dishes as described previously (8). After 10-14 days of selection on mycophenolic acid (MPA) and MMC the cells were fixed, stained and clones were counted.

Unscheduled DNA Synthesis.

Two days after seeding in medium without MPA and MMC the cells were exposed to UV light (16 J/m^2) and incubated in thymidine-free, Ham's F10 medium supplemented with ^3H-thymidine (10 µCi/ml; specific activity 50 Ci/mmol) and 5% dialyzed fetal calf serum After Bouin fixation the preparations were processed for autoradiography (Kodak AR10 stripping film), exposed for 1 week at 4°C, developed and stained with Giemsa solution. For each preparation the number of grains per fixed square of 50 nuclei was counted.

RNA preparation and Northern blotting.

Total RNA was isolated from adult mice (Balb C), size fractionated on 1% agarose gels and after transfer to nitrocellulose filters hybridized to mouse ERCC-1 cDNA probes following standard procedures (18). Probes were labeled using the random primer technique (19).

cDNA cloning

A mouse brain cDNA library was prepared in phage λgt10 and screened with a human ERCC-1 cDNA probe as reported earlier (20).

Plasmid Constructions and Sequencing.

Routine protocols were used for plasmid isolation, subcloning and ligation (18). Nucleotide sequences were determined by the chemical cleavage (21) or chain termination method (22). Oligonucleotides for specific priming and Tab-linker mutagenesis were made with an Applied Biosystems DNA synthesizer. Human and mouse ERCC-1 cDNA plasmids were constructed as follows:

pTZME. Both EcoRI inserts of λgt10 mouse ERCC-1 cDNA were subcloned in pTZ19R (Pharmacia) yielding pM4a-2, harboring the 5'

cDNA part and pM4a-1 containing the remaining 3′ half. The artificial 5′ EcoRI site of pM4a-2 was removed by very short Bal31 treatment starting from the adjacent SstI site in the polylinker. The retained HindIII and PstI site of the polylinker were used to insert the HindIII-PstI fragment of pcDX (23) harboring the SV40 early promoter. Finally the EcoRI fragment of pM4a-1 was subcloned behind the 5′ cDNA part yielding pTZME.

pcDEMP1 and pcDEMP2. The 0.42 kb SmaI fragment of human ERCC-1 cDNA clone pcDE (14) was subcloned in pSP65 in which the unique PstI site was deleted giving pSPSma. This plasmid was linearized with PstI or KpnI (unique ERCC-1 cDNA sites) and 0.5 µg ligated overnight at 4°C to 4 µg of either a PstI-Tab linker (5′ GCTGCA 3′) or a Kpn Tab-linker (5′ GCGTAC 3′) in 10 µl (24). After ethanol precipitation (to eliminate excess unligated linker) the DNA was kinased and ligated to generate circular molecules. The insertion of a single Tab-linker in PstI and KpnI site was confirmed by sequence analysis. The PstI and KpnI mutated inserts of pSPSma were subsequently recloned in pcDE yielding pcDEMP-1 and pcDEMP-2 respectively.

pcDEBgl. pcDE (14) was digested with BglII, treated with Klenow DNA polymerase to create blunt ends and religated.

pcDEASt. The StuI fragment of pcDE was deleted by StuI digestion and religation yielded pcDEΔStu with a single StuI site and ERCC-1 sequences 3′ of codon 214 (14). Construct pcDEΔStu was linearized by StuI/AvaI double digestion which releases a ± 60 bp StuI-AvaI fragment. After klenow treatment to fill in the AvaI site the earlier deleted StuI fragment was inserted again yielding pcDEASt.

RESULTS

Characterization of mouse ERCC-1 cDNA.

Southern blot hybridization with DNA digests of various vertebrates indicated that the ERCC-1 cDNA was strongly conserved in evolution. Under reduced stringency conditions specific hybridization was found with mammalian, reptile, avian and fish DNA and very weakly with DNA of Drosophila (data not shown) whereas no hybridization was found with DNA from Trypanosomes, yeast and E.coli . This indicated that it should be possible to

isolate the mouse ERCC-1 cDNA from a mouse library using human ERCC-1 probes. A full length mouse ERCC-1 cDNA clone (designated λcDME) was isolated from a brain λgt10 cDNA library using human ERCC-1 cDNA as a probe. To establish whether this clone encoded a functional ERCC-1 protein the cDNA insert of λcDME was released with EcoRI, subcloned in a SV40 based mammalian expression cartridge (see Materials and Methods) and transfected to CHO 43-3B cells. The results of this experiment, in which human ERCC-1 cDNA construct pcDE (14) served as a positive control, are shown in Table 1. Using the mouse ERCC-1 cDNA construct pTZME, stable MMC resistant transformants could be generated with a transfection frequency (not shown) similar to pcDE. Furthermore, mouse cDNA transformants displayed levels of unscheduled DNA synthesis (UDS) in the wild type range (see Table 1), indicating that the isolated mouse cDNA harbors a functional ERCC-1 gene.
The complete nucleotide sequence of λcDME and predicted amino acid sequence are given in Figure 1. The mouse ERCC-1 cDNA appears to encode a protein of 298 amino acids and deduced molecular weight of 32970 Dalton. The alignment of the mouse ERCC-1 protein with its human homologue, the postulated functional domains (12) and the homology with the yeast RAD10 and

TABLE 1
Test for functional mouse ERCC-1 gene.

Cell line	Transfected DNA	MMC resistant clones	UDS*)
43-3B	pcDE	+	25 ± 1
43-3B	pTZME	+	27 ± 1
43-3B	-DNA	-	6 ± 1
CHO-9			19 ± 1**)
HeLa			33 ± 1**)

*) expressed as average number of grains (±SEM) per fixed square in 50 nuclei.
**) UDS of untransfected cells.
Mouse ERCC-1 cDNA was constructed behind the SV40 early promoter (yielding pTZME) and cotransfected with pSV3gptH to 43-3B cells followed by selection on mycophenolic acid and mitomycin-C (MMC). To substantiate repair proficiency of transformants UV-induced unscheduled DNA synthesis (UDS) was determined.

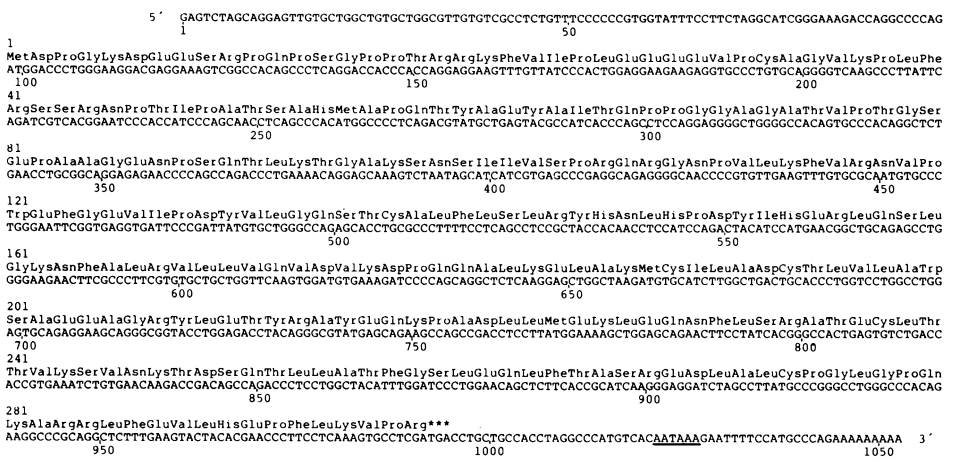

Figure. 1.
Nucleotide sequence and translated amino acids (in three letter code) of the mouse ERCC-1 cDNA. Amino acids are numbered on the left and nucleotides are numbered below the sequence. The polyadenylation signal AATAAA is underlined.

E.coli uvrA and uvrC repair proteins is shown in Figure 2A. Despite the differences between both mammalian proteins their extent of homology with the yeast and bacterial proteins is comparable. The positions of the conserved and non-conserved amino acid changes between the mouse and human protein are schematically depicted in Figure 2B. The overall homology between both proteins is 85%. However, it is striking that the majority (>70%) of the amino acid and nucleotide substitutions are concentrated in the N-terminal part of ERCC-1. Of the first 100 amino acids 70% are homologous whereas the region from 100 to 200 and 200 to 298 have similarities of 97% and 89% respectively. Two amino acid changes are found in the postulated NLS domain, 3 in the suggested ADP-monoribosylation site and none in the potential DNA-binding domain. The 2 conservative substitutions in the NLS (Ala→Thr and Lys→Arg) are at positions which have been shown not to be critical in the SV40 T-antigen NLS (25). Therefore these changes are not expected to abolish a potential NLS function of this domain. The high degree of sequence conservation in the putative DNA-binding domain supports the idea that this part of the protein is very important for its function and cannot

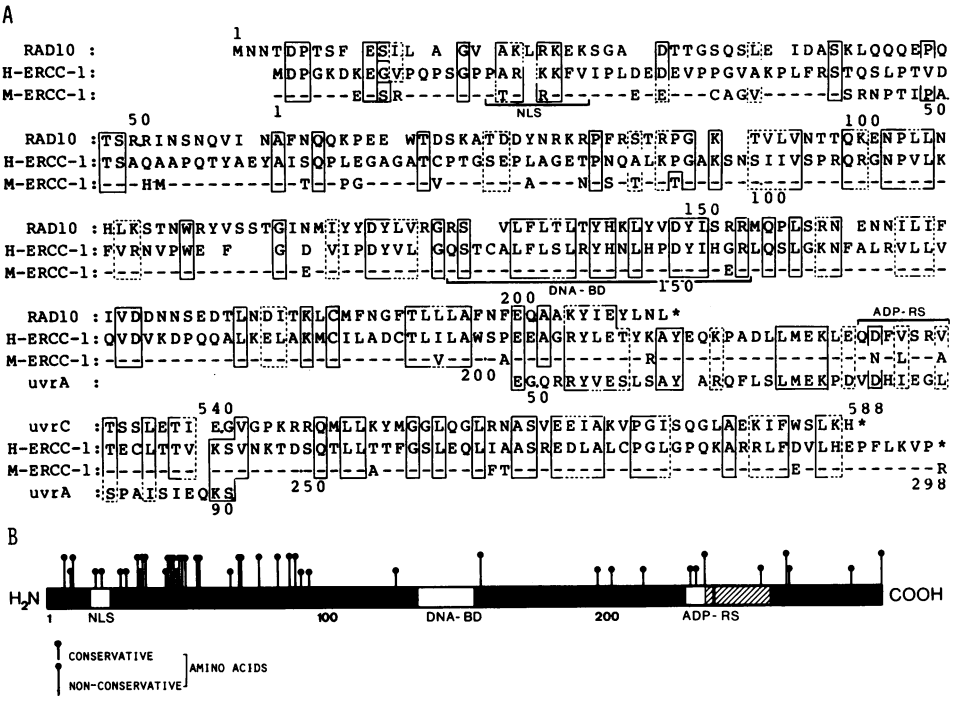

Figure. 2.
A: Alignment of mouse (M) and human (H) ERCC-1 amino acids (in one letter code) with yeast RAD10 and parts of E.coli uvrA and uvrC proteins. Of the mouse protein (horizontal bars) only differences with human ERCC-1 are depicted. Identical amino acids are boxed with solid lines and physico-chemically related residues (K, R; D, E and V, I, L) are indicated by open boxes between the microbial and mammalian proteins. Previously postulated domains (12) are shown by black lines: NLS, nuclear location signal; DNA-BD, DNA binding domain; ADP-RS, ADP-monoribosylation site. Asterisks indicate stopcodons. The RAD10, H-ERCC-1, uvrA and uvrC amino acid sequence and numbering are as described (14, 41, 42, 43 respectively).
B: Schematical presentation of conserved and non-conserved amino acid changes in mouse and human ERCC-1 protein. Conserved amino acids are grouped as follows: A, S, T, P and G; N, D, E and Q; H, R and K; M, L, I and V; F, Y and W. The postulated NLS, DNA-BD and ADP-RS are depicted as boxes (see also Fig.2A). The shaded region corresponds to human exon VIII which was previously found to be subject to alternative splicing (14).

tolerate many amino acid alterations. The 2 conservative amino acid changes in the region to which we have tentatively assigned an ADP-monoribosylation site are at positions not incompatible

with such a function. However, the non-conservative substitution at position 235 (Val→Ala) neighbours the arginine residue that is the actual site of ADP-ribosylation by cholera toxin in the consensus sequence of this domain (26). Hence, the difference between the mouse and human protein at this position makes it less likely that ADP-monoribosylation plays an important role in ERCC-1 protein processing.

At the nucleotide level the similarity of the coding regions of mouse and human ERCC-1 is 82% and of all base changes 70% are at 'wobble' base positions. With respect to the 5' untranslated region it is worth noting that at position -3 upstream of the translation initiation site a conserved C-residue is located which is highly exceptional for eukaryotic mRNAs (27).

Partial characterization of the mouse ERCC-1 promoter region.

The mouse ERCC-1 gene was isolated from a genomic EMBL-3 library using mouse cDNA as a probe. Four overlapping EMBL-3 clones, shown in Figure 3, hybridized to 5' and 3' cDNA probes (not shown) indicating that the mouse ERCC-1 gene has a maximum size of 16-17 kb which is comparable to the previously reported size of the human ERCC-1 gene (28). In order to determine the mouse ERCC-1 promoter sequence a 0.9 kb HindIII-BamHI fragment hybridizing with a 5' cDNA probe was subcloned in pTZ19R, yielding pMHB5. The sequence strategy of this clone, depicted in Figure 3, revealed genomic sequences that were completely identical to the cDNA sequence (Figure 1). Moreover, it was found that the genomic organization of mouse ERCC-1 exon 2 is similar to that of the human gene (not shown). A comparison of the mouse 5' genomic and untranslated cDNA sequence with the corresponding human sequence (28) is presented in Figure 4. A long stretch of homologous nucleotides is found around the transcriptional start site of the human gene (box A, Figure 4), which makes it likely that transcription of the mouse ERCC-1 gene initiates at the corresponding position in box A. This means that deletions or insertions rather than nucleotide substitutions have mainly contributed to the differences in the mouse (113 nucl.) and human (153 nucl.) 5' untranslated regions. We previously showed that the ERCC-1 promoter is located within 170 bp upstream of the transcriptional start site (28). Like the human ERCC-1 promoter

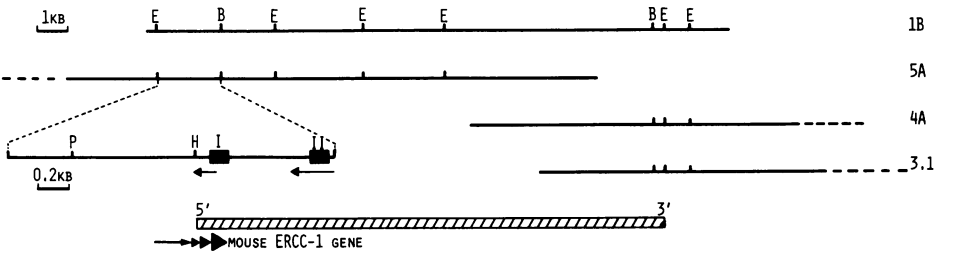

Figure. 3.
Isolation of the mouse ERCC-1 gene. Physical maps are shown of 4 overlapping clones that were isolated from a genomic EMBL-3 library. The position of the ERCC-1 gene (shaded bar) was deduced from hybridization with cDNA probes. Not all restriction sites are given. Non-characterized flanking parts are shown as dashed lines. The 5′ EcoRI-BamHI fragment was studied in detail and sequenced as shown by arrows. Exon I and II are indicated as black boxes. Restriction sites: E, EcoRI; B, BamHI; H, HindIII; P, PstI.

the mouse promoter is lacking clearly identifyable CAAT and TATA boxes as well as GC-rich regions. However, at 60 to 70 bp upstream of the ′cap′-site a domain of 30-40 bp (box B, Figure 4) is coinciding with a region in the human promoter which harbors three CCTCC pentanucleotides that are each spaced by 5 nucleotides. The conservation of this region suggests that it has

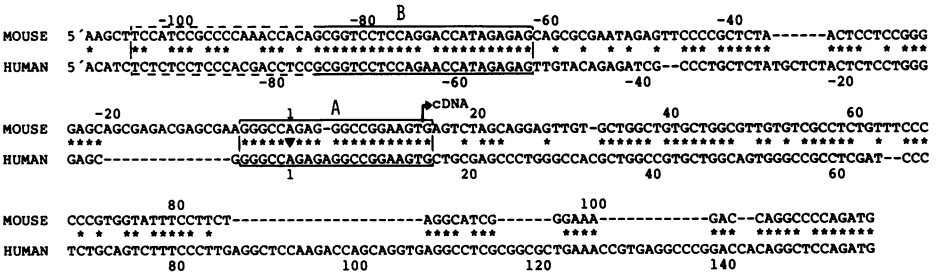

Figure. 4.
Alignment of 5′ mouse and human ERCC-1 sequences upstream of the translational start site (ATG). The mouse sequence is a compilation of genomic and cDNA (Fig. 1) sequence. The human sequence is from Van Duin et al. (14, 28). Boxes A and B show the most conserved regions in the ERCC-1 promoter. Nucleotide numbering is based on transcriptional start site (▲) of the human gene. Gaps are introduced to improve alignment of both sequences.

a regulatory function for ERCC-1 transcription. We have screened the EMBL sequence data base for nucleotide sequences homologous to box B, however, no apparent homology turned up.

Expression of ERCC-1 in mouse organs and developmental stages.

To examine possible regulatory aspects of the ERCC-1 promoter, ERCC-1 expression was investigated by Northern blot analysis of various mouse organs and stages of development. As a control for differences in the amount of RNA the blot was rehybridized with with a probe for the GAPDH mRNA which is considered to be present in relatively constant amounts and has a size very close to that of ERCC-1. As is shown in Figure 5 ERCC-1 transcripts could be detected in all organs (including also liver, stomach, bonemarrow and thymus, not shown) and stages of development examined. The differences in hybridization signal in the different lanes are to a large extent also observed with the GAPDH probe and therefore reflect mainly variation in the quantity of RNA on the filter. From comparison of the hybridization intensities of ERCC-1 and those of other genes (e.g. GAPDH and actin) we infer that ERCC-1 RNA falls in the class of low abundant messengers. Furthermore, in situ hybridization experiments on sections of mouse embryo's did not reveal tissues with a specific high level of ERCC-1 expression (not shown). These results support the conclusion that ERCC-1 is equally expressed at low levels throughout the whole body and at various stages of embryogenesis. The ERCC-1 promoter may thus represent a member of a novel class of promoters specifying a constitutive low expression level.

Mutagenesis of the human ERCC-1 cDNA.

A number of mutant human ERCC-1 cDNAs were constructed (see Materials and Methods) and transfected to CHO 43-3B cells to test for functionality. The results of these experiments are summarized in Figure 6. Construct pcDE-72 is included as negative control, since this clone lacks the alternatively spliced exon VIII which is essential for phenotypic complementation of the mutation in CHO 43-3B cells (14). Similarly all other alterations induced in the C-terminal portion of ERCC-1 appear to be deleterious as well. Construct pcDEASt, that encodes a 'RAD10-like' ERCC-1 protein terminating exactly at the point where (the homology with) the RAD10 gene product stops (see Figure 2A) does

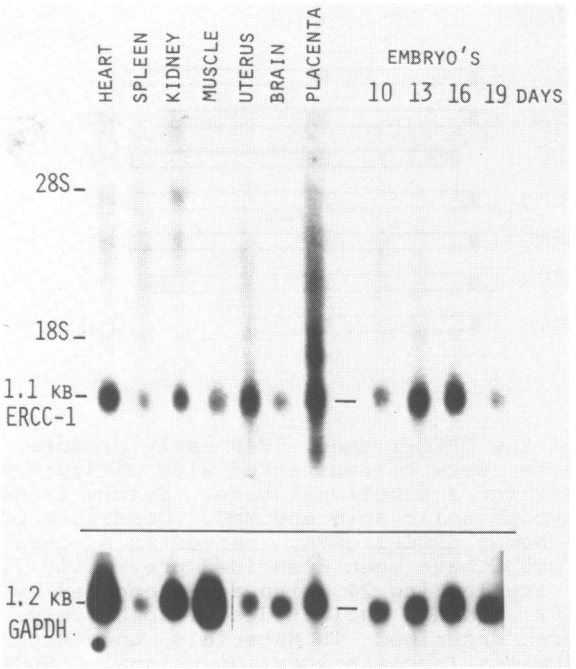

Figure. 5.
Northern blot analysis of mouse organs. Equal amounts (20 µg) of total RNA was size fractionated on a 1% agarose gel and after transfer to nitrocellulose hybridized to ^{32}P-labeled mouse ERCC-1 cDNA (upper part) and a probe for the glyceraldehydephosphate dehydrogenase (GAPDH, 44; lower part). The position of ribosomal subunits and ERCC-1 and GAPDH transcripts are indicated on the left.

not display detectable correcting activity. The same holds for construct pcDEBgl which encodes a truncated protein of 287 amino acids with 17 unrelated C-terminal residues due to a frameshift mutation. These findings are in striking contrast to deletion of the N-terminal ERCC-1 region. We reported previously that a truncated protein lacking the first 54 amino acids (encoded by construct pcD3C) is still able to confer MMC resistance to CHO mutant cells (14). Tab-linker mutagenesis (24) was applied to insert 6 nucleotides in the unique PstI and KpnI site in the coding part of the human ERCC-1 cDNA. The resulting introduction of a leucine and glutamine behind the putative DNA-binding domain (residue 158; construct pcDEMP1) did not affect the repair

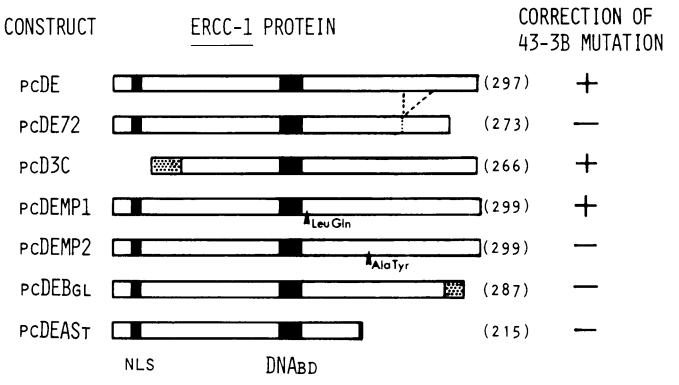

Figure. 6.
Mutagenesis of the ERCC-1 gene. SV40 early promoter driven ERCC-1 cDNA constructs were cotransfected with pSV3gptH to CHO 43-3B cells, to test for a functional gene. Stable transformants were selected on mycophenolic acid and MMC. Construct pcDE, harboring the complete human ERCC-1 cDNA, served as a positive control. pcDE-72 and pcD3C have been described previously (14) and encode proteins that are lacking 24 amino acids encoded by exon VIII and the first 162 coding nucleotides respectively. The other constructs are described in Materials and Methods. Putative functional domains (nuclear location signal, NLS; DNA-binding domain, DNA-BD) that emerged from human and mouse ERCC-1 amino acid comparison are depicted as black boxes. Unrelated amino acid sequences due to frame shifts are shown as dotted areas. Amino acid number is given in parentheses.

function of ERCC-1 in CHO mutants whereas an extra alanine and tyrosine residue distal from amino acid 208 (construct pcDEMP2) inactivated the protein. In conclusion the data presented here are consistent with the notion that the C-terminal part is crucial for ERCC-1 function in contrast to the N-terminus.

DISCUSSION

The complex evolution of the human ERCC-1 gene, resulting in a 'mosaic' type of homology to repair proteins of lower organisms has prompted us to study the gene in other species. Here we describe the isolation and characterization of cDNA and genomic clones of mouse ERCC-1. Transfection of the mouse ERCC-1 cDNA to 43-3B cells conferred UV- and MMC resistance and restored UDS. The induction of a UDS level inbetween that of CHO9 wild type and HeLa cells was also found after transfection of the human ERCC-1

gene and cDNA (8). The finding that UDS is higher in corrected 43-3B cells compared to wild type CHO cells can be explained by differences in nucleotide poolsizes, cell morphology and other factors influencing UDS. Alternatively, it is not excluded that transfected ERCC-1 genes from human and mouse induce a higher UDS level then the endogenous CHO gene.

Southern blot and sequence analysis of mouse genomic clones indicate that the mouse and human gene are similar in size. Preliminary data also suggest a very similar gene organization. The ERCC-1 gene of both mouse and man is driven by an exceptional promoter lacking 'classical' promoter elements. Comparison of the mouse and human promoter region revealed a highly homologous sequence at 50 to 90 bp upstream of the transcriptional start site. We have not noticed this putative promoter motif in other genes The high sequence conservation suggests that ERCC-1 expression is mediated through interaction of transcription factors with this region. However, additional experiments, including 'footprinting' are required to verify this assumption.

In HeLa cells ERCC-1 expression appears to be constitutive and not inducible by UV and MMC (28). We present here that ERCC-1 is transcribed at low levels in all mouse organs and stages of development investigated. Also Northern blot analysis of a number of different human cell lines revealed low constitutive levels of ERCC-1 expression (not shown). Therefore ERCC-1 is probably not only operative in repair of environmentally induced DNA-damage (e.g. UV photo products) but perhaps more importantly in the removal of DNA injuries that are induced at a constitutive rate in all cells and tissues by various intracellular processes.

With respect to ERCC-1 expression at the protein level it is worth noting that rodent and human ERCC-1 mRNAs have a C-residue at position -3 proximal to the translation initiation site. More than 95% of eukaryotic messengers harbor a purine at that position (27) and mutation studies have demonstrated that this dramatically enhances translation efficiency (29, 30). Although the presence of a G-residue at position +4 in the ERCC-1 transcripts might partially compensate for the lack of a purine at -3 (27) it seems likely that ERCC-1 has a conserved low translational efficiency. Yeast DNA excision repair genes also

belong to a category of lowly expressed genes (both at RNA and protein level) as deduced from codon usage and translation initiation consensus (31, 32). It will be of interest to see whether the recently isolated human repair genes XRCC-1 (L.Thompson, pers. comm.) and ERCC-2 (9) are subject to the same mode of translational control.

The cloning and sequence analysis of the mouse ERCC-1 gene has yielded instructive information that can be used to elucidate the function of the protein. Comparison of mouse and human amino acid sequences shows that the N-terminal protein part is much less conserved than the rest of the protein. This is in accordance with the pattern of similarity with the yeast RAD10 protein, which showed a high level of homology in the middle part of ERCC-1 (corresponding with the C-terminal half of RAD10) but only barely detectable similarity between the N-termini of both proteins. The apparent reduced evolutionary pressure for sequence conservation of the 5´ portion of ERCC-1 fits also nicely with the transfection results of the ´decapitated´ cDNA construct pcD3C (Figure 6) which demonstrated that a large N-terminal segment can be omitted without affecting the correcting potential of the ERCC-1 protein. This underlines the idea that ERCC-1 and the yeast RAD10 protein operate in a related step in the intricate excision repair process.

Notwithstanding this extensive homology, one important difference remains between the mammalian and yeast protein: a C-terminal extension of ERCC-1, which appears to be essential for proper functioning of the protein in CHO cells. The finding of this region in the mouse ERCC-1 protein indicates that it was present before the evolutionary lines to mouse and man diverged (65-80 million years ago).

Which evolutionary events may have caused the remarkable difference with the yeast gene? It is not excluded that a primordial ERCC-1 gene has lost its C-terminus to generate a RAD10-like version. Alternatively, C-terminal sequences might have been added to an ancestral RAD10-like gene yielding the ERCC-1-like gene structure. The fact that the extra region of ERCC-1 displays homology with prokaryotic repair proteins may be the result of convergent evolution. Another interesting

possibility is that the ERCC-1 gene has acquired functional domains from prokaryotic genes that originally resided in mitochondrial DNA but have migrated to the nucleus in the course of evolution. We are currently trying to isolate the Drosophila ERCC-1/RAD10 homologue which will hopefully shed more light on the evolution of ERCC-1.

Evolution has provided mouse ERCC-1 with one extra C-terminal amino acid compared to the human protein. Despite this minor difference our mutation studies indicate that the C-terminal part of ERCC-1 seems to be very important for its repair function. This region of ERCC-1 displays significant homology with the C-terminus of the E.coli uvrC protein (13). Interestingly, it appears that a mutation in this part of uvrC also leads to inactivation (33). Taken together these data support the idea that ERCC-1 and uvrC share a similar important function.

Several nuclear proteins are provided with positively charged domains that direct active transport through the nuclear membrane (34-38). A consensus sequence for this nuclear location signal (NLS) has emerged from detailed studies of Kalderon et al. (24, 38) and Colledge et al. (39). Amino acids 16 to 23 are conserved between mouse and man and show structural homology with a SV40 T-antigen NLS (12). It is therefore somewhat unexpected that the N-terminal ERCC-1 part appears to be non-essential. However, several explanations can be put forward for this observation. First, the size of the truncated ERCC-1 protein should allow it to enter the nucleus also by passive diffusion (40). Although probably less efficient, this process might still ensure sufficient ERCC-1 levels in the nucleus to permit correction of the CHO 43-3B mutation. In this context it should be realized that cDNA transformants in general have integrated multiple copies of the ERCC-1 cDNA constructs. Moreover, the cDNA inserts are driven by a SV40 promoter which is expected to accomplish a much higher expression level than the promoter of the unique endogenous ERCC-1 gene. Secondly, it is possible that N-terminal truncated ERCC-1 proteins can only be functional in rapidly dividing cells in culture in which the nuclear membrane is frequently absent, allowing the decapitated ERCC-1 protein to 'sneak in'. Therefore one could speculate that a deletion like in

cDNA construct pcD3C would be more serious in non-proliferating tissues *in vivo*. Finally, it cannot be excluded that ERCC-1 is harboring a second NLS of another type as has been found for polyoma virus large T (34) and the glucocorticoid receptor (35). Currently experiments are underway to investigate the hypothesized NLS in more detail.
Purification of the ERCC-1 protein will be an important prerequisite for studying the function of ERCC-1. The results presented in this paper demonstrate that such studies benefit from detailed analysis of evolutionary related genes.

ACKNOWLEDGEMENTS

We thank Dr. A. Yasui for making construct pcDEASt, Dr. G.C. Grosveld for generously providing the mouse genomic EMBL-library, Dr. F.J.Benham for the GAPDH probe, Dr. M.J.Vaessen for *in situ* hybridization experiments, Drs. M. von Lindern and M. Schilham for mouse organ RNAs and Dr. N.G.J. Jaspers for critically reading the manuscript. Mrs. R.J. Boucke and Mr. J. Fengler are acknowledged for preparing the manuscript. This work was financially supported by EURATOM (contract nr. BI6-141-NL) and MEDIGON, Foundation of Medical Scientific Research in The Netherlands.

* At the recent UCLA meeting on 'Mechanisms and Consequences of DNA damage processing ' in Taos (january 1988) it was decided to renumber the CHO excision deficient complementation groups 1 and 2 to match with the number of the human ERCC-genes correcting their respective defects: ERCC-1, group 1 (formerly group 2); ERCC-2, group 2 (formerly group 1).

REFERENCES

1. 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.
2. Thompson, L.H. and Carrano, A.V. (1983) UCLA symposium on Molecular and Cellular Biology New Series vol. 11, 125-143.
3. Thompson, L.H., Salazar, E.P., Brookman, K.W., Collins, C.C., Stewart, S.A., Busch, D.B. and Weber, C.A. (1987) J.Cell Sci.Suppl. 6, 97- 110.
4. De Weerd-Kastelein, E.A., Keijzer, W. and Bootsma, D. (1972) Nature 238, 80-83.
5. Fischer, E., Keijzer, W., Thielmann, H.W., Popanda, O., Bohnert, E., Edler, L., Jung, E.G. and Bootsma, D. (1985) Mutat.Res. 145, 217-225.

6. Thompson, L.H., Mooney, C.L. and Brookman, K.W. (1985) Mutat. Res. 150, 423-429.
7. Friedberg, E.C. (1985) DNA repair, Freeman & Company, San Francisco.
8. 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.
9. Weber, C.A., Salazar, E.P., Stewart, S.A. and Thompson, L.H. (1988) Mol.Cell.Biol. 8, 1137-1146.
10. Van Duin, M., Janssen, J.H., de Wit, J., Hoeijmakers, J.H.J., Thompson, L.H., Bootsma, D. and Westerveld, A. (1988) Mutat.Res. 193, 123-130.
11. Zdzienicka, M.Z., Roza, L., Westerveld, A., Bootsma, D. and Simons, J.W.I.M. (1987) Mutat.Res. 183, 69-74.
12. 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.
13. Doolittle, R.F., Johnson, M.S., Husain, I., Van Houten, B., Thomas, D.C. and Sancar, A. (1986) Nature 323, 451-453.
14. 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.
15. Jentsch, S., McGrath, J.P. and Varshavsky, A. (1987) Nature 329, 131-134.
16. Ozkaynak, E., Finley, D., Solomon, M.J. and Varshavsky, A. (1987) EMBO J. 6, 1429-1439.
17. Wood, R.D. and Burki, H.J. (1982) Mutat.Res. 95, 505-514.
18. Maniatis, T., Fritsch, E.F. and Sambrook, J.(1982). Molecular cloning, A laboratory Manual (Cold Spring Harbor, New York).
19. Feinberg, A.P. and Vogelstein , B. (1983) Anal.Biochem. 132, 6-13.
20. Meijer, D., Hermans, A., von Lindern, M., van Agthoven, T.,de Klein, A., Mackenbach, P., Grootegoed, A., Talarico, D., Della Valle, G. and Grosveld, G. (1987) EMBO J. 6, 4041-4048.
21. Maxam, A.M. and Gilbert, W. (1980) Meth.Enzymol. 65, 499-560.
22. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J.Mol.Biol. 143, 161-178.
23. Okayama, H. and Berg, P. (1983) Mol.Cell Biol. 3, 280-289.
24. Barany, F. (1985) Gene 37, 111-123.
25. Kalderon, D., Roberts, B.L., Richardson, W.D. and Smith, A.E. (1984) Cell 39, 499-509.
26. Van Dop, C., Tsubokawa, M., Bourne, R. and Ramachandran, J. (1984) J.Biol.Chem. 259, 696-698.
27. Kozak, M. (1984) Nucl.Acids Res. 12, 857-872.
28. Van Duin, M., Koken, M.H.M., Van den Tol, J., Ten Dijke, P., Odijk, H., Westerveld, A.,Bootsma, D. and Hoeijmakers, J.H.J. (1987) Nucl.Acids Res. 15, 9195-9213.
29. Kozak, M. (1986) Cell 44, 283-292.
30. Kozak, M. (1987) Nucl.Acids Res. 15, 8115-8148.
31. Sharp, P.M., Tuohy, T.M.F., Mosurski, K.R. (1986) Nucl. Acids Res. 14, 5125-5143.
32. Hamilton, R., Watanabe, C.K., De Boer, H.A. (1987) Nucl. Acids Res. 15, 3581-3593.
33. Van Sluis, C. and Dubbeld, D. (1983) in: DNA Repair, A Laboratory Manual of Research Procedures vol. 2 (eds. E.C

Friedberg and P.C. Hanawalt) 267-283.
34. Richardson, W.D., Roberts, B.L. and Smith, A.E. (1986) Cell 44, 77-85.
35. Picard, D. and Yamamoto, K.R. (1987) EMBO J. 6, 3333-3340.
36. Burglin, T.R. and De Robertis, E.M. (1987) EMBO J. 6,2617-2625.
37. Stone, J., de Lange, T., Ramsay, G., Jakobovits, E., Bishop, M., Varmus, H. and Lee, W. (1987) Mol.Cell Biol. 7, 1697-1709.
38. Kalderon, D., Richardson, W.D., Markham, A.F. and Smith, A.E. (1984) Nature 311, 33-38.
39. Colledge, W.H., Richardson, W.D., Edge, M.D. and Smith, A.E. (1986) Mol.Cell.Biol. 6, 4136-4139.
40. Paine, P.L., Moore, L.C. and Horowitz, S.B. (1975) Nature 254, 109-114.
41. Reynolds, P., Prakash, L., Dumais, D., Perozzi, G. and Prakash, S. (1985) EMBO J. 4, 3549-3552.
42. Husain, I., Van Houten, B., Thomas, D.C. and Sancar, A. (1986) J. Biol. Chem. 261,4895-4901.
43. Sancar, G.B., Sancar, A. and Rupp, D. (1984) Nucl. Acids Res. 12, 4593-4608.
44. Benham, F.J., Hodgkinson, S. and Davies, K.E. (1984) EMBO J. 3, 2635-2640.