Carcinogenesis vol.12 no.12 pp.2361–2368, 1991

Characterization of the mouse homolog of the *XPBC/ERCC*-3 gene implicated in xeroderma pigmentosum and Cockayne's syndrome

Geert Weeda[1], Libin Ma, Reinier C.A.van Ham, Dirk Bootsma[2], Alex J.van der Eb and Jan H.J.Hoeijmakers[2]

Laboratory for Molecular Carcinogenesis, Sylvius Laboratory, PO Box 9503, 2300 RA Leiden and [2]Department of Cell Biology and Genetics, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands

[1]Present address: Department of Cell Biology and Genetics, Erasmus University, Rotterdam, The Netherlands

The human *XPBC/ERCC-3* DNA repair gene specifically corrects the repair defect of xeroderma pigmentosum (XP) complementation group B and rodent repair mutant cell lines of group 3. The gene encodes a presumed DNA- and chromatin-binding helicase involved in early steps of the excision repair pathway. To study the evolution of this gene, its expression in different tissues and stages of development and to permit the generation of a mouse model of XP by targeted gene replacement in mouse embryonal stem cells, we have isolated the mouse *XPBC/ERCC-3* homolog. Sequence comparison of the predicted protein revealed a 96% amino acid identity with the human gene product. Notably, all postulated functional domains were strictly conserved. The mouse *XPBC/ERCC-3* promoter is—like its human counterpart—devoid of classical promoter elements such as TATA and CAAT boxes and contains several conserved segments with unknown function. One of these conserved regions, consisting in part of a polypyrimidine track, is also present in the *ERCC-1* promoter. The mouse *XPBC/ERCC-3* gene is expressed constitutively at low levels in all tissues examined except for testis, where its expression is significantly enhanced.

Introduction

DNA repair pathways play an important role in protecting genetic material from deleterious effects of DNA damaging agents such as environmental radiation (e.g. UV light, ionizing radiation) and numerous chemicals (see ref. 1 for a comprehensive review). An intensively studied repair process is that of nucleotide-excision repair. In *Escherichia coli* this process constitutes part of the SOS response and includes the concerted action of at least six proteins for which the individual roles have been elucidated in great detail (for review see refs 2–4). However, the mechanism of this pathway in eukaryotic cells is poorly understood. In *Saccharomyces cerevisiae* mutant analysis and gene cloning have revealed the existence of at least 10 genetic loci implicated in excision repair (5,6). In mammals two classes of excision repair mutants can be discerned: at least eight complementation groups within the class of laboratory-induced rodent mutant cells (7,8), and within the category of human excision repair disorder, and seven complementation groups in cells from excision-deficient xeroderma pigmentosum (XP*) patients (designated XP-A to XP-G)† and three groups in the repair disorder Cockayne's syndrome (CS) (13). The autosomal, recessive cancer-prone disease XP is clinically characterized by extreme sensitivity of the skin to sunlight (UV), pigmentation abnormalities and, frequently, mental retardation (see ref. 14 for a review).

Cells from XP patients exhibit extreme photosensitivity and at the biochemical level display a defect in the incision step of the excision-repair pathway (14). CS patients, who have only a subtle defect in excision repair (15), display skeletal deformation, severe mental retardation and photosensitivity but—in contrast to XP—not a dramatic increase of skin cancer. The single patient of the XP group B and a patient recently reassigned to XP group D (formerly thought to comprise a distinct group H—see refs 9–12) display the clinical hallmarks of XP as well as CS. Complementation tests between rodent and XP mutants—although far from complete—until recently have not revealed any overlap between these classes of repair-deficient cell lines (16–18). These findings suggest a considerable biochemical complexity underlying the nucleotide excision repair pathway. Recently, we have cloned the human *ERCC-3* gene (*E*xcision *r*epair *c*ross *c*omplementing rodent repair deficiency) by virtue of its ability to correct CHO mutants belonging to complementation group 3 (19). *ERCC-3* appeared to be the gene defective in XP complementation group B, which at the same time represents CS group C (20). One of the interesting features emerging from the analysis of the human repair genes cloned thus far is striking amino acid sequence conservation of the encoded proteins. The *ERCC-1* gene product was found to be homologous to the *S.cerevisiae* excision-repair protein RAD10 (21). In addition, *ERCC-2* is the homolog of the *S.cerevisiae* excision-repair gene *RAD3* (22), which encodes a DNA-dependent ATPase and helicase (23,24). The human *XPBC/ERCC-3* gene is also thought to specify a DNA- and chromatin-binding helicase (20). Sequence comparison with cloned yeast excision-repair genes has revealed no striking overall homology. To establish further the significance of the putative functional domains in the *XPBC/ERCC-3* protein, to investigate its degree of sequence conservation within mammals and to permit the study of the function of *XPBC/ERCC-3* by reverse genetics in a mouse model system, we have isolated and partially characterized the mouse *XPBC/ERCC-3* gene.

Materials and methods

General procedures

Purification of nucleic acids, restriction digestions, gel electrophoresis, nick translation and filter hybridization were performed according to established procedures (25).

Cell lines and transfection

The UV-sensitive CHO 27-1 cells (26) were cultured in F-10/Dulbecco's minimal essential medium (DMEM) supplemented with 5% newborn calf and 5% fetal calf serum (Gibco), 100 μg/ml streptomycin and 100 units/ml penicillin. Transformed cell lines harboring the *E.coli gpt* dominant marker gene were selected and cultured in XGPT medium as described previously (27). The selection medium was changed every 3–5 days.

*Abbreviations: XP, xeroderma pigmentosum; CS, Cockayne's syndrome; MPA, mycophenolic acid; NLS, nuclear location signal.

†The sole patient until recently comprising the XP-H complementation group belongs to group D (9–12).

UV survival

For UV-survival determinations, exponentially growing cultures were trypsinized and 10^2-10^4 cells were plated onto 6 cm dishes and left to attach for ~12 h. Subsequently, cells were rinsed with PBS and exposed to UV light (254 nm, peak) from a germicidal UV lamp at a fluence rate of 0.5 J/m^2. After cultivation in non-selective medium for 7 days, clones were fixed and stained with Coomassie brilliant blue.

Southern blot analysis

Restriction-endonuclease digestions were carried out as recommended by the manufacturer (Pharmacia). Chromosomal DNA fragments (15 μg) were size-fractionated by gel electrophoresis on a 0.8% agarose gel. After treatment with alkali, the DNA was blotted onto nylon filters (Biotrace) according to the supplier's instruction. Probes were labeled by nick-translation.

RNA preparation and Northern blotting

Total RNA was isolated from adult (BALB/c) mice using the LiCl/urea method (28). RNA was size-fractionated on a 1% agarose–formaldehyde gel, transferred onto nitrocellulose filters and hybridized to a mouse *XPBC/ERCC-3* cDNA probe. The cDNA was labeled by ^{32}P nick-translation (25).

Gene cloning

A mouse brain cDNA library prepared in vector lambda gt10 (29) was screened with a human *XPBC/ERCC-3* cDNA fragment isolated from pCD1 (19). The filters were washed with 1 × SSC (1 × SSC consists of 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% SDS at 65°C. A mouse genomic library cloned into vector EMBL-3 was kindly provided by Dr G.Grosveld, Rotterdam. More than 2 × 10^6 independent clones were screened with a mouse *XPBC/ERCC-3* cDNA probe.

Plasmid construction and sequencing

cDNA and genomic inserts were subcloned into M13mp18, M13mp19 or in pTZ19r (Pharmacia) for sequence analysis by the dideoxy-chain termination method (30), with Sequenase (United States Biochemicals) and M13 or sequence-specific primers. Oligonucleotides were made by an Applied Biosystems DNA synthesiser. The *Eco*RI inserts of the lambda gt10 mouse *XPBC/ERCC-3* cDNA were subcloned into the expression vector pSVL (Pharmacia), yielding plasmid pSM3-1.

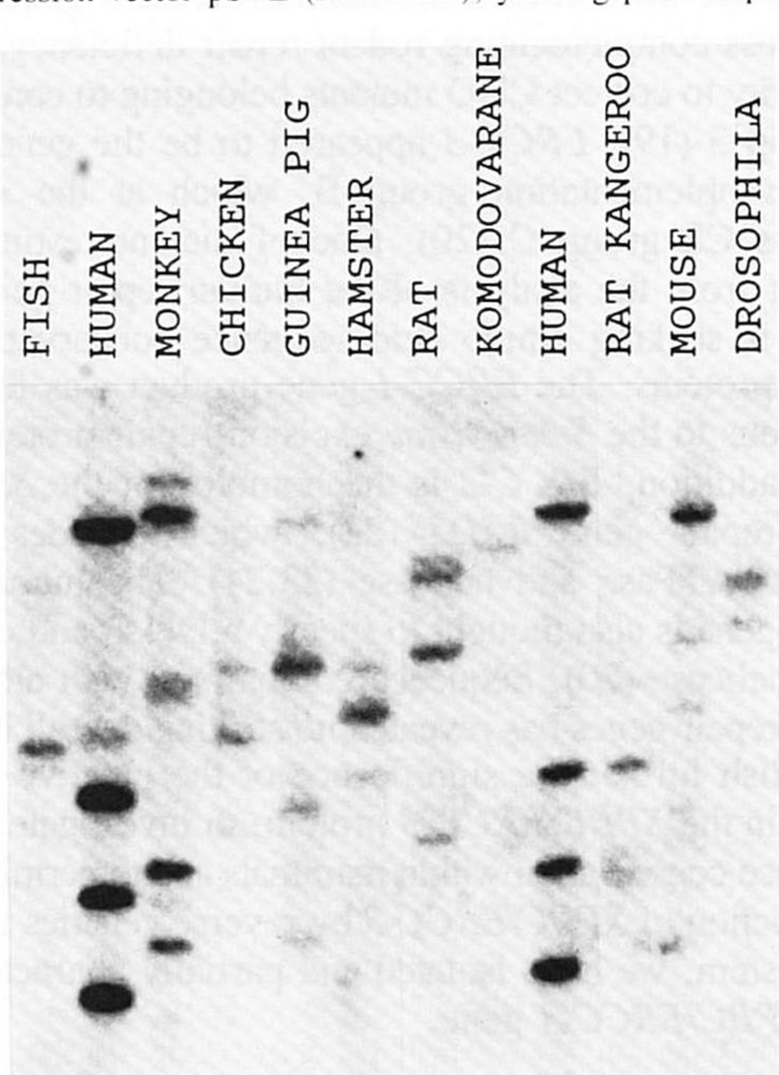


Fig. 1. Southern blot analysis of DNA isolated from various eukaryotic species for the presence of *XPBC/ERCC-3* cross-hybridizing sequences. High mol. wt DNA (15 μg) was digested with *Eco*RI. The DNA was electrophoresed on a 0.8% agarose gel, blotted and hybridized with a ^{32}P-labeled, nick-translated human *XPBC/ERCC-3* cDNA probe (isolated from cDNA clone, pCD1) at 65°C. The filter was washed with 1 × SSC/0.1% SDS at 65°C.

Results

Assessment of the sequence conservation of the XPBC/ERCC-3 gene

To determine whether the *XPBC/ERCC-3* gene is conserved in evolution, Southern blot hybridization with DNA digests of various eukaryotic species was performed. The Southern Zoo-blot presented in Figure 1 shows that under moderately stringent hybridization and washing conditions (1 × SSC, 65°C) specific hybridization was found with DNA of various vertebrates, such as mammals, reptiles, birds, fish and even with invertebrates like *Drosophila melanogaster*. This observation indicates that the *XPBC/ERCC-3* gene is present in many eukaryotes, and that it is strongly conserved.

Isolation and characterization of the mouse XPBC/ERCC-3 cDNA

To isolate the mouse cognate of the *XPBC/ERCC-3*, a mouse brain cDNA library was screened with a human cDNA probe under similar hybridization conditions as used for the Southern blot analysis. Several cDNA clones (designated pM3L-1, pM3L-2 and pM3L-3) were isolated. To establish whether one of these clones encodes a functional *XPBC/ERCC-3* protein, the longest cDNA (pM3L-1, ~2.7 kb) was subcloned in a SV40-based mammalian expression cartridge yielding pSM3-1 (see Materials and methods for details) and transfected into the UV-sensitive, *ERCC-3*-deficient CHO mutant 27-1 together with dominant marker pSV3gptH. The total mycophenolic acid (MPA)-resistant population of 27-1 transformants was tested for its UV resistance. Figure 2 presents the UV survival of wild-type CHO-9, mutant 27-1 and the pooled MPA transformants transfected either with the mouse (pSM3-1) or the human (pSH-1) *XPBC/ERCC-3* cDNA. Clearly, the mouse *XPBC/ERCC-3* cDNA transformants have largely (though not completely) regained wild-type UV resistance, indicating that the isolated mouse cDNA harbors a functional *XPBC/ERCC-3* gene. The mouse cDNA induces a higher level of UV resistance than the human *XPBC/ERCC-3*

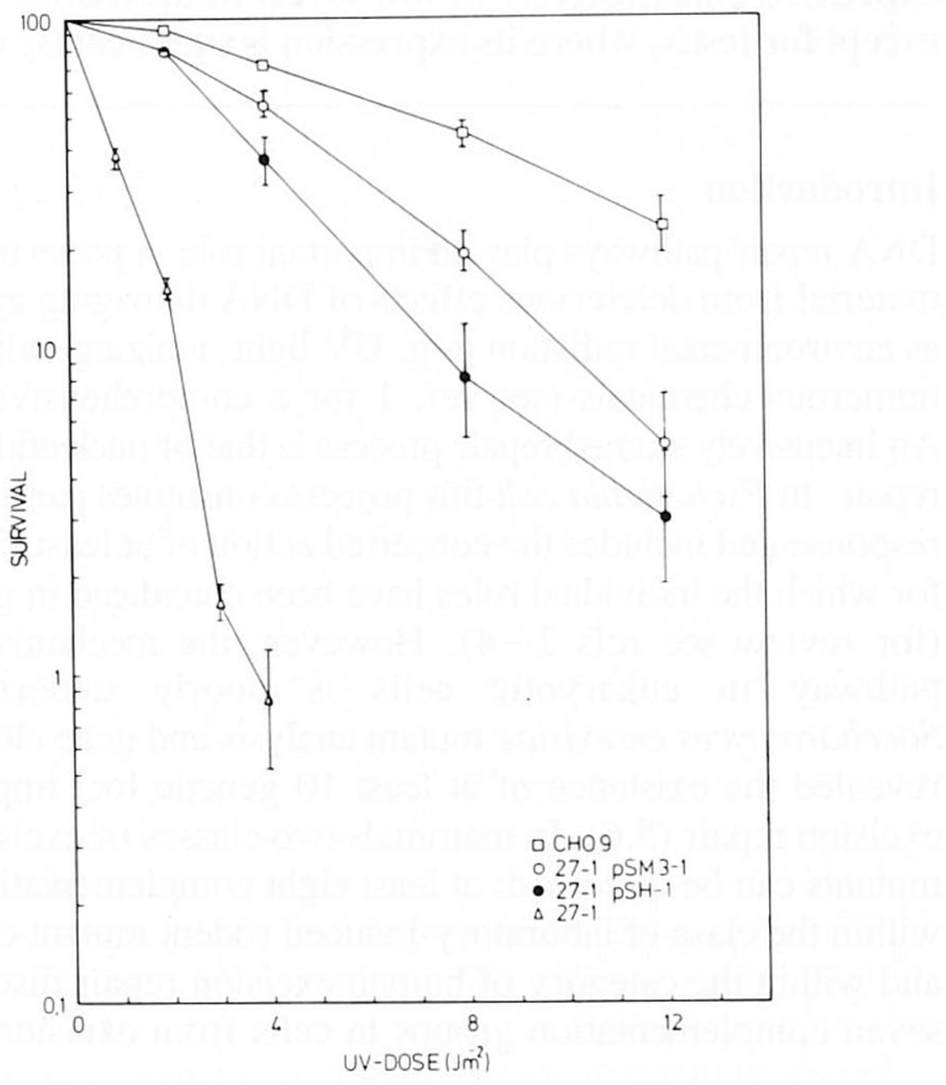


Fig. 2. UV survival of the wild-type parental cell line CHO-9, the UV-sensitive CHO mutant 27-1 and mouse (pSM3-1) and human (pSH-1) *XPBC/ERCC-3* transformants. The curves show the average of two experiments.

cDNA determined in parallel (percentage correction at D37 is 58% for mouse and 40% for human cDNA).

The complete nucleotide sequence of cDNA clone pM3-1 and the predicted amino acid sequence are shown in Figure 3. The first ATG (position 64) starts an open reading frame of 2346 bp. The mouse *XPBC/ERCC-3* cDNA encodes a protein of 783 amino acids (one amino acid longer than the human protein) and a deduced peptide mass of ~89 kd. The alignment of the

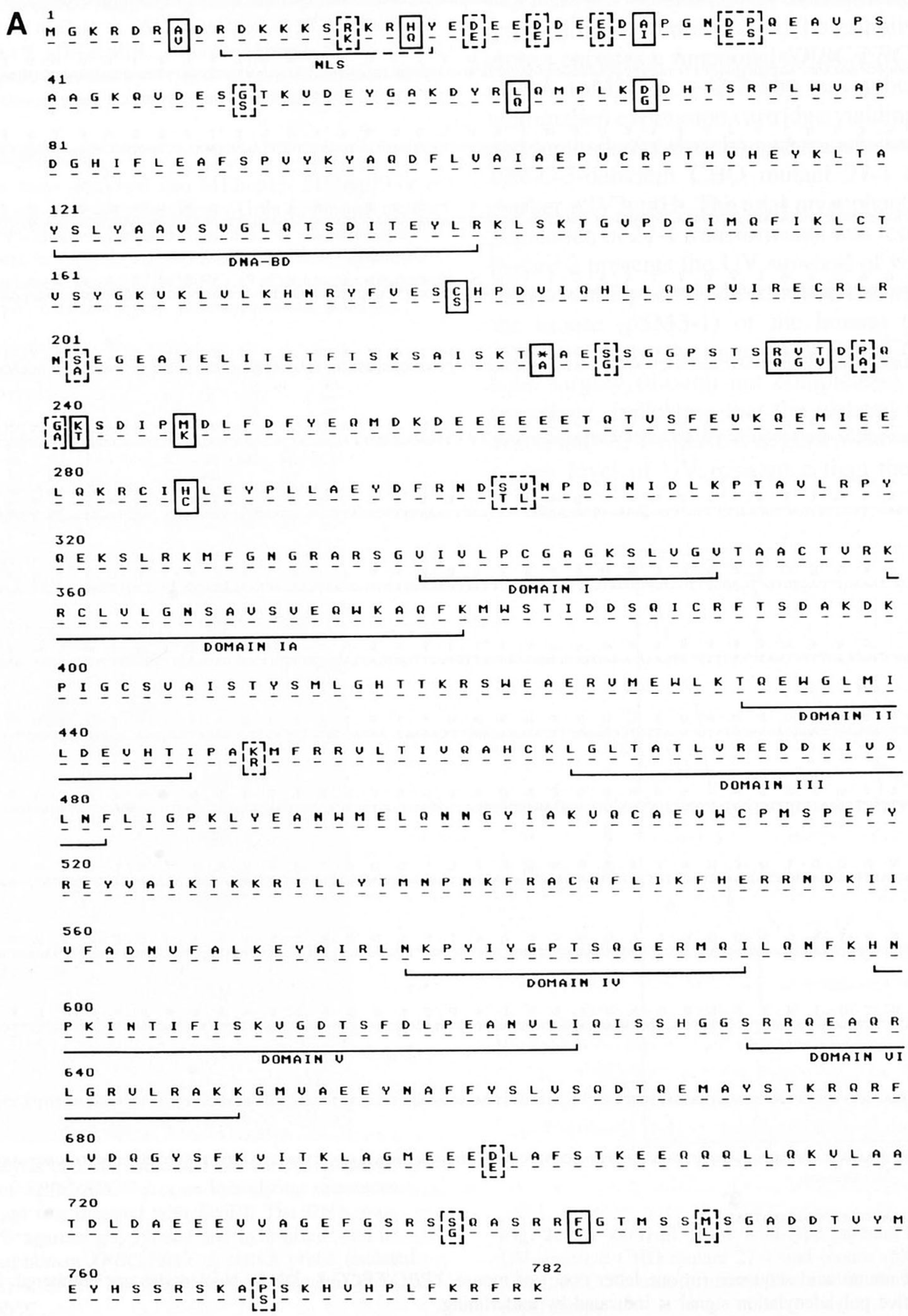
```
5'
                                                               M  G  K  R  D  R  V  D  R  D  K  K  K  S  K  K  R  Q  Y
GCAAGTAACTCCTGCGGAGAGGAGGTGGTGCGGTGTGCCGGCCTGCCTGCATCTAGAGCTGTTATGGGCAAAAGAGATCGAGTGGACCGCGACAAAAAGAAATCCAAGAAGAGGCAGTAT  120

 E  E  E  E  E  D  E  D  D  I  P  G  N  E  S  Q  E  A  V  P  S  A  A  G  K  Q  V  D  E  S  S  T  K  V  D  E  Y  G  A  K
GAAGAGGAAGAGGAAGACGAAGACGACATTCCTGGGAACGAGTCTCAGGAGGCGGTGCCCTCCGCCGCTGGGAAACAGGTGGACGAGTCCAGCACCAAAGTGGATGAGTATGGAGCAAAG  240

 D  Y  R  Q  Q  M  P  L  K  G  D  H  T  S  R  P  L  W  V  A  P  D  G  H  I  F  L  E  A  F  S  P  V  Y  K  Y  A  Q  D  F
GACTACAGACAGCAGATGCCACTAAAGGGTGACCATACCTCCAGGCCCCTCTGGGTGGCTCCTGATGGCCACATTTTCTTGGAAGCCTTCTCTCCAGTTTACAAATATGCCCAAGACTTC  360

 L  V  A  I  A  E  P  V  C  R  P  T  H  V  H  E  Y  K  L  T  A  Y  S  L  Y  A  A  V  S  V  G  L  Q  T  S  D  I  T  E  Y
CTGGTGGCAATTGCAGAGCCGGTGTGCCGGCCTACTCATGTACATGAATACAAGCTAACCGCCTACTCCCTCTATGCAGCGGTCAGTGTTGGGCTGCAAACCAGTGACATTACTGAGTAC  480

 L  R  K  L  S  K  T  G  V  P  D  G  I  I  Q  F  I  K  L  C  T  V  S  Y  G  K  V  K  L  V  L  K  H  N  R  Y  F  V  E  S
CTCAGAAAGCTCAGTAAGACTGGAGTTCCTGATGGAATTATCCAGTTTATTAAGCTGTGCACTGTCAGCTATGGGAAGGTCAAGCTGGTCCTCAAGCACAACAGGTACTTTGTTGAAAGT  600

 S  H  P  D  V  I  Q  H  L  L  Q  D  P  V  I  R  E  C  R  L  R  N  A  E  G  E  A  T  E  L  I  T  E  T  F  T  S  K  S  A
TCCCACCCTGATGTTATCCAGCATCTTCTTCAAGACCCAGTGATCCGGGAATGTCGCTTGAGGAACGCCGAGGGCGAGGCCACCGAACTCATCACAGAGACTTTTACAAGCAAATCTGCT  720

 I  S  K  T  A  A  E  G  S  G  G  P  S  T  S  Q  G  V  D  A  Q  A  T  S  D  I  P  K  D  L  F  D  F  Y  E  Q  M  D  K  D
ATTTCTAAGACGGCGGCGGAAGGCAGTGGCGGGCCTTCTACTTCGCAGGGGGTAGATGCACAGGCCACGTCTGACATACCCAAAGACCTGTTTGATTTTTATGAGCAAATGGACAAGGAT  840

 E  E  E  E  E  E  T  Q  T  V  S  F  E  V  K  Q  E  M  I  E  E  L  Q  K  R  C  I  C  L  E  Y  P  L  L  A  E  Y  D  F  R
GAGGAGGAGGAGGAAGAGACACAGACAGTGTCCTTTGAAGTTAAGCAGGAAATGATCGAGGAGCTGCAGAAGCGCTGCATCTGCTTAGAGTACCCGCTGCTGGCAGAGTATGACTTCCGG  960

 N  D  T  L  N  P  D  I  N  I  D  L  K  P  T  A  V  L  R  P  Y  Q  E  K  S  L  R  K  M  F  G  N  G  R  A  R  S  G  V  I
AATGACACTCTCAATCCTGACATCAACATTGACCTGAAGCCCACAGCCGTACTCAGACCCTATCAGGAGAAGAGCCTGCGGAAGATGTTTGGGAACGGACGAGCACGCTCAGGAGTCATT  1080

 V  L  P  C  G  A  G  K  S  L  V  G  V  T  A  A  C  T  V  R  K  K  R  C  L  V  L  G  N  S  A  V  S  V  E  Q  W  K  A  Q  F
GTTCTTCCCTGTGGTGCTGGGAAGTCCCTGGTCGGCGTGACTGCCGCATGCACTGTCAGAAAGCGCTGTCTCGTCCTGGGCAACTCGGCTGTGTCTGTGGAGCAGTGGAAAGCCCAGTTT  1200

 K  M  W  S  T  I  D  D  S  Q  I  C  R  F  T  S  D  A  K  D  K  P  I  G  C  S  V  A  I  S  T  Y  S  M  L  G  H  T  T  K
AAGATGTGGTCAACCATCGATGACAGCCAGATCTGCCGCTTCACCTCAGATGCCAAGGACAAGCCCATCGGCTGCTCCGTTGCCATTAGCACTTACTCTATGCTGGGCCACACCACCAAA  1320

 R  S  W  E  A  E  R  V  M  E  W  L  K  T  Q  E  W  G  L  M  I  L  D  E  V  H  T  I  P  A  R  M  F  R  R  V  L  T  I  V
AGGTCCTGGGAAGCTGAGAGAGTCATGGAGTGGCTCAAAACCCAGGAGTGGGGCCTCATGATCCTTGACGAGGTGCACACCATTCCAGCCAGGATGTTCCGGCGGGTTCTGACTATTGTG  1440

 Q  A  H  C  K  L  G  L  T  A  T  L  V  R  E  D  D  K  I  V  D  L  N  F  L  I  G  P  K  L  Y  E  A  N  W  M  E  L  Q  N
CAGGCGCACTGTAAGCTTGGTTTGACTGCAACCCTTGTCCGGGAAGATGACAAAATTGTTGACTTAAATTTCCTGATTGGGCCCAAGCTCTATGAAGCCAACTGGATGGAGCTGCAGAAC  1560

 N  G  Y  I  A  K  V  Q  C  A  E  V  W  C  P  M  S  P  E  F  Y  R  E  Y  V  A  I  K  T  K  K  R  I  L  L  Y  T  M  N  P
AATGGGTACATCGCTAAAGTCCAGTGTGCTGAGGTTTGGTGCCCGATGTCTCCCGAGTTCTACCGAGAGTATGTGGCAATCAAAACAAAGAAACGGATCCTGTTGTACACCATGAATCCC  1680

 N  K  F  R  A  C  Q  F  L  I  K  F  H  E  R  R  N  D  K  I  I  V  F  A  D  N  V  F  A  L  K  E  Y  A  I  R  L  N  K  P
AACAAATTCAGAGCCTGCCAGTTTCTGATCAAGTTTCATGAAAGGAGGAATGACAAGATTATTGTCTTTGCTGACAACGTATTTGCCTTGAAGGAATATGCCATTCGGCTGAACAAACCT  1800

 Y  I  Y  G  P  T  S  Q  G  E  R  M  Q  I  L  Q  N  F  K  H  N  P  K  I  N  T  I  F  I  S  K  V  G  D  T  S  F  D  L  P
TATATCTACGGGCCCACGTCCCAGGGAGAACGTATGCAGATTCTCCAGAACTTCAAGCACAACCCCAAAATCAACACCATCTTCATCTCCAAGGTTGGTGACACATCCTTTGATCTGCCA  1920

 E  A  N  V  L  I  Q  I  S  S  H  G  G  S  R  R  Q  E  A  Q  R  L  G  R  V  L  R  A  K  K  G  M  V  A  E  E  Y  N  A  F
GAAGCCAATGTCCTCATTCAGATCTCTTCCCATGGTGGCTCCCGACGGCAGGAAGCCCAGAGACTGGGACGGGTACTCAGAGCCAAGAAAGGGATGGTCGCAGAGGAATACAATGCCTTT  2040

 F  Y  S  L  V  S  Q  D  T  Q  E  M  A  Y  S  T  K  R  Q  R  F  L  V  D  Q  G  Y  S  F  K  V  I  T  K  L  A  G  M  E  E
TTCTACTCCCTGGTATCTCAGGACACACAGGAAATGGCTTATTCTACCAAGCGACAGAGATTCCTAGTGGATCAAGGTTACAGCTTTAAGGTAATCACAAAGCTAGCTGGCATGGAGGAA  2160

 E  E  L  A  F  S  T  K  E  E  Q  Q  Q  L  L  Q  K  V  L  A  A  T  D  L  D  A  E  E  E  V  V  A  G  E  F  G  S  R  S  G
GAGGAGCTGGCGTTCTCCACCAAAGAGGAGCAGCAGCAGCTCCTGCAGAAGGTGCTGGCAGCCACCGACTTGGATGCAGAGGAGGAAGTGGTGGCTGGAGAGTTTGGCTCTAGATCTGGC  2280

 Q  A  S  R  R  C  G  T  M  S  S  L  S  G  A  D  D  T  V  Y  M  E  Y  H  S  S  R  S  K  A  S  S  K  H  V  H  P  L  F  K
CAGGCATCCCGACGCTGTGGCACCATGAGCTCCCTGTCAGGCGCAGACGACACCGTGTACATGGAGTACCACTCCTCCCGAAGCAAGGCCTCCAGCAAGCACGTGCACCCGCTTTTCAAA  2400

 R  F  R  K  *
CGCTTCAGGAAGTGACGCCCACCCTGTGCTCAGGTGAAGACTGGCTGCCTGGATCCGTGTTGAAGAGAGACTTCTTTCTTCACCGGTGTGATCGTTTCTATCCAAGACTGTGTCCTTTCC  2520

TATTGACTGACTGCTGAAAACCAGTCAGTGTGAGGCCTTAACTGCCTTCGTGGGTCCTAGGTTTAAAAAGGAAAGAAAAGCTACTGTTGTTTTTATTCAAACTTGTTAAGAGTGAATCA  2640

                                   3'
TTTCTGTATATAAAAATTATACCACTTAATTAT                                                                                        2673
```

Fig. 3. Nucleotide and deduced amino acid sequence (in one letter code) of mouse *XPBC/ERCC-3* cDNA. Nucleotides are numbered. The asterisk denotes the translation stopcodon. The putative polyadenylation signal is indicated by underlining.

XPBC/ERCC-3 protein with its human homolog, and the postulated functional domains are presented in Figure 4(A). The positions of the conserved and non-conserved amino acid substitutions between the mouse and human protein are schematically depicted in Figure 4(B). The overall degree of identity and similarity between the two proteins is 96 and 98% respectively. It is striking that the majority of the amino acid and nucleotide substitutions are concentrated in three regions: the N- and C-terminal parts of the protein, and between the postulated 'helix−turn−helix' DNA binding motif and the putative nucleotide binding domain, which is the first of a series of seven helicase motifs. It is worth noting that the presumed DNA-binding region and particularly the central part of the protein with the seven consecutive helicase domains are unchanged. At position 224 the mouse *XPBC/ERCC-3* protein harbors an extra alanine residue. One conservative amino acid change (R → K) was found in the postulated nuclear location signal (NLS) at a position which has been shown not to affect the nuclear-targetting function in the SV40 T-antigen NLS (31) which renders the sequence completely consistent with the consensus K K/RxK/R (32). The high degree of sequence conservation in the postulated domains is in agreement with their presumed function. In the 248 bp 3′ non-coding region, the heptanucleotide ATATAAA starting at position 2649, might function as poly(A) addition signal (33). However, as we have not identified a poly(A) tail it is not excluded that the cDNA clones presented in this study lack 3′ sequences carrying additional poly(A) signals.

Cloning of the mouse XPBC/ERCC-3 gene region

The mouse *XPBC/ERCC-3* gene region was isolated from a genomic EMBL-3 library using the full-length mouse cDNA probe. Five partially overlapping EMBL-3 recombinants were isolated from 2×10^6 independent clones (i.e. seven haploid genome equivalents). Figure 5 shows their physical maps. The hybridization and mapping data indicate that ~25 kb of the mouse gene has been cloned but that the 3′ part of the gene is still



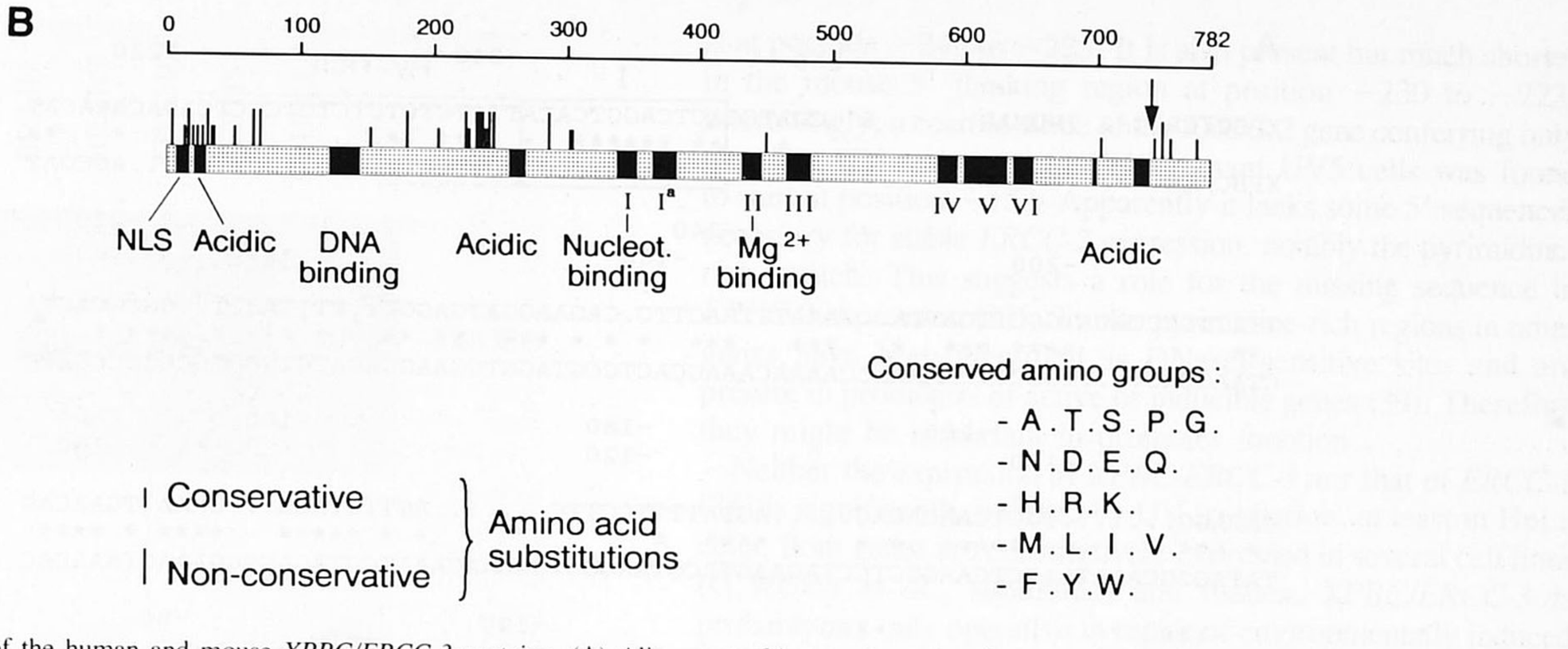


Fig. 4. Comparison of the human and mouse *XPBC/ERCC-3* proteins. (**A**) Alignment of human (upper) and mouse (lower lines) amino acid sequences (in one-letter code). Only differences of the mouse protein from human *XPBC/ERCC-3* have been printed. The mouse sequence is one residue longer than the human which is 782. *Denotes amino acid residue that is absent in the human protein. Non-conservative amino acid changes are boxed with solid lines, and physicochemically related residues (A, S, T, P and G; N, D, E and Q; M, L, I and V; F, Y and W) are indicated by stippled boxes. Previously postulated domains are underlined. NLS, nuclear location signal; DNA-BD, DNA-binding domain; I–VI, helicase domains. (**B**) Schematic presentation of conserved and non-conserved amino acid changes in the mouse and human *XPBC/ERCC-3* protein. Conserved amino acids are grouped as above. The postulated NLS, DNA-BD, acidic domains and helicase domains are depicted as boxes. The site of the XP-B mutation (20), which has previously been determined, is indicated with an arrow.

lacking. This resembles the situation with the human gene, where this part is also strongly underrepresented in all genomic libraries screened (19).

Partial characterization of the mouse XPBC/ERCC-3 promoter region

In order to characterize the mouse *XPBC/ERCC-3* promoter region a 369 bp *Xba*I/*Sma*I fragment hybridizing with a 5′ oligonucleotide primer was subcloned in M13mp8, yielding pM3P-1. Sequence analysis revealed that the genomic sequences were completely identical to the corresponding mouse cDNA sequence (Figure 3). A comparison of the mouse 5′ genomic and untranslated cDNA sequence with the corresponding human region is presented in Figure 6(A). A long stretch of homologous nucleotides is found around the putative transcriptional start site of the human gene (Figure 6A, box B) which makes it likely that transcription of the mouse *XPBC/ERCC-3* gene initiates at the corresponding position in box B. In addition there is a segment around position +40 (box A) which is strongly conserved. For the remaining the 5′ UTR sequences are quite different. This suggests that mainly nucleotide substitutions have contributed to the divergence of the mouse and human 5′ untranslated regions. We have previously shown that the human *XPBC/ERCC-3* promoter is located within 259 bp upstream of the putative transcriptional start site (G.Weeda, L.Ma, A.J.van der Eb and J.H.J.Hoeijmakers, submitted for publication). Like the human *XPBC/ERCC-3* promoter, the corresponding mouse region lacks clearly recognizable CAAT and TATA boxes, as well as SP-1 boxes (Figure 6A). However, at 21–33 bp upstream of the possible mouse 'cap site', one segment (Figure 6A, box C) has retained a high level of sequence homology with the human promoter. Furthermore, at position −231 to −242 a homologous region (Figure 6A, box I) was found that previously has been shown to be present also in the promoter of another human repair gene *ERCC-1* (Figure 6B). The polypyrimidine stretch downstream to the domain conserved in the *XPBC/ERCC-3* gene is present but covers a region of only 8 bp in the mouse promoter. The conservation of these regions, especially the region noted as box I, suggests that they have a regulatory function in the transcription of *XPBC/ERCC-3*. Nucleotide sequence comparison

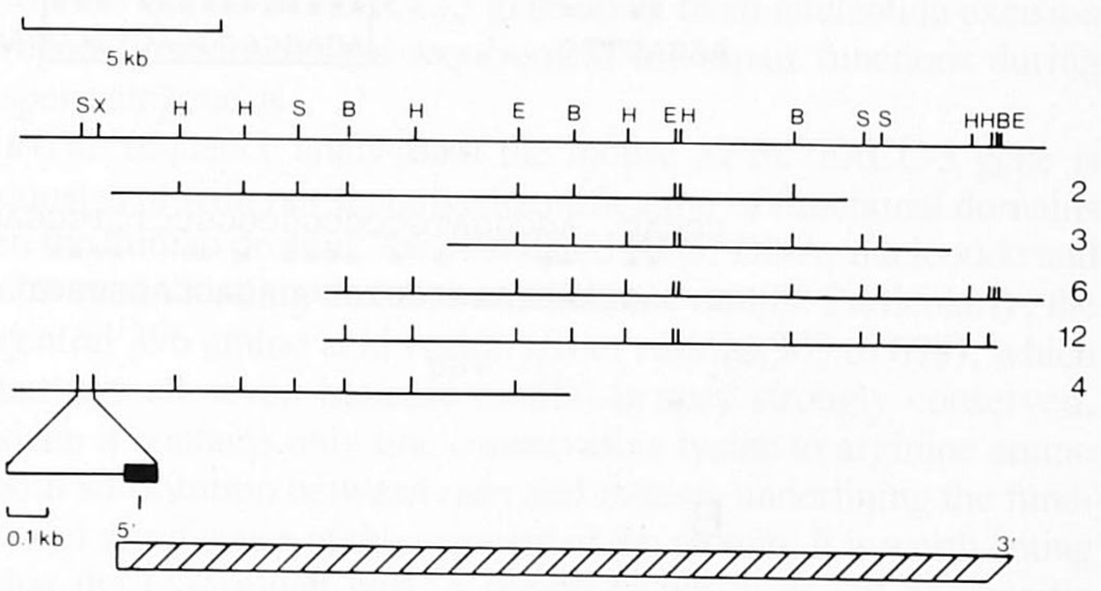
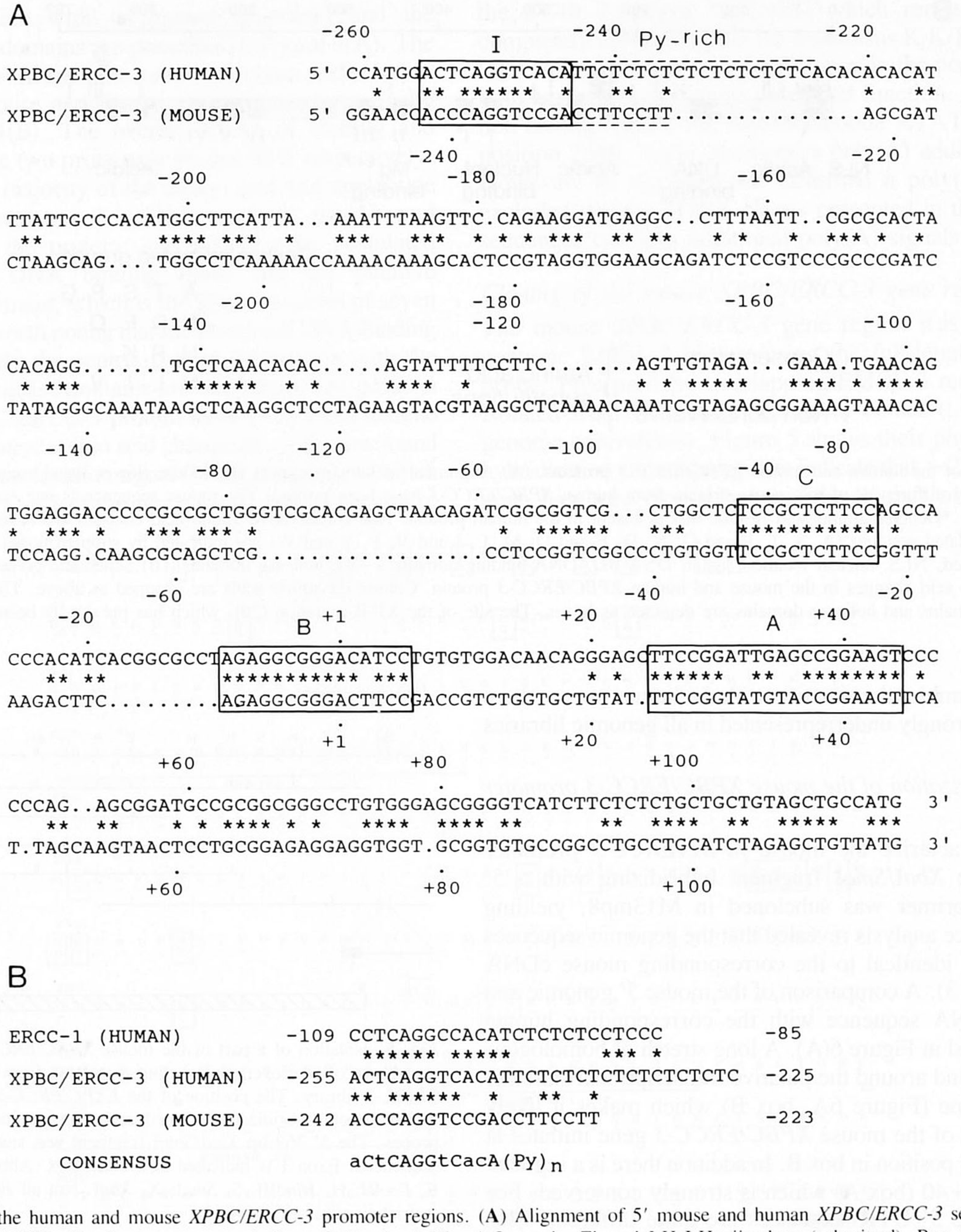


Fig. 5. Isolation of a part of the mouse *XPBC/ERCC-3* gene. Physical maps are shown of five overlapping clones, isolated from a genomic lambda EMBL-3 library. The position of the *EXPC/ERCC-3* gene (hatched bar) was deduced from hybridization experiments with oligonucleotides and cDNA probes. The 5′ 369 bp *Xba*I/*Sma*I fragment was studied in detail and sequenced. Exon 1 is indicated as a black box. Abbreviations: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; S, *Sma*I; X, *Xba*I. Not all *Bam*HI and *Xba*I restriction sites are shown.

of these conserved motifs with the EMBL sequence database did not reveal detectable homology to other promoter elements.

Expression of XPBC/ERCC-3 in mouse organs

The expression of *XPBC/ERCC-3* was investigated by Northern blot analysis of various mouse organs and tissues. As a control for differences in the amount of RNA in each lane the blot was rehybridized with a probe for the human elongation factor I (hEF-1α) (34), which is considered to be present in approximately constant amounts in many tissues (unpublished results). In addition, the Northern blot was rehybridized to probes of two other 'housekeeping' genes with relatively constant amounts of mRNA/cell: GAPDH and Y-actin. Results similar to the hEF-1α hybridization were obtained (data not shown). As shown in Figure 7, *XPBC/ERCC-3* transcripts could be detected in all organs investigated. Most differences in hybridization signal correlate with variations in the amount of RNA on the filter. The *XPBC/ERCC-3* expression levels in testis are, however, significantly increased. Furthermore, similar low levels of the



B

```
ERCC-1 (HUMAN)          -109 CCTCAGGCCACATCTCTCTCCTCCC          -85
                             ****** *****      *** *
XPBC/ERCC-3 (HUMAN)     -255 ACTCAGGTCACATTCTCTCTCTCTCTCTC     -225
                             ** ******  *   **  *
XPBC/ERCC-3 (MOUSE)     -242 ACCCAGGTCCGACCTTCCTT              -223

    CONSENSUS                aCtCAGGtCacA(Py)n
```

Fig. 6. Comparison of the human and mouse *XPBC/ERCC-3* promoter regions. (**A**) Alignment of 5′ mouse and human *XPBC/ERCC-3* sequences upstream of the translation start site (ATG). The human sequence is from G.Weeda, L.Ma, A.J.van der Eb and J.H.J.Hoeijmakers (submitted). Box A, B, C and I represent conserved regions in the *XPBC/ERCC-3* promoter region. Nucleotide numbering is based on the putative transcriptional start site of the human gene. Gaps have been introduced to improve alignment of the sequences. (**B**) Alignment of a part of the 5′ mouse and human *XPBC/ERCC-3* and human *ERCC-1* sequences. The human *ERCC-1* sequence is from ref. (38).

transcript were detected at various states of embryonal development (10−19 days) (data not shown). *XPBC/ERCC-3* expression was also studied in the mouse embryonal carcinoma cell line p19, which can be induced *in vitro* to differentiate to ectodermal, mesodermal and epidermal lineages. A similar basal low level of expression was found in the various cell types, reflecting different directions of early differentiation (data not shown). The *XPBC/ERCC-3* promoter may thus be a member of a class of promoters causing a basal low level of transcription in most tissues.

Discussion

The human *XPBC/ERCC-3* gene is absolutely required for excision of UV-induced dimers and bulky adducts from DNA. It complements the defective incision of UV-irradiated DNA of CHO mutant 27-1 and other members of rodent complementation group 3, and in addition the excision defect in the complex human syndrome XP-B. The predicted amino acid sequence suggests that it encodes a presumed ATP-dependent helicase, which could locally unwind DNA or chromatin structure in the direct vicinity of DNA lesions and thereby facilitate the access of excision repair enzymes to sites of base damage. Alternatively, or in addition, the protein could be involved in releasing a damage-containing oligonucleotide from the DNA by local unwinding of the duplex region after incision. The evolutionary sequence conservation of other excision repair genes, notably *ERCC-1* and *ERCC-2* (21,22), has prompted us to study *XPBC/ERCC-3* in other species.

Transfection of the mouse *XPBC/ERCC-3* cDNA inserted into a mammalian expression vector conferred substantial but not

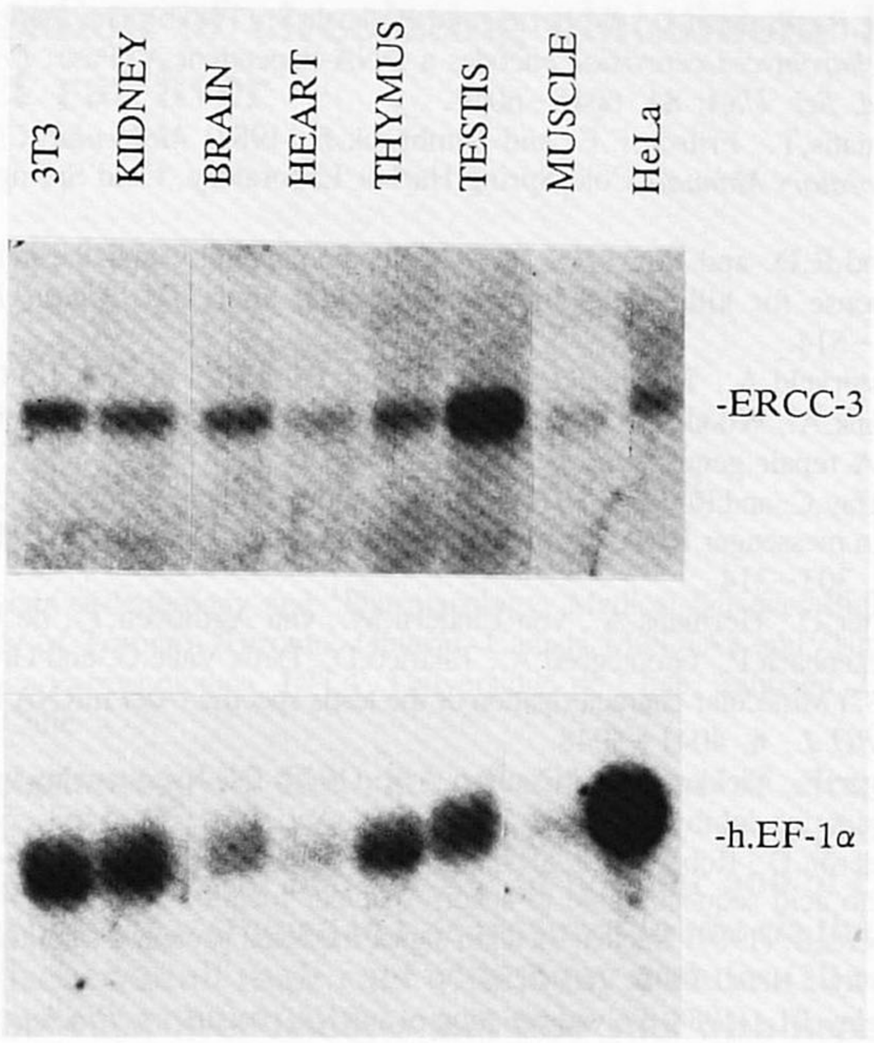


Fig. 7. Northern blot analysis of mouse organs for *XPBC/ERCC-3* RNA expression. Total RNA was size-fractionated on a 1% agarose gel and, after transfer to nitrocellulose, hybridized to ^{32}P-labeled mouse *XPBC/ERCC-3* cDNA. The blot was rehybridized with a 1.4 kb *Pst*I fragment from pICHEF kindly provided by I.Laird (Leiden) carrying the α-subunit of elongation factor 1 (34). The relatively strong hybridization of the hEF-1α probe to the HeLa mRNA is due to the fact that a human hEF-1α probe was used.

complete UV resistance of 27-1 cells to a level intermediate between that of CHO-9 wild-type and 27-1 transfected with the human *XPBC/ERCC-3* cDNA or gene. The incomplete correction by the mouse *XPBC/ERCC-3* gene may be explained by the fact that a heterologous (mouse or human) protein has to function in the context of a Chinese hamster repair machinery. Similar results have been obtained earlier in the case of mutant 43-3B belonging to rodent group 1, after transfection with the human *ERCC-1* gene (27) and in XP-A cells transfected with the mouse *XPAC* gene (35,36). In contrast, complete correction was achieved after transfection of the human *ERCC-2* gene in CHO group 2 mutants (37). Our finding that the mouse *XPBC/ERCC-3* cDNA seems to complement the hamster defect better than the human counterpart is compatible with the idea that the mouse protein fits better in the hamster system than the human gene product.

The human *XPBC/ERCC-3* gene is driven by an exceptional promoter devoid of classical promoter elements (G.Weeda *et al.*, submitted). Comparison of the mouse and human *XPBC/ERCC-3* promoter regions reveal several segments of homologous sequences: one around the putative cap site of the mouse gene and a region at 23−34 bp and at 231−242 bp upstream of the putative cap site. The specific sequence conservation of these fragments suggests that they are implicated in regulation of *XPBC/ERCC-3* expression, e.g. through interaction with transcription factors. However, additional experiments including DNA 'footprinting' and band-shift gel-retardation assays are required to verify this assumption. Comparison of the 5′ flanking sequence of *ERCC-1* and *ERCC-2* with that of the mouse and human *XPBC/ERCC-3* revealed the presence of a pyrimidine-rich stretch in each case: in *ERCC-1* between position −97 to −85 (38) and in *ERCC-2* between position −201 to −168 (22). In the human *XPBC/ERCC-3* gene this pyrimidine-rich region is at position −243 to −225. It is also present but much shorter in the mouse 5′ flanking region at position −230 to −223. Interestingly, a cosmid clone of the *ERCC-2* gene conferring only transient UV resistance to CHO mutant UV5 cells was found to start at position −153. Apparently it lacks some 5′ sequences necessary for stable *ERCC-2* expression, notably the pyrimidine-rich stretch. This suggests a role for the missing sequence in *ERCC-2* expression (22). Similar pyrimidine-rich regions in other genes have been identified as DNaseI-sensitive sites and are present in promoters of active or inducible genes (39). Therefore they might be important in promoter function.

Neither the expression of *XPBC/ERCC-3* nor that of *ERCC-1* (38) is significantly induced by UV irradiation, at least in HeLa cells. Both genes are constitutively expressed in several cell lines (G.Weeda *et al.*, submitted) and tissues. *XPBC/ERCC-3* is probably not only operative in repair of environmentally induced DNA damage (e.g. UV photoproducts) but, perhaps more importantly, in the constitutive removal of DNA injuries induced by various intracellular processes in all cells and tissues. Therefore, it is remarkable that in testis tissue the expression is higher than in the other tissues tested. Interestingly, expression levels of *XPBC/ERCC-3* are induced during the course of spermatogenesis (data not shown). This suggests an additional role for the *XPBC/ERCC-3* gene apart from nucleotide excision repair or an increased requirement for repair functions during spermatogenesis.

The sequence analysis of the mouse *XPBC/ERCC-3* gene is consistent with our tentative identification of functional domains in the human protein: the postulated NLS, DNA, nucleotide and chromatin binding domains and helicase motifs. Particularly, the central 396 amino acid region (from residue 303 to 699), which harbors all seven helicase motifs, is very strongly conserved, since it contains only one conservative lysine to arginine amino acid substitution between man and mouse, underlining the functional significance of this segment of the protein. It is worth noting that the C-terminal part, a region to which as yet no specific functional domain is assigned and which is altered by a frameshift mutation in XP-B (20), is one of the three areas where evolutionary changes occur more frequently than elsewhere. The fact that replacing the C-terminal 41 amino acids by a nonsense sequence completely inactivates excision repair in XP-B strongly suggests that this segment nevertheless has an important role in repair. Apparently, a relatively high level of amino acid substitutions is tolerated in this area without compromising this repair function.

Finally, the overall homology between the mouse and human *XPBC/ERCC-3* gene products is very high; considerably higher than between the human and murine homologs of *ERCC-1* (86% identity, 92% homology; 40) and also higher than the similarity between the corresponding XPAC polypeptides (85% identity, 94% homology; 36). This suggests that the *XPBC/ERCC-3* gene is more strongly conserved in eukaryotic evolution than these two repair proteins and that a (not yet characterized) yeast equivalent must exist. Indeed we have recently identified cross-hybridizing sequences in genomic DNA digests and libraries from *Drosophila, S.pombe* and *S.cerevisiae*. Analysis of these more distant versions of the *XPBC/ERCC-3* will reveal which regions and postulated functional domains are conserved. The cloning of the mouse *XPBC/ERCC-3* homolog should permit the generation of a mouse model of XP and CS by targetted gene replacement in mouse embryonal stem cells to study the function of *XPBC/ERCC-3 in vivo*. These experiments are currently underway.

Acknowledgements

We would like to thank Dr H.van Ormondt for critically reading this manuscript, M.Kuit is acknowledged for the photographical work and Mrs R.Boucke for skilful typing. The work was financially supported by the Netherlands Organization for Advancement of Pure Research (NWO) through the Foundation of Medical Scientific Research (contract no. 900-501-091) and the Dutch Cancer Society (contract no. IKR 90-20).

References

1. Friedberg,E.C. (1985) *DNA Repair*. Freeman and Co., San Francisco.
2. Grossmann,L. and Yeung,A.T. (1990) The Uvr ABC endonuclease of *Escherichia coli*. *Photochem. Photobiol.*, **51**, 749–755.
3. Sancar,A. and Sancar,G.B. (1988) DNA repair enzymes. *Annu. Rev. Biochem.*, **57**, 9–67.
4. Van Houten,B. (1990) Nucleotide excision repair in *Escherichia coli*. *Microbiol. Rev.*, **54**, 18–51.
5. Friedberg,E.C. (1988) DNA repair in the yeast *Saccharomyces cerevisiae*. *Microbiol. Rev.*, **52**, 70–102.
6. Haynes,R.H. and Kunz,B.A. (1981) In Strathern,J., Jones,E. and Broach,J. (eds), *The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 371–414.
7. Collins,A. and Johnson,R.T. (1987) DNA repair mutants in higher eukaryotes. *J. Cell. Suppl.*, **6**, 61–81.
8. Busch,D., Greiner,C., Lewis,K., Ford,R., Adair,A. and Thompson,L.H. (1989) Summary of complementation groups of UV-sensitive CHO cell mutants isolated by large-scale screening. *Mutagenesis*, **4**, 349–354.
9. Johnson,R.T. (1989) Reply to letter by Robbins,J.H. *Hum. Genet.*, **84**, 101.
10. Johnson,R.T., Ellito,G.C., Squires,S. and Joysey,V.C. (1989) Lack of complementation between xeroderma pigmentosum complementation group D and H. *Hum. Genet.*, **81**, 203–210.
11. Robbins,J.H. (1989) No lack of complementation for unscheduled DNA synthesis between xeroderma pigmentosum complementation group D and H. *Hum. Genet.*, **82**, 99–100.
12. Vermeulen,W., Stefanini,M., Giliani,S., Hoeijmakers,J.H.J. and Bootsma,D. (1991) Xeroderma pigmentosum complementation group H falls into complementation group D. *Mutat. Res.*, **255**, 201–208.
13. Lehmann,A.R. (1987) Cockayne's syndrome and trichothiodystrophy: defective repair without cancer. *Cancer Rev.*, **7**, 82–103.
14. Cleaver,J.E. and Kraemer,K.H. (1989) In Scriver,C.R., Beaudet,A.L., Sly,W.S. and Valle,D. (eds), *The Metabolic Basis of Inherited Disease*. McGraw-Hill, New York, Vol. II, pp. 2949–2971.
15. Venema,J., Mullenders,L.H.F., Natarajan,A.T., van Zeeland,A.A. and Mayne,L.V. (1990) The genetic defect in Cockayne's syndrome is associated with a defect in repair of UV-induced DNA damage in transcriptionally active DNA. *Proc. Natl. Acad. Sci. USA*, **87**, 4707–4711.
16. Thompson,L.H., Mooney,C.L. and Brookman,K.W. (1985) Genetic complementation between UV-sensitive CHO mutants and xeroderma pigmentosum fibroblasts. *Mutat. Res.*, **150**, 423–429.
17. Stefanini,M., Keijzer,W., Westerveld,A. and Bootsma,D. (1985) Interspecies complementation analysis between xeroderma pigmentosum and UV-sensitive Chinese hamster cells. *Exp. Cell Res.*, **161**, 373–380.
18. Van Duin,M., Vredeveldt,M.G., Mayne,L., Odijk,H., Vermeulen,W., Klein,B., Weeda,G., Hoeijmakers,J.H.J. and Bootsma,D. (1989) The cloned human DNA excision repair gene *ERCC-1* fails to correct xeroderma pigmentosum complementation groups A through I. *Mutat. Res.*, **217**, 83–92.
19. Weeda,G., van Ham,R.C.A., Masurel,R., Westerveld,A., Odijk,H., de Wit,J., Bootsma,D., van der Eb,A.J. and Hoeijmakers,J.H.J. (1990) Molecular cloning and biological characterization of the human excision repair gene *ERCC-3*. *Mol. Cell. Biol.*, **10**, 2570–2581.
20. Weeda,G., van Ham,R.C.A., Vermeulen,W., Bootsma,D., van der Eb,A.J. and Hoeijmakers,J.H.J. (1990) A presumed DNA helicase encoded by *ERCC-3* is involved in the human repair disorders xeroderma pigmentosum and Cockayne's syndrome. *Cell*, **62**, 777–791.
21. Van Duin,M., de Wit,J., Odijk,H., Westerveld,A., Yasui,A., Koken,M.H.M., Hoeijmakers,J.H.J. and Bootsma,D. (1986) Molecular characterization of the human excision repair gene *ERCC-1*: cDNA cloning and amino acid homology with the yeast DNA repair gene *RAD10*. *Cell*, **44**, 913–923.
22. Weber,C.A., Salazar,E.P., Stewart,S.A. and Thompson,L.H. (1990) *ERCC-2*: cDNA cloning and molecular characterization of a human nucleotide excision repair gene with high homology to yeast *RAD3*. *EMBO J.*, **9**, 1437–1447.
23. Sung,P., Prakash,L., Matson,S.W. and Prakash,S. (1987a) RAD3 protein of *Saccharomyces cerevisiae* is a DNA helicase. *Proc. Natl. Acad. Sci. USA*, **84**, 8951–8955.
24. Sung,P., Prakash,L., Weber,S. and Prakash,S. (1987b) The *RAD3* gene of *Saccharomyces cerevisiae* encodes a DNA-dependent ATPase. *Proc. Natl. Acad. Sci. USA*, **84**, 6045–6049.
25. Maniatis,T., Fritsch,F.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
26. Wood,R.D. and Burki,H.J. (1982) Repair capability and the cellular age response for killing and mutation induction after UV. *Mutat. Res.*, **95**, 505–514.
27. Westerveld,A., Hoeijmakers,J.H.J., van Duin,M., de Wit,J., Odijk,H., Pastink,A., Wood,R.D. and Bootsma,D. (1984) Molecular cloning of a human DNA repair gene. *Nature*, **310**, 425–428.
28. Auffray,C. and Rougeon,F. (1980) Purification of mouse immunoglobin heavy-chain messenger RNAs from total myeloma tumor DNA. *Eur. J. Biochem.*, **107**, 303–314.
29. 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) Molecular characterization of the testis specific *c-abl* mRNA in mouse, *EMBO J.*, **6**, 4041–4048.
30. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
31. Kalderon,D., Roberts,B.L., Richardson,W.D. and Smith,A.E. (1984) A short amino acid sequence able to specify nuclear location. *Cell*, **39**, 499–509.
32. Roberts,B. (1989) Nuclear location signal-mediated protein transport. *Biochim. Biophys. Acta*, **1008**, 263–280.
33. Manley,J.L. (1988) Polyadenylation of mRNA precursors. *Biochim. Biophys. Acta*, **950**, 1–12.
34. Brands,J.H.G.M., Maassen,J.A., van Hemert,F.J., Amons,R. and Möller,W. (1986) The primary structure of the α-subunit of human elongation factor 1. *Eur. J. Biochem.*, **155**, 167–171.
35. Tanaka,K., Satokata,I., Ogita,Z., Uchida,T. and Okada,Y. (1989) Molecular cloning of a mouse DNA repair gene that complements the defect of group-A xeroderma pigmentosum. *Proc. Natl. Acad. Sci. USA*, **86**, 5512–5516.
36. Tanaka,K., Miura,N., Satokato,I., Miyamato,I., Yoshida,M.C., Satoh,Y., Kondo,S., Yasui,A., Okayama,H. and Okada,Y. (1990) cDNA sequence of the xeroderma pigmentosum group A gene encoding a protein with a zinc finger structure. *Nature*, **348**, 73–76.
37. Weber,C.A., Salazar,E.P., Stewart,S.A. and Thompson,L.H. (1988) Molecular cloning and biological characterization of a human gene *ERCC-2*, which corrects the nucleotide excision repair defect in CHO UV5 cells. *Mol. Cell. Biol.*, **8**, 1137–1146.
38. Van Duin,M., Koken,M.H.M., Van den Tol, Ten Dijke,P., Odijk,H., Westerveld,A., Bootsma,D. and Hoeijmakers,J.H.J. (1987) Genomic characterization of the human DNA excision repair gene *ERCC-1*. *Nucleic Acids Res.*, **15**, 9195–9213.
39. Elgin,S.C.R. (1988) The formation and function of DNAse I hypersensitive sites in the process of gene activation. *J. Biol. Chem.*, **263**, 19259–19262.
40. Van Duin,M., van den Tol,J., Warmerdam,P., Odijk,H., Meijer,D., Westeveld,A., Bootsma,D. and Hoeijmakers,J.H.J. (1988) Evolution and mutagenesis of the mammalian excision repair gene *ERCC-1*. *Nucleic Acids Res.*, **16**, 5305–5322.

Received on July 8, 1991; revised on August 28, 1991; accepted on September 6, 1991