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

Geert Weedaw, Libin Ma, Reinier C.A. van Ham, Dirk Bootsm, Alex J. van der Eb and Jan H.J. Hoejmakers

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 polyuridylidine 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 (Excision repair cross complementing 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 gfp 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 10^7 - 10^8 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 0.75 J/cm^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 instructions. 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 32P 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) and 0.1% SDS at 65°C. A mouse genomic library cloned into vector EMBL-3 was kindly provided by Dr. G. Grosfeld, 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 synthesizer. The EcoRI 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 EcoRI. The DNA was electrophoresed on a 0.8% agarose gel, blotted and hybridized with a 32P-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.

**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.

---

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 Zoor-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 pSV3gpH. 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
Characterization of the mouse homolog of XPBC/ERCC-3
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
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-targeting 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 × 10⁶ 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–VII, 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, which has previously been determined, is indicated with an arrow.

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 XbaI/SmaI fragment was studied in detail and sequenced. Exon I is indicated as a black box. Abbreviations: B. BamHI; E, EcoRI; H, HindIII; S, SmaI; X, XbaI. Not all BamHI and XbaI 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 blots were 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
transcript were detected at various stages 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 conservatism 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
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 testsis 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 targeted 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 Ornoudt for critically reading this manuscript. M.Kuit is acknowledged for the photographic 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


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