Nucleic Acids Research, Vol. 19, No. 22 6301–6308

Structure and expression of the human *XPBC/ERCC-3* gene involved in DNA repair disorders xeroderma pigmentosum and Cockayne's syndrome

Geert Weeda*, Libin Ma, Reinier C.A.van Ham, Alex J.van der Eb and Jan H.J.Hoeijmakers[1]
Laboratory for Molecular Carcinogenesis, Sylvius Laboratory, PO Box 9503, 2300 RA Leiden and [1]Department of Cell Biology and Genetics, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands

Received July 8, 1991; Accepted October 10, 1991

GenBank accession no. M31899

ABSTRACT

The human *XPBC/ERCC-3* was cloned by virtue of its ability to correct the excision repair defect of UV-sensitive rodent mutants of complementation group 3. The gene appeared to be in addition implicated in the human, cancer prone repair disorder xeroderma pigmentosum group B, which is also associated with Cockayne's syndrome. Here we present the genomic architecture of the gene and its expression. The *XPBC/ERCC-3* gene consists of at least 14 exons spread over approximately 45 kb. Notably, the donor splice site of the third exon contains a GC instead of the canonical GT dinucleotide. The promoter region, first exon and intron comprise a CpG island with several putative GC boxes. The promoter was confined to a region of 260 bp upstream of the presumed cap site and acts bidirectionally. Like the promoter of another excision repair gene, *ERCC-1*, it lacks classical promoter elements such as CAAT and TATA boxes, but it shares with *ERCC-1* a hitherto unknown 12 nucleotide sequence element, preceeding a polypyrimidine track. Despite the presence of (AU)-rich elements in the 3'-untranslated region, which are thought to be associated with short mRNA half-life actinomycin-D experiments indicate that the mRNA is very stable (t 1/2 > 3h). Southern blot analysis revealed the presence of *XPBC/ERCC-3* cross-hybridizing fragments elsewhere in the genome, which may belong to a related gene.

INTRODUCTION

To cope with DNA lesions induced by physical and chemical agents all living organisms have acquired a complex network of DNA-repair pathways (see ref. 1 for a comprehensive review). The best studied repair process is that of nucleotide-excision repair. In contrast to *E. coli* (see for recent overviews: ref. 2–4) little is known about the molecular mechanism of the nucleotide excision repair pathway in eukaryotes. In mammals two classes of excision repair-deficient mutants can be discerned: a minimum of 8 complementation groups within the class of laboratory-induced rodent mutant cells (5), and within the category of human mutants at least 7 complementation groups in cells from excision-deficient xeroderma pigmentosum (XP) patients (designated XP-A to XP-G—the sole patient comprising group H falls into complementation group D [6–9]) and three groups in the repair disorder Cockayne's syndrome (CS) (see refs. 10 and 11 for recent reviews). This extensive genetic heterogeneity suggests a considerable biochemical complexity underlying the nucleotide excision repair pathway. The impact of nucleotide excision repair mechanisms at the level of the organism is illustrated by the clinical manifestations of these syndromes. Both XP and CS are rare, autosomal recessive disorders characterized by hypersensitivity of the skin to UV exposure and frequently neurological defects. XP patients, in addition, present abnormal pigmentation and other skin defects in sun-exposed parts and a predisposition to skin cancer (see 12 for an extensive review). CS patients, who have only a subtle defect in excision repair (13) display skeletal deformation and severe mental retardation but not a dramatic increase in incidence of skin cancer. Efforts in several laboratories to isolate XP-correcting genes by DNA-mediated gene transfer (14,15) are hampered by the relatively poor transfection properties of most SV40-immortalized human fibroblast lines (16). Recently, Tanaka and co-workers described the cloning of the XP-A correcting gene (XPA-C) after large scale transfection experiments to an SV40-immortalized fibroblast line (17,18). In contrast, considerable progress has been made in cloning human DNA repair genes utilizing UV-sensitive Chinese hamster cells as recipients for DNA-mediated gene transfer. The genes correcting the rodent repair defects are termed Excision-Repair Cross-Complementing rodent repair deficiency or *ERCC* genes, where the number refers to the rodent complementation group, that is corrected by the human gene. In this way the human *ERCC-1, 2, 3, 5* and *6* genes have been isolated (19–23), as well as the *XRCC-1* gene that corrects the X-ray sensitivity of

* To whom correspondence should be addressed at Department of Cell Biology and Genetics, Erasmus University, Rotterdam, The Netherlands

CHO mutant EM-9 (24). The importance of the category of rodent mutants as a model for human repair disorders is stressed by our recent finding that a mutation in the *ERCC-3* gene underlies the inborn defect in XP complementation group B. This group represents a very rare conjunction of XP as well as CS. The XPB-correcting *XPBC/ERCC-3* gene encodes a putative DNA helicase (25). In this report we present the architecture of the *XPBC/ERCC-3* gene, its pattern of expression and studies on its promoter.

MATERIALS AND METHODS

General procedures and Nucleic Acids

Purification of nucleic acids, restriction enzym digestions, gel electrophoresis, nick translation and filter hybridization were performed according to established procedures (26). The fragments used for nick translation were: a 1.3-kb *Pst* I fragment from pRGAPDH-13 (27), carrying a rat GAPDH cDNA; the 1.7-kb *Eco*RI-*Cla*I fragment from pHM-1 (28) carrying exon 3 and 3′ flanking sequences of the human *c-myc* gene. DNA sequences were determined by the dideoxy-chain termination method (32) with Sequenase (United States Biochemicals) or Taq-Track (Promega), and with either M13-or sequence-specific oligonucleotides as primers. Oligonucleotide primers were synthesized in an Applied Biosystem DNA syntheziser.

Cell culture and transfection

The UV-sensitive CHO cell line 27-1 (29) and HeLa cells were grown in DMEM/F10 (1:1) medium supplemented with 5% newborn and 5% fetal-calf serum and antibiotics. Human primary fibroblasts (VH10) cells were cultured in the same medium supplemented with 10% fetal-calf serum. For promoter studies *XPBC/ERCC-3* DNA constructs (2−5 μg) were cotransfected with the dominant marker pSV3gptH (1−2 μg) to 27-1 cells as described previously (21). After approximately two weeks of selection in XGPT medium, cells were reseeded and UV-irradiated with 12 J/m^2 (254 nm peak, at a fluence rate of 0.5 J/m^2); the surviving clones were fixed and counted. Transient transfection experiments were carried out using HeLa TK$^-$ cells as recipients and (10 μg) CAT plasmids as described before (30). CAT activity was assayed as detailed earlier (31).

UV-induction

HeLa cells were grown to near-confluency and after rinsing with PBS (phosphate-buffered saline), UV-irradiated (254 nm peak) and subsequently incubated with culture medium for various periods of time, after which they were harvested and used for poly (A)$^+$ RNA isolation (26).

Determination of mRNA stability

To determine the stability of *XPBC/ERCC-3* mRNA, actinomycin-D was added to exponentially growing cultures of HeLa TK$^-$ cells to a concentration of 5 μg/ml. At several time points after addition of actinomycin-D cells were harvested and total cellular RNA isolated.

XPBC/ERCC-3 promoter constructs

The genomic 4.2-kb *Pst*I fragment of cos1 containing *XPBC/ERCC-3* 5′ sequences, was subcloned into pTZ18 yielding pHEP (see Fig. 2 for relevant restriction sites of the *XPBC/ERCC-3* 5′) region. The *XPBC/ERCC-3* cDNA clone pCD1 has been described previously (21). From pHEP, various *XPBC/ERCC-3* 5′ segments were isolated, using restriction sites indicated in Fig. 5A and cloned in front of the *XPBC/ERCC-3* cDNA. This yielded the minigene constructs pHEP-1 (containing the approximately 1 kb *Hind*III/*Sst*II fragment) and pHEP-3 (containing the 378 bp *Nco*I fragment of which the ATG start codon was removed by Mung-Bean exonuclease). This was confirmed by sequence analysis. Similarly constructs were made in which the *XPBC/ERCC-3* cDNA was replaced with the CAT gene derived from vector pBA-CAT (33). These pHEP-CAT constructs are indicated in Fig. 5A.

RESULTS

Architecture of the XPBC/ERCC-3 gene

Cosmids cos1, cos2 and cos8, harbouring the major part of the human *XPBC/ERCC-3* gene were isolated from a cosmid library originating from a repair-proficient secondary transformant of CHO mutant 27-1. Transfection of the three cosmid clones separately or together to 27-1 cells did not result in the generation of UV-resistant transformants, in contrast to cDNA clones. Detailed physical maps were prepared for the cloned human inserts of the cosmids (21, Fig. 1A). Comparison of chromosomal DNA from the secondary transformants and HeLa cells with the cloned cosmid inserts revealed that the human insert of cos8 differed in one area not covered by cos 1 and 2 from the DNA in the secondary transformant from which the cosmid library was derived and from HeLa DNA. This rearrangement affected a region that was 'coinherited' by all independent transformants, analyzed and therefore likely belonged to the *XPBC/ERCC-3* gene (21 and data not shown). The absence of this segment provides a reasonable explanation for the consistent inability of the cosmids to correct the repair defect of 27-1 cells.

The size of the gene and its location on cos1, cos2 and cos8 was determined in several ways: i. by systematic comparison of specific parts of the cloned *XPBC/ERCC-3* region with the DNA of independent genomic 27-1 transformants using Southern blot analysis ii. By hybridization of different probes of a full-length *XPBC/ERCC-3* cDNA to the inserts of cos1 and cos 8. From these results we deduced that the *XPBC/ERCC-3* gene covers a region of circa 45 kb. of which a segment of approximately 9 kb. constituting the 3′ end of the *XPBC/ERCC-3* gene, is not present in the insert of cos8 due to a rearrangement. A detailed physical map of the *XPBC/ERCC-3* region on cos1, cos2 and cos8 is presented in Fig. 1A. The borders of the gene and the positions of the exons were determined and the intron-exon junctions sequenced. The first intron/exon has an extremely high C+G content (80%; data not shown) and appears to constitute part of a CpG-rich island. A clustering of sites for the restriction enzymes *Bss*HII, *Sst*II, *Sma*I, *Nae*I and *Nar*I with one or more CpG dinucleotides in their recognition sites characteristic for CpG islands (35) occurs in this area, as shown in Fig. 2. The sequence between 70 to 90 bp upstream of the translation start site has the potential to form a stem-loop structure with a calculated ΔG of −15.4 kcal/mol based on estimates of RNA secondary structure stability (36) (see Fig. 3).

As shown in Fig. 3 and 4A, all sequences around the intron-exon borders are consistent with the consensus donor and acceptor splicing signals (37), with the notable exception of the 5′ splice-donor site of the third intron where instead of the customary GT a GC dinucleotide is observed. Identical sequence data were obtained from two independently isolated genomic cosmid clones, ruling out the possibility of a cloning artifact. Potential branch-

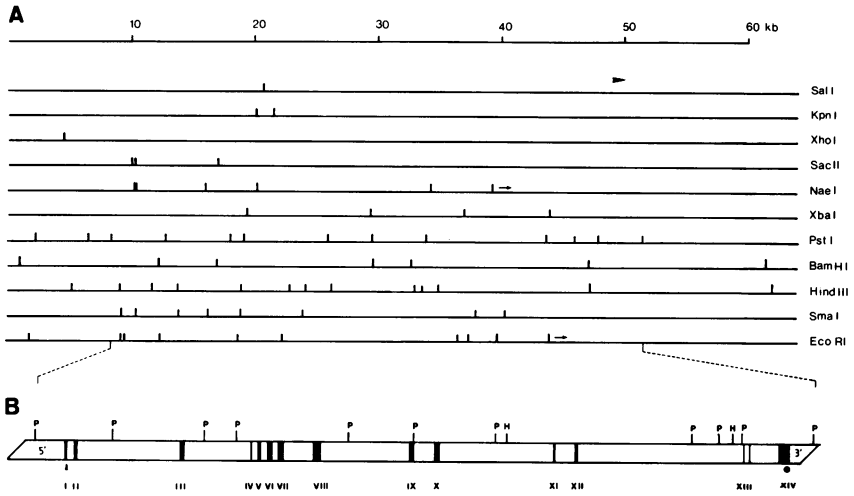

Fig. 1. Genomic organization of the *XPBC/ERCC-3* gene. A. Detailed physical map. The arrowhead (below the kb scale) refers to the genomic region that is not present in cos8 and hitherto refractory to cloning (21). The presence of additional small exon-containing fragments in the missing area is not excluded. Small arrows (top) define the regions for which not all restriction-enzyme cleavage sites have been mapped. Restriction fragments smaller than 1 kb are not indicated except for the 5′ region of the *XPBC-ERCC-3* gene. B. Intron-exon structure. Exons (filled boxes) are indicated with Roman numbers. The exon or exons in the region of which the corresponding genomic part not has been cloned can not be precisely mapped and are depicted with an open box. The distance from exon XIII to exon XIV is fixed, however, the exact location within the 3′ *Pst*I fragment is not known. The dot denotes the polyadenylation signal AATAAA. Abbreviations: P: *Pst*I and H: *Hind*III

sites are present at 30−50 bp upstream of the splice-acceptor sites (Fig. 4A). The deduced transcriptional orientation and the genomic organization are depicted in Fig. 1B. The cloned genomic DNA from cos1, cos2, and cos8 appeared to harbor exons 1 to 12 (corresponding to position 2040 in the cDNA). In order to isolate parts of the remaining 3′ region that was sofar refractory to cloning, we amplified the 3′ part of the HeLa *XPBC/ERCC-3* gene by means of the polymerase-chain reaction (PCR) using cDNA-derived oligonucleotide primers covering the cDNA sequence between positions 2158 and 2750. The amplification resulted in a fragment of approximately 2.1 kb, containing the last exon and 153 bp from the preceding one. Using amplimers based on more 5′ cDNA sequences that where absent in cos8, we did not succeed in amplifiying the missing genomic fragment. This may be explained by the presence of a large intron in the region between the oligonucleotide primers used. From the location of the known exons and their borders it can be deduced that there must be at least one exon between nucleotide 2040 and 2158 which is missing in our genomic clones.

Therefore, the human *XPBC/ERCC-3* gene consists of a minimum of 14 exons ranging in size from 50 to 439 bp spread over a region of approximately 45 kb (Fig. 1A). At the 3′ end of the gene a 309-bp non-coding region (including the stop codon TGA) contains the polyadenylation signal AATAAA (38) which is situated 28 bp 5′ to the poly (A) addition site (Fig. 3). At positions 2694 and 2705 two AUUUA motifs are located which in other genes have been implicated in determining mRNA stability (see below) (39).

The XPBC/ERCC-3 promoter

The nucleotide sequence of the region upstream of the first exon is shown in Fig. 3. To determine the position of the start site of the mRNA S1 analysis and primer-extension was attempted, but did not reveal a major cap site, probably due to the high GC-content of the entire region (70%). However, transfection of cDNA expression plasmids into human cells revealed a transcript

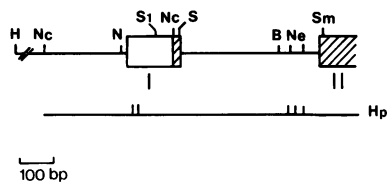

Fig. 2. Fine map of the 5′ end of the *XPBC/ERCC-3* gene. Exons I and II are indicated by boxes. The non-coding region is denoted by an open box, the coding sequence by a hatched boxes. The genomic fragments *Hind*III/*Sac*II (site S1); *Hind*III/*Nar*I and *Nco*I/*Nco*I were used for the minigene constructs and the promoter CAT plasmids as described in Materials and Methods. Abbreviations: B: *Bss*HII; H: *Hind*III; Hp: *Hpa*II; Ne: *Nae*I (2 sites close together); N: *Nar*I; Nc: *Nco*I; S: *Sac*II; Sm: *Sma*I.

of the same size as the endogenous *XPBC/ERCC-3* mRNA, indicating that the cDNA must be (nearly) full-length (data not shown). Moreover, 17 and 19 nucleotides upstream of the start of the longest cDNA clone pCD1, two putative cap sites were found matching with the loosely defined consensus for transcription initiation PyPyCAPyPyPy (starting on the A residue) (Fig.3; 40). A comparison of the human 5′ *XPBC/ERCC-3* genomic sequence with the corresponding mouse sequence indicated a long stretch of homologous nucleotides around this region (Weeda et al. unpublished results). Therefore, we think that this conserved segment is a likely candidate for representing the transcription initiation site, however, further experiments are required to prove this presumption. The 5′ upstream region lacks the canonical TATA and CAAT promoter signals but harbors a reverse sequence motif matching the consensus SP1-binding site (G/T)GGGCGG(G/A)(G/A)(C/T) (41) at positions −89 to −81 (indicated in Fig. 3 with 'GC'). A noteworthy feature of the 5′ flanking region is the presence of a pyrimidine-rich stretch at position −243/−225 (Fig. 3, dotted line, Fig. 4B). Upstream

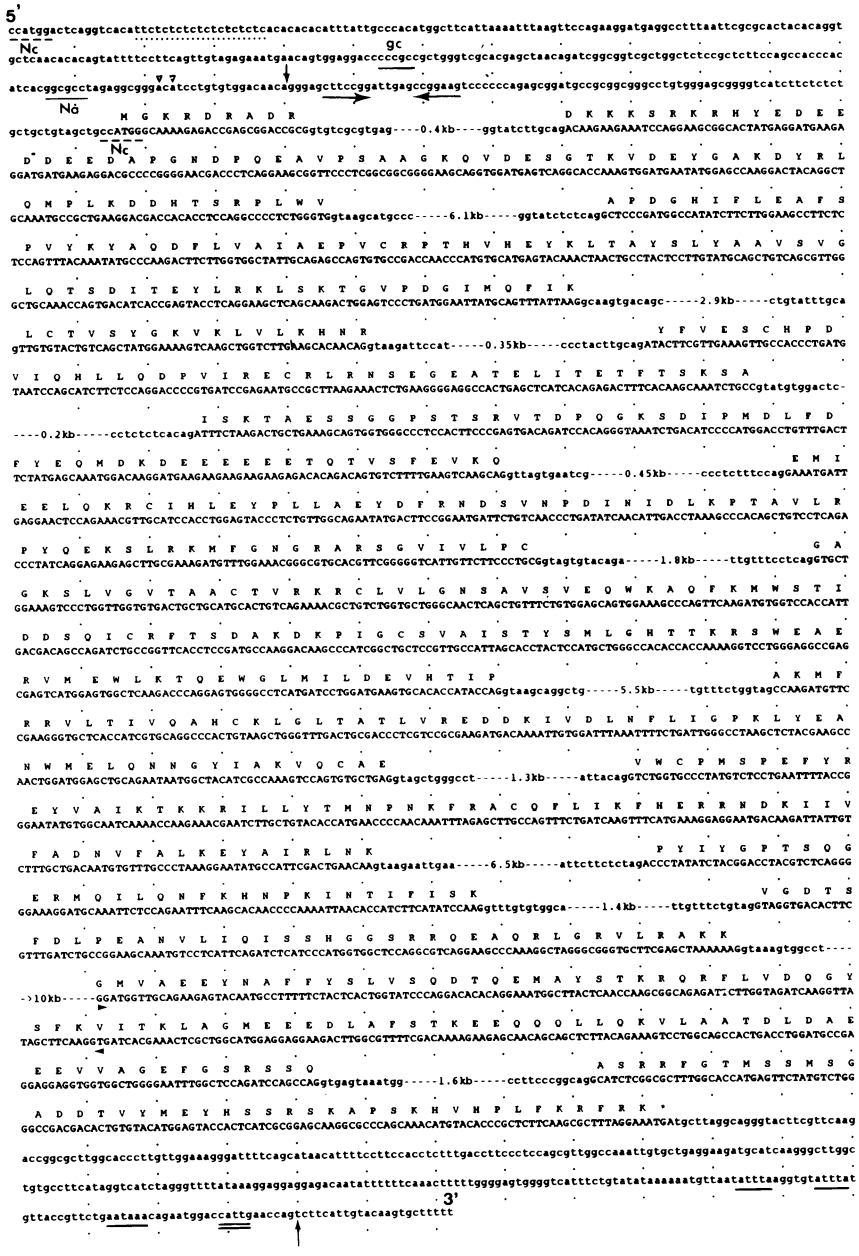

Fig. 3. Nucleotide sequence of the 14 exons in the *XPBC/ERCC-3* gene including intron-exon junctions, and 5′ and 3′ flanking genomic sequences. Coding sequences are depicted in upper case. The deduced amino acid sequence is indicated above the nucleotide sequence in one-letter code. A pyriminide-rich stretch in the 5′ region is indicated by a dotted line, the putative transcriptional start sites by open triangles. The vertical arrow in the 5′ region denotes the start of the longest cDNA clone. The *Nco*I (Nc) and *Nar*I (Na) restriction-sites used for promoter constructs are indicated. The G/C-rich box (potential SP-1 binding site) is indicated by (GC) and single underlining, the inverted repeat is indicated by two large arrows, the (AU)-rich sequences and polyadenylation signal are underlined. The CAYTG region is indicated by double underlining. The vertical arrow in the 3′ region below the sequence indicates the polyadenylation site as determined by sequence analysis. 3′ *XPBC/ERCC-3* flanking sequences are determined from DNA clones obtained by inverted PCR. The region between the arrowheads has not been cloned at the level of the genome and may contain extra introns.

of this polypyrimidine stretch a sequence motif: CTCAGGT/CCACA (−255 to −244) (Box I) is located, which is also present in front of a polypyrimidine stretch in the 5′ region of the *ERCC-1* gene (42) (see fig. 4B). The conservation of this region suggests that it has a function in regulation of transcription of both genes. We have searched in the EMBL sequence data base for nucleotide sequences homologous to this region, but to no avail.

To verify whether the *XPBC/ERCC-3* 5′ flanking sequences can drive transcription, several constructs were made by inserting *XPBC/ERCC-3* 5′ fragments in front of the chloramphenicol acetyltransferase gene (CAT) of vector pBACAT (see Materials and Methods for details and Fig. 5A). The resulting constructs were tested for transient expression by transfection into HeLa TK$^-$ cells and assay for CAT activity 48h after transfection. As shown in Fig. 5B two constructs, pHEP-cat1 and pHEP-cat3 induced approximately the same CAT activity. A deletion clone (pHEP-cat5) ending at position −14 did not show any CAT activity, which is in agreement with our tentative assignment of the cap site. The findings imply that the region 259 bp upstream of the putative cap site harbours promoter activity. As a negative

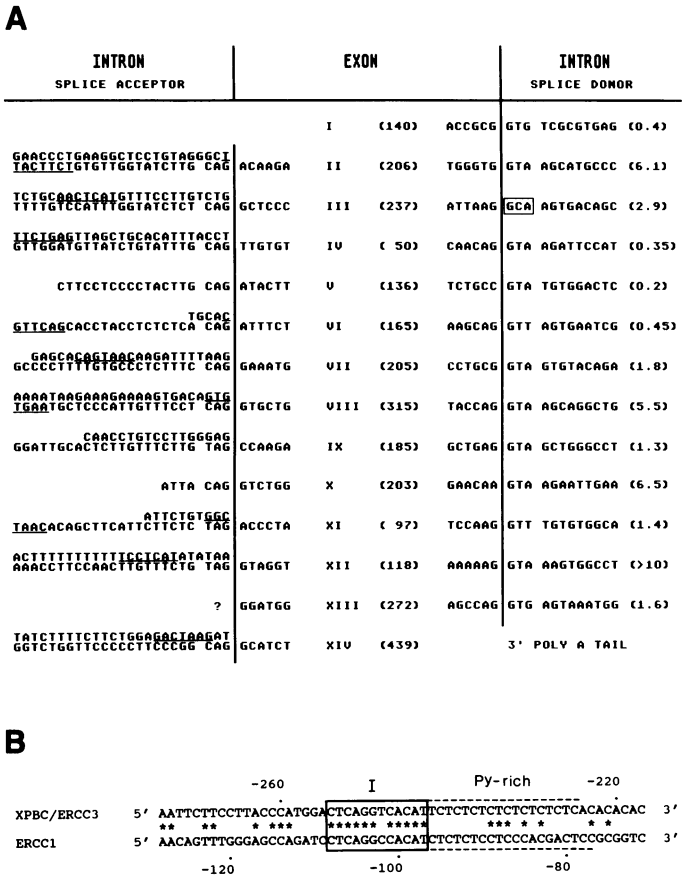

Fig. 4. A. Structural organization of the *XPBC/ERCC-3* gene. The nucleotide sequence of each intron-exon junction is shown. Vertical lines represent intron-exon borders, Splice acceptor and donor sequences are all except one in accordance with the reported consensus sequence $(Y)_nNCAG/GT$ and AAG/GTA respectively (40). The third splice donor site harbouring an unusual GC dinucleotide is boxed. Exon numbering is as in Fig. 1B. The size of each intron and exon (in kb and bp, respectively) is given between brackets. Potential branch sequences matching the loosely defined (YNYTRAR) consensus, if present are underlined. B. Alignment of a part of the 5' human *XPBC/ERCC-3* and *ERCC-1* regions (42). The homopurine-homopyrimidine tract (Py) and the homologous region (box I) are indicated. Nucleotide numbering is based on the (putative) transcriptional start sites of both genes.

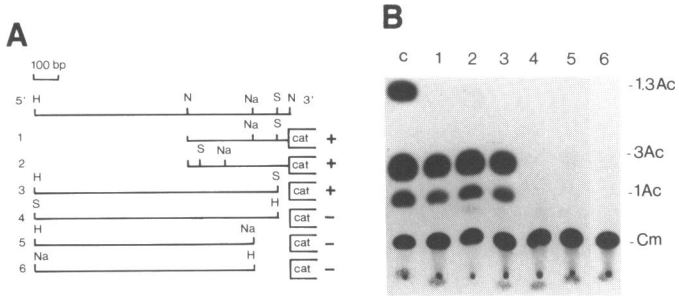

Fig. 5. Promoter activity of 5' *XPBC/ERCC-3* flanking sequences. A. The transfected promoter-CAT constructs are shown schematically. CAT activity is indicated by + or −. 1:pHEP-cat1; 2:pHEP-cat2; 3:pHEP-cat3; 4:pHEP-cat4; 5:pHEP-cat5; 6:pHEP-cat6. Abbreviations: H: *Hind*III; Na: *Nar*I; N: *Nco*I; S: *Sst*II. B. Analysis of the human *XPBC/ERCC-3* promoter in HeLa TK^- cells. CAT-expression plasmids containing the various promoter fragments were transfected into HeLa Tk^- cells. After normalization of the protein content CAT activity was measured. The μg promoter CAT plasmid was used for each transfection. For the RSVCAT control (lane c: pRSVCAT) half of the lysate was used. Numbering corresponds with Fig. 4A.

Table I.

Transfected DNA		fraction of transformant surviving UV-dose*
pcD1	cDNA + SV40 early promoter	31
pcD1-SV	cDNA without promoter	0.5
pHEP-1	cDNA + 1.1 kb promoter fragment	20
pHEP-3	cDNA + 0.4 kb promoter fragment	20

*average of 5 dishes (ca. 100−200 colonies per dish)
Identification of the *XPBC/ERCC-3* promoter region. *XPBC/ERCC-3* promoter activity was determined by counting the number of UV-resistant clones after co-transfection of *XPBC/ERCC-3* minigene constructs and the dominant marker pSV3gptH into 27-1 cells, reseeding the colonies and irradiation with 12 J/m^2. The relative survival value (%) was determined by dividing the number of colonies in UV-irradiated dishes by the corresponding number of unirradiated dishes.

control, the promoter fragments of pHEB-cat1, −3 and −5 in the inverted orientation (yielding pHEP-cat2, −4 and −6 respectively) were transfected into HeLa TK^- cells. Surprisingly, pHEP-cat2 induced also clear CAT activity (Fig. 4A, lane 3). This suggests that the 376-bp *Nco*I fragment also includes a promoter and cap site in the reverse direction.

Further proof that the promoter of the *XPBC/ERCC-3* gene is located between the presumed cap site and −259 was obtained using plasmids containing genomic fragments in front of the *XPBC/ERCC-3* cDNA (see Materials and Methods) and co-transfection of these constructs to the UV-sensitive 27-1 hamster cells with the dominant selection marker plasmid pSV3gptH. Transformants were selected with mycophenolic acid for expression of the co-transfected *E. coli* gpt gene and with UV light for *XPBC/ERCC-3* expression. The results summarized in Table I show that pCD1, harboring *XPBC/ERCC-3* cDNA under the direction of the strong SV40 early promoter yielded a comparable number of UV-resistant clones as *XPBC/ERCC-3* constructs driven by the endogenous promoter.

Regulation of *XPBC/ERCC-3* gene expression

Northern blot analysis of various human cell lines revealed a low level of *XPBC/ERCC-3* transcription (21, and unpublished results). To investigate whether the *XPBC/ERCC-3* promoter is induced by DNA-damaging agents, Northern blot analysis was performed on poly(A)$^+$ RNA of HeLa cells at several time points after UV-irradiation (1 J/m^2). Hybridization with the probe for the γ-actin gene indicated that approximately equal amounts of poly (A)$^+$ RNA were loaded in each lane. As an internal control for UV-induction, the filter was rehybridized with a metallothionein IIa (MTIIa) cDNA probe. It has been reported that MTIIa expression is strongly induced by UV light and the phorbol ester, TPA treatment (43). From Fig. 6 it is evident that UV irradiation with a UV dose of 1 J/m^2 (or 10 J/m^2; data not shown) did not result in significant changes of *XPBC/ERCC-3* transcription during a period of 12 h. Although HeLa cells are of epidermal origin, they are transformed and may have lost a UV-inducible response present *in vivo*. Therefore, we have also analyzed the effect of UV-irradiation on *ERCC-3* expression in short term cultured skin keratinocytes. However, also in this case no significant UV-induction was observed (data not shown).

To examine whether expression of *XPBC/ERCC-3* fluctuates with the proliferative state and/or with the stage in the cell-cycle, primary human diploid fibroblasts were synchronized by growth arrest at high cell density and stimulated to proliferation by

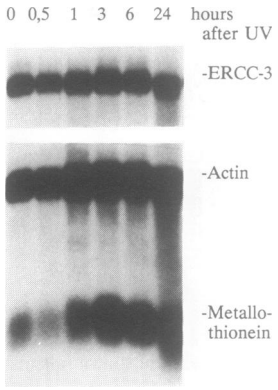

Fig. 6. Effect of UV irradiation on *XPBC/ERCC-3* transcription. Poly(A)$^+$ RNA (approximately 7.5 μg) of exponentially growing UV-irradiated (1 J/m^2) HeLa cells was size-fractionated on a 1% agarose gel and after blotting to nitrocellulose hybridized to a ^{32}P-labeled, nick-translated *XPBC/ERCC-3* cDNA probe, a γ-actin cDNA probe and to a metallotionein IIA cDNA probe.

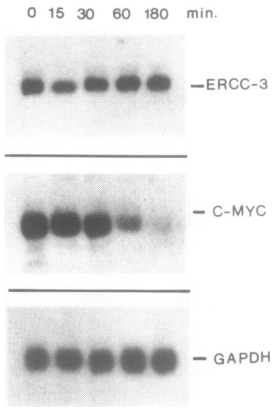

Fig. 7. Effect of actinomycin-D treatment of the *XPBC/ERCC-3* mRNA stability. Total cytoplasmic RNA (50 μg) was isolated at the time points indicated from HeLa cells which were grown in the presence of 5 μg/ml actinomycin-D. After size fractionation of equal amounts of RNA, the RNA was blotted onto nitrocellulose filters. The filter was hybridized with an *XPBC/ERCC-3* cDNA probe. The same filter was rehybridized with a human *c-myc* genomic probe and GAPDH cDNA probe.

reseeding in medium containing 15% fetal-calf serum. Northern blot analysis showed that the *XPBC/ERCC-3* transcription levels are constant and not significantly induced during the S-phase and M-phase (data not shown).

An additional region of interest is the 3'-UTR region of the gene. 3' untranslated sequences of several transiently expressed genes have been implicated in mRNA stability. Particularly an (AU)-rich region with the core sequence AUUUA was found to induce mRNA lability (39). Such elements are present in the 3' untranslated region of the *XPBC/ERCC-3* sequence shown in Fig. 3. This prompted us to investigate the stability of cytoplasmic *XPBC/ERCC-3* mRNA. To that end actinomycin-D was added to HeLa TK$^-$ cells, and cytoplasmic RNA was isolated after several time intervals to determine the half-life of *XPBC/ERCC-3* mRNA. Northern blot analysis (Fig. 7) shows that even after 3 h of treatment with actinomycin-D no significant differences in *XPBC/ERCC-3* mRNA amounts could be detected. As a

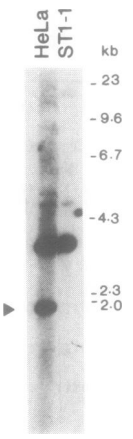

Fig. 8. Southern blot analysis of *XPBC/ERCC-3* genomic seqeunces. High-molecular weight DNA (15 μg) from HeLa cells and a secondary repair-proficient transformant (ST1−1) was digested with *Pst*I and size-fractionated on a 0.8% agarose gel. The DNA was blotted and hybridized with a 375 bp 3' *XPBC/ERCC-3* cDNA probe harboring a part of the last exon. The cross-hybridizing genomic restriction fragment is indicated with an triangle.

positive control on the effect of actinomycin-D, the same filter was also hybridized with a human *c-myc* probe, which has been reported to have a short half-life of 15−30 min (39). A decrease in the relative amount of *c-myc* mRNA was readily observed on the same filter. The addition of inhibitors of protein synthesis like cycloheximide did not influence the *XPBC/ERCC-3* mRNA stability (data not shown).

Cross-hybridising genomic sequence outside the *XPBC/ERCC-3* locus

Southern blot hybridization of human genomic DNA with *XPBC/ERCC-3* cDNA produces a complex pattern of hybridizing bands. Systematic comparison by Southern blot analysis revealed that in Hela genomic DNA extra fragments hybridized that can not be accounted for by the map and are absent in DNA of repair proficient transformant ST1-1 with an intact copy of the gene. An example using a 3' *XPBC/ERCC-3* cDNA probe is shown in Fig. 8. This 375 bp probe hybridizes to a 3.7 kb genomic *Pst*I fragment. The hybridization pattern of genomic DNA from HeLa cells revealed the presence of an additional restriction fragment of approximately 2.1 kb not present in DNA of ST1-1 nor in DNA of other secondary UV-resistant 27-1 transformants. Similar results were obtained using 5' *XPBC/ERCC-3* cDNA probes. Hybridization to other human DNA's (the human lymphoblastoid cell line; GM1855), gave identical results (data not shown), indicating that this finding can not be explained by fragment length polymorphisms. We concluded that a cross-hybridizing sequence exists elsewhere in the human genome.

DISCUSSION

In this paper the physical organization and expression of the human *XPBC/ERCC-3* is presented. The gene is approximately 45 kb long and consists of at least 14 exons. Virtually all 5' and 3' splice boundaries in functional genes obey the GT-AG rule (37). The importance of the conserved GT dinucleotide in the splice donor sequence has been examined in naturally occurring mutations (44−46) and after site-specific mutagenesis of GT to GG in 5' splice sites of an intron in adenovirus or in the rabbit

β-globin (47–48). These studies showed that the presence of the GT dinucleotide is a prerequisite for efficient RNA splicing. However, a few active splice sites violating the GT-dinucleotide rule have been found (49–52). In these rare instances a GC was present instead of a GT. In our case, the *XPBC/ERCC-3* gene has been cloned from a UV-resistant 27-1 genomic transformant which originally received and since then retained only 1 copy of the (active) gene per cell. Therefore, we conclude that this -GC- splice donor must be functional *in vivo*.

Functional domains of proteins are in many cases encoded by discrete exons (53–55). In this respect it is worth noting that exon 2 of the *XPBC/ERCC-3* gene codes for the postulated nuclear location signal, whereas the presumed DNA-binding domain is part of exon 3. The 7 motifs with striking homology to conserved helicase domains are encoded by separate exons, except for domains I and V.

Using a functional assay we have shown that the *XPBC/ERCC-3* promoter is located within a stretch of 259 nucleotides proximal to the putative transcription start site. The *XPBC/ERCC-3* promoter lacks the classical TATA and CAAT boxes which are typically located in front of many eukaryotic genes between positions −30 to −20 and −80, respectively (40). However most 'housekeeping' genes are driven by promoters that lack such elements but instead contain multiple GC boxes (56–59). These GC boxes have the potential to bind transcription factor SP1 (60). It has been shown that SP1 can bind asymmetrically to an SP1 box (61) and can promote bidirectional transcription (62–63). The presence of this SP1 recognition sequence in the *XPBC/ERCC-3* promoter could account for the bidirectional functioning of this DNA segment. The presence of a cluster of CpG dinucleotides was used as an indicator of a CpG-rich island. It is now well established that CpG islands are often found associated with promoter regions of genes and that nuclear factors specifically bind to these sequences (64, 65).

Comparison of the sequence of the human *XPBC/ERCC-3* promoter with that of *ERCC-1* revealed a homologous segment consisting of two domains, a pyrimidine rich region preceeded by a 12 nucleotide sequence (Fig. 4B). Regions with alternating CT residues can adopt triple-helical structures *in vitro* (66) and have been reported for promoters of constitutive or inducible genes (67). Recently, footprint analysis has shown that in the Drosophila heat-shock genes hsp70 and hsp20 such a pyrimidine-rich region binds a transacting factor (68) which appears to be similar or closely related to the described 'GAGA' transcription factor (69). In addition, upstream of this pyrimidine-rich region a motif is found that is highly similar to a stretch in the *ERCC-1* promoter (box I). The possible involvement of these elements in the basal transcription levels of *ERCC-1* and *XPBC/ERCC-3* is currently under investigation.

Northern blot analysis of poly(A)$^+$RNA from different cell lines indicates low levels of *XPBC/ERCC-3* transcripts. Low constitutive or basal expression of repair proteins has also been observed in *E. coli* and yeast (70). In bacteria efficient removal of DNA lesions is carried out by an excision repair system that is part of the DNA damage-inducible SOS response. In yeast, agents that elicit this response induce the excision repair gene *RAD2* whereas transcription of *RAD1*, *RAD3* and *RAD10* is not substantially affected by DNA damage (71). In addition, other DNA damage inducible genes have been described (72 and references therein), including DNA polymerase-β (73), DNA ligase (74). *XPBC/ERCC-3* expression does not seem to be significantly induced by UV-light in HeLa cells and keratinocytes, although we cannot exclude a 2–3 fold induction which is difficult to detect in Northern blot analysis.

In both *E. coli* and Drosophila evidence is found for a partial overlap between the cellular responses to heat-shock and DNA damaging agents (75–77). Therefore we analysed *XPBC/ERCC-3* transcription also after heat-shock induction in HeLa cells. Although Hsp70 transcripts were clearly induced, the cellular level of *XPBC/ERCC-3* transcripts was unaffected (unpublished results).

At positions 2694 and 2705 in the 3′ UTR of the *XPBC/ERCC-3* cDNA, (AU)-rich regions are found similar to those described in 3′ untranslated regions of several oncogenes and transiently expressed genes. These regions are thought to be involved in determining mRNA stability (78). The mRNA for *c-Fos* seems to be more stable in constructs lacking this signal (79) whereas the addition of this short sequence destabilized previously stable messengers (80). The destabilizing effect of (AU)-rich sequences could be due to interaction with the poly(A)-protein complex (81). Recently, it has been shown that a trans-acting factor can bind to this AUUUA motif (82). However, based on our actinomycin-D experiments we conclude that *XPBC/ERCC-3* mRNA is not subject to this type of post-transcriptional regulation, notwithstanding the occurrence of similar sequence elements.

The organization of the human *XPBC/ERCC-3* gene has been elucidated. Remarkably, at least one additional genomic *PstI* fragment cross-hybridizes to *XPBC/ERCC-3* cDNA sequences which is absent in DNA of secondary UV-resistant transformants of 27-1 cells and hence, is not part of the *XPBC/ERCC-3* gene. Northern blots hybridized to different *XPBC/ERCC-3* cDNA probes show no cross-hybridization to other mRNAs, however, we cannot exclude the existence of a very low-expressed mRNA. This could implicate the presence of either a pseudogene or a gene with sequence homology. Further experiments are underway to discriminate between these two possibilities.

ACKNOWLEDGEMENTS

We would like to thank Dr. H. van Ormondt for critically reading this manuscript, Prof. D. Bootsma for his stimulating interest and support, and R. Masurel for technical assistance, M. Kuit for photography and Mrs. R.J. Boucke for skilful typing. We thank Dr. P. Herrlich (Karlsruhe, FRG) for providing the MTIIA cDNA clone and S. Gibson and C. Backendorf for gift of RNA from (UV)-exposed keratinocytes. 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 EURATOM (contract no. BJ6-141-NL) and by the Dutch Cancer Society (IKR 90-20).

REFERENCES

1. Friedberg, E.C. (1985) *DNA Repair.* (Freeman and Company, San Francisco).
2. Sancar, A. and Sancar, G.B. (1988) *Ann. Rev. Biochem.*, **57**, 29–67.
3. Van Houten, B. (1990) *Microbiol. Rev.*, **54**, 18–51.
4. Grossman, L. and Yeung, A.T. (1990) *Photochem. and Photobiol.*, **51**, 749–755.
5. Busch, D., Greiner, C., Lewis, K., Ford, R., Adair, G. and Thompson, L.H. (1989) *Mutagenesis*, **4**, 349–354.
6. Johnson, R.T., Ellito, G.C., Squires, S. and Joysey, V.C. (1989) *Hum. Genet.*, **81**, 203–210.
7. Robbins, J.H. (1989) *Hum. Genet.*, **84**, 99–100.
8. Johnson, R.T. (1989) *Hum. Genet.*, **84**, 101.

9. Vermeulen, W., Stefanini, M., Giliani, S., Hoeijmakers, J.H.J., and Bootsma, D. (1991) *Mutat. Res.*, **255,** 201–208.
10. Lehmann, A.R. and Dean, S.W. (1989) In: *'Handbook of Experimental Pharmacology'* (eds. C.S. Cooper and P.L. Garver), vol. 94/II. pp. 71–101, Springer, Berlin.
11. Hoeijmakers, J.H.J. and Bootsma, D. (1990) *Cancer Cells*, **2,** 311–320.
12. Cleaver, J.E. and Kraemer, K.H. (1989). In: *'The Metabolic Basis of Inherited Disease'* (Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D.) (eds), vol. II. McGraw-Hill, Inc. chapter 120, pp. 2949–2971.
13. Venema, J., Mullenders, L.H.F., Natarajan, A.T., van Zeeland, A.A. and Mayne L.V. (1990) *Proc. Natl. Acad. Sci. USA*, **87,** 4707–4711.
14. Teitz, T., Naiman, T., Avissar, S.S., Bar, S., Okayama, H. and Canaani, D. (1987) *Proc. Natl. Acad. Sci. USA*, **84,** 8801–8804.
15. Arrand, J.E., Bone, N.M. and Johnson, R.T. (1989) *Proc. Natl. Acad. Sci USA*, **86,** 6997–7001.
16. Hoeijmakers, J.H.J., Odijk, H. and Westerveld, A. (1987) *Exp. Cell Res.*, **169,** 111–119.
17. Tanaka, K., Satokata, I., Ogita, Z., Uchida, T. and Okada, Y. (1989) *Proc. Natl. Acad. Sci. USA*, **84,** 8801–8804.
18. Tanaka, K., Miura, N., Satokata, I., Miyamato, I., Yoshida, M.C., Satoh, Y., Kondo, S., Yasui, A., Okayama, H., and Okada, Y. (1990) *Nature*, **348,** 73–76.
19. Westerveld, A., Hoeijmakers, J.H.J., van Duin, M., de Wit, J., Odijk, H., Pastink, A., Wood, R.D., Bootsma, D. (1984) *Nature (London)*, **310,** 425–428.
20. Weber, C.A., Salazar, E.P., Stewart, S.A. and Thompson, L.H, (1988) *Mol. Cell. Biol.*, **8,** 1137–1146.
21. 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). *Mol. Cell. Biol.*, **10,** 2570–2581.
22. Mudgett, J.S. and MacInnes, M.A. (1990) *Genomics*, **8,** 623–633.
23. Troelstra, C., Odijk, H., de Wit, J., Bootsma, D. and Hoeijmakers, J.H.J. (1990). *Mol. Cell. Biol.*, **10,** 5806–5815.
24. Thompson, L.H., Brookman, K.W., Jones, N.J., Allen, S.A. and Carrano, A.V. (1990) *Mol. Cell. Biol.*, **10,** 6160–6171.
25. Weeda, G., van Ham, R.C.A., Vermeulen, W., Bootsma, D., van der Eb, A.J. and Hoeijmakers, J.H.J. (1990) *Cell*, **62,** 777–791.
26. Maniatis, T., Fritsch, F.F. and Sambrook, J. (1982). In: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
27. Fort, P., Marty, L., Piechacyzk, M., El Sabrouty, S., Dani, C., Jeanteur, P. and Blanchard, J.M. (1985). *Nucl. Acids Res.*, **13,** 1431–1442.
28. Adams, J.M., Geronddakis, W., Webbe, E., Corcoran, L.M. and Cory, S. (1983) *Proc. Natl. Acad. Sci. USA*, **80,** 1982–1986.
29. Wood, R.D. and Burki, H.J. (1982) *Mutat. Res.*, **95,** 505–514.
30. Angel, P., Baumann, I., Stein, B., Delius, H., Rahmsdorf, H.J. and Herrlich, P. (1987) *Mol. Cell Biol.*, **7,** 2256–2266.
31. Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982) *Mol. Cell. Biol.*, **2,** 1044–1051.
32. Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74,** 5463–5467.
33. Dery, C.V., Herrmann, C.H. and Mathews, M.B. (1987) *Oncogene*, **2,** 15–23.
34. Auffray, C. and Rougeon, F. (1980) *Eur. J. Biochem.*, **107,** 303–314.
35. Lindsay, S., and Bird, A.P. (1987) *Nature (London)*, **327,** 336–338.
36. Tinoko, I.Jr., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) *Nature New Biol.*, **246,** 40–41.
37. Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P. (1986). *Ann. Rev. Biochem.*, **55,** 1119–1150.
38. Wickens, M. and Stephenson, P. (1984) *Science*, **226,** 1045–1051.
39. Shaw, G. and Kamen, R. (1986) *Cell*, **46,** 659–667.
40. Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.*, **50,** 349–383.
41. Kadonaga, J.T., Jones, K.A. and Tjian, R. (1986) *Trends in Biochem. Sci.*, **11,** 20–23.
42. Van Duin, M., Koken, M., van den Tol, J., ten Dijke, P., Odijk, H., Westerveld, A., Bootsma, D. and Hoeijmakers, J.H.J. (1987) *Nucleic Acids Res.*, **15,** 9195–9213.
43. Angel, P., Pöting, A., Mallick, U., Ramsdorf, H.J., Schorp, M. and Herrlich, P. (1986) *Mol. Cell. Biol.*, **6,** 1760–1766.
44. Felber, B.K., Orkin, S.H. and Hamer, D.H. (1982) *Cell*, **29,** 895–902.
45. Treisman, R., Proudfoot, N.J., Shandar, M. and Maniatis, T. (1982) *Cell*, **29,** 903–911.
46. King, C.R. and Piatigorsky, J. (1983) *Cell*, **32,** 707–712.
47. Montell, C., Fisher, E.F., Caruthers, M.H. and Berk, A.J. (1982) *Nature (London)*, **295,** 380–384.
48. Wieringa, B., Meyer, F., Reiser, J. and Weissman, C. (1983). *Nature (London)*, **301,** 38–42.
49. Avvedimento, V.E., Vogeli, G., Yamada, Y., Maizel, J.V. Jr., Fastan, I. and Crombrugghe, B. (1980) *Cell*, **21,** 689–696.
50. Ebril, C. and Niesing, J. (1983) *EMBO J.*, **2,** 1339–1343.
51. Dodgson, J.B. and Engels, J.D. (1983) *J. Biol. Chem.*, **258,** 4623–4629.
52. Broderick, T.P., Schaff, D.A., Bertino, A.M., Dulsh, M.K., Tischfield, J.A. and Stambrook, P.J. (1987) *Proc. Natl. Acad. Sci. USA*, **84,** 3349–3353.
53. Gilbert, W. (1978) *Nature (London)*, **271,** 501.
54. Blake, C.C.F. (1978) *Nature (London)*, **273,** 267.
55. Traut, T.W. (1988) *Proc. Natl. Acad. Sci. USA*, **85,** 2944–2948.
56. Melton, D.W., Konecki, D.S., Brennand, J. and Caskey, T. (1984) *Proc. Natl. Acad. Sci. USA*, **81,** 2147–2151.
57. Singer-Sam, J. Keith, D.H., Tani, K., Simmer, R.L., Shively, L., Lindsay, S., Yoshida, A. and Riggs, A.D. (1984) *Gene*, **32,** 409–417.
58. Reynolds, G.A., Basu, S.K., Osborne, T.F., Chin, D.J., Gil, G., Brown, M.S., Goldstein, J.L. and Luskey, K.L. (1984) *Cell*, **38,** 275–285.
59. Valerio, D., Duyvesteyn, M.G.C., Dekker, B.M.M., Weeda, G., Berkvens, Th.M., van der Voorn, L., van Ormondt, H. and van der Eb, A.J. (1985) *EMBO J.*, **4,** 437–443.
60. Dynan, W.S., Sazer, S., Tjian, R. and Schimke, R.T. (1986) *Nature (London)*, **319,** 246–248.
61. Gidoni, D., Dynan, W.S. and Tjian, R. (1984) *Nature (London)*, **312,** 409–413.
62. Dynan, W.S., Saffer, J.D., Lee, W.S. and Tjian, R. (1985). *Proc. Natl. Acad. Sci. USA*, **82,** 4915–4919.
63. Gidoni, D., Kadonaga, J.T., Barrera-Saldana, H., Takahashi, K., Chambon, P. and Tjian, R. (1985) *Science*, **230,** 511–517.
64. Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L. and Bird, A.P. (1989) *Cell*, **58,** 499–507.
65. Gardiner-Garden, M. and Frommer, M. (1987) *J. Mol. Biol.*, **196,** 261–282.
66. Htun, H. and Dahlberg, J.E. (1988) *Science*, **241,** 1791–1795.
67. Elgin, S.C.R. (1988) *J. Biol. Chem.*, **263,** 19259–19262.
68. Gilmour, D.S., Thomas, G.H. and Elgin, S.C.R. (1989) *Science*, **245,** 1487–1490.
69. Biggin, M.D. and Tjian, R. (1988) *Cell*, **53,** 699–711.
70. Friedberg, E.C. (1988) *Microbiol. Rev.*, **52,** 70–102.
71. Robinson, G.W., Nicolet, C.M., Kalainikov, D. and Friedberg, E.C. (1986) *Proc. Natl. Acad. Sci. USA*, **83,** 1842–1846.
72. Mai, S., Stein, B., van den Berg, S., Kaina, B., Lücke-Huele, C., Ponta, H., Rahmsdorf, H.J., Kraemer, M., Gebel, S., Herrlich, P. (1989) *J. Cell Science*, **94,** 609–615.
73. Fornace, A.J. Jr., Zmudzka, B., Hollander, M.C. and Wilson, S.H. (1989) *Mol. Cell. Biol.*, **9,** 851–853.
74. Mezzina, M., Nocentini, S. and Sarasin, A. (1982) *Biochemie*, **64,** 743–748.
75. Krueger, J.H. and Walker, G.C. (1984) *Proc. Natl. Acad. Sci. USA*, **81,** 1499–1503.
76. McClanahan, T. and McEntee, K. (1985) *Mol. Cell. Biol.*, **6,** 90–96.
77. Vivino, A.A., Smith, M.D. and Minton, K.W. (1986) *Mol. Cell. Biol.*, **6,** 4767–4769.
78. Schuler, G.D. and Cole, M.D. (1988) *Cell*, **55,** 1115–1122.
79. Wilson, T. and Treisman, R. (1988) *Nature (London)*, **336,** 396–399.
80. Kabnick, K.S. and Hausman, D.E. (1988). *Mol. Cell. Biol.* **8,** 3244–3250.
81. Brewer, G. and Ross, J. (1988). *Mol. Cell. Biol.* **8,** 1697–1708.
82. Malter, J.S. (1989). *Science* **246,** 664–665.