Molecular and functional analysis of the *XPBC/ERCC-3* promoter: transcription activity is dependent on the integrity of an Sp1-binding site

Libin Ma, Geert Weeda[1], Aart.G.Jochemsen, Dirk Bootsma[1], Jan H.J.Hoeijmakers[1] and Alex J.van der Eb*
Laboratory for Molecular Carcinogenesis, Medical Genetics Center, Sylvius Laboratories, University of Leiden, Wassenaarseweg 72, 2333 AL Leiden, and [1]Department of Cell Biology and Genetics, Medical Genetics Center, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands

ABSTRACT

The human *XPBC/ERCC-3* gene, which corrects the excision-repair defect in xeroderma pigmentosum group B cells and the UV-sensitive CHO mutant 27-1 cells, appears to be expressed constitutively in various cell types and tissues. We have analysed the structure and functionality of the *XPBC/ERCC-3* promoter. Transcription of the *XPBC/ERCC-3* gene is initiated from heterogeneous sites, with a major startpoint mapped at position − 54 (relative to the translation start codon ATG). The promoter region does not possess classical TATA and CAAT elements, but it is GC-rich and contains three putative Sp1-binding sites. In addition, there are two elements related to the cyclic AMP (cAMP)-response element (CRE) and the 12-O-tetradecanoyl phorbol-13-acetate-response element (TRE) in the 5′-flanking region. Transient expression analysis of *XPBC/ERCC-3* promoter-CAT chimeric plasmids revealed that a 127-bp fragment, spanning position − 129 to − 3, is minimally required for the promoter activity. Transcription of the *XPBC/ERCC-3* promoter depends on the integrity of a putative Sp1-binding site in close proximity to the major cap site. Band shift assays showed that this putative Sp1-binding site can interact specifically with a nuclear factor, most likely transcription factor Sp1 (or an Sp1-like factor) *in vitro*.

INTRODUCTION

The nucleotide-excision repair pathway is one of the major components of DNA repair systems in prokaryotes and eukaryotes. This pathway removes a wide range of lesions, including UV-induced photo-products as well as bulky chemical adducts (see ref. 1 for a general review on DNA repair). Defects in the excision repair system may have serious consequences for the organism, such as the human cancer-prone syndrome

xeroderma pigmentosum (XP), an autosomal, recessive hereditary disease in which most patients have a defect in the early steps of the excision repair pathway (2). The molecular mechanism and the regulation of the excision repair system have been extensively studied in *Escherichia coli* and there is good evidence that some of the excision-repair genes (*uvrA*, *uvrB*, and *uvrD*) constitute part of the inducible SOS regulon that responds to DNA damage (3). In the yeast *Saccharomyces cerevisiae*, expression of the *RAD2*, *RAD7*, and *RAD23* genes, which function in excision repair, is DNA-damage-inducible (4−6), whereas transcription of other excision-repair genes, the *RAD1*, *RAD3*, and *RAD10* genes, is not substantially affected by DNA damage (4). The mechanism as well as the regulation of mammalian excision repair systems are poorly understood. Recently, genes correcting the excision-repair defect in XP group A and B cells, the *XPAC* gene (XP-A-correcting gene) (7) and the *XPBC/ERCC-3* gene (8, 9), respectively, have been cloned. Several other human excision-repair genes, the *ERCC* (excision repair cross-complementing rodent repair deficiency) genes, including *ERCC-1*, *ERCC-2*, *ERCC-5*, and *ERCC-6* were also cloned (10−13).

The human *XPBC/ERCC-3* gene encodes a putative DNA helicase and was initially cloned on the basis of its ability to correct the excision-repair defect in CHO mutant 27-1 cells (8). Subsequently, it was discovered that it complemented the excision-repair defect in XP-B cells. A C→A transversion in the splice acceptor site of the last *XPBC/ERCC-3* intron, causing a 4-bp insertion in the mRNA and an inactivating frameshift at the protein level, was found to underlie the molecular defect in XP group B cells (9). The genomic structure of the human *XPBC/ERCC-3* gene has been established in a great detail (14). The gene has a length of approximately 45 kb and consists of at least 14 exons (14). The *XPBC/ERCC-3* gene appears to be constitutively expressed in various cell types and tissues, similar to another DNA-repair gene *ERCC-1* (14, 15). Initial sequence comparison of the *XPBC/ERCC-3* 5′-flanking region with the

ERCC-1 promoter revealed two regions of conservation, a pyrimidine-rich stretch preceded by a 12-nucleotide motif (box I) (14, 15).

We initially have identified a 376-bp genomic fragment containing the promoter of the *XPBC/ERCC-3* gene (14). Here we report further studies regarding the *XPBC/ERCC-3* promoter, in which we define the major transcription initiation site of the *XPBC/ERCC-3* gene and identify a 127-bp sequence as the functional promoter. Our results suggest that transcription activity of the *XPBC/ERCC-3* promoter depends on the integrity of an Sp1-binding site in close proximity to the major cap site.

MATERIALS AND METHODS

Cell culture, DNA transfection, and CAT assay

HeLa TK⁻ cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 2% fetal calf serum and 8% newborn calf serum, 100 units/ml of penicillin, and 50 µg/ml of streptomycin. Primary human foreskin fibroblasts (VH10 cells) and SV40-transformed VH10 cells were cultured in a 1:1 mixture of DMEM and Ham's F10 supplemented with 10% fetal calf serum and antibiotics. For transfection, cells were seeded at a density of 1×10^6 cells/10-cm dish 24 h before transfection. The cells were transfected with 10 µg of DNA from each promoter-CAT construct, using the calcium-phosphate precipitation method (16). Approximately 48 h after transfection, cells were harvested and a protein extract was prepared by three repeated freeze-thaw cycles of 5 min in dry ice, followed by 5 min at 37°C, in 100 µl of 250 mM Tris-Cl (pH 7.8). Samples normalized for protein content (100 µg) were assayed for CAT activity by previously described procedures (17). CAT activity was quantified by liquid scintillation counting of excised sections of thin layer chromatography plates and expressed as percent conversion of [¹⁴C]chloramphenicol to its acetylated derivatives.

RNA isolation and in vitro transcription

Cytoplasmic RNA was isolated by the NP-40 method (18). In vitro transcription was performed according to the manufacturer's instruction (Promega). The template was plasmid pHEP-3 (Fig. 1B) in which the *XPBC/ERCC-3* 5'-flanking sequence (from −370 to −1) was fused to the *XPBC/ERCC-3* cDNA (9). (For convenience, we use a numbering system in which +1 denotes the A in the initiation codon ATG).

Primer extension and RNase protection

Primer extension was carried out according to the published protocol (18). A 20-mer oligonucleotide 5'-dCCCATGGCAGC-TACAGCAGC-3' extending from +5 to −15 was used as a primer. RNase protection was performed as described previously (19). For detection of the endogenous *XPBC/ERCC-3* transcripts, a plasmid designated pHEP-4 was made in which a 459-bp fragment containing the 5' flanking region, exon 1, and a portion of exon 2 of *XPBC/ERCC-3* was cloned into the vector pEP30 (a T7 and SP6 promoter-containing plasmid provided by P.W. Laird). Plasmid pHEP-4 was created by subcloning the *Sma*I (+89)-*Eco*RI (in the polyliker region of the vector) fragment from pHEP-3 into pEP30. A second construct, designated pEP-cat3, was generated to analyse the *XPBC/ERCC-3* promoter-CAT RNA. Plasmid pEP-cat3 was made by inserting a 533-bp *Stu*I (−280)-*Eco*RI (in the CAT gene) fragment from p(−370/−3)CAT (the plasmid described below) into the pEP30.

Construction of promoter-CAT reporter plasmids

Plasmid pHEP (14) was used as a source of all subcloned genomic fragments. Plasmid pHEP-5 was created by the subcloning into pEP30 of a 1.2-kb *Sma*I (−809)-*Sma*I fragment derived from pHEP, which contains the 5'-flanking region, the first exon, the first intron, and a portion of the second exon of *XPBC/ERCC-3*. Plasmid p(−370/−3)CAT was constructed by subcloning the 368-bp Mung-Bean-nuclease-treated *Nco*I-*Nco*I fragment (position −370 to −3) into the pBA-CAT (20) at the *Hind*III site (blunt-ended with the Klenow fragment of DNA polymerase I). Plasmids p(−809/−3)CAT and p(−341/−3)CAT were generated by excising the 319-bp *Sst*II (−49)-*Xho*I (in the polylinker region of the vector) fragment of p(−370/−3)CAT and replacing it with a 759-bp *Sst*II (−49)-*Sal*I (at the polylinker region of the vector) fragment from pHEP-5 and a 290-bp *Sst*II-*Xho*I fragment amplified by the polymerase chain reaction, respectively. The 290-bp *Sst*II-*Xho*I fragment was amplified with an *XPBC/ERCC-3*-specific primer 5'-dGCTCGAGCTCACAC-ACACAT-3' and a CAT-specific primer 5'-dTACGATGCCA-TTGGG-3' and then digested with *Sst*II and *Xho*I. Plasmids p(−280/−3)CAT, p(−169/−3)CAT, p(−129/−3)CAT, p(−119/−3)CAT, and p(−110/−3)CAT were created by Bal31 deletions. Such deletions were obtained from p(−809/−3)CAT by linearizing it at the *Stu*I site (at −280) of the *XPBC/ERCC-3* promoter sequence and digesting with Bal31 exonuclease (4 units/20-µg of DNA) for 0 to 2 min at 30°C. After digestion with *Bam*HI (in the polylinker region of the vector) which removed the *XPBC/ERCC-3* sequences upstream of −280, the DNA fractions containing the deleted plasmids were purified by electrophoresis through low-melting temperature agarose gels, treated with the Klenow fragment to repair the ends, and recircularized. All constructs were verified by DNA sequencing analysis.

DNA sequencing and computer analyses

Nucleotide sequences were determined by the chain termination method (21). Sequence data were analysed on a VAX computer using the software computer package of the Genetics Computer Group (University of Wisconsin, Madison, WI).

Band shift assay

Band shift experiments were performed as described previously (22). Crude nuclear extracts from HeLa TK⁻ cells were kindly provided by H. van Dam, Leiden, and were prepared as described (22). Seven µg of nuclear extract was incubated with 10 fmol of labeled probe and 2 µg of poly(dI-dC).poly(dI-dC) at room temperature for 30 min. Samples were loaded onto 4% native polyacrylamide gels prepared in 25 mM Tris, 1 mM EDTA, and 190 mM glycine. Electrophoresis was performed at room temperature for 2 h at 15 mA.

Oligonucleotides used either as probes or as competitors in the band shift assays were obtained by annealing of 5'-agctAGAG-GCGGGACATCCT-3' and 5'-gatcAGGATGTCCCGCCTCT-3' (erccSp1), 5'-agctCGGGGCGGGGCGGGGC-3' and 5'-gatcG-CCCCGCCCCGCCCCG-3' (consSp1), and 5'-agctCCCGTG-ACGTCACCCG-3' and 5'-gatcCGGGTGACGTCACGGG-3' (fibCRE) (23). The probes were labeled with the Klenow fragment of DNA polymerase I. The lower-case letters indicate the nucleotides incorporated into the probes in the labeling reaction.

RESULTS

Mapping of the major transcription initiation site of the XPBC/ERCC-3 gene

As a first step to delineate the location of *cis*-acting regulatory elements, we determined the transcription initiation site of the *XPBC/ERCC-3* gene both by primer extension and by RNase protection.

First, the 5' end of *XPBC/ERCC-3* transcript was localized by primer extension. A 5'-end-labeled 20-mer oligonucleotide complementary to sequences spanning position +5 to −15 was hybridized to cytoplasmic RNA isolated from HeLa cells and extended using MoMuLV reverse transcriptase. (For simplicity, we use a numbering system in which +1 denotes the A in the presumed translation initiation codon ATG). As can be seen in Fig. 1A, multiple initiation sites were found. The major primer-extended product, surrounded by a cluster of minor primer
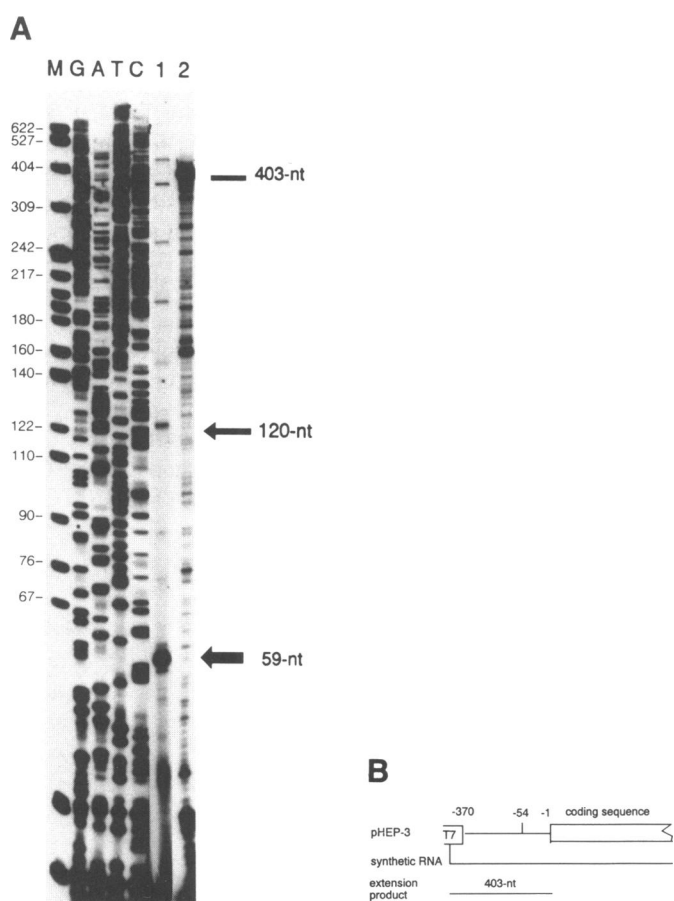
extension stops, has a length of 59 nt corresponding to position −54 (lane 1). In addition, a series of minor stops was found further upstream, one of these (120 nt, corresponding to position −115) could represent the cap site of the published cDNA sequence (9).

To rule out the possibility that the major observed stop in the primer extension reaction had been caused by pausing or premature termination of the reverse transcriptase reaction, we employed a synthetic RNA molecule as template for the primer extension analysis. This RNA molecule was transcribed from the plasmid template pHEP-3, in which the 5'-flanking region of *XPBC/ERCC-3* was ligated to the *XPBC/ERCC-3* cDNA (Fig. 1B). The synthetic RNA was subjected to primer extension with the 20-mer oligonucleotide as described above. As shown in Fig. 1A (lane 2), the most abundant extension product (403 nt, including 28 nt of the plasmid linker sequences) was precisely the size expected for the RNA molecule generated. No strong stops were observed surrounding the major cap site (position −54). This analysis validates the efficiency of the reverse transcription reaction and suggests that neither the high G+C
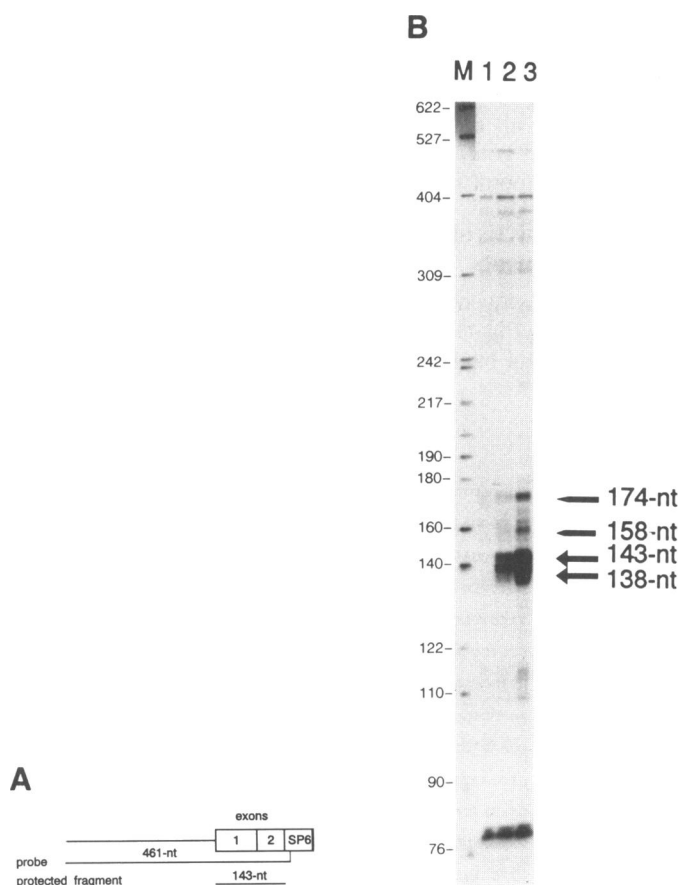

Fig. 1. Primer extension analysis of the transcription initiation sites of the human *XPBC/ERCC-3* gene. **A.** A 20-mer 5'-end-labeled oligonucleotide extending from +5 to −15 was annealed to 20 μg of HeLa cell RNA (lane 1) and extended upstream with Murine Leukaemia Virus reverse transcriptase. An RNA molecule generated by *in vitro* transcription from the plasmid template, pHEP-3, was included in the same reaction in order to verify the efficiency of the reverse transcription reaction (lane 2). **B.** Diagram of the plasmid template pHEP-3, the RNA generated, and the expected extension product (403 nt). The resulting products were analysed on a 6% polyacrylamide/8 M urea gel next to a sequencing ladder (lanes G, A, T, and C). Numbers alongside the gel indicate the sizes of labeled marker fragments (pBR322, digested with *Hpa*II) (lane M). The major extended product obtained with HeLa cell RNA has a length of 59 nt and is indicated by a large arrow. A minor one with a length of 120 nt is indicated by a small arrow. Additional minor products with size greater than 120 nt are also observed.

Fig. 2. RNase protection mapping of *XPBC/ERCC-3* transcripts in HeLa cells. **A.** Schematic representation of the plasmid template, the ^{32}P-labeled RNA probe derived from the template, and the expected protected RNA fragment for HeLa cell transcripts. The uniformly labeled antisense RNA probe was transcribed from the chimeric plasmid pHEP-4, in which the *XPBC/ERCC-3* 5'-flanking sequences (from −370) were fused to the cDNA sequences extending to +89 inside exon 2. **B.** Electrophoretic analysis of protected fragments obtained by the use of 20 μg (lane 2) and 40 μg (lane 3) HeLa cell RNA, or 20 μg tRNA (lane 1) for RNase protection. The two most abundant RNase-resistant bands (with lengths of 143 nt and 138 nt) are indicated by the large arrows. The two minor bands are indicated by the small arrows. Lane M contains the molecular size markers (see Fig. 1A).

content nor a possible secondary structure of the mRNA can account for the major stop observed in our primer extension analysis.

The major cap site as determined by primer extension was confirmed by RNase protection experiment. The results of this analysis are shown in Fig. 2. Fig. 2A is a diagrammatic representation of the plasmid template, the anti-sense RNA probe, and the band expected to be protected. The probe was hybridized with various amounts of HeLa cell RNA and digested with a combination of RNases A and T1. Two major RNase-resistant bands, with lengths of 138 and 143 nucleotides, were detected (Fig. 2B, lanes 2 and 3). Calculated from the position of initiator codon ATG, they localize the 5' ends of the *XPBC/ERCC-3* transcripts to positions −49 and −54, respectively. In addition, two minor protected bands (158 nt and 174 nt) were observed when the amount of RNA was raised to 50 μg (Fig. 2B, lane 3); the significance of these bands is not clear.

Cytoplasmic RNAs from primary VH10 fibroblasts, VH10 cells (24 h) after UV-irradiation (16 J/m^2), and SV40-transformed VH10 cells were also examined by primer extension analysis. The lengths and distribution pattern of the extension products in all RNAs differed little from those observed for HeLa cell RNA (data not shown), which indicated that the choice of the major transcription initiation site does not seem to vary in a cell-type-specific or a DNA-damage-dependent manner.

Considering all data, we conclude that transcription of the *XPBC/ERCC-3* gene is initiated from heterogeneous sites with the major startpoint being located at position −54.

Structural analysis of the 5'-flanking region discerns putative promoter elements

The nucleotide sequence of the *XPBC/ERCC-3* 5'-flanking region is presented in Fig. 3. Comparison of the sequence with those in the DNA data banks (EMBL, GCG version 6.0) revealed a 290-bp *Alu* repeat sequence from −702 to −413 showing approximately 84% similarity to a consensus *Alu* repeat (24). A short segment of *Alu*-like sequence was found from −836 to −717. The element between −836 to −594 has the potential to form a DNA cruciform, as predicted by computer analysis (data not shown).

The *XPBC/ERCC-3* promoter does not contain a canonical TATA nor a CAAT box at their characteristic positions relative to the transcription initiation site (25). The *XPBC/ERCC-3* promoter region also lacks any identifiable housekeeping initiator protein (HIP1)-binding sequences (26). The sequences immediately surrounding the major cap site show some resemblance to the 'initiator' element (27). Like many other genes in which transcription is initiated from heterogeneous sites, the *XPBC/ERCC-3* promoter region (between −200 to −1) is GC-rich (67%) and comprises a total of 18 CpG dinucleotides. There are several C/G clusters which might interact with transcription factor ETF, a recently described nuclear factor that binds to promoters without a TATA box (28). Furthermore, there are two elements (position −805 to −796, GAGGCGGACG; and position −122 to −113, GAGGCGGGAC) similar to the recognition and binding site of transcription factor Sp1 (29, 30). According to the numbering used in this study, the previously described putative Sp1-binding site (CCCCCGCCGC; 14) is at position −204 to −195. The pyrimidine-rich stretch and the conserved motif (box I) which have been found in the promoters of *XPBC/ERCC-3* and *ERCC-1*, another human DNA excision-repair gene (14, 15), are at position −357 to −339 and −369 to −358, respectively.

Several ATTA boxes, which have been implicated in the binding of homeobox proteins in vertebrates and in *Drosophila melanogaster* (31), are present between position −442 to −272. We also noticed the existence of a sequence motif (TCGCCC-AGGC) resembling the consensus AP-2 site (32) at position −678 to −669. The sequence from −484 to −478 (TGACCTCA)


```
                              5'-CTTAAGGCCT TCAACTGATT
-880  TGGATGAGGC CCACCCGCAT TATTGAGGGT AATCTGCTTT ATTTAGGCTG AGACATGAGA
-820  ATGGCATGAA CCCGGGAGGC GGACGTTGCA GTGAGCTGAA ATCGCGCCAC TGCATTCCAG
                    Sp1
-760  CCTGGGCGAC AGGGCGAGAC TCCGTCTCAA AAAAAAAAAA AAAAAAAGAA AAAAATTTTT
-700  TTTGAGACGA ACTCTCGGCC TGTCGCCCAG GCTGTAGTGC AGTGGTGCAA TCTCAGCTCA
                              AP-2
-640  GTGCAGCCTT TACCTTCTGG GTTCAAGAGA TTCTCCCATC TCAGCCTCCC AAGTAGCTTA
-580  GACTACAGGC GTGTGCCACC ATGCCTGGCT ACTTTTTTTT GTATTTTTCG CAGAGATGGG
-520  GTTTCACCAT GTTGGCCAGG CTGGCCTCGA ACTCCTGACC TCAAGTGATC CACCCACCTC
                                        CRE/TRE-related
-460  AGCCTCCCAA AGTGCTGGAA TTACAGGCGT GAGCCACCGT GCTCAGCCAA ATTTTTTTTA
-400  ATTAAAAAAA AGAATTCTTC CTTACCCATG GACTCAGGTC ACATTCTCTC TCTCTCTCTC
                                        inverted CRE/TRE
-340  TCACACACAC ATTTATTGCC CACATGGCTT CATTAAAATT TAAGTTCCAG AAGGATGAGG
-280  CCTTTAATTC GCGCACTACA CAGGTGCTCA ACACACAGTA TTTTCCTTCA GTTGTAGAGA
-220  AATGAACAGT GGAGGACCCC CGCCGCTGGG TCGCACGAGC TAACAGATCG GCGGTCGCTG
                               Sp1
-160  GCTCTCCGCT CTTCCAGCCA CCCACATCAC GGCGCCTAGA GGCGGGACAT CCTGTGTGGA
                                                       Sp1
-100  CAACAGGGAG CTTCCGGATT GAGCCGGAAG TCCCCCCAGA GCGGATCGCG CGGCGGGCCT
                                                   *-54
- 40  GTGGGAGCGG GGTCATCTTC TCTCTGCTGC TGTAGCTGCC ATGGGCAAAA-3'
                                                -1 +1
```
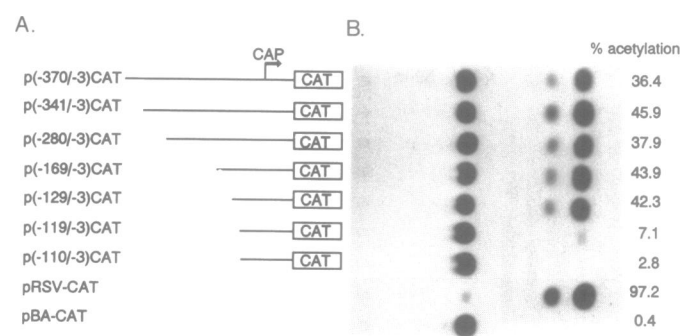
Fig. 4. Transient-expression analysis of the 5' deletion mutants of *XPBC/ERCC-3* promoter-CAT plasmids in HeLa TK⁻ cells. **A.** Diagrammatic representation of *XPBC/ERCC-3* promoter-CAT chimeric constructs. The major transcription initiation site determined in this study is indicated by an arrow. Numbers correspond to the endpoints of the deletion mutants. **B.** Result of a representative CAT assay, showing the relative activity of various 5' promoter deletions in HeLa TK⁻ cells. Cells were transiently transfected with 10 μg of each *XPBC/ERCC-3* promoter-CAT plasmid. After 48 h, cells were harvested and total cell extracts were prepared. CAT activity was measured, quantified and expressed as the percentage of conversion of chloramphenicol to its acetylated derivatives. The results of three independent experiments were averaged to give the reported values of CAT activity for each CAT-construct tested. One hundred μg of protein was used to assay CAT activity. For pRSV-CAT, 2 μg of plasmid was used for transfection and 20 μg of protein for CAT assay.

shows a high degree of similarity to the consensus cyclic AMP (cAMP)-responsive element (CRE; TGACGTCA) and the 12-O-tetradecanoyl phorbol-13-acetate-responsive element (TRE; TG-A[C/G]TCA) (33—35). For convenience, we will refer to this motif as a CRE/TRE-related element. Interestingly, the sequence from −367 to −360 in box I (TCAGGTCA) forms an almost perfect inverted repeat with the CRE/TRE-related element. For this reason we will refer to this sequence as an inverted CRE/TRE-related element.

The pertinence of these putative elements for transcription regulation can be confirmed only by functional mapping of the promoter region.

Functional analysis defines the regulatory regions in the XPBC/ERCC-3 promoter

To delineate the sequence elements required for *XPBC/ERCC-3* gene transcription, we constructed a series of chimeric plasmids containing various 5′-progressive deletion fragments fused upstream to the bacterial CAT gene (Fig. 4A). The 5′ deletion mutants were created either by deleting restriction fragments or by Bal31 exonuclease digestion. All of the 5′ deletion fragments had the same 3′ endpoint at position −3 and were cloned into the promoter-less pBA-CAT plasmid (20). These constructs were transiently transfected into HeLa TK⁻ cells and assayed for CAT enzyme activity. In addition, pRSV-CAT containing the Rous sarcoma virus (RSV) long terminal repeat (LTR) and pBA-CAT, the parent plasmid, were included as constitutive positive and negative controls, respectively. Fig. 4B shows the 5′ deletion analysis through the *XPBC/ERCC-3* promoter.

Construct p(−809/−3)CAT, containing 807 bp of the *XPBC/ERCC-3* sequence, exhibited a promoter activity comparable to that of the p(−370/−3)CAT (data not shown), revealing that the putative AP-2 element and the CRE/TRE-related element are not directly involved in the basal promoter activity. Further deletion of the inverted CRE/TRE-related element and the pyrimidine-rich stretch (construct p(−341/−3)CAT) had, unexpectedly, no significant influence on the basal promoter activity. Analysis of constructs p(−280/−3)CAT, p(−169/−3)CAT, and p(−129/−3)CAT made clear that sequences up to −129 are dispensable for the basal promoter activity. In marked contrast, deletion of only 10 bp between −129 and −119 (construct p(−119/−3)CAT), which overlaps with the putative Sp1-binding element in close proximity to the major cap site, drastically reduced the expression level (> 6-fold), suggesting that this region is absolutely required for full promoter function. Although further deletion between −119 to −110 (construct p(−110/−3)CAT), reduced the CAT activity to one-fifteenth of that mediated by p(−129/−3)CAT, there still was some CAT activity to be detected, indicating that the remaining region, too, might contain some weak promoter activity.

Comparison of the levels of CAT enzyme activity (over a concentration range of both plasmid and cellular-extract protein) mediated by p(−370/−3)CAT and by pRSV-CAT in transient assays in HeLa TK⁻ cells, showed that in HeLa cells the *XPBC/ERCC-3* promoter has approximately 1% of the strength of the RSV-LTR (data not shown).

Plasmids p(−370/−3)CAT and p(−129/−3)CAT were also transfected into primary VH10 and SV40-transformed VH10 fibroblasts. In spite of the low transfection efficiency, both constructs were equally active in human fibroblasts (data not shown).

Taken together, these results demonstrated that a 127-bp fragment, from −129 to −3, is minimally required for basal promoter activity of the *XPBC/ERCC-3* gene. Furthermore, transcription activity of the *XPBC/ERCC-3* promoter is very much dependent on the integrity of the putative Sp1-binding site in close proximity to the major cap site.

To verify that in transient assays the major cap site is properly used, RNase protection was performed with cytoplasmic RNA isolated from transiently transfected HeLa cells. Fig. 5A presents a schematic diagram of the plasmid template, the RNA probe, and the products expected to resist RNase digestion. The results of RNase protection are shown in Fig. 5B. The pRSV-CAT transcript (lane 4) protected a fragment of 254 nt, precisely the size expected for the 5′ terminus of the CAT gene. Transcripts from p(−370/−3)CAT (lane 5) and p(−129/−3)CAT (lanes 6) protected fragments with the same length of 308 nt. This fragment
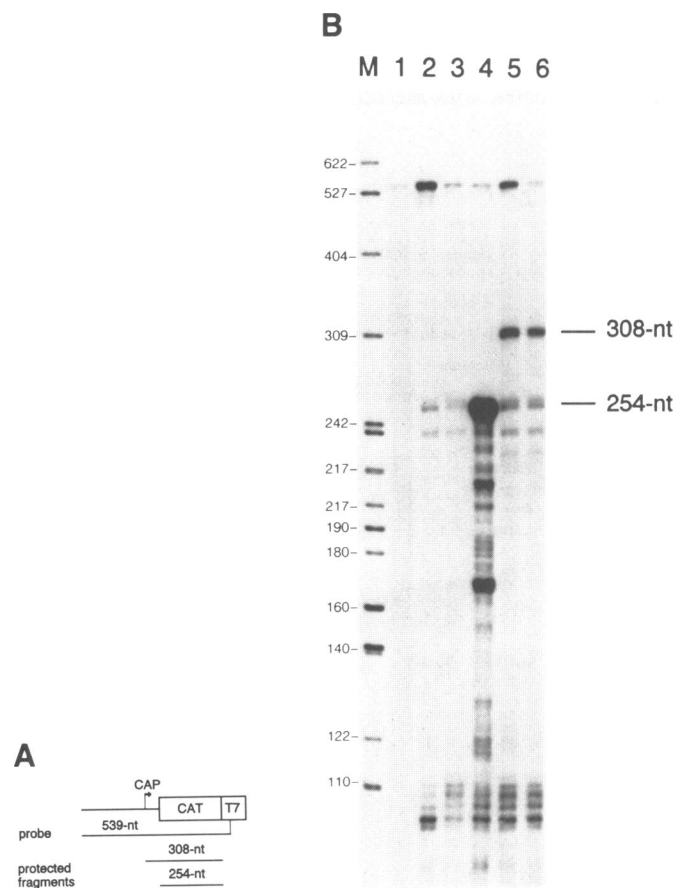

Fig. 5. RNase protection analysis of transiently transfected promoter-CAT constructs. **A.** Structure of the antisense RNA probe used in RNase protection. A 533-nt *StuI-EcoRI* fragment, which contains the *XPBC/ERCC-3* promoter-CAT sequence from position −280 (to −3) of the *XPBC/ERCC-3* promoter sequence to the *EcoRI* site in the CAT gene, was cloned into pEP30 and a uniformly labeled antisense RNA probe was generated by T7 polymerase transcription. The expected protected fragments for *XPBC/ERCC-3* promoter-CAT transcripts (308-nt) and for the pRSV-CAT transcript (254-nt) are indicated. **B.** Cytoplasmic RNA was prepared from HeLa TK⁻ cells 48 h after transfection with constructs p(−370/−3)CAT, p(−129/−3)CAT, and the control pRSV-CAT, respectively. Twenty μg of cytoplasmic RNA was hybridized with the uniformly labeled RNA probe, as shown in (A), and digested with RNases A and T1. Lanes: 1, input probe; 2, tRNA control; 3, RNA isolated from nontransfected HeLa cells as a negative control; 4, 5, and 6, RNA isolated from pRSV-CAT, p(−370/−3)CAT, and p(−129/−3)CAT transfected HeLa cells, respectively.

defines a transcription initiation site which is in complete agreement with the major transcription startpoint of the endogenous *XPBC/ERCC-3* gene as described above.

The putative Sp1-binding element interacts specifically with a nuclear factor *in vitro*

Our transient transfection assays established that the putative Sp1-binding site in close proximity to the major cap site is essential for the basal *XPBC/ERCC-3* promoter activity. Sp1 was one of the first mammalian transcriptional activators to be described, and it binds specifically to a GC-box found in many viral and cellular promoters (29, 30). To determine whether the putative Sp1-binding site in close proximity to the major cap site of *XPBC/ERCC-3* is able to interact with transcription factor(s), we conducted a gel-mobility-shift assay. A ^{32}P-labeled double-stranded oligonucleotide, erccSp1, covering the DNA region from −123 to −108 in the *XPBC/ERCC-3* promoter (Fig. 6A) was incubated with a crude nuclear extract from HeLa TK$^-$ cells and then analysed by gel electrophoresis. As shown in Fig. 6B (lane 1), it formed two complexes in the presence of the nonspecific competitor poly(dI-dC)-poly(dI-dC). Both complexes were also detected with the oligonucleotide consSp1 (Fig. 6B,

A XPBC/ERCC-3 promoter sequence:

```
        -129      -119      -110
          *         *         *
ACGGCGCCTAGAGGCGGGACATCCTGTGTG
TGCCGCGGATCTCCGCCCTGTAGGACACAC

erccSp1:
      agctAGAGGCGGGACATCCTgatc
      tcgaTCTCCGCCCTGTAGGActag

consSp1:
      agctCGGGGCGGGGCGGGGCgatc
      tcgaGCCCCGCCCCGCCCCGctag
```

B

Probe ⇒	erccSp1						consSp1					
Competitor ⇒ (molar excess)	erccSp1		consSp1		CRE		erccSp1		consSp1		CRE	
	10x	100x	10x	100x	100x		10x	100x	10x	100x	100x	
Lane ⇒	1	2	3	4	5	6	7	8	9	10	11	12

Fig. 6. Band shift assays of the *XPBC/ERCC-3* Sp1 element. **A.** The oligonucleotide probe, erccSp1, used for band shift assay is aligned below the segment of the *XPBC/ERCC-3* promoter from −132 to −103. The 5′ endpoints of three deletions within this region (−129, −119, and −110) are also indicated. An oligonucleotide, consSp1, containing a consensus Sp1-binding site was used as the sequence-specific competitor. The lower-case letters indicate the nt incorporated into the probes in the labeling reaction. **B.** Oligonucleotide probes erccSp1 (lanes 1 to 6) and consSp1 (lanes 7 to 12) were incubated with 7 µg crude nuclear extracts from HeLa TK$^-$ cells in the absence or presence (either 10- or 100-fold molar excess) of unlabeled oligonucleotide competitors as indicated above the lanes. The protein-DNA complexes formed with erccSp1 and consSp1 are indicated.

lane 7) which contains a high-affinity binding consensus of nuclear factor Sp1 (Fig. 6A) (30). The upper complex resolves into two distinct bands on lower exposures (data not shown) which are consistent with the existence of two forms of Sp1 (105 and 95 kDa) (29, 36). The lower complex, which showed binding affinity and a competition pattern identical to that of Sp1, might represent a degradation or posttranslational product of Sp1 (37, 38) or a factor distinct from Sp1 as suggested by Refs. (39, 40).

To determine the specificity and affinity of the binding, we carried out a series of competition experiments. An unrelated oligonucleotide, fibCRE (see Materials and Methods), which contains the CRE sequence of the fibronectin promoter (23), was included as a negative control. The complex formed with labeled fibCRE was almost completely disrupted by a 10-fold molar excess of unlabeled fibCRE, whereas a 100-fold excess of either unlabeled erccSp1 or unlabeled consSp1 was unable to do so (data not shown). The DNA-protein complex formed with the *XPBC/ERCC-3* Sp1 element (lane 1) could be disrupted either by an excess of unlabeled erccSp1 oligonucleotide (lanes 2 and 3) or by an excess of consSp1 oligonucleotide (lanes 4 and 5) and vice versa (lanes 7 to 11), whereas fibCRE oligonucleotide could compete for none of the complexes (lanes 6 and 12); this established a relatedness in the binding specificity between the *XPBC/ERCC-3* Sp1 element and the consensus Sp1 element. It seems that the *XPBC/ERCC-3* Sp1 element has a slightly lower affinity in protein binding than the consensus Sp1 element, since a 100-fold excess of the consSp1 oligonucleotide could abolish the complexes formed with the *XPBC/ERCC-3* Sp1 element almost completely (lane 5), whereas a 100-fold excess of the erccSp1 oligonucleotide had only a slight effect on the retardation complexes formed with the consensus Sp1 element (lane 9).
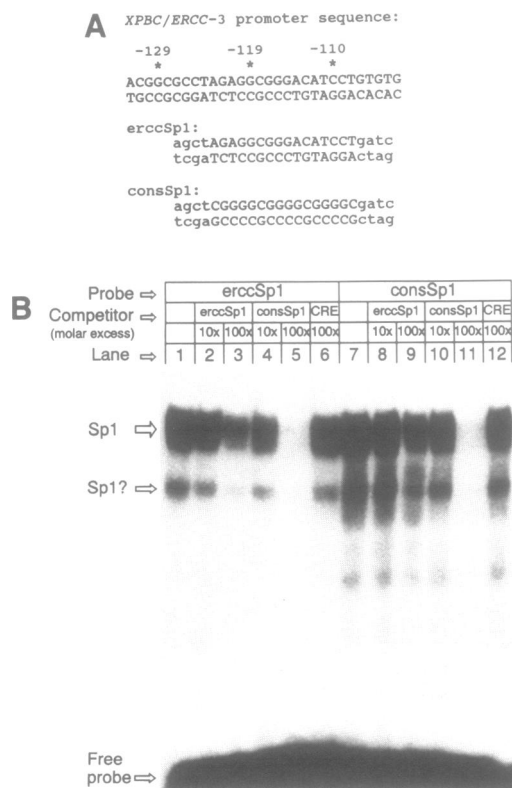
DISCUSSION

The human *XPBC/ERCC-3* gene, encoding a putative DNA helicase, is involved in an early step of nucleotide-excision repair and turned out to be the XP-B-correcting gene. Here we present a structural and functional analysis of the *XPBC/ERCC-3* promoter, the first mammalian DNA-repair gene promoter to be examined. The major transcription initiation site of the *XPBC/ERCC-3* gene has been determined by a combination of primer extension and RNase protection. Deletion analysis of the promoter-CAT reporter genes localizes the functional promoter region and reveals that transcription activity of the *XPBC/ERCC-3* promoter depends on the integrity of a putative Sp1-binding site. This putative Sp1-binding site can interact specifically with a nuclear factor, most likely transcription factor Sp1 or an Sp1-like factor, *in vitro*.

The *XPBC/ERCC-3* upstream region resembles the promoter sequences of some so-called 'housekeeping' genes (41), such as those encoding adenosine deaminase (42), hypoxanthine phos-phoribosyl transferase (HPRT) (43), and 3-hydroxy-3-methyl-glutaryl coenzyme A (HMG CoA) reductase (44), as well as the 5′-regions of some growth-related genes, such as those encoding the rat transforming growth factor (45), the epidermal-growth-factor receptor (46), and c-Harvey ras (47). Lack of typical TATA and CAAT boxes, but a high G+C content, multiple GC-boxes, and a heterogeneity of transcription initiation sites appear to be the characteristic features of the promoter regions of such genes (41). The *XPBC/ERCC-3* promoter also lacks canonical TATA and CAAT boxes, is GC-rich (from −200 to −1, 67% G+C), and contains three putative Sp1-binding sites. *XPBC/*

ERCC-3 mRNA initiates from multiple sites; the major transcription startpoint is located at 54 nt upstream of the presumed initiator codon ATG. Additional minor initiation sites extending beyond this region are also observed, and one of these (at −115) could be consistent with the cap site previously presumed by us (14) and account for the 5′ end of the published *XPBC/ERCC-3* cDNA clones (9). The choice of the major cap site does not seem to vary in a cell-type-specific or DNA-damage-dependent manner. In transient assays with *XPBC/ERCC-3* promoter-CAT constructs, we observed only one single initiation site, corresponding to the major startpoint of the endogenous gene. We assume that other distal *cis*-acting elements are required for the heterogeneity of transcription initiation which are absent in our promoter-CAT constructs.

Using a series of 5′3′ deletion mutants, we have shown that the promoter element required for the full level of expression in transient assays resides within a stretch of *XPBC/ERCC-3* sequence extending from −129 to −3 (construct p(−129/−3)CAT). Further deletion studies indicated that the integrity of a putative Sp1-binding site in close proximity to the major cap site is a prerequisite for active transcription of the *XPBC/ERCC-3* promoter. Furthermore, this Sp1 element is well conserved between mouse and human *XPBC/ERCC-3* promoters (48), supporting our conclusion that transcription factor Sp1 is important for regulation of the *XPBC/ERCC-3* gene.

Band shift assays demonstrated that the putative Sp1-binding site in close proximity to the major cap site of *XPBC/ERCC-3* can interact specifically with a nuclear factor *in vitro*. An oligonucleotide containing a consensus Sp1-binding site specifically competed with the putative *XPBC/ERCC-3* Sp1 element in forming DNA-protein complexes and vice versa. Therefore, it is most likely that the protein which complexes to the *XPBC/ERCC-3* Sp1 element represents transcription factor Sp1. The competition studies indicated that the *XPBC/ERCC-3* Sp1 element might have a lower affinity for Sp1 binding than the consensus Sp1-binding site. However, binding affinity is not the only determinant for Sp1 activation. Earlier work has established that in promoters containing multiple Sp1-binding elements the most important one is usually the one closest to the gene, approximately 40−70 nucleotides upstream of the transcription initiation site (30). Even weak Sp1-binding sites, such as GC-box I in the herpes simplex virus I (HSV) thymidine kinase (TK) promoter (49) and that in the SV40 early promoter (50), might well be very important for transcriptional activation. Recently, transcription factors other than the Sp1 factor which also interact with the GC-box have been identified (40, 51). We presently have no evidence for or against the possible involvement of other transcription factors than Sp1 itself interacting with the *XPBC/ERCC-3* Sp1 element.

Polypurine-polypyrimidine sequences have recently attracted much attention since they can form an unusual hairpin triplex structure (52). The function of the polypyrimidine motif is unknown at present. The pyrimidine-rich stretch in the *ERCC-2* promoter has been suggested to play a role in stable *ERCC-2* expression, since a cosmid lacking this sequence can only confer transient UV-resistance to UV5 cells (11). It has been suggested that the pyrimidine-rich stretch conserved between the promoters of *XPBC/ERCC-3* and *ERCC-1* may also play a role in regulation of the expression of *XPBC/ERCC-3* as well as *ERCC-1* (14). Deletion of the pyrimidine-rich stretch, however, does not affect the *XPBC/ERCC-3* promoter activity in transient assays. We do not know whether this element has any other direct effect on the

local chromatin structure at the *XPBC/ERCC-3* locus, for example, in the regulation of the *XPBC/ERCC-3* gene expression.

Finally, the existence of CRE/TRE-related elements and a putative AP-2 element in the *XPBC/ERCC-3* promoter makes it interesting to investigate whether cAMP/TPA and/or UV-light is able to induce the expression of the *XPBC/ERCC-3* gene. In mammalian cells, DNA damage leads to activation of the UV response with a putative role to protect cells against the effects of DNA damage. Interestingly, several mammalian phorbol esters-inducible genes are also induced by DNA damage (53). The UV- and phorbol-ester-response element of the collagenase promoter coincides with the AP-1 (c-Fos/c-Jun) binding site (22, 33). Recent studies with the *c-jun* gene identified AP-1 binding elements as likely regulators of gene activation during the mammalian UV response; however, the data suggested that DNA damage activates AP-1 by a mechanism that differs from that by TPA (54). Our preliminary data have demonstrated the CRE/TRE-related element as well as the inverted one in the *XPBC/ERCC-3* promoter can interact with nuclear factors *in vitro* (data not shown). Since the *XPBC/ERCC-3* gene is one of the few mammalian DNA excision repair genes that have been characterized so far, investigation of the possible induction by UV as well as by TPA and/or cAMP might shed some light on the question whether the mammalian DNA excision repair pathway is a component of the mammalian UV response.

REFERENCES

1. Friedberg, E.C. (1985) *DNA Repair*, W.H. Freeman & Co., San Francisco.
2. Cleaver, J.E. and Kraemer, K.H. (1989) In *The metabolic basis for inherited disease* (Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D., eds) Vol. 2, pp. 2949−2971, McGraw-Hill Book Co., New York.
3. Walker, G.C. (1985) *Annu. Rev. Biochem.*, **54**, 425−457.
4. Robinson, G.W., Nicolet, C.M., Kalainov, D. and Friedberg, E.C. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1842−1846.
5. Jones, J.S., Prakash, L. and Prakash, S. (1990) *Nucleic Acids Res.*, **18**, 3281−3285.
6. Madura, K. and Prakash, S. (1990) *Nucleic Acids Res.*, **18**, 4737−4742.
7. Tanaka, K., Miura, N., Satokata, I., Miyamoto, I., Yoshida, M.C., Satoh, Y., Kondo, S., Yasui, A., Okayama, H. and Okada, Y. (1990) *Nature* **348**, 73−76.
8. 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.
9. 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.
10. Van Duin, M., de Wit, J., Odijk, H., Westerveld, A., Yasui, A., Koken, M.H.M., Hoeijmakers, J.H.J. and Bootsma, D. (1986) *Cell*, **44**, 913−923.
11. Weber, C.A., Salazar, E.P., Stewart, S.A. and Thompson, L.H. (1990) *EMBO J.*, **9**, 1437−1447.
12. Mudgett, J.S. and MacInnes, M.A. (1990) *Genomics*, **8**, 623−633.
13. Troelstra. C., Odijk, H., de Wit, J., Westerveld, A., Thompson, L.H., Bootsma, D. and Hoeijmakers, J.H.J. (1990) *Mol. Cell. Biol.*, **11**, 5806−5813.
14. Weeda, G., Ma, L., van Ham, R.C.A., van der Eb, A.J. and Hoeijmakers, J.H.J. (1991) *Nucleic Acids Res.*, **19**, 6301−6308.

15. Van Duin, M., Koken, M.H.M., van den Tol, J., ten Dijke, P., Odijk, H., Westerveld, A., Bootsma, D. and Hoeijmakers, J.H.J. (1987) *Nucleic Acids Res.,* **15**, 9195−9213.

16. Van der Eb, A.J. and Graham, F.L. (1980) *Methods Enzymol.,* **65**, 826−839.

17. Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982) *Mol. Cell. Biol.,* **2**, 1044−1051.

18. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

19. Laird-Offringa. I.A., de Wit, C.L., Elfferich, P. and van der Eb, A.J. (1990) *Mol. Cell. Biol.,* **10**, 6132−6140.

20. Dery, C.V., Herrmann, C.H. and Mathews, M. B. (1987) *Oncogene,* **2**, 15−23.

21. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA,* **74**, 5463−5467.

22. Stein, B., Rahmsdorf, H.J., Steffen, A., Litfin, M. and Herrlich, P. (1989) *Mol. Cell. Biol.,* **9**, 5169−5181.

23. Dean, D.C., Blakeley, M.S., Newby, R.F., Ghazal, P., Hennighausen, L. and Bourgeois, S. (1989) *Mol. Cell. Biol.,* **9**, 1498−1506.

24. Jurka, J., and Smith, T. (1988) *Proc. Natl. Acad. Sci. USA,* **85**, 4775−4778.

25. Jones N.C., Rigby, P.W. and Ziff, E.B. (1988) *Gene Dev.,* **2**, 267−281.

26. Means, A.L. and Farnham, P.J. (1990) *Mol. Cell. Biol.,* **10**, 653−661.

27. Smale, S.T. and Baltimore, D. (1989) *Cell,* **57**, 103−113.

28. Kageyama, R., Merlino, G.T. and Pastan, I. (1989) *J. Biol. Chem.,* **264**, 15508−15514.

29. Briggs, M.R., Kadonaga, J.T., Bell, S.P. and Tjian, R. (1986) *Science,* **234**, 47−52.

30. Kadonaga, J.T., Jones, K.A. and Tjian, R. (1986) *Trends Biochem. Sci.,* **11**, 20−23.

31. Biggin, M.D. and Tjian, R. (1989) *Trends Genet.,* **5**, 377−383.

32. Imagawa, M., Chiu, R. and Karin, M. (1987) *Cell,* **51**, 251−260.

33. Angel,P., Imagawa, M., Chiu, R., Stein, B., Imbra, R.J., Rahmsdorf, H.J., Jonat, C., Herrlich, P. and Karin, M. (1987) *Cell,* **49**, 729−739.

34. Roesler, W.J., Vandenbark, G.R. and Hanson, R.W. (1988) *J. Biol. Chem.,* **263**, 9063−9066.

35. Ziff, E.B. (1990) *Trends Genet.,* **6**, 69−72.

36. Kadonaga, J. T., Carner, K.R., Masiarz, F.R. and Tjian, R. (1987) *Cell,* **51**, 1079−1090.

37. Dawson, P.A., Hofmann, S.L., van der Westhuyzen, D.R., Sudhof, T.C., Brown, M.S. and Goldstein, J.L. (1988) *J. Biol. Chem.,* **263**, 3372−3379.

38. Letovsky, J. and Dynan, W.S. (1989) *Nucleic Acids Res.,* **17**, 2639−2653.

39. Xiao, J.-H., Davidson, I., Macchi, M., Rosales, R., Vigneron, M., Staub, A. and Chambon, P. (1987) *Genes Dev.,* **1**, 794−807.

40. Huang, H.C., Sundseth, R. and Hansen, U. (1990) *Genes Dev.,* **4**, 287−298.

41. Dynan, W.S. (1986) *Trends Genet.,* **2**, 196−197.

42. Valerio, D., Duyvesteyn, M.G.C., Dekker, B.M.M., Weeda, G., Berkvens, T.M., van der Voorn, L., van Ormondt, H. and van der Eb, A.J. (1985) *EMBO J.,* **4**, 437−443.

43. Melton, D.W., Konecki, D.S., Brennand, J. and Caskey, C.T. (1984) *Proc. Natl. Acad. Sci. USA,* **81**, 2147−2151.

44. 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.

45. Blasband, A.J., Rogers, K.T., Chen, X., Azizkhan, J.C. and Lee, D.C. (1990) *Mol. Cell. Biol.,* **10**, 2111−2121.

46. Ishii, S., Xu, Y.-H., Stratton, R.H., Roe, B.A., Merlino, G.T. and Pastan, I. (1985) *Proc. Natl. Acad. Sci. USA,* **82**, 4920−4924.

47. Ishii, S., Merlino, G.T. and Pastan, I. (1985) *Science,* **230**, 1378−1381.

48. Weeda, G., Ma, L., van Ham, R.C.A., Bootsma, D., van der Eb, A.J. and Hoeijmakers. J.H.J. (1991) Carcinogenesis, in press.

49. Jones, K.A., Yamamoto, K.R. and Tjian, R. (1985) *Cell,* **42**, 55129−572.

50. Gidoni, D., Kadanoga, J.T., Barrera-Saldaña, H., Takahashi, K., Chambon, P. and Tjian, R. (1985) *Science,* **230**, 511−517.

51. Merchant, J.L., Demediuk, B. and Brand, S.J. (1991) *Mol. Cell. Biol.* **11**, 2686−2696.

52. Htun, H. and Dahlberg, J.E. (1988) *Science,* **241**, 1791−1795.

53. Kaina, B., Stein, B., Schönthal, A., Rahmsdorf, H.J., Ponta, H. and Herrlich, P. (1989) In *DNA repair mechanisms and their biological implications in mammalian cells* (Lambert, M.W., and Laval, J., eds) pp.149−165, Plenum Publ. Corp., New York.

54. Devary, Y., Gottlieb, R.A., Lau, L.F. and Karin, M. (1991) *Mol. Cell. Biol.,* **11**, 2804−2811.