Nucleic Acids Research, 1993, Vol. 21, No. 17 4111–4118

A 127 kDa component of a UV-damaged DNA-binding complex, which is defective in some xeroderma pigmentosum group E patients, is homologous to a slime mold protein

Masashi Takao, Marija Abramic'+, Malcolm Moos Jr[2], Vesna Rapic' Otrin, John C.Wootton[1], Mary McLenigan, Arthur S.Levine and Miroslava Protic'*

Section on DNA Replication, Repair and Mutagenesis, National Institute of Child Health and Human Development, and [1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health and [2]Laboratory of Developmental Biology, Division of Cellular and Gene Therapy, Food and Drug Administration, Bethesda, MD 20892, USA

Received April 6, 1993; Revised and Accepted June 30, 1993 GenBank accession no. L20216

ABSTRACT

A cDNA which encodes a ~127 kDa UV-damaged DNA-binding (UV-DDB) protein with high affinity for (6–4)pyrimidine dimers [Abramic', M., Levine, A.S. & Protic', M., J. Biol. Chem. 266:22493–22500, 1991] has been isolated from a monkey cell cDNA library. The presence of this protein in complexes bound to UV-damaged DNA was confirmed by immunoblotting. The human cognate of the UV-DDB gene was localized to chromosome 11. UV-DDB mRNA was expressed in all human tissues examined, including cells from two patients with xeroderma pigmentosum (group E) that are deficient in UV-DDB activity, which suggests that the binding defect in these cells may reside in a dysfunctional UV-DDB protein. Database searches have revealed significant homology of the UV-DDB protein sequence with partial sequences of yet uncharacterized proteins from *Dictyostelium discoideum* (44% identity over 529 amino acids) and *Oryza sativa* (54% identity over 74 residues). According to our results, the UV-DDB polypeptide belongs to a highly conserved, structurally novel family of proteins that may be involved in the early steps of the UV response, e.g., DNA damage recognition.

INTRODUCTION

Ultraviolet light (UV) is a major environmental mutagen and carcinogen. UV-induced skin cancer is the most frequently occurring malignancy among Caucasians, and may account for more than 600,000 new cases in the United States annually (1). Two major mutagenic photoproducts, cyclobutane pyrimidine dimers and (6-4) pyrimidine dimers, are formed in UV-irradiated DNA *in vitro* and *in vivo*, and both lesions have been implicated in neoplastic transformation (2–4). Efficient removal of UV photoproducts from DNA is mediated by DNA repair enzymes. In prokaryotes, the initial recognition of UV-induced DNA lesions is accomplished through the interaction of specific damage-recognition proteins and modified DNA. A damaged DNA-binding protein complex, $UvrA_2B$, which is involved in *E.coli* excision repair, recognizes a broad spectrum of chemically and physically induced DNA lesions, including both types of pyrimidine dimers (5).

Studies on the human sun-sensitive, cancer-prone disease, xeroderma pigmentosum (XP), which comprises seven genetic complementation groups (A–G) with various degrees of DNA repair deficiency, suggest that a multienzyme complex is required for efficient recognition and incision steps of excision repair in eukaryotes (6). Genes or cDNAs encoding putative DNA repair proteins that are defective in XP group A–D and XP group G cells have been isolated from normal human cells, and found specifically to complement UV survival and DNA repair in these XP cells (reviewed in 7; 8–10). While two of these XP genes (XPBC and XPDC) code for putative DNA helicases, the product of the XP A gene is a DNA-binding protein (11). Whether these proteins have higher affinity for UV-damaged than for intact DNA is equivocal; conflicting results on the XP A protein have been reported (11, 12). Several mammalian DNA-binding proteins that recognize a specific chemical or physical DNA lesion have recently been purified and cloned, but their role in DNA repair has yet to be established (reviewed in 13).

In an attempt to isolate a mammalian DNA-binding protein(s) with high affinity for UV-irradiated DNA, we and others have

* To whom correspondence should be addressed at: NIH, NICHD, Bldg. 6, Rm. 1A-15, Bethesda, MD 20892, USA

+ Present address: Department of Organic Chemistry and Biochemistry, Institute Rudjer Boškovic', Zagreb, Croatia

identified a DNA-binding complex (termed UV-DDB), which apparently recognizes (6-4) pyrimidine dimers *in vitro* but does not bind either to a TT cyclobutane pyrimidine dimer or to a bulky chemical lesion (13–17). Further studies have implicated the UV-DDB complex in the repair of UV-induced photoproducts on the basis of the regulation of UV-DDB activity by UV light, and the absence of binding activity in some XP group E patients (13, 14, 18–21).

As a first step in elucidating the role of UV-DDB protein in DNA repair, we now describe the isolation of a full length cDNA encoding the monkey ~127 kDa UV-DDB polypeptide which we previously purified (15). The N-terminal part of the translated UV-DDB sequence shows no significant similarity to sequences in data banks, while the C-terminal part is homologous to the partial sequences of yet uncharacterized proteins from slime mold and rice. Using anti-UV-DDB antibodies, we have identified the UV-DDB polypeptide as a component of complexes formed between UV-damaged DNA and nuclear extracts from monkey and normal human cells.

MATERIALS AND METHODS

Large scale purification of the UV-DDB protein and microsequence analysis

UV-DDB protein was purified from UV-treated African green monkey kidney cells as reported (15) with the following modifications: 1) TC-7 cells, a clone of CV-1 cells, were used as a source of UV-DDB protein because they have about 2-fold higher UV-DDB activity after UV irradiation than parental cells (14); 2) the cells were grown to high density, and harvested 3–5 days after UV irradiation. This resulted in twice as many cells per cm^2 of surface at the time of harvesting when compared to the original procedure; 3) two chromatographic steps in the purification protocol were omitted, and all buffers adjusted to 5–10 mM CHAPS and 0.3 M NaCl to reduce the loss of protein during concentration and dialysis. Nuclear extract was loaded on a column of hydroxylapatite, and UV-DDB-active fractions were eluted with 0.1–0.3 M KPO_4 buffer, pH 7.3. Fractions were pooled, concentrated and applied to a column of UV-irradiated DNA-cellulose. UV-DDB activity was eluted with 1.0 to 2.0 M NaCl. Fractions were separately concentrated in collodion bags (molecular weight cutoff 25,000) and analyzed on SDS-polyacrylamide gel electrophoresis using a PhastSystem (Pharmacia LKB Biotechnology). All UV-DDB-active fractions revealed the presence of a ~126 kDa protein.

One aliquot of the 1.5 M NaCl UV-DDB fraction was dialyzed at room temperature against 10 mM Tris–HCl buffer, pH 8.0, 0.1% SDS. Proteins were separated on an analytical SDS-polyacrylamide gel, and electrotransfered to a nitrocellulose membrane. UV-DDB protein was digested *in situ* with trypsin (22, 23). Peptide samples were diluted into 8 M guanidine chloride and passed through a 0.22 μm Ultrafree-MC filter (Millipore). Peptides were separated by high pressure liquid chromatography (Hewlett-Packard Model 1090 M equipped with a column oven) maintained at 60°C using a Vydac 214TP52 column and elution solvents as described (24). Fractions were collected manually, immediately placed on dry ice, and stored at −80°C. The peptides were analyzed with an Applied Biosystems Model 477A sequencer equipped with a Model 120A PTH Analyzer. The modifications to the standard Normal-1 chemistry recommended by Tempst and Riviere (25) and Speicher (26) were used. Data analysis was aided by the Applied Biosystems Model 610A software.

Oligonucleotides

Oligonucleotides were synthesized with an Applied Biosystems 380B DNA synthesizer or custom made by Oligos Etc., Inc. (Wilsonville, OR) and Genosys Biotechnologies, Inc. (The Woodlands, TX). Peptide-specific PCR primers were designed based on the sequences of tryptic peptides of purified UV-DDB. For an amino acid with 2 codon-degeneracy, a mixture of the two codons was used in the oligonucleotide sequence. For an amino acid with 3 or 4 codon-degeneracy, inosine (I) was incorporated. For leucine or arginine, YTI or RGI was used, respectively, and for serine, two primers including AGY or TCI were synthesized separately.

Isolation of a partial UV-DDB cDNA by PCR

Poly $(A)^+$RNA was prepared from UV-irradiated TC-7 cells, and reverse transcribed from the peptide-specific primers with a Fast Track™ and a cDNA Cycle™, respectively (Invitrogen, San Diego, CA). cDNA sequences were amplified by PCR with all possible combinations of primers. The PCR products were analyzed by Southern hybridization with an appropriate internal oligonucleotide probe to detect authentic fragments. The fragment obtained was analysed by sequencing and used as a DNA probe for screening cDNA libraries.

Screening of cDNA libraries and cloning of the 5′ region

Oligo (dT)-primed unidirectional cDNA was synthesized using poly$(A)^+$ RNA from UV-irradiated TC-7 cells and ligated to a plasmid vector. Transformants of *E.coli* DH10B were screened by means of colony hybridization using the DNA probe described above. Cloning of the UV-DDB cDNA 5′ region was performed by amplifying fragments from the DNA pool of a monkey cDNA library (Stratagene Inc., La Jolla, CA). An aliquot of the DNA pool was mixed with a pair composed of the gene-specific primer (5′-GTAAATGTTTCCTAGGCCACTACATA GGAG-3′) and vector-specific primer, and subjected to PCR amplification. PCR products were analysed by Southern hybridization using as the probe an appropriate restriction fragment from the previously cloned cDNA.

DNA sequencing

The cDNA obtained by library-screening was subcloned into pTZ19R or pGEM7Zf$^+$ using appropriate restriction sites, and sequenced with Sequenase V2.0 using a 7-deaza-dGTP Reagent™ (US Biochemical Corporation, Cleveland, OH). Some regions were sequenced by the use of cDNA-specific synthetic oligonucleotides. PCR fragments were sequenced either directly using a DNA Cycle Sequencing System™ (Gibco-BRL, Gaithersburg, MD), or after subcloning into the appropriate plasmids. In the latter case, at least four clones were sequenced to avoid misinterpretation of a sequence which might be mutated during the PCR. Both strands were sequenced throughout the entire region.

Northern hybridization

Total RNA was isolated from human and monkey cells according to Chirgwin (27), except that cesium chloride was replaced with cesium trifluoroacetate (Pharmacia). Skin fibroblasts from two patients with XP group E, XP82TO and XP95TO, were a

generous gift from S.Kondo, Japan. The origin of other cells, growth conditions and UV irradiation were as described (14, 16). Northern analysis was performed as recomended (Oncor, Gaithersburg, MD). DNA probes were labeled with ^{33}P-alpha-ATP (47.3 TBq/mmol; DuPont/NEN) using a Prime-a-Gene^{R} Labeling System (Promega Corp., Madison, WI).

Gel retardation assay and Western blotting

The DNA-binding assay was carried out as described previously (14, 15) using a UV-irradiated 60-mer oligonucleotide (13). Western analysis was done using chemiluminescent detection as recomended (Tropix, Bedford, MA). Polyclonal antibodies were raised in rabbits against the recombinant C-terminal domain of UV-DDB (amino acids 776 to 1140). To prepare this antigen, a fragment of UV-DDB cDNA (nucleotides 2412 to 3775) was amplified by PCR using Pfu polymerase (Stratagene), subcloned into the pRSETB vector (Invitrogen), and introduced into *E.coli* strain BL21[pLysE](28). The recombinant UV-DDB domain was overproduced by IPTG induction, and purified by Ni-chelate affinity chromatography (Quiagen, Chatsworth, CA) and SDS-polyacrylamide gel electrophoresis. The sensitivity of antibody detection, at a dilution of 1000 to 3000, was ~10 ng of purified recombinant protein.

Chromosomal localization

The chromosomal location of the human UV-DDB gene was determined by PCR analysis of DNA from a panel of 25 human-hamster hybrid cells (Bios Laboratories, New Haven, CT). The PCR was carried out with AmpliWax PCR Gems^{TM} applying the 'Hot Start' technique (Perkin-Elmer Corp., Norwalk, CT).

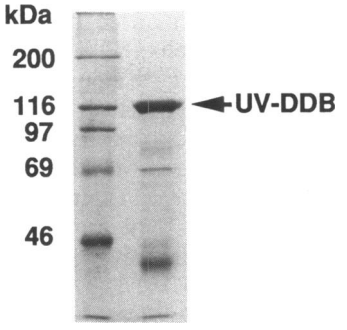

Figure 1. Analytical SDS-polyacrylamide gel electrophoresis of UV-DDB activity eluted with 1.5 N NaCl from UV-irradiated DNA-cellulose. Coomassie blue staining. kDa, molecular size markers.

RESULTS

Simplified isolation of UV-DDB protein for microsequencing

With a modified purification procedure, we obtained ~580 mg of nuclear extract proteins from 74 mL packed TC-7 cells which yielded ~640 μg of ~120 kDa UV-DDB polypeptide, presuming ~70% purity (Fig. 1). This is about a 5-fold improvement in the yield of UV-DDB protein with respect to our original procedure (15). The other major protein bands visible on the same SDS-polyacrylamide gel migrate as polypeptides of ~70 and ~40 kDa. A ~40 kDa protein band was present in all UV-DDB-active fractions eluted from UV-irradiated DNA-cellulose column and also in the UV-DDB MonoS Peak III purified previously (15). These ~70 and ~40 kDa proteins have not been analysed further. The UV-DDB polypeptide was digested with trypsin which yielded ~50 peptide peaks on reverse-phase high-pressure liquid chromatography, and 20 of these peaks were microsequenced.

Primary sequence of the UV-DDB protein from primate cells

We cloned a full-length cDNA encoding the ~127 kDa UV-DDB protein. Amino acid sequences of tryptic peptides obtained by microsequencing were used to design oligonucleotide primers. Among possible primer combinations, one primer set corresponding to sequences of peptides NVDSNEQG (in sense) and EMLGGEIIPY (in antisense) could initiate amplification of a 720 bp fragment (designated CV720) from the first strand cDNA pool. The CV720 fragment hybridized to an internal oligonucleotide probe whose sequence was later confirmed to be a part of CV720. A clone (pCV82) with a 3.3 kb-long insert was isolated from 7×10^5 recombinants of a monkey cDNA library which was screened with CV720 as a probe. Northern analysis revealed a unique UV-DDB mRNA in monkey cells of about 4.4 kb (see below). Therefore, the missing 5′ region was amplified from a lambda cDNA library. Using a gene-specific primer derived from the sequence 267 bp downstream of the 5′-end of the pCV82 insert, and a vector-specific primer, we obtained a 1.3 kb fragment (pCVU15) whose sequence partially overlapped the sequence of pCV82. Using another gene-specific primer designed from near the 5′-end of the pCVU15 insert did not reveal a cDNA with any further upstream sequence.

The entire cDNA sequence, generated from pCV82 and pCVU15 inserts, is 4199 nucleotides long (GenBank accession number L20216). A first ATG appears at nucleotide 87 and confers a single continuous open reading frame coding for a polypeptide of 1140 residues. We assigned this ATG to the initiation codon of the gene because, i) the deduced amino acid sequence has a molecular weight of 126,967 daltons, consistent with the apparent size of denatured UV-DDB protein (Fig. 1,

Table 1. UV-DDB peptides identified by microsequencing

38–LEIYVVTAEGLRP–52	570–LPSFELL–578
158–LEELHVIDVK–169	579–EMLGGEIIP–589
169–LYGCQAP–177	628–VTLGTQPTVLR–640
335–LNVDSNEQGSYVVAMETF–354	713–TVPLY–719
369–QGQGQ–375	823–DPNTYFIVGTAMVY–838
391–NGIGIHEHASI–403	847–IVVFQYSDGK–858
419–ETDDTLVLSFVGQ–433	917–GDFILVGDLM–928
484–ALVSEWK–492	1081–TEPATGFIDG–1092
514–ALYYLQIHPQ–525	

Figure 2. Alignment of the C-terminal domain of UV-DDB with translated cDNA sequences of uncharacterized proteins from slime mold, rice and human myoblasts (ORF-399). Positions of identical amino acids in adjacent (*) or non-adjacent (|) sequences; −, gaps in the sequence. A serine-rich cluster is underlined.

and Ref. 15), ii) the cDNA sequences of monkey and human UV-DDB show perfect identity after but not before the first ATG (Takao, unpublished), and iii) the sequence around the ATG codon (tagacATGt) is acceptable for the initiation of eukaryotic translation (29). The cDNA is consistent with the human frequency of codon usages, with the percent of G+C bases being 40−65. The polyadenylation signal AATAAA is located 21 nucleotides upstream from the poly(A) tail.

Seventeen peptide sequences (Table 1) obtained by microsequencing of the monkey UV-DDB protein match the deduced protein sequence of the isolated cDNA. The predicted pI of the translated protein is 4.9, which is close to the experimentally determined pI of ~5.3 for the monkey UV-DDB activity (Protic', unpublished). Searches of the NCBI non-redundant amino acid and nucleotide sequence databases using the BLAST series of programs (30−32) revealed that the C-terminal part of the translated UV-DDB sequence is clearly homologous (44% identity over 529 amino acids) to the translated partial cDNA sequence of an uncharacterized protein from a slime mold, *Dictyostelium discoideum* (Fig. 2, and S.Alexander, personal communication)(33). Within the same region of UV-DDB protein, three additional amino acid matches with various degrees of homology were found: a 74 amino acid match with 54% identity to the translated partial cDNA sequence of an uncharacterized protein from rice, *Oryza sativa* (Fig. 2; 34), a 361 amino acid match with 24% identity to the complete cDNA sequence of an uncharacterized protein from human myoblasts (Fig. 2; 37), and a 31 amino acid match (residues 627−659 with 42% identity and a Poisson probability of 2.4×10^{-6}; 35) to a POU domain protein from the zebrafish, *Brachyodanio rerio* (not shown; 36). The conserved short amino acid region (amino acids 405−436) of the POU domain protein is located just upstream of the bipartite DNA-binding domain, and may not be involved in DNA-binding (T.Johansen, personal communication).

Analysis of the UV-DDB sequence for local amino acid compositional complexity (38) located a significant low-complexity serine-rich cluster at residues 755−775 with a putative serine/threonine phosphorylation site and potential protease cleavage sites (Fig. 2). This cluster might act as an interdomain linker, a target for covalent post-translational modifications, or a protease-sensitive site. This view of a two-domain model for the UV-DDB protein is consistent with the predominant bands of ~80 kDa and ~40 kDa obtained after digestion of native UV-DDB protein with trypsin, thermolysin or proteinase K (Abramic' and Protic', unpublished).

The N-terminal region of the translated UV-DDB sequence showed no statistically significant sequence similarity by the criteria of Karlin and Altschul (35) and Altschul *et al.*(30) in database searches (NCBI non-redundant databases of June 1993) by pairwise BLAST alignments. Specific searches were also made using highly-diagnostic weight matrix discriminators (39) predictive of several well-recognized motifs found in DNA-binding proteins (helix-loop-helix, helix-turn-helix, homeobox, POU-domain, basic region, leucine zipper, four classes of zinc-finger). These searches were all negative, and no similarities were found to any other proteins in the transcription factors database, TFD (40). This result suggests that if the UV-DDB protein has a DNA-binding function, the protein structure involved is a new type or an atypical variant of some established motif.

Searches of the NCBI databases with the 3'-untranslated nucleotide sequence of the UV-DDB cDNA revealed near-identical matches to two independent, previously-unidentified human partial cDNAs (ESTs) from the collections of Adams et al.(41)(EST 00220 from the Stratagene hippocampus library, catalog number 936205) and Khan et al.(42) (EST IB748 from the human infant total brain library, Bento Soares). The occurrence of two independent clones in the small EST collections available to date is consistent with a relatively high abundance of the UV-DDB homolog mRNA in human brain (see below).

Expression of UV-DDB mRNA in human cells and tissues

Using a probe prepared from the pCV82 insert, expression of UV-DDB mRNA was examined by Northern analysis in several human tissues (Fig. 3A) and cell lines (Fig. 3B). A unique RNA of ~4.4 kb was detected in all tissues examined with the greatest abundance, relative to beta-actin mRNA, in placental and brain tissues. Similarly, primary skin fibroblasts from three XP E patients, with (XP95TO) or without (XP82TO and GM02415)

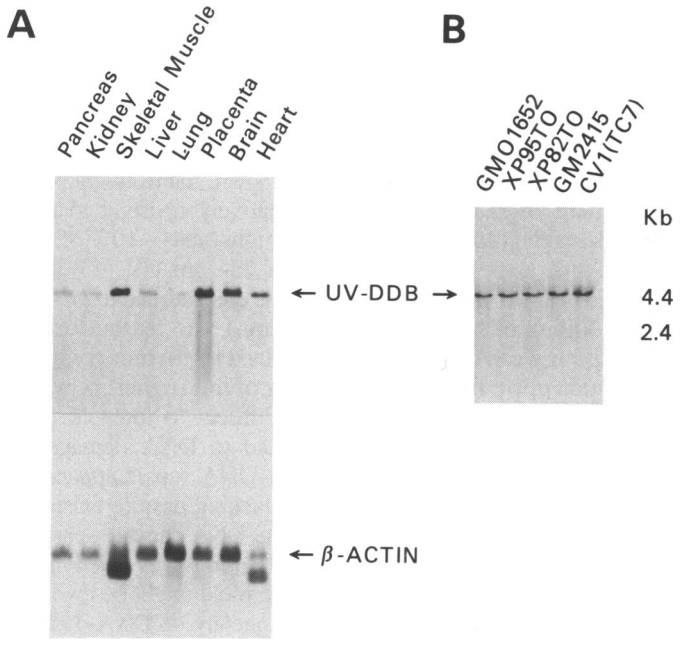

Figure 3. Expression of UV-DDB mRNA in (**A**) human tissues and (**B**) skin fibroblasts from normal human (GM01652) and XP group E patients (XP95TO, XP82TO, and GM2415), and monkey kidney TC-7 cells. Kb, position of RNA molecular size markers. Five μg of total RNA (B) or 2 μg of poly A$^+$ RNA (A)(MTN Blot, Clontech) were fractionated on a formaldehyde-agarose gel, transferred to a nylon membrane and hybridized with a cloned UV-DDB cDNA insert or a beta-actin probe.

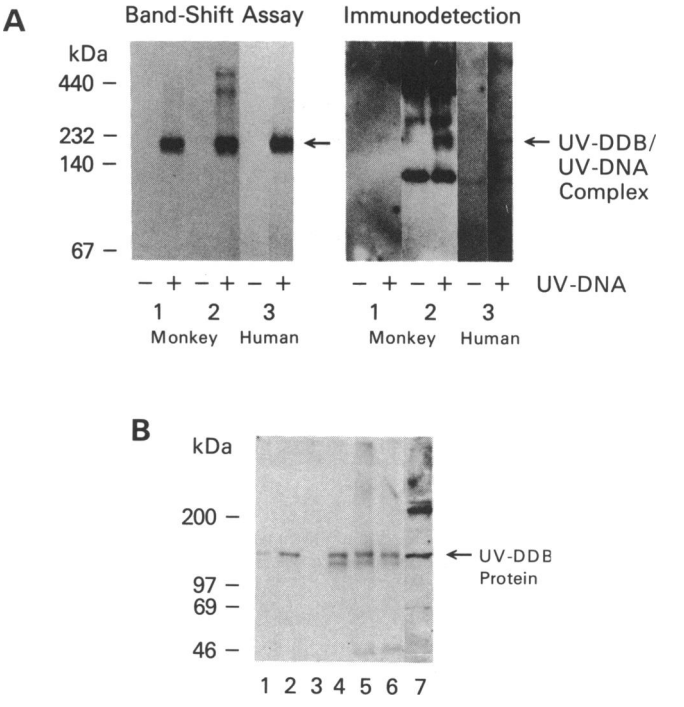

Figure 4. A) Immunodetection of UV-DDB polypeptide as a component of UV-damaged DNA-binding complexes from monkey and human cells. Aliquots of DNA-binding reactions were run on 4% polyacrylamide gels for 90 min (not shown) and on 4–20 % polyacrylamide gels (samples 1 and 2 for 3175 V-h, sample 3 for 1550 V-h). Resolved complexes were electrotransfered from gels to a PVDF membrane, and radiolabeled DNA probe was washed off with the blocking buffer containing 0.5 M NaCl. Membranes were exposed to X-ray films before and after washing. Autoradiograms of the gradient gels before (left) and autofluorograms of the same gels (right) after hybridization with UV-DDB specific antibodies. Visible bands on the left panels are complexes of UV-damaged DNA and proteins. Free probe, which was visible on the control 4% homogeneous gels (not shown), migrated out of the gels. Nuclear extract from UV-treated TC-7 cells (lane 1); 1.0 M NaCl eluate from UV-irradiated DNA-cellulose of TC-7 nuclear extract (lane 2); whole-cell extract from normal human lymphoblastoid cells, GM01953A (lane 3). Binding reactions were without (- UV-DNA) or with (+ UV-DNA) UV-irradiated DNA probe. **B)** Immunodetection of UV-DDB polypeptide in various fractions and nuclear extract from UV-irradiated monkey cells. Autofluorograms of UV-DDB immunoreactive bands after separation on a 6% SDS-polyacrylamide gel and Western transfer to a PVDF mambrane. MonoS UV-DDB peak I (lane 1) and peak III (lane 2) were purified previously (13); 0.7 M (lane 3), 1.0 M (lane 4), 1.5 M (lane 5) and 2.0 M (lane 6) NaCl eluates from UV-irradiated DNA-cellulose; nuclear extract (lane 7). Sample buffer contained 50 mM Tris–HCl, pH 6.8, 1% SDS, 10% glycerol, and 0.65 M beta-mercaptothanol. kDa, position of molecular size markers.

UV-DDB activity in a gel retardation assay, have UV-DDB mRNA indistinguishable in size and abundance from that found in normal human fibroblasts or monkey kidney cells. While a correlation between the level of UV-DDB mRNA expression and the capacity for excision repair in various human tissues cannot be established at the present time because of the lack of DNA repair data for these human samples, our results with the XP group E cells (which are ~50% repair-deficient) suggest that the level of expression of the UV-DDB gene is not a good indicator of the repair capacity.

UV-DDB polypeptide is a component of UV-damaged DNA-binding complexes

The presence of UV-DDB protein in complexes which retard UV-damaged DNA was examined by native polyacrylamide gel electrophoresis and immunoblotting (Fig. 4A). With the monkey UV-DDB fraction from a UV-DNA cellulose column as well as with nuclear extracts from monkey and normal human cells, a new immunoreactive UV-DDB protein band is visible at the position of retarded UV-irradiated DNA. Such a band is not visible in samples without added UV-damaged DNA. The size of the UV-DDB/UV-DNA complex is about ~200 kDa which is close to the size of a UV-DDB homodimer (15). There is no detectable binding of UV-DNA to UV-DDB monomers (that are visible in samples 2 and 3 just below the 140 kDa marker), which is consistent with our earlier finding that the UV-DDB monomer is an inactive form of the UV-DDB protein (15). Anti-UV-DDB antibodies detected a major UV-DDB polypeptide as a doublet of ~130 kDa, and two additional bands at ~200 kDa and ~50 kDa after resolution of the UV-DDB fractions and nuclear extract on an SDS-polyacrylamide gel (Fig. 4B). When the same samples were run on a urea-SDS gel, a single strong UV-DDB band was visible at ~130 kDa, perhaps the result of further solubilization of the UV-DDB protein and its migration to the lower molecular weight position (not shown).

Chromosomal location of the gene for human UV-DDB

A partial cDNA clone of the human homolog of UV-DDB has been isolated from a HeLa cDNA library (Takao, unpublished). Several base changes and small deletions/insertions were found in the 3′-noncoding region of the human cDNA when compared to the monkey sequence. This sequence information enabled the design of PCR primers (5′-CAGCTGCCCCCAGAGCCACA-3′ and 5′-ACCCATGGCTTGGCAGTCAG-3′) to amplify specifically a segment of the human UV-DDB gene. With this

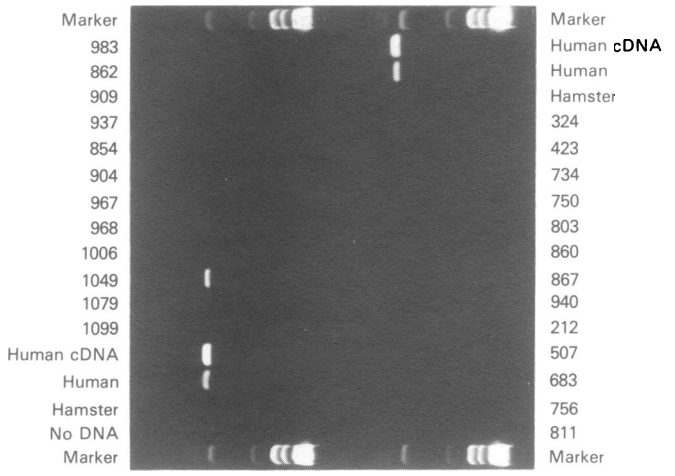

Figure 5. Chromosomal localization of the human UV-DDB gene. Agarose gel electro-phoresis of amplification products specific for human UV-DDB after staining with ethidium bromide.

primer pair, a single product of 425 bp was obtained from both a human cDNA clone and genomic DNA (Fig. 5). Under the same conditions, no detectable amplified products were obtained with monkey or chinese hamster genomic DNA. PCR analysis of the DNA from a panel of 25 human-hamster hybrid cell lines showed that 2 cell lines produced the characteristic 425 bp fragment (a strong signal in cell line 1049 and about an 8-fold weaker signal in cell line 683), implying that the gene for the human UV-DDB must be present in a human chromosome common to these two cell lines but not in any chromosomes in the other 23 cell lines. Analysis of the chromosomal content of this panel showed concordance of the human UV-DDB gene with chromosome 11. The weaker signal detected in cell line 683 (visible on the overexposed gel) is the result of a low percentage of hamster cells that have retained chromosome 11 (Bios Laboratories, personal communication).

DISCUSSION

Using protein microsequence information and PCR technology, we have cloned a cDNA which encodes a component of a primate DNA-binding complex with high affinity for double-stranded UV-irradiated DNA. The predicted molecular weight (127 kDa) of the cDNA-encoded protein is in agreement with our experimentally determined value for the purified UV-DDB protein. The amino acid sequences of 17 UV-DDB tryptic peptides align perfectly to the sequence of translated cDNA downstream from the assigned translation initiation codon. The 4.2 kb size of the monkey cDNA clone we isolated is close to the ~4.4 kb size of the human and monkey UV-DDB mRNA that was detected by Northern analysis. Antibodies raised against the recombinant UV-DDB protein hybridize to a ~127 kDa polypeptide from fractions of purified monkey UV-DDB protein as well as from nuclear extracts from monkey and human cells. Taken together, our results demonstrate that the cDNA which we have isolated encodes the UV-DDB protein.

The sequence of UV-DDB protein does not reveal any known DNA-binding- or other structural motifs that would offer additional information on its biochemical or cellular function. However, the UV-DDB protein appears to be a member of a structurally novel, highly conserved family of proteins: A partial sequence of a UV-DDB homolog, which has been recently isolated from the slime mold, is 44% identical to the monkey UV-DDB (33, and S. Alexander, personal communication). The cellular function and biochemical properties of the UV-DDB homolog in the slime mold are not yet known. Interestingly, wild-type strains of *D. discoideum* are relatively resistant to most DNA-damaging agents, including UV light, and ~10% of cells survive after exposure to 155 J/m^2 of 254 nm UV (43).(This UV dose is about 10-fold greater than the dose needed to produce similar killing of human cells in culture.) The reason for this UV resistance can only be partly explained by asymmetry in the distribution of pyrimidines in the DNA of this organism, which may account for a shielding effect of three- to fourfold. The additional resistance of the slime mold to DNA damage is, perhaps, provided by a very efficient DNA repair apparatus, conclusive evidence for which is still lacking despite existence of an impressive collection of slime mold DNA damage-sensitive mutants assigned to 11 complementation groups (reviewed in 44; 45). The UV-DDB homolog in the slime mold might be involved in DNA damage-recognition, and consequently, in DNA-damage repair or DNA-damage tolerance, the functions proposed for the mammalian UV-DDB activity (see below and ref. 14, 13, 16, 17, 46).

A partial cDNA sequence of an uncharacterized protein from rice may be a part of the plant UV-DDB homolog because it is 54% and 38% identical to the C-terminal region of the monkey UV-DDB protein and the UV-DDB homolog from slime mold, respectively. In contrast, the human protein from myoblasts (Fig. 2, ORF-399), although structurally related, shows only 26−29% identity to the same C-terminal region of the three proteins. This percent of identity decreases to 24 when the entire sequence of the protein is taken into account. Therefore, this novel human protein may be a member of another family of proteins that are structurally related but not homologs of the UV-DDB protein. Future studies are needed to unveil the cellular function and relation of these proteins.

A significant homology was found between the short stretch of residues in the UV-DDB protein and the POU domain protein from zebrafish. This protein is a member of the POU domain family of DNA-binding proteins implicated in transcriptional activation and stimulation of initiation of DNA replication (36). However, the region of homology between UV-DDB and the POU domain protein is unlikely to be involved in binding to DNA because it is located just upstream from the bipartite DNA-binding domain, POU-specific domain and POU-homeodomain. The zebrafish POU domain protein mRNA is expressed ubiquitously during embryonic development which is similar to our finding that UV-DDB mRNA is expressed in all human tissues tested. Ubiquitous tissue expression and evolutionary conservation are characteristics of proteins involved in essential cellular functions such as transcription, replication, DNA repair, and the stress response, as well as proteins involved in cellular organizing structures such as the cytoskeleton.

Several mammalian genes implicated in nucleotide excision repair show striking sequence similarity to their homologs from yeast and Drosophila, and some of them are also essential for cell growth (reviewed in 7, 47 and 48). While most of the polypeptides encoded by mammalian repair genes have not yet been characterized, the gene products of their homologs have been identified as DNA helicases, and surprisingly, potential

transcriptional and translational regulators. Recently, one of these helicases (XPBC/ERCC3 gene product) has been identified as part of the basal transcription factor TFIIH, further linking an association of nucleotide excision repair and transcription (reviewed in 48). Two human DNA-binding proteins that recognize either apurinic/apyrimidinic sites in DNA or cisplatinum DNA-crosslinks have been implicated in transcriptional activation, but their role in DNA repair is still uncertain (reviewed in 13). It will be of interest to determine whether the primate UV-DDB protein belongs to a class of such DNA-binding proteins, and if the previously observed regulation of UV-DDB activity by UV light is related to a specific UV-induced activation/repression of genes by the UV-DDB protein (14, 16, 18).

The absence of UV-DDB activity (also refered to as 'XPE-binding factor'; 19) in cells from 3 of 12 XP E patients, and its presence in cells from more than 20 non-XP E individuals tested, suggests a significant association between deficiency of UV-DDB activity and the XP E phenotype (18–21). Patients belonging to XP group E have characteristic clinical signs of XP genodermatoses, with a high incidence of skin tumors (49–51). However, XP E cells show only a moderate defect in survival and DNA repair synthesis after UV irradiation, which is in contrast to the severe UV hypersensitivity and excision repair deficiency found in several other XP groups. XP E cells are apparently able to remove both types of pyrimidine dimers, but with slower kinetics than normal human cells (52–54). UV-DDB protein from primate cells binds with high affinity to (6-4)pyrimidine dimers and has no detectable enzyme activity toward UV-damaged DNA (13–17, 19, 55). This UV-DDB protein might also bind to some cyclobutane pyrimidine dimers but not the major UV-photoproduct, TT cyclobutane dimers. Therefore, a defect in recognition of some UV lesions, which may alter the rate but not the overall repair of UV-induced photoproducts, might explain the relatively mild phenotype and DNA repair deficiency of XP group E patients. Our analysis of cells from two XP E patients that are defective in UV-DDB activity has failed to demonstrate any gross alteration in size or in abundance of UV-DDB mRNA. These results suggest that another gene product may play a role in the XP group E defect or that structural alterations in the UV-DDB polypeptide may occur. Interestingly, the expression of another XP complementing gene, XPGC (9), has been found recently to be unaltered in one XP group G cell line which suggests that the increased UV sensitivity of these XP-G cells is likely to be due to a minor rather than gross alteration of a normally active XPGC allele.

Recently, Hwang and Chu (55) have reported purification of a 125 kDa DNA-binding protein ('XPE-binding factor') from human placenta that has high affinity for UV-irradiated DNA and shows striking biochemical similarity (i.e., behavior on chromatographic matrices, substrate specificity, requirement for free thiols, similar size of the denatured protein, etc.) to the UV-DDB protein which we isolated from monkey kidney cells (15). With UV-DDB-specific antibodies in hand and UV-DDB sequence information, it should now be possible to test whether XPE-binding factor and UV-DDB protein are true functional and structural homologs.

ACKNOWLEDGEMENTS

We thank K.H.Kraemer, J.Hurwitz, and P.Herrlich for critical reading of the manuscript and helpful discussions.

REFERENCES

1. Scotto, J. (1989) *Sunlight, Ultraviolet Radiation, and the Skin*, NIH Consensus Development Conference, pp. 61.
2. Brash, D.E., Rudolph, J.A., Simon, J.A., Lin, A., McKenna, G.J., Baden, H.P., Halperin, A.J. & Ponten, J. (1991) *Proc. Natl. Acad. Sci. USA 88*, 10124–10128.
3. Mitchell, D.L. & Nairn, R.S. (1989) *Photochem. Photobiol. 49*, 805–819.
4. Protic'-Sabljic', M., Tuteja, N., Munson, P.J., Hauser, J., Kraemer, K.H. & Dixon, K. (1986) *Mol. Cell. Biol. 6*, 3349–3356.
5. Van Houten, B. (1990) *Microbiol. Rev. 54*, 18–51.
6. Cleaver, J.E. & Kraemer, K.H. (1989) in: Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D. (Eds.), *The Metabolic Basis of Inherited Disease*, Vol. 2, McGraw-Hill, New York, pp. 2949–2971.
7. Hoeijmakers, J.H.J. (1991) *J. Cell Sci. 100*, 687–691.
8. Legerski, R. & Peterson, C. (1992) *Nature 359*, 70–73.
9. Scherly, D., Nouspikel, T., Corlet, J., Ucla, C., Bairoch, A. & Clarkson, S.G. (1993) *Nature 363*, 182–185.
10. O'Donovan, A. & Wood, R.D. (1993) *Nature 363*, 185–188.
11. Robins, P., Jones, C.J., Biggerstaff, M., Lindahl, T. & Wood, R.D. (1991) *The EMBO J. 10*, 3913–3921.
12. Eker, A.P.M., Vermeulen, W., Miura, N., Tanaka, K., Jaspers, N.G.J., Hoeijmakers, J.H.J. & Bootsma, D. (1992) *Mutation Res. 274*, 211–224.
13. Protic', M. & Levine, A.S. (1993) *Electrophoresis*, in press.
14. Hirschfeld, S., Levine, A.S., Ozato, K. & Protic', M. (1990) *Mol. Cell. Biol. 10*, 2041–2048.
15. Abramic', M., Levine, A.S. & Protic', M. (1991) *J. Biol. Chem. 266*, 22493–22500.
16. McLenigan, M., Levine, A.S. & Protic', M. (1993) *Photochem. Photobiol. 57*, 655–662.
17. Treiber, D.K., Chen, Z. & Essigmann, J.M. (1992) *Nucleic Acids Res. 20*, 5805–5810.
18. Protic', M., Hirschfeld, S., Tsang, A.P., McLenigan, M., Dixon K. & Levine A.S. (1992) in: Rossman, T. (Ed.), *Induced Effects of Genotoxic Agents in Eukaryotic Cells*, Hemisphere Publishing Corp., Washington, pp. 21–37.
19. Chu, G. & Chang, E. (1988) *Science 242*, 564–567.
20. Keeney, S., Wein, H. & Linn, S. (1992) *Mutat. Res. 273*, 49–56.
21. Kataoka, H. & Fujiwara, Y. (1991) *Biochem. Biophys. Res. Commun. 175*, 1139–1143.
22. Aebersold, R.E., Leavitt, J., Saavedra, R.A., Hood, L.E. & Kent, S.B.H. (1987) *Proc. natl. Acad. Sci. USA 84*, 6970–6974.
23. Tempst, P., Link, A.J., Riviere, L.R., Fleming, M. & Elicone, C. (1990) *Electrophoresis 11*, 537–553.
24. Stone, K.L. & Williams, K.R. (1986) *J. Chromatography 359*, 203–212.
25. Tempst, P. & Riviere, L.R. (1989) *Anal. Biochem. 183*, 290–300.
26. Speicher, D.W. (1989) in: Hugli, T. (Ed.), *Techniques in Protein Chemistry*, Academic Press, San Diego, pp. 24–35.
27. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.T. & Rutter, W.J. (1979) *Biochemistry 18*, 5294–5299.
28. Studier, F.W., Rosenberg, A.H., Dunn, J.J. & Dubendorff (1990) *Methods Enzymol. 185*, pp. 60–89.
29. Kozak, M. (1986) *Cell 44*, 283–292.
30. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) *J. Mol. Biol. 215*, 403–410.
31. Altschul, S.F. & Lipman, D.J. (1990) *Proc. Natl. Acad. Sci. USA 87*, 5509–5513.
32. Gish, W. & States, D.J. (1993) *Nature Genetics* (in the press).
33. Sydow, L., Alexander, H. & Alexander, S.J. (1992) Genbank accession number X65937.
34. Sasaki, T. & Minobe, Y. (1993) Genbank accession number D15449.
35. Karlin, S. & Altschul, S.F. (1990) *Proc. Natl. Acad. Sci. USA 87*, 2264–2268.
36. Johansen, T., Moens, U., Holm, T., Fjose, A. & Krauss, S. (1993) *Nucl. Acids Res. 21*, 475–483.
37. Nomura, N., Miyajima, N., Kawarabayashi, Y. & Tabata, S. (1993) Genbank accession number D13642.
38. Wootton, J.C. & Federhen, S. (1993) *Computers and Chemistry* (in the press).
39. Akrigg, D., Attwood, T.K., Bleasby, A.J., Findlay, J.B.C., North, A.C.T., Maughan, N.A., Parry-Smith, D.J., Perkins, D.N. & Wootton, J.C. (1992) *Comput. Appl. Biosci. 8*, 295–296.
40. Ghosh, D. (1992) *Nucl. Acids Res. 20* Suppl., 2091–2093.
41. Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R. & Venter, J.C. (1991) *Science 252*, 1651–1656.
42. Khan, A.S., Wilcox, A.S., Polymeropoulos, M.H., Hopkins, J.A., Stevens,

T.J., Robinson, M., Orpana, A.K. & Sikela, J.M. (1992) *Nature Genetics 2*, 180–185.
43. Welker, D.L. & Deering, R.A. (1978) *J. Gen. Microbiol. 109*, 11–23.
44. Deering, R.A. (1988) in: Friedberg, E.C. & Hanawalt, P.C. (Eds.), *DNA Repair*, Vol. 3, Marcel Dekker, New York, pp. 39–76.
45. Bronner, C.E., Welker, D.L. & Deering, R.A. (1992) *Mutation Res. 274*, 187–200.
46. Chu, G. & Chang, E.(1990) *Proc. Natl. Acad. Sci. USA 87*, 3324–3327.
47. Friedberg, E.C. (1992) *Cell 71*, 887–889.
48. Bootsma, D. & Hoeijmakers, J.H.J. (1993) *Nature 363*, 114–115.
49. Thielmann, H.W., Popanda, O., Edler, L. & Jung, E.G. (1991) *Cancer Res. 51*, 3456–3470.
50. Kondo, S., Fukuro, S., Mamada, A., Kawada, A., Satoh, Y. & Fujiwara, Y. (1988) *J. Invest. Dermatol. 90*, 152–157.
51. Kondo S., Mamada, A., Miyamoto, C., Keong, C.-H., Satoh, Y. & Fujiwara, Y. (1989) *Photodermatology 6*, 89–95.
52. Mitchell, D.L. (1988) *Photochem. Photobiol. 48*, 51–57.
53. Zelle, B. & Lohman, P.H.M. (1979) *Mutat. Res. 62*, 363–368.
54. Roza, L., Vermeulen, W., Bergen Henegouwen, J.B.A., Eker, A.P.M., Jaspers, N.G.J., Lohman, P.H.M. & Hoeijmakers, J.H.J. (1990) *Cancer Res. 50*, 1905–1910.
55. Hwang, B.J. & Chu, G. (1993) *Biochemistry 32*, 1657–1666.