MOLECULAR AND CELLULAR BIOLOGY, June 1995, p. 3206–3216 Vol. 15, No. 6
0270-7306/95/$04.00+0

Molecular Cloning and Expression of Human cDNAs Encoding a Novel DNA Ligase IV and DNA Ligase III, an Enzyme Active in DNA Repair and Recombination

YING-FEI WEI,[1] PETER ROBINS,[2] KENNETH CARTER,[1] KEITH CALDECOTT,[2] DARRYL J. C. PAPPIN,[3] GUO-LIANG YU,[1] RUI-PING WANG,[1] BRENDA K. SHELL,[1] RACHEL A. NASH,[2] PRIMO SCHÄR,[2] DEBORAH E. BARNES,[2] WILLIAM A. HASELTINE,[1] AND TOMAS LINDAHL[2]*

Human Genome Sciences, Inc., Rockville, Maryland 20850-3338,[1] and Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, Hertfordshire EN6 3LD,[2] and Imperial Cancer Research Fund, London WC2 3PX,[3] United Kingdom

Received 9 November 1994/Returned for modification 15 December 1994/Accepted 23 February 1995

Three distinct DNA ligases, I to III, have been found previously in mammalian cells, but a cloned cDNA has been identified only for DNA ligase I, an essential enzyme active in DNA replication. A short peptide sequence conserved close to the C terminus of all known eukaryotic DNA ligases was used to search for additional homologous sequences in human cDNA libraries. Two different incomplete cDNA clones that showed partial homology to the conserved peptide were identified. Full-length cDNAs were obtained and expressed by in vitro transcription and translation. The 103-kDa product of one cDNA clone formed a characteristic complex with the XRCC1 DNA repair protein and was identical with the previously described DNA ligase III. DNA ligase III appears closely related to the smaller DNA ligase II. The 96-kDa in vitro translation product of the second cDNA clone was also shown to be an ATP-dependent DNA ligase. A fourth DNA ligase (DNA ligase IV) has been purified from human cells and shown to be identical to the 96-kDa DNA ligase by unique agreement between mass spectrometry data on tryptic peptides from the purified enzyme and the predicted open reading frame of the cloned cDNA. The amino acid sequences of DNA ligases III and IV share a related active-site motif and several short regions of homology with DNA ligase I, other DNA ligases, and RNA capping enzymes. DNA ligases III and IV are encoded by distinct genes located on human chromosomes 17q11.2-12 and 13q33-34, respectively.

DNA strand breaks and gaps are generated transiently during replication, repair, and recombination. In mammalian cell nuclei, rejoining of such strand interruptions depends on several different DNA polymerases and DNA ligases. The occurrence of three different DNA ligases was established previously by biochemical and immunological characterization of purified enzymes (33, 41). However, the interrelationship among these proteins was unclear, as a cDNA clone has been available only for DNA ligase I, the major enzyme of this type in proliferating cells (2). The main function of DNA ligase I appears to be the joining of Okazaki fragments during lagging-strand DNA replication (18, 31, 43). Clues as to the physiological role of DNA ligase III have come from its physical interaction in a high-salt-resistant complex with another nuclear protein, the XRCC1 gene product (5, 20). The *XRCC1* gene encodes a 70-kDa protein, which by itself does not appear to join DNA strand breaks (5, 20, 39). However, mutant rodent cells deficient in XRCC1 protein exhibit reduced DNA ligase III activity, defective strand break repair, and an anomalously high level of sister chromatid exchanges; are hypersensitive to simple alkylating agents and ionizing radiation; and have an altered mutation spectrum (with an increased proportion of deletions) after exposure to ethyl methanesulfonate (5, 20, 28, 39). These data indicate that *xrcc1* mutant cells are defective in base excision-repair and strongly suggest that both DNA ligase III and XRCC1 are active in this process (8). In a separate development, a purified mammalian protein fraction active in repair and recombination processes in vitro was shown to contain a DNA ligase with the properties of DNA ligase III, but no detectable amounts of DNA ligase I (13). The role of the distinct enzyme, DNA ligase II, remains unclear, although an observed increase in DNA ligase II activity during meiotic prophase suggests a role in meiotic recombination (11). Comparison of ^{32}P-adenylylated DNA ligases II and III by partial or complete proteolytic cleavage patterns indicated that these two enzymes share extensive amino acid sequence similarity or identity in regions flanking their active sites but that they are quite different from DNA ligase I (33). Neither DNA ligase I, II, nor III is exclusively a mitochondrial enzyme.

The human cDNA encoding DNA ligase I was cloned by hybridization to synthetic oligonucleotides deduced from the partial amino acid sequence of the purified bovine enzyme and by functional complementation of a *Saccharomyces cerevisiae cdc9* conditional lethal DNA ligase mutant (2). The predicted amino acid sequence of the human DNA ligase I cDNA shows homology to other eukaryotic DNA ligases, having 83% identity to murine DNA ligase I (36) and 44 and 37% identity to the *Schizosaccharomyces pombe* cdc17$^+$ and *S. cerevisiae* CDC9 gene products, respectively (2), and it shows limited homology to other microbial DNA ligases. The gene encoding DNA ligase I in human cells has been assigned to chromosome 19q13.2-13.3 (27). No cDNA clones representing DNA ligase II or III were obtained during cDNA screening procedures for DNA ligase I. However, it remained possible that limited sequence homology between the enzymes might occur within discrete shared motifs corresponding to short arrays of essential amino acid residues; such conserved sequence homologies could provide a strategy for isolating relevant cDNA clones. In the present work, we describe the molecular cloning, sequenc-

* Corresponding author.

ing, and mapping of both the DNA ligase III cDNA and a cDNA encoding a functional enzyme that is identical to a previously unrecognized fourth DNA ligase isolated from human cell nuclei, DNA ligase IV. This sets the stage for investigation of the roles of these enzymes in cellular DNA replication, repair, and recombination processes.

MATERIALS AND METHODS

Isolation of full-length cDNA clones and DNA sequencing. Approximately 10^6 recombinants representing a human cDNA library were screened with a radiolabeled partial cDNA probe by standard procedures (21). Full-length DNA ligase III and DNA ligase IV cDNAs were isolated from a HeLa library (Life Technologies) and a prostate library (Human Grenome Sciences [HGS]), respectively. The insert size of positive clones was verified by PCR with gene-specific and vector-specific primers. DNA sequencing was carried out on both strands at the core sequencing facility of HGS by using an automated ABI373 DNA Analysis System (Applied Biosystems Inc.).

In vitro transcription-translation of cDNA clones. Putative full-length cDNA clones were subcloned as follows: (i) DNA ligase III was subcloned as a *Sal*I-*Not*I restriction fragment into the multiple cloning site of pSPORT (Life Technologies), with the 5′ end proximal to the T7 promoter, and (ii) DNA ligase IV was subcloned as an *Eco*RI-*Xho*I restriction fragment into the multiple cloning site of pBluescript SK (Stratagene), with the 5′ end proximal to the T3 promoter. The DNA ligase III or IV plasmid constructs (1 μg) were linearized with either *Not*I or *Xho*I (New England Biolabs), downstream of the cDNA inserts, and then transcribed and capped at 37°C for 30 min with T7 or T3 RNA polymerase and the mCAP RNA capping kit (Stratagene). The reactions were terminated by incubation with 10 U of RNase-free DNase at 37°C for 5 min. Following phenol-chloroform extraction and ethanol precipitation, the in vitro transcription products were resuspended in 20 μl of 10 mM Tris HCl–1 mM EDTA, pH 8.0 (TE). The transcripts (0 to 5 μl, made up to a final volume of 5 μl in each case with water) were translated in 20 μl of rabbit reticulocyte lysate (Amersham) at 30°C for 90 min. In order to radiolabel the products of in vitro translation, reaction mixtures were supplemented with 20 μCi [^{35}S]methionine (3,000 Ci mmol^{-1}; Amersham). Translations were terminated by incubation with 5 μl of a solution containing 400 μg of RNase A ml^{-1} and 50 mM EDTA at 37°C for 15 min (final volume, 30 μl). Samples (5 μl) of mixtures from translations carried out in the presence of [^{35}S]methionine were analyzed by electrophoresis in sodium dodecyl sulfate (SDS)–7.5% polyacrylamide gels and autoradiography. Nonradiolabeled translation products were assayed for ability to form protein-adenylate complexes after removal of ATP by chromatography through spun 1-ml columns of Sephadex G-50 (Pharmacia) equilibrated with TE.

DNA ligase assays. Samples (5 μl) from in vitro translations were adenylylated in reaction mixtures (30 μl) containing 60 mM Tris HCl (pH 8.0), 10 mM $MgCl_2$, 50 μg of bovine serum albumin (BSA) ml^{-1}, 5 mM dithiothreitol (DTT), and 1 μCi of [α-^{32}P]ATP (3,000 Ci mmol^{-1}; Amersham) at 20°C for 10 min and then analyzed by electrophoresis in SDS–7.5% polyacrylamide gels and autoradiography. In order to monitor transfer of [^{32}P]AMP from protein-adenylate to a nicked DNA substrate, 5-μl samples from adenylylation reactions were incubated for further time periods with or without 500 ng of nonradiolabeled oligo(dT)$_{16}$-poly(dA), as described previously (41). The ability to transfer [^{32}P]AMP from enzyme-adenylate to the hybrid substrate, oligo(dT)-poly(rA) or oligo(rA)-poly(dT), differentiates DNA ligases I, II, and III (41). However, both these hybrid substrates were rapidly degraded by an RNase H activity upon incubation in the reticulocyte lysate, even when mixtures were used directly without termination of translation reactions by addition of RNase A.

Affinity purification with XRCC1-His protein. Mixtures for in vitro translation reactions carried out in the presence of [^{35}S]methionine were made up to 100 μl by the addition of 50 mM Tris HCl (pH 7.5)–50 mM NaCl–0.1 mM EDTA–1 mM DTT–4% (vol/vol) glycerol (buffer A) and applied to a 1-ml spun column of Sephadex G-50 (Pharmacia) equilibrated with buffer A. A 50-μl sample of the eluate was incubated at 20°C for 20 min with or without the addition of 3 μg (~2% of total protein) of recombinant C-terminally histidine-tagged XRCC1 protein (XRCC1-His) (4a). The mixture was diluted to 100 μl with buffer A, added to a 25-μl bed volume of Ni-nitrilotriacetic acid agarose (Qiagen), and incubated on ice for 20 min with frequent agitation. A 50-μl sample of the suspension was retained (load). The agarose beads were recovered from the remainder of the suspension, and the supernatant of nonadsorbed material was retained. The beads were washed four times in 100 μl and then once in 50 μl (final wash) of buffer A containing 25 mM imidazole. XRCC1-His was eluted with 2 50-μl volumes of buffer A containing 200 mM imidazole. Aliquots (25 μl) of the samples taken at each stage were analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and autoradiography. Recovery of ^{35}S-labeled protein in the 200 mM imidazole eluate was quantitated by using a phosphorimager (Molecular Dynamics); recovery of XRCC1-His was monitored by Coomassie blue staining of SDS-polyacrylamide gels prior to autoradiography.

Far Western blotting analysis. Samples from translation reactions, carried out with or without the addition of transcript, were separated by SDS-PAGE and transferred to nitrocellulose alongside 10 μg of HeLa cell nuclear extract (23). Proteins were analyzed by far Western blotting (22) with a ^{32}P-phosphorylated XRCC1 protein probe. Briefly, XRCC1-His recombinant protein (1 μg) was phosphorylated in a reaction mixture (30 μl) containing 25 mM Tris HCl (pH 7.5), 10 mM $MgCl_2$, 25 mM NaCl, 1 mM DTT, 50 μCi of [γ-^{32}P]ATP (3,000 Ci mmol^{-1}; Amersham), and 0.6 mU of casein kinase II (Boehringer Mannheim) at 20°C for 25 min. Unincorporated nucleotide was removed by application to a 1-ml spun column of Sephadex G-50 (Pharmacia) equilibrated with 1 mg of BSA ml^{-1} in TE. Blots were incubated at 4°C for 6 h with radiolabeled XRCC1 protein (1 μg, 10^7 cpm) at a final concentration of 50 ng/ml. Washed and dried blots were analyzed by using a phosphorimager.

Purification of DNA ligases from mammalian tissues and cell lines. Bovine DNA ligases I and II were isolated essentially as described previously (40, 41). DNA ligase I was purified from calf thymus, and DNA ligase II was purified from calf liver. DNA ligases III and IV were isolated from a nuclear extract (23) prepared from 60 liters of exponentially growing HeLa cells. These two human enzymes initially copurified by the column chromatography procedures described for bovine DNA ligase III (41) but were separated during further purification. A detailed comparison of the two 100-kDa DNA ligases and further biochemical characterization of DNA ligase IV will be described elsewhere.

Mass spectrometry of tryptic peptides from DNA ligase IV. Following concentration in a Centricon-30 filtration unit (Amicon), an aliquot (0.5% volume) of the purified protein fraction containing full-length HeLa DNA ligase IV (20 μg of total protein at 0.2 mg ml^{-1}) was adenylylated with [α-^{32}P]ATP (see above) before being added back to the rest of the protein. After electrophoresis through an SDS–7.5% polyacrylamide gel, protein was transferred in 10 mM 3-[cyclohexylamino]-1-propanesulfonic acid (pH 11.0)–10% methanol to an Immobilon-P membrane (Millipore). The membrane was dried under a vacuum and stained with 0.005% sulforhodamine B (Sigma) in 30% methanol–0.2% acetic acid (29). When DNA ligase-[^{32}P]AMP was added to the protein sample as a marker, the desired DNA ligase band could be identified by aligning the autoradiograph with the blot. The band was then excised and recovered for protease digestion and mass spectrometry. DNA ligase IV was digested with trypsin, and the molecular masses of the resultant peptides were determined by using a LaserMAT mass spectrometer. The fragment masses were used to screen the molecular weight search (MOWSE) peptide mass database, as described previously (30, 34).

Northern (RNA) hybridization analysis. Human tissues were provided by the Cooperative Human Tissue Network, funded by the U.S. National Cancer Institute. The procedure for isolation of human total RNA was modified from the RNAzol B method of Biotecx Laboratories, Inc. Briefly, 2 g of tissue was mixed with 40 ml of RNAzol B solution and homogenized on a Polytron PT3000 homogenizer (Brinkmann) at 22,000 rpm for 2 to 3 min. A 5-ml volume of chloroform was added to the homogenate, after which the mixture was vortexed for 15 s and incubated on ice for 5 min. The suspension was centrifuged at 12,000 × *g* at 4°C for 15 min, and the upper phase was mixed with an equal volume of isopropanol, incubated at 4°C for 5 min to precipitate RNA, and centrifuged again at 12,000 × *g* for 15 min. The RNA pellet was resuspended in 2 ml of water, extracted twice with phenol-chloroform, ethanol precipitated, and finally resuspended in water. A 15-μg amount of total RNA was electrophoresed in a 1% agarose–formaldehyde gel, transferred to a nylon membrane, and hybridized with a radiolabeled partial DNA ligase III or IV cDNA probe (21). Blots were washed three times with 0.1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% SDS at 65°C for 1 h and subjected to autoradiography.

Isolation of partial genomic clones and FISH analysis. Genomic clones of human DNA ligase III or IV were isolated from a lambda FIX II library (Stratagene) by using standard procedures (21). Clones containing a portion of the human DNA ligase IV gene, or a full-length DNA ligase III cDNA, were nick translated with digoxigenin-dUTP (Boehringer Mannheim), and fluorescence in situ hybridization (FISH) was carried out as detailed by Johnson et al. (15). Individual chromosomes were counterstained with DAPI (4′,6-diamidino-2-phenylindole), and color digital images, containing both DAPI and gene signals, were recorded by using a triple-band pass filter set (Chroma Technology, Inc., Brattleboro, Vt.) in combination with a charged coupled-device camera (Photometrics, Inc., Tucson, Ariz.) and variable excitation wavelength filters (14). Images were analyzed by using the ISEE software package (Inovision Corp., Durham, N.C.).

Nucleotide sequence accession numbers. The sequence of the HGS473238 cDNA, encoding DNA ligase III (see Results), has been deposited in the EMBL database under accession number X84740. The sequence of the HGS340221 cDNA, encoding DNA ligase IV (see Results), has been deposited in the EMBL database under accession number X83441.

RESULTS

Isolation of cDNA clones with partial homology to other DNA ligases. All eukaryotic DNA ligases cloned to date show partial sequence homology with mammalian DNA ligase I, particularly in a region flanking a conserved lysine residue within a -K-DG-R- active-site motif that is essential for formation of an enzyme-adenylate reaction intermediate (42). It

would be anticipated that the other two DNA ligases previously described to be present in mammalian cells, DNA ligases II and III, share partial sequence homology with this active-site region, as they utilize a reaction mechanism identical to that of other DNA ligases. Nevertheless, analysis of active-site peptides obtained after digestion of the ^{32}P-adenylylated enzymes with reagent proteases indicated that the sequence of the active-site region of DNA ligase I is distinctly different from that of DNA ligase II or III (33). Aside from the active-site motif, a highly conserved 16-amino-acid sequence of unknown function is located near the carboxyl termini of DNA ligases from vaccinia virus, *S. cerevisiae*, and *S. pombe* (38), as well as human DNA ligase I (2). A rabbit polyclonal antiserum was raised against a synthetic consensus peptide corresponding to this conserved sequence (40). The antiserum clearly identified DNA ligase I of human or bovine origin but did not cross-react detectably with small amounts of bovine DNA ligase II or III (40, 41). However, with larger quantities of more highly purified enzyme preparations, a weak cross-reaction with both bovine DNA ligase II and bovine DNA ligase III was detected (data not shown), indicating that a peptide sequence with partial homology to the consensus sequence occurs in mammalian DNA ligases II and III. Furthermore, by analogy with DNA ligase I, this peptide would be expected to be C terminal to the active site in DNA ligases II and III, an advantage in cDNA screening strategies, as such sequences are more frequently represented in oligo(dT)-primed cDNA libraries.

We searched a current database of human cDNA sequences identified as expressed sequence tags (ESTs) (1) for sequences with amino acid homology with the active-site region of DNA ligase I or with the C-terminal consensus sequence, NH_2-GIS LRFPRFTRIREDK-COOH (40). Ten ESTs were identified with the latter sequence: seven were identical to human DNA ligase I, another clone (approximately 1.5 kb) was unique and showed identity with the search peptide at 13 of 16 residues, and the remaining two clones (approximately 2 kb) were identical to each other and showed identity with the search peptide at 10 of 16 residues. In database searches, both the 1.5- and 2-kb clones showed the greatest amino acid sequence homology with vaccinia virus DNA ligase (data not shown). Human cDNA libraries were screened by hybridization with the EST clones in order to obtain full-length cDNAs. Clones HGS 473238 and HGS340221, putative full-length cDNAs corresponding to the 1.5- and 2-kb ESTs, respectively, were analyzed further.

Clone HGS473238 encodes a polypeptide with the properties of DNA ligase III. In vitro transcription of the HGS473238 cDNA insert with T7 RNA polymerase and translation of the capped message in a rabbit reticulocyte lysate supplemented with [^{35}S]methionine yielded a major radioactively labeled polypeptide with an apparent molecular mass of 100 kDa. This polypeptide comigrated with a 100-kDa adenylylated band containing partially purified DNA ligase III (and IV) upon SDS-PAGE but showed electrophoretic mobility quite different from that of authentic mammalian DNA ligase II (Fig. 1). We examined whether the 100-kDa in vitro translation product interacted with human XRCC1 protein; this was shown previously to be a characteristic of DNA ligase III (5). XRCC1-His was incubated with the [^{35}S]methionine-labeled in vitro translation product of the HGS473238 cDNA clone to allow formation of XRCC1-protein complexes, after which nickel-agarose beads were added to affinity bind XRCC1-His. The agarose beads were washed to remove nonspecifically associated polypeptides prior to elution of XRCC1-His with 200 mM imidazole. Eluted XRCC1-His protein fractions were examined for cofractionation of associated radiolabeled polypeptides. XRCC1-His bound the product of clone HGS473238, as indicated by the partial depletion of radiolabeled polypeptides from the nonadsorbed fraction (Fig. 2B, lane 2) and their recovery together with XRCC1-His in the imidazole eluate (Fig. 2A and B, lanes 4 and 5). Recovery of radiolabeled polypeptides was dependent on addition of XRCC1-His (Fig. 2C). Approximately 50% of the full-length 100-kDa translation product, and as much as 90% of some of the truncated polypeptides, was recovered with XRCC1-His. In contrast, the separate 100-kDa in vitro translation product of the HGS 340221 clone (see below) did not interact significantly with XRCC1-His; less than 3% of the full-length product was recovered (Fig. 2D). Furthermore, when the HGS473238 clone was translated in the absence of [^{35}S]methionine prior to XRCC1-His affinity purification of the product, a ~100-kDa polypeptide recovered in the 200 mM imidazole eluate generated a protein-[^{32}P]AMP complex upon incubation with [α-^{32}P]ATP in a standard DNA ligase assay (data not shown). Similar results were observed when in vitro translation products were examined by far Western blotting with a radiolabeled XRCC1-His probe (Fig. 2E). The XRCC1 probe detected a 100-kDa polypeptide in HeLa cell nuclear extract. The product of clone HGS473238 was also detected by the XRCC1 probe, but the product of clone HGS340221 was not. These results indicate that the HGS473238 cDNA clone encodes a 100-kDa polypeptide with the properties of the previously characterized protein, DNA ligase III.

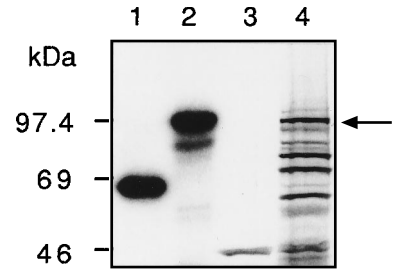


FIG. 1. In vitro transcription-translation of full-length DNA ligase III cDNA. The HGS473238 cDNA was transcribed with T7 RNA polymerase, the capped message was translated in a rabbit reticulocyte lysate supplemented with [^{35}S]methionine, and radiolabeled products were analyzed by SDS-PAGE and autoradiography. Identical translation reactions were carried out both with (lane 4) and without (lane 3) addition of the transcript. The major 100-kDa translation product is indicated by an arrow. The following DNA ligases partially purified from mammalian cells were labeled with [α-^{32}P]ATP and applied to the same gel to allow molecular mass comparisons: bovine DNA ligase II (70 kDa, lane 1) and human DNA ligases III and IV (100 kDa, lane 2). An active fragment (87 kDa) is also visible (lane 2). The positions of ^{14}C-methylated protein size markers (Amersham) are indicated.

Full-length cDNA clone encoding human DNA ligase III. The longest open reading frame of the HGS473238 cDNA, encoding DNA ligase III, extends from bp 73 to 3099 (3,027 bp) within the cDNA clone and would encode a polypeptide of 1,009 amino acids, a molecular mass of ~115 kDa. The next downstream ATG at bp 334 occurs in a typical translation start consensus (16) and defines an open reading frame of 2,766 bp (922 amino acids); the protein produced in this case would be approximately 103 kDa, consistent with both the observed molecular mass of the in vitro translation product and the apparent molecular mass of authentic DNA ligase III purified from HeLa cells by standard chromatographic procedures. This would indicate that HGS473238 represents a full-length cDNA clone. Furthermore, a 5′-truncated cDNA clone lacking the first 78 bp of HGS473238 (and the first ATG codon) produced an in vitro translation product with electrophoretic mobility

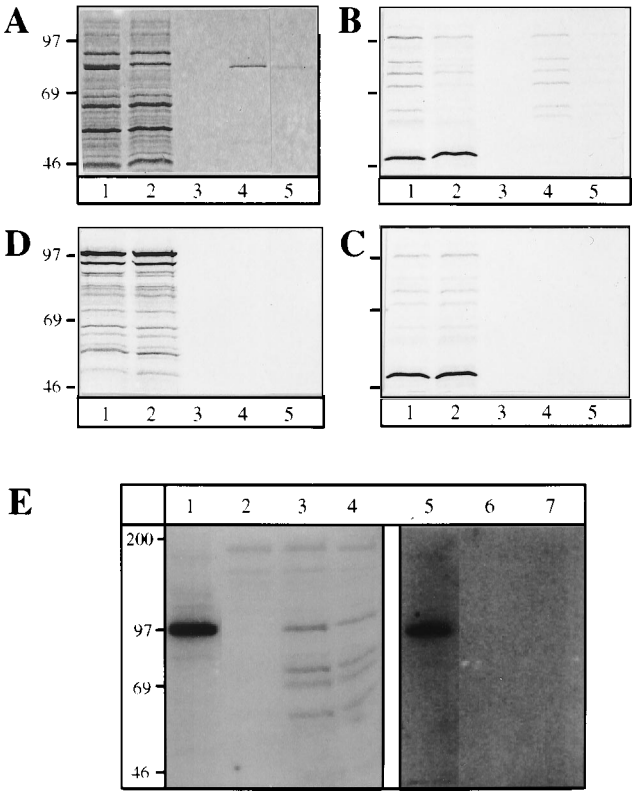


FIG. 2. Interaction of in vitro-translated DNA ligase III with XRCC1. (A to D) Affinity purification. The in vitro transcripts of HGS473238 (A to C) and HGS340221 (D) were translated in the presence of [^{35}S]methionine, and the protein products were incubated with (A, B, and D) and without (C) XRCC1-His protein. Recovery of ^{35}S-labeled protein during affinity purification of XRCC1-His on nickel-agarose beads was monitored by SDS-PAGE and autoradiography (B to D); recovery of XRCC1-His protein was monitored by Coomassie blue staining (A). Lanes 1, load onto beads (see Materials and Methods); lanes 2, nonadsorbed material; lanes 3, 25 mM imidazole final wash; lanes 4, first 200 mM imidazole eluate; lanes 5, second 200 mM imidazole eluate. (E) Far Western blotting. HeLa cell nuclear extract or samples from in vitro transcription-translation of HGS473238 (lanes 1 to 4) or HGS340221 (lanes 5 to 7) were analyzed by SDS-PAGE, far Western blotting with a ^{32}P-phosphorylated XRCC1 probe, and autoradiography. Lanes 1 and 5, 10 μg of HeLa cell nuclear extract; lanes 3 and 4, 10- and 5-μl samples from translation reactions with the HGS473238 transcript; lane 7, 10-μl sample from translation reaction with HGS340221 transcript; lanes 2 and 6, 10-μl samples from negative (no added transcript) control reactions carried out in parallel with the HGS473238 and HGS340221 translations, respectively. Molecular mass markers are as described for Fig. 1.

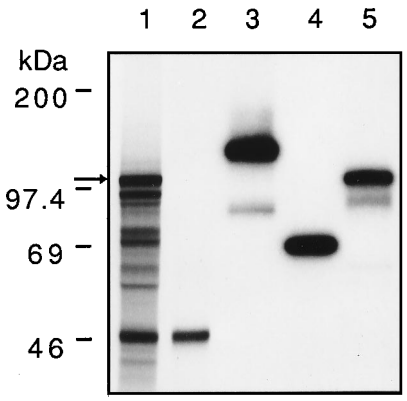


FIG. 3. In vitro transcription-translation of full-length DNA ligase IV cDNA. The HGS340221 cDNA was transcribed with T3 RNA polymerase, the capped message was translated in a rabbit reticulocyte lysate supplemented with [^{35}S]methionine, and radiolabeled products were analyzed by SDS-PAGE and autoradiography. Identical translation reactions were carried out both with (lane 1) and without (lane 2) addition of the transcript. The major 100-kDa translation product is indicated by an arrow. The following DNA ligases partially purified from mammalian cells were labeled with [α-^{32}P]ATP and applied to the same gel to allow molecular mass comparisons: bovine DNA ligase I (125 kDa, lane 3), bovine DNA ligase II (70 kDa, lane 4), and human DNA ligases III and IV (100 kDa, lane 5). Active fragments (85 and 87 kDa) are also visible (lanes 3 and 5). Molecular mass markers are as described for Fig. 1.

identical to that of the product encoded by HGS473238, in support of assignment of the ATG at bp 334 as the translation initiation codon. With an open reading frame of 2,766 bp, the 5′ untranslated region (UTR) would be 333 bp in length and contains an in-frame stop codon. At the 3′ end of the cDNA clone, the 3′ UTR extends for 283 bp beyond the open reading frame before a consensus -AATAAA- polyadenylation signal (32) at bp 3382, 20 bp before the poly(A) tail.

Clone HGS340221 encodes a distinct polypeptide with the enzymatic properties of a DNA ligase. In vitro transcription of the HGS340221 cDNA insert with T3 RNA polymerase and translation of the capped message in a rabbit reticulocyte lysate supplemented with [^{35}S]methionine yielded a major radioactively labeled polypeptide with an apparent molecular mass of 100 kDa. This polypeptide comigrated with a 100-kDa adenylylated band containing partially purified DNA ligase IV (and III) upon SDS-PAGE but showed electrophoretic mobility quite different from that of authentic mammalian DNA ligases I and II (Fig. 3). Since this gene product was distinct from the previously known DNA ligases I, II, and III, it became important to demonstrate that the protein acted as a DNA ligase in standard enzyme assays. Eukaryotic and viral DNA ligases interact with ATP to form a covalent enzyme-AMP intermediate as the first step of a ligation reaction; the size of the radioactively labeled reaction intermediate can be estimated by SDS-PAGE following incubation of the enzyme with [α-^{32}P]ATP. The adenylate group can be specifically removed from the complex by incubation with nicked DNA, since the AMP moiety is transferred to the 5′ phosphate at a strand break. When the in vitro transcript of HGS340221 was translated in the absence of [^{35}S]methionine and incubated with [α-^{32}P]ATP in a standard assay to monitor formation of a DNA ligase-AMP reaction intermediate, a 100-kDa ^{32}P-labeled product was identified by SDS-PAGE (Fig. 4, lane 1). Although other small polypeptides in the reticulocyte lysate became radioactively labeled upon incubation with the [α-^{32}P]ATP, generation of the 100-kDa product was dependent upon addition of the HGS340221 transcript to translation reaction mixtures (Fig. 4, lanes 1 and 6). Furthermore, the radiolabeled AMP residue was rapidly lost from the 100-kDa polypeptide upon incubation with the synthetic nicked DNA substrate, oligo(dT)$_{16}$-poly(dA), but remained as an enzyme-AMP complex under identical incubation conditions in the absence of the substrate (Fig. 4, lanes 2 to 5). The rapid dissociation of the enzyme-AMP complex in the presence of nicked DNA is characteristic of a DNA ligase.

Full-length cDNA clone encoding human DNA ligase IV. In vitro transcription-translation experiments indicated that human cDNA clone HGS340221 encodes a 100-kDa polypeptide that acts as a DNA ligase in standard in vitro assays. We have shown that human cells contain two DNA ligases with molecular masses of approximately 100 kDa as estimated by SDS-PAGE. One of these, previously identified as DNA ligase III, is encoded by the HGS473238 cDNA. The other is a previously unidentified, fourth DNA ligase activity from human cells; the further biochemical characterization of this enzyme, DNA li-

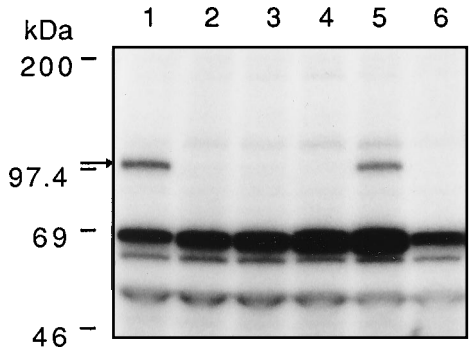


FIG. 4. Formation of DNA ligase IV-adenylate and discharging of the complex upon incubation with a nicked DNA substrate. The in vitro transcript of HGS340221 was translated in the absence of [^{35}S]methionine and then incubated with [α-^{32}P]ATP. Identical translation reactions were carried out both with (lane 1) and without (lane 6) addition of the transcript. The transcript-dependent 100-kDa adenylylated polypeptide is indicated by an arrow. Adenylylated translation products were incubated with (lanes 2 to 4) and without (lane 5) oligo(dT)$_{16}$-poly(dA) for 1 min (lane 2), 2 min (lane 3), or 5 min (lanes 4 and 5). Radiolabeled polypeptides were analyzed by SDS-PAGE and autoradiography. Molecular mass markers are as described for Fig. 1.

gase IV, will be described elsewhere. In this study, DNA ligase IV has been isolated from HeLa cell nuclei by standard methods of enzyme purification and digested with trypsin, and the resultant peptides have been analyzed by mass spectrometry. Proteins produce unique fingerprints of peptide masses following enzymatic digestion, and so it is possible to determine whether an isolated protein shows identity with a known sequence by comparison against a peptide mass database (30). All the tryptic peptide masses of DNA ligase IV purified from HeLa cell nuclei matched those of tryptic peptides predicted from the open reading frame in the HGS340221 clone (Table 1) and did not do so with any other protein in the current database of 84,000 proteins. These data confirm that the protein encoded by the HGS340221 cDNA is identical to human DNA ligase IV.

The longest open reading frame extends from bp 274 to 3006 (2,733 bp) within the cDNA clone and would encode a polypeptide of 911 amino acids. An ATG at bp 274 defines the longest open reading frame, but a pyrimidine-rich stretch immediately preceding this ATG places it within a highly unfavorable sequence context for translation initiation (16). Thus, it seems unlikely that this first ATG is used in vivo. Further downstream at bp 475, a putative initiation ATG occurs in a translation start consensus sequence (16) and defines an open reading frame of 2,532 bp (844 amino acids). The protein produced would be approximately 96 kDa, consistent with the observed molecular mass of the in vitro translation product. Since the translated polypeptide in turn showed the same apparent molecular mass as did authentic DNA ligase IV purified from HeLa cells by standard chromatographic procedures, this would indicate that HGS340221 represents a full-length cDNA clone. The 14 tryptic peptides of DNA ligase IV analyzed by mass spectrometry were matched with expected sequences, and evenly distributed, over the range of amino acid residues 3 to 833 of the 844-residue open reading frame product. No peptide was detected that would correspond to a sequence in the open reading frame between bp 274 and 475, in support of the assignment of the latter ATG sequence as the translation initiation codon. With an open reading frame of 2,532 bp, the 5′ UTR would be 474 bp in length and contain four in-frame stop codons. The 3′ UTR at the 3′ end of the cDNA clone was sequenced for 319 bp beyond the open reading frame, 3,325 bp from the 5′ end of the clone. Despite the consensus -AATAAA- polyadenylation signals (32) at bp 3247 and 3318, the cDNA apparently extends further 3′ for some 700 bp.

TABLE 1. Correspondence of human DNA ligase IV tryptic peptides identified by mass spectrometry with those predicted from the open reading frame of the HGS340221 cDNA clone

Peptide mass (Da)	Peptide sequence	Corresponding amino acid residues from cDNA sequence[a]
805	EEGIMVK	359–365
825	LGHETLR	310–316
981	LILPQLER	3–10
1,016	SFVPWQPR	667–674
1,033	YWKPFHR	458–464
1,105	DALKLLNYR	38–46
1,152	HQSFYIETK	198–206
1,179	ESWVTDSIDK	824–833
1,352	MQMHKDGDVYK	212–222
1,600	NEVIDALNEAIDKR	345–358
1,629	HLYIGGDDEPQEKK	547–560
1,714	VGSGCTMKELYDLGLK	439–454
1,749	RYEILSSIFTPIPGR	318–332
2,302	EYDCYGDSYFIDTDLNQLK	690–708

[a] The open reading frame encoded by the HGS340221 cDNA is shown in Fig. 5.

Homology between human DNA ligases III and IV, other ATP-dependent DNA ligases, and RNA guanylyltransferases. The predicted amino acid sequences encoded by the human DNA ligase III and IV cDNAs are shown in Fig. 5, aligned with human DNA ligase I. At optimum alignment, there is extensive amino acid sequence homology among the three enzymes. DNA ligases III and IV show identity at 13 and 10, respectively, of the 16 amino acid residues that define the consensus peptide conserved in other ATP-dependent DNA ligases, which was the probe used in the initial search of ESTs to identify further potential DNA ligase clones. The DNA ligase III and IV sequences are also identical at 8 and 7, respectively, of 12 residues (−2 to +9) flanking the active-site lysine of DNA ligase I, and both contain the minimum active-site consensus for all ATP-dependent DNA ligases, -K-DG-R-, with Lys-421 (DNA ligase III) and Lys-206 (DNA ligase IV) being the putative active lysines.

The positions of these two highly conserved motifs within the predicted amino acid sequences of human DNA ligases I, III, and IV are indicated in Fig. 6. Although their amino acid sequences are not colinear at optimum alignment, DNA ligases I and IV differ by just 1 amino acid in the size of the region between the two motifs, while DNA ligases I and III differ by 9 amino acids. The 3′ flanking motif is located 37 amino acids from the C terminus of DNA ligase I, whereas the DNA ligase III sequence extends a further 195 residues and the DNA ligase IV sequence extends another 323 residues beyond this motif. DNA ligase IV shows no homology to either DNA ligase I or III, or to any other protein sequences in the databases, within this extended C-terminal region, while the C terminus of DNA ligase III shows weak homology to several proteins in the databases, including ~20% identity (45% similarity) to a 144-amino-acid sequence within the C-terminal one-quarter of both human and murine XRCC1 (data not shown). In their N-terminal regions, DNA ligases I, III, and IV show very limited sequence homology beyond about 30 residues upstream of their active sites, and DNA ligase I has an extended hydrophilic N-terminal region with no homology to DNA li-

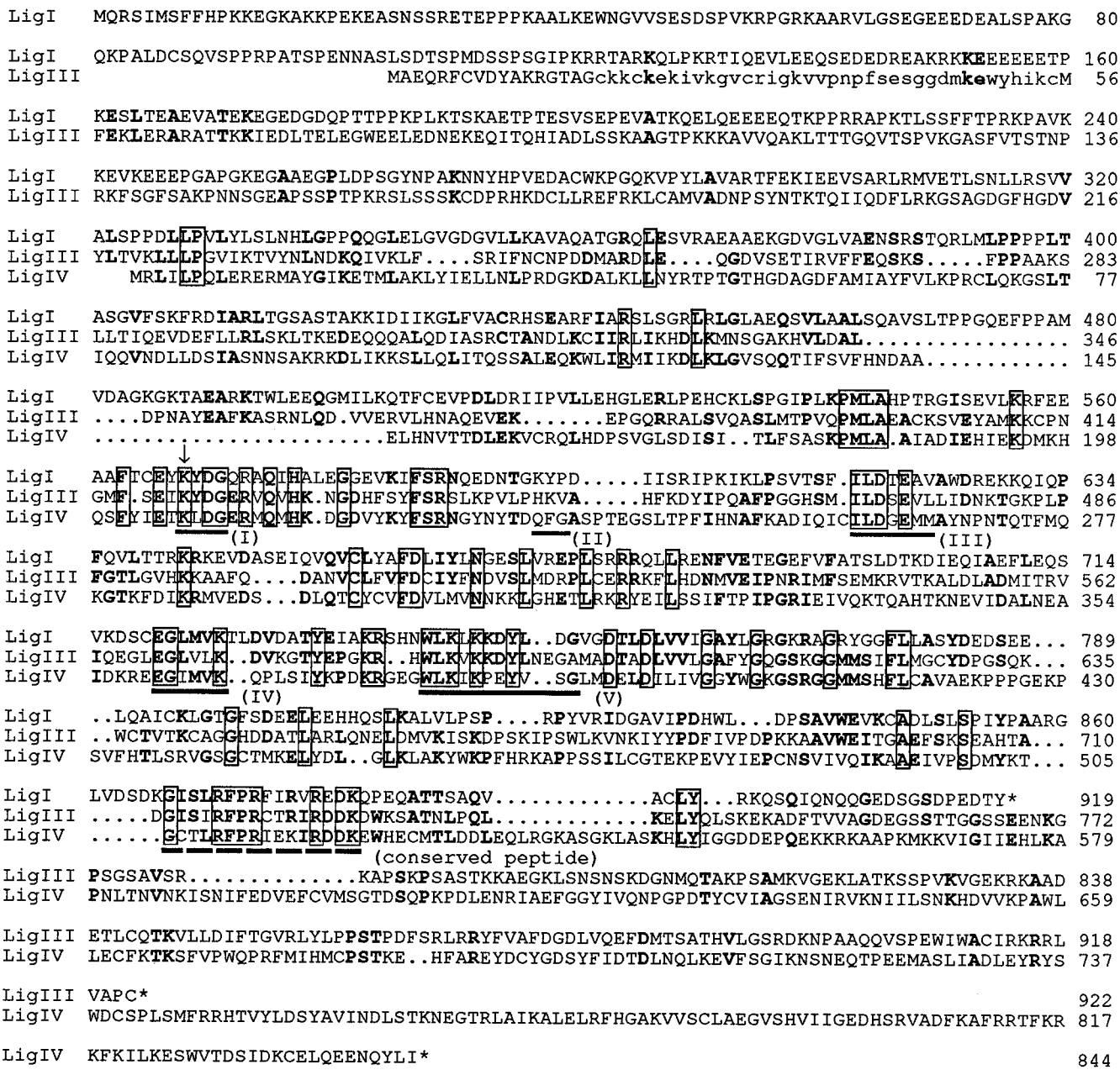


FIG. 5. Amino acid sequences encoded by DNA ligase III and IV cDNAs and alignment with human DNA ligase I. The predicted amino acid sequences of DNA ligase III (LigIII) and DNA ligase IV (LigIV) were aligned with that of DNA ligase I (LigI) by using the Pileup and Bestfit programs in the Genetics Computer Group package, version 7 (9a). Amino acid residues that are identical in at least two of the three sequences at optimal alignment are indicated in boldface type; residues conserved in all three sequences are boxed. Solid lines denote peptides with partial homology to five motifs (I to V) that are conserved between ATP-dependent DNA ligases and RNA capping enzymes (37), motif I corresponds to the DNA ligase active site, and the position of the reactive lysine residue is marked by a vertical arrow. The broken line denotes the conserved peptide found in ATP-dependent DNA ligases that was used to identify DNA ligase-specific ESTs. Lowercase letters indicate the putative zinc finger at residues 18 to 55 in the sequence encoded by DNA ligase III cDNA.

gase III or IV; this is shown schematically in Fig. 6. The N-terminal 112 amino acids of the DNA ligase III cDNA show approximately 30% identity (50 to 60% similarity) to residues 3 to 107, and also residues 108 to 217, of human poly(ADP-ribose) polymerase (PARP) (6). The same two regions of identity are also found with the murine, bovine, avian, and *Drosophila* PARP homologs and contain two evolutionarily conserved zinc fingers within the DNA-binding domain of PARP. The position of the putative zinc finger in the sequence encoded by the open reading frame of the DNA ligase III cDNA is indicated in Fig. 5; alignment with the two zinc fingers of PARP is shown in Fig. 7.

Within the region of partial sequence homology among DNA ligases I, III, and IV, the homology occurs in discrete short stretches of amino acid residues. These largely overlap with five motifs previously shown to be conserved, in the same order and with nearly identical spacing, between ATP-dependent DNA ligases and yeast and viral RNA capping enzymes (37). These conserved regions (I to V) are indicated in Fig. 5; the active-site consensus corresponds to region I. The highly

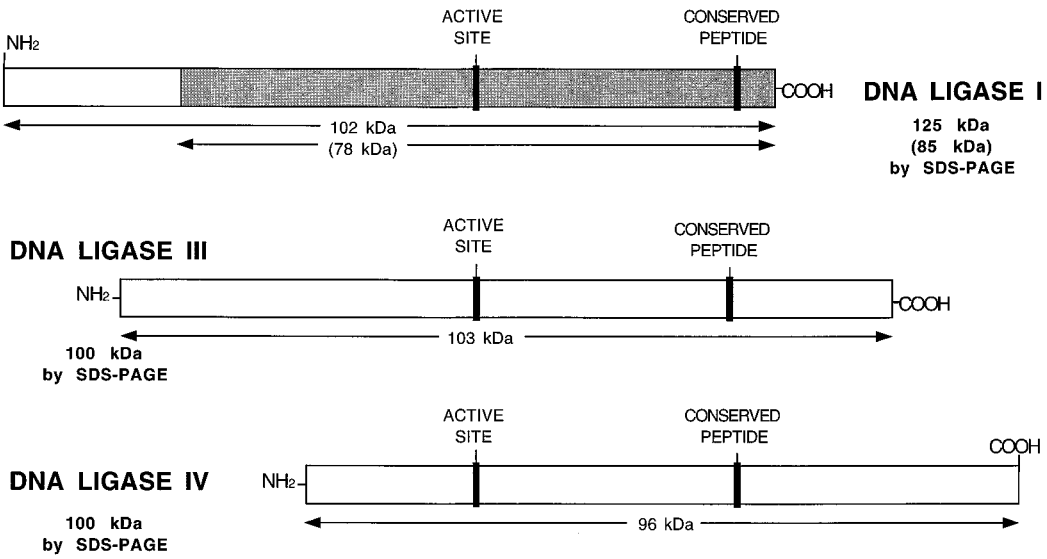


FIG. 6. Schematic representation of DNA ligases I, III, and IV aligned at their active sites. The active-site motif and the conserved peptide (as indicated in Fig. 5) are indicated by solid bars. The shaded box represents the C-terminal catalytic domain of DNA ligase I (40). The predicted molecular masses of the polypeptides encoded by the respective open reading frames are indicated below the boxes, while figures to the side in boldface type give the apparent molecular masses of the proteins as estimated by SDS-PAGE; figures in parentheses refer to the catalytic fragment of DNA ligase I. The three full-length enzymes are of similar size, but DNA ligase I migrates anomalously slowly during SDS-PAGE because of its hydrophilic N-terminal region (2).

conserved motif flanking the 3′ boundary of the region of homology between DNA ligases I, III, and IV is unique to ATP-dependent DNA ligases and is not found in the RNA capping enzymes. Similarly to vaccinia virus DNA ligase, human DNA ligases III and IV do not contain the region II motif which is present in the capping enzymes, *S. pombe* and *S. cerevisiae* DNA ligases, and murine and human DNA ligase I (36, 37).

One of the most striking features to emerge from these sequence comparisons is the near identity of peptides within the predicted amino acid sequence encoded by the DNA ligase III cDNA with sequenced tryptic peptides from the 70-kDa bovine DNA ligase II protein (44). These tryptic peptides span the region between the active site and conserved DNA ligase-specific motif and are also highly homologous (65% identity) to the corresponding region of the vaccinia virus DNA ligase (44). At optimum alignment, the entire predicted amino acid sequence of the 63-kDa vaccinia virus DNA ligase shows 55% homology with amino acids 173 to 744 of the product of the DNA ligase III open reading frame. The sequence 411-(K) CPNGMFSEIKYDGERVQVH(K)-431 encoded by the DNA ligase III cDNA, with Lys-421 the putative active lysine, is identical to the active-site tryptic peptide identified in the purified bovine DNA ligase II protein (44) and different from that of DNA ligase I (42).

Human DNA ligases III and IV are transcribed from unique genes. The cDNA clone obtained for DNA ligase III identified a single mRNA of approximately 4.2 kb by Northern hybridization analysis (Fig. 8). Similarly, the ~4-kb DNA ligase IV cDNA identified a unique mRNA of corresponding size (Fig. 9). Analysis of the distribution in tissue of the DNA ligase III and IV transcripts indicated that both are low-abundance mRNAs which are most easily detected in testis, thymus, and prostate tissues. Both messages are also seen in heart tissue (Fig. 8 and 9). As defects in human DNA repair enzymes have been implicated in genetic disease (25), it was important to determine the precise chromosomal location of the human genes encoding the unique DNA ligase III and IV transcripts. For DNA ligase III, the cDNA clone was used for FISH to human chromosome metaphase spreads (17). Approximately 20 spreads were analyzed, and most gave a doublet signal characteristic of genuine hybridization on at least one chromosome 17. A doublet signal was not detected on any other chromosome. Detailed analysis of 19 individual chromosomes, by a combination of fractional length measurements and fluorescence banding, combined with high-resolution image analysis indicated that human DNA ligase III is located within bands 17q11.2 to 17q12 (Fig. 10). Genomic DNA fragments encoding part of the DNA ligase IV gene were cloned and used for mapping of this gene. Approximately 30 spreads were analyzed and gave a doublet signal, while such a signal was not detected on any other chromosome. Detailed analysis of 22 individual spreads (as described above) indicated that the human DNA ligase IV gene is positioned within bands 13q33 to 13q34 (Fig. 11).

DISCUSSION

A human cDNA clone that encodes DNA ligase III has been isolated. DNA ligase III differs from other DNA ligases in its specific interaction with the XRCC1 protein (5, 20) that is involved in the repair of DNA single-strand breaks (39). Such

PARP zinc finger 1:	21	CKKCSESIPKDSLRMAIMVQSPMFD--GKVPHWYHFSC	56
DNA LIGASE III:	18	CKKCKEKIVKGVCRIGKVVPNPFSESGGDMKEWYHIKC	55
PARP zinc finger 2:	125	CKGCMEKIEKGQVRLSKKMVDPEKPQLGMIDRWYHPGC	162

FIG. 7. Alignment of the putative zinc finger in the sequence encoded by DNA ligase III cDNA with the two zinc fingers in human PARP. Amino acid residues are shown in single-letter code; zinc-coordinating residues are shown in outline typeface. Two gaps have been introduced in the first PARP finger.

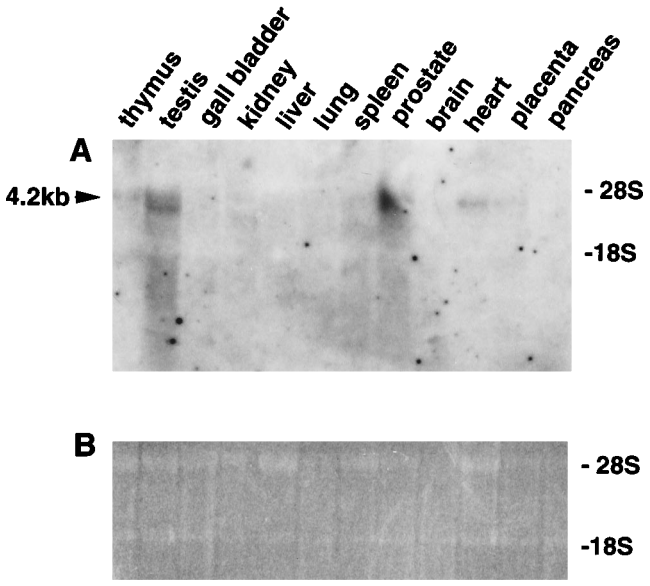


FIG. 8. Distribution of human DNA ligase III mRNA in tissues. A 15-μg amount of total RNA was electrophoresed on a 1% agarose–formaldehyde gel and analyzed by Northern hybridization with a radiolabeled partial cDNA probe. (A) Autoradiograph of the Northern hybridization blot showing transcription of the gene in 12 human tissues. (B) Ethidium bromide staining of the gel as a control for equivalent loading of the different tissue samples. The positions of the 28S and 18S rRNA bands are indicated.

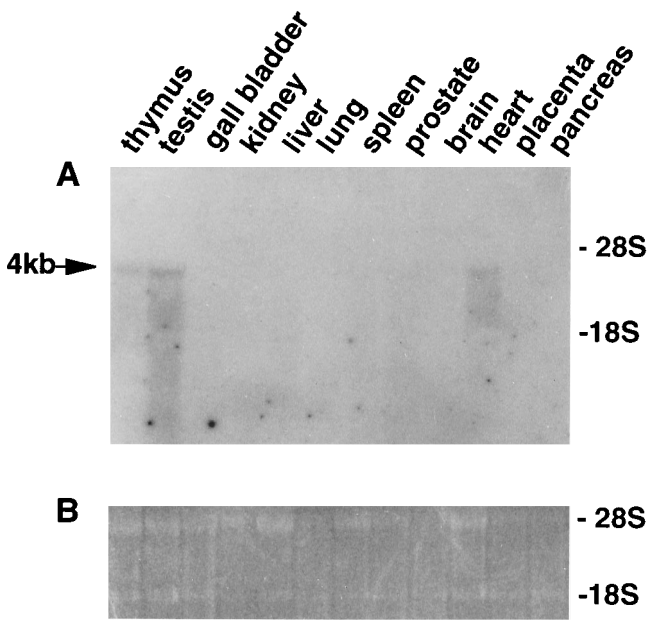


FIG. 9. Distribution of human DNA ligase IV mRNA in tissues. Experimental details and panels are as described for Fig. 8.

protein-protein interactions may confer functional individuality on DNA ligases and explain the requirement for several different enzymes of this type in cell nuclei. In agreement with previous data (5), in vitro-translated DNA ligase III could be recovered by affinity chromatography on nickel-agarose following addition of XRCC1-His protein to the reticulocyte lysate and also could be detected by far Western blotting with an XRCC1 probe. Intriguingly, the C termini of DNA ligase III and XRCC1 share weak amino acid sequence homology. More striking is the presence of a putative zinc finger near the N terminus of DNA ligase III (Fig. 5 and 7) and homology of this region with both zinc fingers in the DNA-binding domain of another protein acting at strand breaks, PARP (6). Mutation analysis has indicated that activation of poly(ADP-ribose) synthesis by binding of PARP to DNA single-strand breaks requires both zinc fingers, while the zinc finger closer to the N terminus of PARP is necessary and sufficient for activation by double-strand breaks (12). The interaction of DNA ligase III with XRCC1 and its activity in DNA repair can be investigated in more detail now that both cDNAs are cloned (39). Meanwhile, the relatively high levels of DNA ligase III transcript observed in thymus and testis tissues would support a proposed role for DNA ligase III in recombination processes (13).

Comparison of the electrophoretic properties of partial or complete tryptic peptides containing the active sites of DNA ligases I, II, and III indicated (i) that the active sites of DNA ligases II and III were identical to each other and different from that of DNA ligase I and (ii) sequence similarity or identity in an extended region flanking the active sites of DNA ligases II and III (33). Thus, the near identity of amino acid residues 421 to 723 of the 100-kDa polypeptide encoded by the human HGS473238 cDNA with tryptic peptides from the 70-kDa bovine DNA ligase II protein (44) is consistent with HGS473238 encoding DNA ligase III. Despite this region of extensive amino acid sequence identity, Northern hybridization analysis with a cDNA probe containing this portion of the DNA ligase III cDNA did not identify an additional related transcript. This indicates that a prospective separate mRNA encoding DNA ligase II in human cells would be of either lower abundance and/or divergent nucleotide sequence compared with DNA ligase III, such that it would not be detected under the hybridization conditions employed here. Similarly, FISH mapping of the human DNA ligase III gene with a complete cDNA probe gave a single signal. These data indicate that DNA ligase III is encoded by a unique gene, producing a single transcript. The N terminus of the DNA ligase II protein is blocked to Edman degradation, indicating that it is a primary translation product (41, 44). However, no putative DNA ligase II cDNA was identified either in the initial search for DNA ligase-related ESTs or during isolation of a full-length DNA ligase III cDNA. The relationship between DNA ligase III and DNA ligase II remains unclear, but it could reflect alternative RNA splicing, alternative translational start codons in the same mRNA, or two closely related genes (33). The high degree of homology of both these mammalian enzymes to the 63-kDa vaccinia virus DNA ligase suggests recruitment of one of these host DNA ligases by the virus.

We have identified a fourth DNA ligase in human cells. This enzyme has the same electrophoretic mobility as the 100-kDa DNA ligase III, and the two proteins copurify initially through several column chromatography steps. We have obtained mass spectrometry data for several peptides from this fourth mammalian DNA ligase, which we call DNA ligase IV. Furthermore, a human cDNA clone that encodes a functional DNA ligase identified as DNA ligase IV by unique agreement of the mass spectrometry data for tryptic peptides from the purified human enzyme with the open reading frame of the cDNA has been isolated. DNA ligase IV shows partial peptide sequence homology with other ATP-dependent DNA ligases, including a consensus active site. An unusual feature of DNA ligase IV, in comparison with all other ATP-dependent DNA ligases, is an extended C-terminal region of 323 amino acid residues that shows no detectable homology to other proteins in the data-

FIG. 10. FISH mapping of the human DNA ligase III gene. A cDNA clone representing the human DNA ligase III gene was hybridized to normal human male chromosomes. (A) Chromosome spread from a single cell showing hybridization to the q arm of each chromosome 17 (arrows). (B) Representative example of a single DAPI-stained (blue) chromosome containing the human DNA ligase III signal photographed by using normal contrast, showing the outline of the chromosome (left), or high contrast, to accentuate the DAPI banding pattern (right). Most gene signals appeared to be within bands q11.2 to q12. (C) Assignment of band position based on detailed analysis of images from 12 individual chromosomes.

bases (Fig. 5 and 6). The further characterization of this novel DNA ligase and the availability of the cognate cDNA clone will allow elucidation of its role in the cell nucleus. The existence of this previously unrecognized human DNA ligase complicates preliminary analysis of crude or partly purified enzyme fractions by SDS-PAGE, because an observed 100-kDa ligase-adenylate complex could be due to either DNA ligase III or IV, or both.

The major parts of the DNA ligase III and IV open reading frames show clear partial homology with the catalytic domain of DNA ligase I (Fig. 5 and 6) and contain the short sequence motifs common to ATP-dependent DNA ligases and RNA capping enzymes. The two groups of enzymes exhibit similar mechanisms of covalent catalysis (19, 37), and the motifs were shown to be essential for catalytic function in nucleotidyl transfer from GTP to the 5′ end of an RNA transcript via an enzyme-GMP reaction intermediate (37), suggesting that the general strategy for interaction with the nucleic acid substrate has been retained. The highly conserved motif that was initially used to identify DNA ligase III and IV cDNAs is unique to ATP-dependent DNA ligases and is not found in the RNA capping enzymes, suggesting that it is required specifically for the joining of nicked DNA.

Most DNA replication and repair enzymes in mammalian cell nuclei, including DNA polymerases α, β, δ, and ε, have direct counterparts in the yeasts *S. cerevisiae* and *S. pombe*. In

FIG. 11. FISH mapping of the human DNA ligase IV gene. A genomic clone containing a portion of the human DNA ligase IV gene was hybridized to normal human male chromosomes. (A) Chromosome spread from a single cell showing hybridization to the q arm of each chromosome 13 (arrows). (B) Representative example of a single DAPI-stained (blue) chromosome containing the human DNA ligase IV signal photographed by using normal contrast, showing the outline of the chromosome (left), or high contrast, to accentuate the DAPI banding pattern (right). Most gene signals appeared to be at the border between bands q33 and q34. (C) Assignment of the band position based on detailed analysis of images from 22 individual chromosomes.

contrast, the abundant enzymes in mammalian cell nuclei that interact specifically with DNA strand interruptions, PARP and DNA-dependent protein kinase, have not been detected in yeast cells, nor has p53, which is elevated in response to DNA strand breaks (26). So far, only a single DNA ligase, which resembles mammalian DNA ligase I, has been found in either *S. cerevisiae* or *S. pombe*, and so it is surprising that a family of distinct nuclear DNA ligases occurs in human cells. Possibly, the responses to DNA strand breaks may differ markedly between lower and higher eukaryotes.

A young female with an inherited molecular defect in DNA ligase I who exhibited severe immunosuppression, stunted growth, and lymphoma, as well as cellular hypersensitivity to DNA-damaging agents, has been identified (3, 45). Cells from the patient showed defective rejoining of Okazaki fragments during DNA replication and excessive gap filling during DNA excision-repair (31). The availability of cDNAs encoding DNA ligases III and IV allows for additional studies of possible DNA ligation deficiencies in a variety of human chromosome instability syndromes. Localization of the genes encoding DNA ligases III and IV to chromosomes 17 and 13 precludes their direct association with either ataxia telangiectasia or Bloom's syndrome, for which the disease loci have been mapped to chromosomes 11 and 15, respectively (9, 10), but several variants and related rare syndromes exist that have not been mapped and for which the genetic defect is at present unknown. Few inherited diseases have been mapped to chromosome bands 13q33 to 13q34 or 17q11.2 to 17q12, although the DNA repair-deficient xeroderma pigmentosum group G has been mapped to 13q33 (35). Chromosomal abnormalities in band 13q34 occur sporadically in several types of neoplasia (24) and are often seen as secondary chromosomal defects in Burkitt's lymphoma and acute lymphocytic leukemia (4, 24). Similarly, various neoplasias are associated with chromosomal abnormalities in the 17q11-12 region, the most common being the translocation between chromosomes 15 and 17 in acute myeloid leukemia subtype M3, which disrupts the retinoic acid receptor alpha gene (7). Chromosomal abnormalities in this region are also frequently reported for both acute myeloid and lymphoblastic leukemias and are seen sporadically in several other cancers (24). As defects in several human DNA repair enzymes have been linked to malignancy, it is possible that the abnormalities seen in the 17q11-12 region and at 13q34 in neoplastic tissue may be associated with alterations in DNA ligase III or IV.

REFERENCES

1. **Adams, M. D., J. M. Kelley, J. D. Gocayne, M. Dubnick, M. H. Polymeropoulos, J. Xiao, C. R. Merril, A. Wu, B. Olde, R. F. Moreno, A. R. Kerlavage, W. R. McCombie, and J. C. Venter.** 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science **252:**1651–1656.
2. **Barnes, D. E., L. H. Johnston, K.-I. Kodama, A. E. Tomkinson, D. D. Lasko, and T. Lindahl.** 1990. Human DNA ligase I cDNA: cloning and functional expression in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **87:**6679–6683.
3. **Barnes, D. E., A. E. Tomkinson, A. R. Lehmann, A. D. B. Webster, and T. Lindahl.** 1992. Mutations in the DNA ligase I gene of an individual with immunodeficiencies and cellular hypersensitivity to DNA-damaging agents. Cell **69:**495–503.
4. **Berger, R., M. Le Coniat, J. Derre, D. Vecchione, and S. J. Chen.** 1989. Secondary nonrandom chromosomal abnormalities of band 13q34 in Burkitt lymphoma-leukemia. Genes Chromosomes Cancer **1:**115–118.

4a. **Caldecott, K.** Unpublished data.

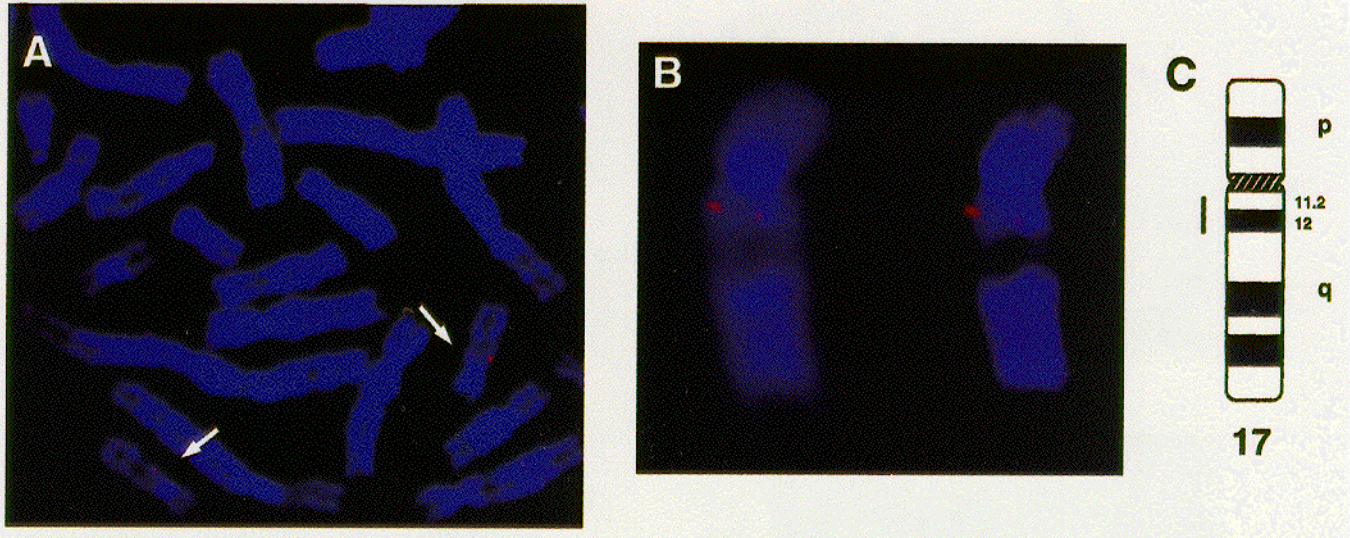
5. **Caldecott, K. W., C. K. McKeown, J. D. Tucker, S. Ljungquist, and L. H. Thompson.** 1994. An interaction between the mammalian DNA repair protein XRCC1 and DNA ligase III. Mol. Cell. Biol. **14:**68–76.
6. **de Murcia, G., and J. M. de Murcia.** 1994. Poly(ADP-ribose) polymerase: a molecular nick-sensor. Trends Biochem. Sci. **19:**172–176.
7. **de Thé, H., C. Chomienne, M. Lanotte, L. Degos, and A. Dejean.** 1990. The t(15;17) translocation of acute promyelocytic leukemia fuses the retinoic acid receptor alpha gene to a novel transcribed locus. Nature (London) **347:**558–561.
8. **Dianov, G., and T. Lindahl.** 1994. Reconstitution of the DNA base excision-repair pathway. Curr. Biol. **4:**1069–1076.
9. **Gatti, R. A., I. Berkel, E. Boder, G. Braedt, P. Charmley, P. Concannon, F. Ersoy, et al.** 1988. Localization of an ataxia-telangiectasia gene to chromosome 11q22-23. Nature (London) **336:**577–580.

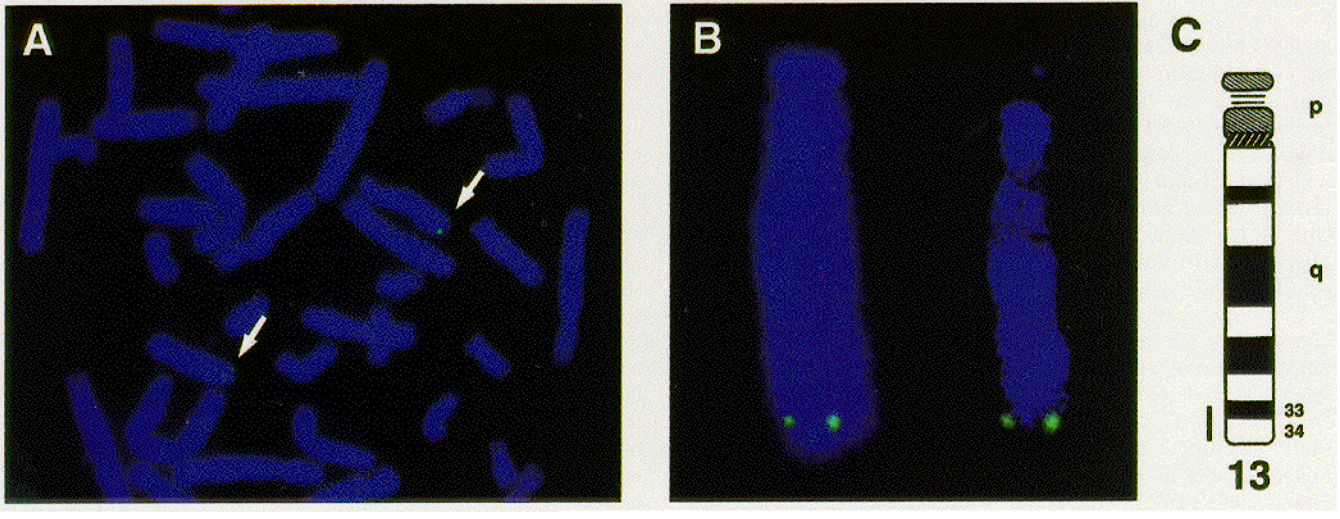
9a. **Genetics Computer Group.** 1991. Program manual of the GCG package, version 7. Genetics Computer Group, Madison, Wis.

10. **German, J., A. M. Roe, M. F. Leppert, and N. A. Ellis.** 1994. Bloom syndrome: an analysis of consanguineous families assigns the locus mutated to chromosome band 15q26.1. Proc. Natl. Acad. Sci. USA **91:**6669–6673.
11. **Higashitani, A., S. Tabata, H. Endo, and Y. Hotta.** 1990. Purification of DNA ligases from mouse testis and their behaviour during meiosis. Cell Struct. Funct. **15:**67–72.
12. **Ikejima, M., S. Noguchi, R. Yamashita, T. Ogura, T. Sugimura, D. M. Gill, and M. Miwa.** 1990. The zinc fingers of poly(ADP-ribose) polymerase are differentially required for the recognition of DNA breaks and nicks and the consequent enzyme activation. J. Biol. Chem. **265:**21907–21913.
13. **Jessberger, R., V. Podust, U. Hübscher, and P. Berg.** 1993. A mammalian protein complex that repairs double-strand breaks and deletions by recombination. J. Biol. Chem. **268:**15070–15079.
14. **Johnson, C. V., J. A. McNeil, K. C. Carter, and J. B. Lawrence.** 1991. A simple, rapid technique for precise mapping of multiple sequences in two colors using a single optical filter set. Genetic Anal. Tech. Appl. **8:**75–76.
15. **Johnson, C. V., R. H. Singer, and J. B. Lawrence.** 1991. Fluorescent detection of nuclear RNA and DNA: implications for genome organization. Methods Cell Biol. **35:**73–99.
16. **Kozak, M.** 1987. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. **15:**8125–8148.
17. **Lawrence, J. B., C. A. Villnave, and R. H. Singer.** 1988. Sensitive high-resolution chromatin and chromosome mapping in situ: presence and orientation of two closely integrated copies of EBV in a lymphoma line. Cell **52:**51–61.
18. **Li, C., J. Goodchild, and E. F. Baril.** 1994. DNA ligase I is associated with the 21 S complex of enzymes for DNA synthesis in HeLa cells. Nucleic Acids Res. **22:**632–638.
19. **Lindahl, T., and D. E. Barnes.** 1992. Mammalian DNA ligases. Annu. Rev. Biochem. **61:**251–281.
20. **Ljungquist, S., K. Kenne, L. Olsson, and M. Sandström.** 1994. Altered DNA ligase III activity in the CHO EM9 mutant. Mutat. Res. **314:**177–186.
21. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
22. **Masson, N., H. C. Hurst, and K. A. W. Lee.** 1993. Identification of proteins that interact with CREB during differentiation of F9 embryonal carcinoma cells. Nucleic Acids Res. **21:**1163–1169.
23. **Masutani, C., K. Sugasawa, J. Yanagisawa, T. Sonoyama, M. Ui, T. Enomoto, K. Takio, K. Tanaka, P. J. van der Spek, D. Bootsma, J. H. J. Hoeijmakers, and F. Hanaoka.** 1994. Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast Rad23. EMBO J. **13:**1831–1843.
24. **Mitelman, F.** 1991. Catalog of chromosome aberrations in cancer, 4th ed. Wiley Liss, New York.
25. **Modrich, P.** 1995. Mismatch repair, genetic stability and tumour avoidance. Philos. Trans. R. Soc. Lond. B **347:**89–95.
26. **Nelson, W. G., and M. B. Kastan.** 1994. DNA strand breaks: the DNA template alterations that trigger p53-dependent DNA damage response pathways. Mol. Cell. Biol. **14:**1815–1823.
27. **Noguiez, P., D. E. Barnes, H. W. Mohrenweiser, and T. Lindahl.** 1992. Structure of the human DNA ligase I gene. Nucleic Acids Res. **20:**3845–3850.
28. **Op het Veld, C. W., M. Z. Zdzienicka, H. Vrieling, P. H. M. Lohman, and A. A. van Zeeland.** 1994. Molecular analysis of ethyl methanesulfonate-induced mutations at the *hprt* gene in the ethyl methanesulfonate-sensitive Chinese hamster cell line EM-C11 and its parental line CH09. Cancer Res. **54:**3001–3006.
29. **Pappin, D. J. C., J. M. Coull, and H. Koster.** 1990. Solid phase sequence analysis of proteins electroblotted or spotted onto polyvinylidene difluoride membranes. Anal. Biochem. **187:**10–19.
30. **Pappin, D. J. C., P. Hojrup, and A. J. Bleasby.** 1993. Rapid identification of proteins by peptide-mass fingerprinting. Curr. Biol. **3:**327–332.
31. **Prigent, C., M. S. Satoh, G. Daly, D. E. Barnes, and T. Lindahl.** 1994. Aberrant DNA repair and DNA replication due to an inherited enzymatic defect in human DNA ligase I. Mol. Cell. Biol. **14:**310–317.
32. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3′ Non-coding region sequences in eukaryotic messenger RNA. Nature (London) **263:**211–214.
33. **Roberts, E., R. A. Nash, P. Robins, and T. Lindahl.** 1994. Different active sites of mammalian DNA ligases I and II. J. Biol. Chem. **269:**3789–3792.
34. **Robins, P., D. J. C. Pappin, R. D. Wood, and T. Lindahl.** 1994. Structural and functional homology between mammalian DNase IV and the 5′ nuclease

domain of *Escherichia coli* DNA polymerase I. J. Biol. Chem. **269:**28535–28538.

35. **Samec, S., T. A. Jones, J. Corlet, D. Scherly, D. Sheer, R. D. Wood, and S. G. Clarkson.** 1994. The human gene for xeroderma pigmentosum complementation group G (XPG) maps to 13q33 by fluorescence *in situ* hybridization. Genomics **21:**283–285.
36. **Savini, E., G. Biamonti, G. Ciarrocchi, and A. Montecucco.** 1994. Cloning and sequence analysis of a cDNA coding for the murine DNA ligase I enzyme. Gene **144:**253–257.
37. **Shuman, S., Y. Liu, and B. Schwer.** 1994. Covalent catalysis in nucleotidyl transfer reactions: essential motifs in *S. cerevisiae* RNA capping enzyme are conserved in *S. pombe* and viral capping enzymes and among polynucleotide ligases. Proc. Natl. Acad. Sci. USA **91:**12046–12050.
38. **Smith, G. L., Y. S. Chan, and S. M. Kerr.** 1989. Transcriptional mapping and nucleotide sequence of a vaccinia virus gene encoding a polypeptide with extensive homology to DNA ligases. Nucleic Acids Res. **17:**9051–9062.
39. **Thompson, L. H., K. W. Brookman, N. J. Jones, S. A. Allen, and A. V. Carrano.** 1990. Molecular cloning of the human *XRCC1* gene, which corrects defective DNA strand break repair and sister chromatid exchange. Mol. Cell. Biol. **10:**6160–6171.
40. **Tomkinson, A. E., D. D. Lasko, G. Daly, and T. Lindahl.** 1990. Mammalian DNA ligases: catalytic domain and size of DNA ligase I. J. Biol. Chem. **265:**12611–12617.
41. **Tomkinson, A. E., E. Roberts, G. Daly, N. F. Totty, and T. Lindahl.** 1991. Three distinct DNA ligases in mammalian cells. J. Biol. Chem. **266:**21728–21735.
42. **Tomkinson, A. E., N. F. Totty, M. Ginsburg, and T. Lindahl.** 1991. Location of the active site for enzyme-adenylate formation in DNA ligases. Proc. Natl. Acad. Sci. USA **88:**400–404.
43. **Waga, S., G. Bauer, and B. Stillman.** 1994. Reconstitution of complete SV40 DNA replication with purified replication factors. J. Biol. Chem. **269:**10923–10934.
44. **Wang, Y.-C. J., W. A. Burkhart, Z. B. Mackey, M. B. Moyer, W. Ramos, I. Husain, J. Chen, J. M. Besterman, and A. E. Tomkinson.** 1994. Mammalian DNA ligase II is highly homologous with vaccinia DNA ligase. J. Biol. Chem. **269:**31923–31928.
45. **Webster, A. D. B., D. E. Barnes, C. F. Arlett, A. R. Lehmann, and T. Lindahl.** 1992. Growth retardation and immunodeficiency in a patient with mutations in the DNA ligase I gene. Lancet **339:**1508–1509.


A
B
C
p
11.2
12
q
17


A
B
C
p
q
33
34
13