Cloning, Expression, and Chromosomal Localization of the 140-Kilodalton Subunit of Replication Factor C from Mice and Humans

BRUNO LUCKOW,[1]† FRED BUNZ,[2] BRUCE STILLMAN,[2] PETER LICHTER,[3]
AND GÜNTHER SCHÜTZ[1]*

Division Molecular Biology of the Cell I[1] and Division Organisation of Complex Genomes,[3] German Cancer Research Center, 69120 Heidelberg, Germany, and Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724[2]

We have isolated a full-length mouse cDNA encoding a lysine-rich protein of 1,131 amino acids with a calculated molecular mass of 126 kDa. The protein binds in a sequence-unspecific manner to DNA, is localized exclusively in the nucleus, and contains a putative ATP binding site and a stretch of 80 amino acids with homology to the carboxy terminus of prokaryotic DNA ligases. On the basis of the following facts, we conclude that the isolated cDNA encodes the 140-kDa subunit of mouse replication factor C (mRFC140). (i) The sequence around the ATP binding site shows significant homology to three small subunits of human replication factor C. (ii) Polyclonal antibodies raised against the protein encoded by this cDNA cross-react with the 140-kDa subunit of purified human replication factor C (hRFC140) and recognize in mouse cell extracts an authentic protein with an apparent molecular mass of 130 kDa. (iii) Sequence comparison with a human cDNA isolated by using tryptic peptide sequence information from purified hRFC140 revealed 83% identity of the encoded proteins. The mRFC140 gene is ubiquitously expressed, and two mRNAs approximately 5.0 and 4.5 kb long have been detected. The gene was mapped by in situ hybridization to mouse chromosome 5, and its human homolog was mapped to chromosome 4 (p13-p14).

During the last several years, considerable progress has been made in the understanding of DNA replication in eukaryotes as a result of the development of cell-free in vitro replication systems, which allowed the identification and characterization of proteins involved in this fundamental process (see references 17, 21, 30, 46, and 49 for reviews). Most researchers have used the simian virus 40 in vitro replication system (26). It is ideally suited for studies of eukaryotic DNA replication, because all necessary factors except the virus-encoded large T antigen are supplied by the host cell. Fractionation of cellular extracts has allowed the identification of several cellular factors which are involved in simian virus 40 replication and which are also believed to be essential for the replication of cellular DNA (references 51 and 57 and references therein). Among these is an accessory protein called replication factor C (RFC), which is required for the elongation of primed templates by DNA polymerases δ and ε (24, 52, 58). RFC (also called activator 1) is a multisubunit protein complex which has been purified from human cells (24, 52), calf thymus (40), and the yeast *Saccharomyces cerevisiae* (12, 58). Purified RFC preparations from HeLa cells contain five discrete subunits with apparent molecular masses of 145, 40, 38, 37, and 36.5 kDa (24). The complex has an intrinsic DNA-dependent ATPase activity, which can be further stimulated by proliferating cell nuclear antigen and replication protein A (24, 53). RFC binds specifically to the primer-template junction (24, 53, 54), and this structure-specific DNA binding activity has been localized by UV cross-linking studies to the 140-kDa subunit of RFC

(54), though a weaker DNA binding activity has been reported to be present in the 37-kDa subunit (39). ATP is bound by the 40-kDa subunit (54). RFC corresponds functionally to the *Escherichia coli* polymerase III γ complex and the phage T4 genes 44 and 62, which encode accessory proteins required for the formation of the DNA polymerase holoenzyme (6, 53). When the protein sequence of the 40-kDa subunit of RFC became available, a significant sequence homology to the T4 gp44 protein could be demonstrated (6). The homology with the τ and γ subunits of *E. coli* polymerase III was weaker (37). So far, only cDNAs encoding the four small subunits of human RFC (hRFC36, hRFC37, hRFC38, and hRFC40 [6, 7, 37]) have been described.

In this report, we describe the fortuitous isolation of a cDNA clone encoding the 140-kDa subunit of mouse RFC (mRFC140) while screening a λgt11 expression library with an oligonucleotide representing a liver cell-specific element of the cyclic AMP-inducible enhancer of the tyrosine aminotransferase gene (3). The identity of the isolated clone was established by the following criteria: (i) local sequence homology to three small subunits of human RFC; (ii) cross-reactivity of antibodies raised against the protein encoded by this cDNA with the 140-kDa subunit of purified hRFC140; and (iii) overall sequence homology to a cDNA clone encoding hRFC140. A full-length cDNA clone from mice and partial cDNA and genomic clones from humans were isolated, and the sequences were determined. The expression pattern of the mRFC140 gene was analyzed by Northern (RNA) blot and RNase protection experiments. Part of the mouse protein was expressed in *E. coli* and used to raise antibodies, which were used in Western blot (immunoblot) and indirect immunofluorescence experiments. Furthermore, the chromosomal localizations of the mouse and human RFC140 genes were determined.

* Corresponding author. Mailing address: German Cancer Research Center, Division Molecular Biology of the Cell I (0115), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. Phone: 49/6221/423411. Fax: 49/6221/423404.
† Present address: Medizinische Poliklinik, Arbeitsgruppe Klinische Biochemie, 80336 Munich, Germany.

MATERIALS AND METHODS

Construction of cDNA expression libraries. Five micrograms of mouse liver poly(A)$^+$ RNA primed with 0.74 μg of p(dN)$_6$ was converted to double-stranded cDNA by using the You-Prime cDNA synthesis kit (Pharmacia). The cDNA was size fractionated on a 1% agarose gel, isolated by isotachophoresis (33), and ligated into λgt11 (18). In vitro packaging reactions were performed with Gigapack II Gold (Stratagene). A second mouse liver cDNA expression library was prepared by the method of Gubler and Hoffman (14). Seven micrograms of poly(A)$^+$ RNA was reverse transcribed into first-strand cDNA, using 7 μg of p(dT)$_{12-18}$ and 1,400 U of Superscript (Bethesda Research Laboratories). The second strand of the cDNA was synthesized with premade solutions from a cDNA synthesis kit (Pharmacia). The cDNA was size fractionated on a Bio-Gel A50m column (18) and ligated into λgt11. The random and oligo(dT)-primed cDNA libraries contained 5 × 10^5 and 7.5 × 10^5 primary recombinant phages.

Screening of cDNA expression libraries with DNA binding sites and isolation of additional mRFC140 cDNA clones. The two mouse liver cDNA expression libraries were plated on *E. coli* Y1090 at low density (10^5 PFU/24- by 24-cm dish [NunC]) and screened as described by Vinson et al. (55), with the following modifications. As the probe, we used a DNA fragment containing five copies of oligonucleotide BIII (3), the sequence of which is derived from the enhancer at −3.6 kb in the rat *tat* gene. This fragment was cloned into the *Bam*HI site of pBLCAT2 (32), and the resulting plasmid was termed p5HBV21A. To generate a probe for screening, 30 μg (10 pmol) of p5HBV21A DNA was irreversibly denatured by NaOH. After neutralization, 20 pmol of oligonucleotide tk (5′-CGTGTTCGAATTCGCCA-3′) was annealed adjacent to the *Bam*HI site of p5HBV21A and extended with Klenow polymerase, dCTP, dGTP, dTTP, and 500 μCi of [α-^{32}P]dATP (3,000 Ci/mmol). After a chase with cold dATP, the labeled probe fragment was excised from the plasmid with *Bam*HI and *Xba*I, separated from the vector on a 5% native polyacrylamide gel, and isolated by isotachophoresis. Nitrocellulose replica filters (BA85; Schleicher & Schüll) were prepared as described previously (55), blocked for 2 h, and incubated with the radiolabeled probe and poly(dI-dC) (5 μg/ml) overnight on a shaking platform at 4°C. Five filters (22 by 22 cm) were incubated simultaneously in one large dish (24 by 24 cm; Nunc) in 250 ml of binding buffer. To isolate additional mRFC140 cDNA clones, a ^{32}P-labeled RNA probe was generated from pmTCTTT14 as described previously (34) and used to rescreen the two cDNA libraries described above. Hybridization and washing conditions were as described by Church and Gilbert (9). The sequence of oligonucleotide BIII is

5′-GATCTGCTGCTCTTTGATCTG-3′
3′-ACGACGAGAAACTAGACCTAG-5′

Isolation of hRFC140 cDNA and genomic clones. Approximately 3 × 10^6 PFU of a human liver cDNA library in λZAPII (Stratagene), kindly provided by S. Ruppert, was screened with the 2.2-kb *Not*I-*Eco*RI fragment from pmTCTTT2. The probe was labeled with ^{32}P by the method of Feinberg and Vogelstein (11), and hybridizations were performed under high-stringency conditions as described above for mRFC140. Four independent clones, designated phTCTTT15, -21, -24, and -25, were isolated. Approximately 1.25 × 10^6 PFU of a human genomic library in λEMBL3, kindly provided by R. Leube (Heidelberg, Germany), was screened with radiolabeled RNA probes generated from phTCTTT24 and phTCTTT25. Generation of the probe and hybridization and washing conditions were as de-

scribed above for mRFC140 cDNA clones. Three independent clones were isolated. DNA from phage λhgTCTTT6-3 was sequenced directly with a CircumVent sequencing kit (New England Biolabs) in combination with an hRFC140-specific primer.

DNA sequencing and computer analyses. All mRFC140 cDNA phage inserts were subcloned as *Not*I fragments into pBluescriptM13+ (Stratagene). Sequencing reactions were performed with alkaline-denatured double-stranded plasmid DNA as described previously (33, 48). All computer analyses of DNA and protein sequences were performed with HUSAR, a program package based on the Genetics Computer Group programs from the University of Wisconsin (10).

Isolation of RNA, Northern blot analysis, and RNase protection experiments. Total RNA was isolated by the method of Chomczynski and Sacchi (8) from various mouse tissues, which had been frozen in liquid nitrogen and stored at −70°C. Poly(A)$^+$ RNA was isolated with oligo(dT)-cellulose (Boehringer) as described previously (42). Northern blotting was performed as described previously (2). As the probe, we used a radiolabeled antisense RNA transcribed by T7 RNA polymerase from pmTCTTT8 cleaved with *Hin*dIII. RNase protection experiments were performed essentially as described elsewhere (1), using samples of 50 μg of total RNA and a hybridization temperature of 54°C. As the probe, we used a 209-nucleotide (nt)-long radiolabeled antisense RNA transcribed by T7 RNA polymerase from pmTCTTT8 digested with *Bam*HI. The protected fragment had a length of 175 bp.

Overexpression of mRFC140 protein in *E. coli.* Plasmid pRSET14H was transformed into *E. coli* INVαF′ (Invitrogen). An overnight liquid culture was diluted 1:100 with fresh LB medium and further incubated until the optical density at 600 nm reached 0.3. Isopropythiogalactopyranoside (IPTG) was added to 1 mM, and the culture was shaken for another 60 min at 37°C. At this point, the cells were infected with M13/T7 (multiplicity of infection of 7.5), a recombinant M13 phage carrying a copy of T7 RNA polymerase (Invitrogen). After another incubation for 3 h, the bacteria were harvested and the protein was isolated. The overexpressed protein formed inclusion bodies, which were isolated essentially as described previously (35).

Immunization of rabbits. *E. coli*-expressed pRSET14H protein was bound to a Ni^{2+}-nitrilotriacetic acid agarose column (Diagen) under denaturing conditions (8 M urea) and eluted with a buffer containing 8 M urea and 30 to 100 mM imidazole (20). For immunization, 500 μg of pure protein was dissolved in 250 μl of 8 M urea and mixed with 250 μl of Freund's complete adjuvant (Sigma). The emulsion was injected in several aliquots subcutaneously into a rabbit. After 5 weeks, the animal received a second injection of 250 μg of protein in 250 μl of 8 M urea emulsified with 250 μl of Freund's incomplete adjuvant (Sigma). Blood was taken from the ear vein 8 days after the second injection. Serum preparation and affinity purification of antibodies were performed as described previously (15, 45). Altogether, four rabbits were immunized.

Extract preparation, Western blotting, and indirect immunofluorescence. Whole cell protein extracts from mouse NIH 3T3 fibroblasts obtained by using a high-salt lysis buffer were prepared as described previously (15). Human RFC protein was purified from 293 cells essentially as described previously (52). Proteins were separated on either 6 or 15% polyacrylamide-sodium dodecyl sulfate (SDS) gels and transferred to nitrocellulose filters (BA85; Schleicher & Schüll), using a semidry blotting apparatus (42). Blocking of filters, incubation with primary and secondary antibodies, and washing steps were performed essentially as described previously (15). For West-

ern blots with mouse whole cell extracts, rabbit preimmune or anti-mRFC140 immune serum was diluted 1:3,000 and the anti-rabbit immunoglobulin G (IgG) alkaline phosphatase conjugate (Promega) was diluted 1:7,500. Filter strips with purified human RFC protein were incubated either with rabbit preimmune or anti-mRFC140 immune serum (diluted 1:2,000) or the monoclonal antibody RFC6 raised against hRFC140 (1 ml of hybridoma supernatant). Detection of immune complexes via enhanced chemiluminescence using species-specific secondary antibodies conjugated with horseradish peroxidase was performed as instructed by the supplier (Amersham). Indirect immunofluorescence was performed essentially as described previously (15), using a 1:50 dilution of rabbit preimmune serum or affinity-purified immune serum. Detection was performed with a 1:200 dilution of an anti-rabbit IgG Texas red conjugate.

Plasmid constructions. The full-length cDNA clone pmTCTTTV was constructed by combining the 5′ half of the cDNA present in pmTCTTT2 with the 3′ half present in pmTCTTT11 via the unique NdeI site at position 2773. To generate pRSET14H, the 2.65-kb HindIII fragment from pmTCTTT14, which encodes 753 amino acids from the carboxy terminus of mRFC140, was ligated into pRSETB (Invitrogen) cut with HindIII. This vector contains a T7 RNA polymerase promoter and allows the construction of fusion proteins with a polyhistidine stretch at the amino terminus. These features guarantee high-level expression in suitable E. coli strains (47) and a simple one-step purification of the recombinant protein (16).

In situ hybridization. DNA probes were labeled by nick translation using biotin- or digoxigenin-modified nucleotides (27). Mouse metaphase spreads were prepared from the spleen of a female BALB/c animal, and human chromosomes were prepared from peripheral blood lymphocytes, using standard procedures of cell stimulation, mitotic arrest, hypotonic treatment, and methanol-acetic acid fixation. In situ hybridization was carried out as described in detail elsewhere (28). Hybridized probe was detected via fluorescein isothiocyanate (FITC) or rhodamine, and chromosomes were banded by 4;6-diamidino-2-phenylindole (DAPI). Digitized images were taken separately for DAPI, FITC, and rhodamine by using a cooled charge-coupled device (CCD) camera. To assess the chromosomal localization of the hybridized probe, images were carefully realigned and DAPI images were overlaid electronically with the FITC or rhodamine images.

Nucleotide sequence accession number. The sequence reported in this publication has been submitted to the EMBL nucleotide sequence data base (accession number X72711 [MMREPCFC]).

RESULTS

Isolation of mRFC140 cDNA clones. Two mouse liver cDNA expression libraries primed either by random hexamers or oligo(dT) were screened with a ^{32}P-labeled DNA fragment by the method of Vinson et al. (55). The DNA fragment used for screening contained five copies of a liver cell-specific element, which was derived from the enhancer located 3.6 kb upstream from the transcription start site in the rat tyrosine aminotransferase gene (3). This sequence has been identified in the meantime as an HNF4 binding site (36). Approximately 1.25 × 10⁶ primary recombinant phages were screened. A single clone with a 3.1-kb cDNA insert (pmTCTTT14; Fig. 1A) was isolated. This cDNA was used as a hybridization probe to rescreen the two cDNA expression libraries. Ten additional cDNA clones ranging in size from 1.2 to 4.0 kb were isolated
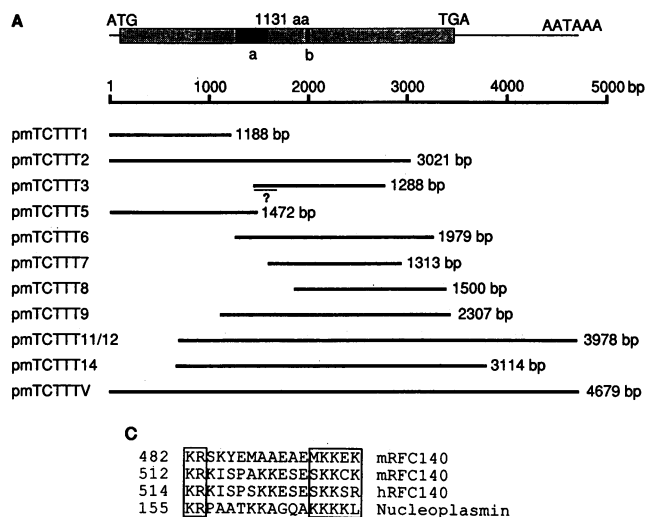

FIG. 1. Nucleotide sequence and deduced amino acid sequence of mRFC140 cDNA. (A) Diagram indicating lengths and relative positions of isolated cDNA clones together with a schematic representation of the full-length cDNA. The coding region is symbolized by a box; 5′ and 3′ untranslated regions are represented by thin lines. Sections marked a and b indicate the positions of the DNA ligase homology and the ATP binding site, respectively. The region of pmTCTTT3 labeled with a question mark represents a 193-bp-long DNA fragment of unknown origin. aa, amino acids. (B) Nucleotide and deduced amino acid sequences of the full-length mRFC140 cDNA. Polyadenylation signals are present at positions 3916 and 4662. (C) Comparison of bipartite nuclear localization signals present in mRFC140, hRFC140, and Xenopus nucleoplasmin. The position of the first amino acid of the nuclear localization signals is indicated. The two basic domains are boxed.

(Fig. 1A). With the exception of pmTCTTT11 and -12, all clones were isolated from the random-primed cDNA library.

Nucleotide sequence of mRFC140 cDNA. All cDNA inserts were subcloned into plasmid vectors, and the complete nucleotide sequence of pmTCTTT14 was determined. Since this clone did not represent a full-length cDNA (Fig. 1A), suitable portions from pmTCTTT2 and pmTCTTT11 were also sequenced. Altogether, the cDNA clones span 4,679 bp. The nucleotide and deduced amino acid sequences are shown in Fig. 1B. The first ATG codon at position 3, which is followed by a stop codon at position 21, is located too close to the 5′ end of the mRNA to be recognized by the ribosome (22). Furthermore, this ATG is present only in pmTCTTT2. cDNA clones pmTCTTT1 and pmTCTTT5 start at positions 4 and 25, respectively. The second ATG at position 80, which is not preceded by an in-frame stop codon, starts an open reading frame of 3,393 bp and is flanked by sequences that fit well to the optimal translation start signal (22). The 3′ untranslated region stretches 1,204 bp and contains two AATAAA polyadenylation signals (positions 3916 and 4662). A full-length cDNA clone designated pmTCTTTV was constructed by combining 5′ sequences from pmTCTTT2 with 3′ sequences from pmTCTTT11 (Fig. 1A). A computer search of the EMBL and GenBank data bases with the full-length cDNA sequence revealed significant homology to a 4.5-kb-long human cDNA encoding a novel DNA-binding protein of unknown function (31). A protein sequence comparison revealed that this cDNA encodes the human homolog of the large subunit of mouse RFC (5).

Amino acid sequence of mRFC140 protein. The open read-

B

```
   1  CAATGGATCCCGAGCCCCGATAACGAATTACCCCAGCCACGGGCCTGCCCACCTGCTCCTGTTCGGTCGTGGGGCTGCGATGGACATTCGGAAATTCTTTGGGGTTATATCAAGTGGAAA  120
   1                                                                        M  D  I  R  K  F  F  G  V  I  S  S  G  K   14

 121  AAAGCCTGTAAACGAGACAGTAAAGAATGAGAAGACAAAAGCTTCGGAGGGAACTGTCAAAGGAAAAAAGGGAGTCAAGGAGGCCAAGGTCAATAACTCCGGTAAAGAGGACGCCTCCAA  240
  15   K  P  V  N  E  T  V  K  N  E  K  T  K  A  S  E  G  T  V  K  G  K  K  G  V  K  E  A  K  V  N  N  S  G  K  E  D  A  S  K   54

 241  ACCAAAGCAGCACAGCAAGAAGAAGAGGATCATCTATGACTCAGACTCAGAGTCAGAGGAGACAGTACAGGTAAAAAATGCTAAAAAGAAATCAGAAAAATTGTCACTGTCTTATAAACC  360
  55   P  K  Q  H  S  K  K  K  R  I  I  Y  D  S  D  S  E  S  E  E  T  V  Q  V  K  N  A  K  K  K  S  E  K  L  S  L  S  Y  K  P   94

 361  TGGTAAAGTCTCTCAGAAGGATCCTGTCACCTACGTCTCTGAGACAGATGAAGATGATGACTTTGTATGTAAAAAGGCAGCCTCCAAATCAAAAGAGAATGGAGTATCTACAAATAGTTA  480
  95   G  K  V  S  Q  K  D  P  V  T  Y  V  S  E  T  D  E  D  D  D  F  V  C  K  K  A  A  S  K  S  K  E  N  G  V  S  T  N  S  Y   134

 481  CCTTGGAACATCAAACGTGAAAAAAAATGAAGAAAACGTTAAGACTAAGAACAAGCCGTTATCTCCAATAAAACTCACACCAACGTCAGTGCTCGACTATTTTGGAACTGAAAGTGTCCA  600
 135   L  G  T  S  N  V  K  K  N  E  E  N  V  K  T  K  N  K  P  L  S  P  I  K  L  T  P  T  S  V  L  D  Y  F  G  T  E  S  V  Q   174

 601  GAGATCTGGGAAGAAGATGGTGACAAGCAAAAGGAAAGAATCTTCTCAAAACACAGAGGATTCCAGATTAAATGATGAGGCCATCGCCAAGCAGCTGCAGCTGGATGAAGATGCAGAGCT  720
 175   R  S  G  K  K  M  V  T  S  K  R  K  E  S  S  Q  N  T  E  D  S  R  L  N  D  E  A  I  A  K  Q  L  Q  L  D  E  D  A  E  L   214

 721  GGAGAGGCAGTTGCATGAAGATGAAGAATTTGCAAGAACACTGGCCTTATTGGATGAAGAACCCAAGATCAAAAAGGCCCGAAAGGATTCTGAAGAGGGAGAAGAATCATTTTCATCTGT  840
 215   E  R  Q  L  H  E  D  E  E  F  A  R  T  L  A  L  L  D  E  E  P  K  I  K  K  A  R  K  D  S  E  E  G  E  E  S  F  S  S  V   254

 841  CCAAGATGATTTAAGCAAAGCAGAAAAGCAGAAAAGCCCTAATAAAGCCGAGCTTTTCTCAACTGCAAGAAAGACCTACAGTCCTGCTAAGCATGGCAAGGGGAGGGCTTCAGAAGATGC  960
 255   Q  D  D  L  S  K  A  E  K  Q  K  S  P  N  K  A  E  L  F  S  T  A  R  K  T  Y  S  P  A  K  H  G  K  G  R  A  S  E  D  A   294

 961  TAAGCAGCCTTGCAAATCAGCTCACCGGAAGGAAGCCTGTTCCTCACCCAAGGCCAGTGCCAAACTGGCGCTTATGAAAGCAAAAGAAGAAAGTTCTTACAACGAAACAGAGCTGCTGGC  1080
 295   K  Q  P  C  K  S  A  H  R  K  E  A  C  S  S  P  K  A  S  A  K  L  A  L  M  K  A  K  E  E  S  S  Y  N  E  T  E  L  L  A   334

1081  TGCAAGAAGAAAAGAAAGCGCCACTGAACCCAAAGGAGAGAAAACAACTCCTAAGAAAACGAAAGTTTCTCCAACTAAGAGAGAGTCTGTAAGCCCAGAAGACTCTGAAAAGAAGCGCAC  1200
 335   A  R  R  K  E  S  A  T  E  P  K  G  E  K  T  T  P  K  K  T  K  V  S  P  T  K  R  E  S  V  S  P  E  D  S  E  K  K  R  T   374

1201  TAATTATCAAGCTTATCGAAGCTACTTGAACCGAGAAGGTCCCAAAGCTCTGGGCTCCAAAGAAATACCCAAGGGAGCTGAAAATTGCTTGGAAGGCCTGACGTTTGTGATCACAGGAGT  1320
 375   N  Y  Q  A  Y  R  S  Y  L  N  R  E  G  P  K  A  L  G  S  K  E  I  P  K  G  A  E  N  C  L  E  G  L  T  F  V  I  T  G  V   414

1321  GCTGGAGTCCATTGAACGAGACGAAGCCAAGTCTCTAATTGAACGTTATGGGGGGAAAGTAACAGGAAACGTGAGCAAGAAAACCAACTACCTCGTCATGGGCCGCGACAGTGGGCAGTC  1440
 415   L  E  S  I  E  R  D  E  A  K  S  L  I  E  R  Y  G  G  K  V  T  G  N  V  S  K  K  T  N  Y  L  V  M  G  R  D  S  G  Q  S   454

1441  CAAGAGTGACAAGGCAGCAGCTCTGGGAACAAAAATCCTCGATGAAGACGGCCTGTTGGATCTGATTCGAACTATGCCAGGCAAGAGATCCAAGTATGAGATGGCTGCTGAGGCTGAGAT  1560
 455   K  S  D  K  A  A  A  L  G  T  K  I  L  D  E  D  G  L  L  D  L  I  R  T  M  P  G  K  R  S  K  Y  E  M  A  A  E  A  E  M   494

1561  GAAGAAAGAAAAGTCCAAATTGGAGCGAACACCCCAAAAAAATGATCAAGGAAAAAGAAAAATTAGTCCAGCCAAGAAGGAGTCAGAGTCTAAAAAGTGCAAGCTGACTCTCCTAAAGAA  1680
 495   K  K  E  K  S  K  L  E  R  T  P  Q  K  N  D  Q  G  K  R  K  I  S  P  A  K  K  E  S  E  S  K  K  C  K  L  T  L  L  K  N   534

1681  CAGTCCTATGAAGGCAGTGAAGAAGGAGGCAAGCACGTGCCCAAGGGGCCTGGATGTCAAGGAGACACATGGCAACCGCAGTAGCAACAAAGAGGAATGTCTGCTCTGGGTGGATAAGTA  1800
 535   S  P  M  K  A  V  K  K  E  A  S  T  C  P  R  G  L  D  V  K  E  T  H  G  N  R  S  S  N  K  E  E  C  L  L  W  V  D  K  Y   574

1801  CAAGCCCGCCTCACTCAAGAACATAATCGGACAGCAGGGTGACCAGAGCTGTGCCAACAAATTACTACGTTGGCTCAGAAACTGGCACAAGAGTTCTCCTGAAGAGAAAAAACATGCAGC  1920
 575   K  P  A  S  L  K  N  I  I  G  Q  Q  G  D  Q  S  C  A  N  K  L  L  R  W  L  R  N  W  H  K  S  S  P  E  E  K  K  H  A  A   614

1921  AAAGTTTGGCAAACTTGCCAGCAAAGATGATGGCTCCAGTTTCAAGGCAGCTCTGCTGTCTGGCCCTCCAGGTGTTGGCAAAACCACTACAGCTTCTCTTGTGTGTCAGGAATTGGGCTA  2040
 615   K  F  G  K  L  A  S  K  D  D  G  S  S  F  K  A  A  L  L  S  G  P  P  G  V  G  K  T  T  T  A  S  L  V  C  Q  E  L  G  Y   654

2041  CAGTTATGTGGAACTAAATGCAAGTGATACTCGGAGTAAGAACAGTTTGAAAGCAGTTGTTGCTGAATCCTTGAACAACACCAGCATCAAAGGCTTTTATACAAGTGGAGCGGCTCCTTC  2160
 655   S  Y  V  E  L  N  A  S  D  T  R  S  K  N  S  L  K  A  V  V  A  E  S  L  N  N  T  S  I  K  G  F  Y  T  S  G  A  A  P  S   694

2161  CGTGAGTGCGAGGCACGCCCTCATCATGGATGAGGTCGATGGCATGGCAGGCAATGAAGACAGGGGTGGAATTCAGGAGCTAATAGGCCTGATAAAGCATACTAAAATTCCCATCATTTG  2280
 695   V  S  A  R  H  A  L  I  M  D  E  V  D  G  M  A  G  N  E  D  R  G  G  I  Q  E  L  I  G  L  I  K  H  T  K  I  P  I  I  C   734

2281  TATGTGCAATGATGAAACCACCCCAAAATTCGCTCATTGGTTCATTATTGTTTTGATCTTCGTTTCCAAAGACCTCGAGTTGAACAGATTAAGAGTGCAATGCTGTCTATTGCATTTAA  2400
 735   M  C  N  D  R  N  H  P  K  I  R  S  L  V  H  Y  C  F  D  L  R  F  Q  R  P  R  V  E  Q  I  K  S  A  M  L  S  I  A  F  K   774

2401  AGAGGGTTTAAAAATCCCCCCTCCAGCTATGAATGAAATAATTTTGGGAGCTAATCAAGATGTCAGACAGGTTTTACACAATCTCAGTATGTGGTGTGCACAAAGTAAGGCCCTGACCTA  2520
 775   E  G  L  K  I  P  P  P  A  M  N  E  I  I  L  G  A  N  Q  D  V  R  Q  V  L  H  N  L  S  M  W  C  A  Q  S  K  A  L  T  Y   814

2521  TGACCAAGCCAAGGCTGATTCTCAGAGGGCTAAGAAGGATATCAGACTGGGCCCATTTGATGTCACTCGGAAAGTGTTTGCAGCTGGAGAGGAGACTGCACACATGTCACTGATGGACAA  2640
 815   D  Q  A  K  A  D  S  Q  R  A  K  K  D  I  R  L  G  P  F  D  V  T  R  K  V  F  A  A  G  E  E  T  A  H  M  S  L  M  D  K   854

2641  GTCAGATCTCTTCTTCCATGACTATTCCATAGCACCCCTCTTTGTCCAGGAGAACTACCTGCATGTAAAGCCTGTGGCTGCAGGAGGTGACATGAAAAAGCACTTGATGCTTTTAAGCCG  2760
 855   S  D  L  F  F  H  D  Y  S  I  A  P  L  F  V  Q  E  N  Y  L  H  V  K  P  V  A  A  G  G  D  M  K  K  H  L  M  L  L  S  R   894

2761  GGCAGCAGATAGCATATGTGACGGTGACTTAGTGGACAATCAGATCCGGAGTAAGCAAAACTGGAGTCTTCTGCCCACCCAGGCCATCTATGCCAGTGTTCTTCCTGGAGAGTTGATGCG  2880
 895   A  A  D  S  I  C  D  G  D  L  V  D  N  Q  I  R  S  K  Q  N  W  S  L  L  P  T  Q  A  I  Y  A  S  V  L  P  G  E  L  M  R   934

2881  GGGCTACATGACTCAGTTTCCGAGCTTCCCGAGCTGGCTGGGGAAGCACTCGTCTACGGGCAAACATGATCGGATTGTTCAGGACCTATCCTTGCACATGAGTCTTAGAACTTACTCCAG  3000
 935   G  Y  M  T  Q  F  P  S  F  P  S  W  L  G  K  H  S  S  T  G  K  H  D  R  I  V  Q  D  L  S  L  H  M  S  L  R  T  Y  S  S   974

3001  CAAAAGGACAGTGAACATGGATTACCTGTCACACATCAGAGACGCACTTGTACGACCCTTGACCTCACAAGGGGTAGAAGGGGCTCAGCATGTCATCAAACTTATGGACACATACTATCT  3120
 975   K  R  T  V  N  M  D  Y  L  S  H  I  R  D  A  L  V  R  P  L  T  S  Q  G  V  E  G  A  Q  H  V  I  K  L  M  D  T  Y  Y  L   1014

3121  GATGAAAGAAGATTTTGAGAACATCATGGAAGTTAGCAGCTGGGGGGGCAAGCCCAGTGCCTTTTCCAAGCTGGATCCCAAGGTGAAAGCGGCCTTCACAAGAGCCTACAATAAGGAGGC  3240
1015   M  K  E  D  F  E  N  I  M  E  V  S  S  W  G  G  K  P  S  A  F  S  K  L  D  P  K  V  K  A  A  F  T  R  A  Y  N  K  E  A   1054

3241  CCACCTGACCCCATACTCACTTCAGGTAGTAAAGACATCGAGGCTCAGCACGGGTCCAGCACTGGATTCTGAGTACAGTGAGGAGTTCCAGGAAGATGACACTCAGTCTGAGAAAGAGCA  3360
1055   H  L  T  P  Y  S  L  Q  V  V  K  T  S  R  L  S  T  G  P  A  L  D  S  E  Y  S  E  E  F  Q  E  D  D  T  Q  S  E  K  E  Q   1094

3361  GGATGCAGTGGAGACTGACGCCATGATCAAGAAAAAGACAAGATCTTCAAAACCTTCAAAGTCAGAAAGAGAGAAGGAGTCCAAAAAAGGAAAAGGAAAGAATTGGAAGAAATGAAATCA  3480
1095   D  A  V  E  T  D  A  M  I  K  K  K  T  R  S  S  K  P  S  K  S  E  R  E  K  E  S  K  K  G  K  G  K  N  W  K  K  *   1132

3481  CTGTTTAGTACCAATAACTGCTGTTAACTTACCCTCTCGGCCAGTCCAGCTGGCCTAGAGCCTGAGAGCCTCACCCTTTGCCAGAGTAACAAGTTTGGTGTAGCACAGCGGGGTGGCCTG  3600

3601  GTGGTCCCATTATAAAGGACGGTACAGCTAGAAGGTCTAACTAGCCAGCTCTCCTGCACCTCTTATAGAGGTGGGTAGGACTGTTCCAGCCCTGCAGTGTCCCTGCCTAACTGTTAGATT  3720

3721  GCTGGGCTTCAAAGTGACTGTGTAGAATCAAAACTAGAAAGTACACAGGCATGGAGTCAGAGTTTAGTTAACCATAGCAAGACGGGACGCAGTGGGCATCTTTTGTAGCTTAAGTTTATC  3840

3841  CTAATGGCCTTTAGAGGAACAATGCTCAATTTTTTTTTTTTTTAAACATGGTAGTTATTTTATGGTGAAATTCTGTAAATAAATCTTAATCCAAAGCAGCTACCACCTGGAGCCGAGTATTC  3960

3961  TTGGTGTATTATCAGGTATTTTGCAAAGCGTGCATCAGGAATCAGATATCCACGTAGCGCATGGGTCACTGTCACCTTGAGTAGCAGACTGTTATGCTGTATATTACTTTCCTTATTAAT  4080

4081  TTGGTTGGATTACAGTATTCAGCAGTTCTTACTAACTCTGTGAATTGCAGTTCTGCAGGTGAAGGGTCAGAGCTCAAATTCCCAGCATCCCCTTGTCCTGACCACATTGCTGCTATAGGT  4200

4201  AGTCTGAGAGAAGCCACTTGGCTAGTCAGGGTGTGCACACATGGGACTGCTACACATTCCTGTCCCCATGCAGTCTTAGCCATCAGGTTGAAACCAGTAGTGGAGTGTAGGCTGGGTCAG  4320

4321  ACCCACAGGCTCATTCTGGCCGGGAGATGCTGTCCTGTGATGGAGTCTATAAGCCCTACTGGAATTCCAGATTGTATCTAGAACTGTGCTTGTTGAGGACCCTGGGCCTTGGCTGTAACC  4440

4441  TGGCCTTTAGGTGTTTAGTTCTTTTTAAATCAGAAAGTGCTCAGAGGTCTATGCTGCCTCTGAGAGGTATATGCCAAAAAGATTATAAAGGAAGGGCCAAGCTCTGTCTTTTTGAGTTCT  4560

4561  CATGGCCTTGTTGGGAGAAGAAACAGAAGTCCTGGAGACATTGAAAGTGTATTAACCAGCATGAAAGATGTGTTTCGTGGAGGTCAGTCGGTTATTATGGAAATAAAACACTGCCTTAA  4679
```

FIG. 1—*Continued.*

```
401  ENCLEGLTFVITGVLESIERDEAKSLIERYGGKVTGNVSK      mRFC140
404  ENCLEGLIFVITGVLESIERDEAKSLIERYGGKVTGNVSK      hRFC140
595  DSPFAGKTVVLTGSLSQMSRDDAKARLVELGAKVAGSVSK      E.coli Ligase
647  SSPLSGKIIVFTGSLQKITRDEAKRQAENLGAKVASSVSK      Z.mobilis Ligase
591  GEALKGLTFVITGHLSR.PREEVKALLRRLGAKVTDSVSR      T.thermophilus Ligase
```

```
441  KTNYLVMGRDSGQSKSDKAAALGTKILDEDGLLDLIRTMP      mRFC140
444  KTNYLVMGRDSGQSKSDKAAALGTKIIDEDGLLNLIRTMP      hRFC140
635  KTD.LVIAGEAAGSKLAKAQELGIEVIDEAEMLRLLGS*.      E.coli Ligase
687  KTN.LVVAGEAAGSKLSKAKELDISIIDEDRWHRIVENGG      Z.mobilis Ligase
630  KTSYLVVG.ENPQSKLEKARALGVPTLTEEELYRLLEART      T.thermophilus Ligase
```

FIG. 2. Sequence homology of mRFC140 and hRFC140 to pro-karyotic DNA ligases. An alignment of the conserved region of mRFC140, hRFC140, and the carboxy terminus of the DNA ligases from E. coli, T. thermophilus, and Zymomonas mobilis is shown. Amino acids identical in at least four of the five proteins are boxed. For each sequence, the position of the first amino acid is indicated. Sequences for the three DNA ligases are taken from references 19, 23, and 44.

```
563  NKEECLLWVDKYKPASLKNIIGQQGDQSCANKLLRWLRNWHKSSPEEKKH    mRFC140
579  NKVENLLWVDKYKPTSLKTIIGQQGDQSCANKLLRWLRNWQKSSSEDKKH    hRFC140
31   AGHYELPWVEKYRPVKLNEIVGNEDTMSRLEVFAREGN............    hRFC40
33   KKAKPVPWVEKYRPKCVDEVAFQEEVVAVLKKSLEGAD............    hRFC37
18   TKIRNLPWVEKYRPQTLNDLISHQDILSTIQKFINEDR............    hRFC36
```

```
613  AAKFGKLASKDDGSSFKAALLSGPPGVGKTTTASLVCQELGYS......Y    mRFC140
629  AAKFGKFSGKDDGSSFKAALLSGPPGVGKTTTASLVCQELGYS......Y    hRFC140
69   ................VPNIIIAGPGTGKTTSILCLARALLGALKD..AM    hRFC40
71   ................LPNLLFYGPPGTGKTSTILAAARELFGPELFRLRV    hRFC37
56   ................LPHLLLYGPPGTGKTSTILACAKQLYKDKEFGSMV    hRFC36
```

```
657  VELNASDTRSKNSLKAVVAESLNNTSIKGFYTSGAAPSVSARHALIMDEV    mRFC140
673  VELNASDTRSKSSLKAIVAESLNNTSIKGFYSNGAASSVSTKHALIMDEV    hRFC140
102  LELNASNDRGIDVVRNKIKMFAQQKVTLPKGRH.........KIIILDEA    hRFC40
106  LELNASDERGIQVVREKVKNFAQLTVSGSRSDGKPCPPF...KIVILDEA    hRFC37
91   LELNASDDRGIDIIRGPILSFASTRTIFKKGF..........KLVILDEA    hRFC36
```

```
707  DGMAGNEDRGGIQELIGLIKHTKIPIICMCNDRNHPKIRSLVHYCFDLRF    mRFC140
723  DGMAGNEDRGGIQELIGLIKHTKIPIICMCNDRNHPKIRSLVHYCFDLRF    hRFC140
143  DSMTDGAQQALRRTMEIYSKTTRFALACNASDKIIEPIQSRCAVLRYTKL    hRFC40
153  DSMTSAAQAALRRTMEKESKTTRFCLICNYVSRIIEPLTSRCSKFRFKPL    hRFC37
131  DAMTQDAQNALRRVIEKFTENTRFCLICNYLSKIIPALQSRCTRFRFGPL    hRFC36
```

FIG. 3. Sequence comparison of the ATP binding sites in different RFC subunits. An alignment of the ATP binding sites present in mRFC140, hRFC140, and three smaller subunits from human RFC (hRFC40, hRFC37, and hRFC36) is shown. Amino acids identical in at least four of the five proteins are boxed. For each sequence, the position of the first amino acid is indicated.

ing frame from positions 80 to 3472 encodes a protein of 1,131 amino acids with a calculated molecular mass of 125,966 Da (Fig. 1B). Because of its high content of basic amino acids (13.4% lysine, 4.6% arginine, and 1.9% histidine), the protein has a calculated isoelectric point of 10.13. The protein lacks known DNA binding motifs such as zinc fingers or a leucine zipper (see reference 38 for a review), but it contains two sequences which fulfill the criteria for a bipartite nuclear localization signal (41) (Fig. 1C). The presence of more than 70 putative phosphorylation sites for protein kinase A, protein kinase C, casein kinase II, and tyrosine kinases suggests that mRFC140 is a phosphoprotein. A computer search of the EMBL and GenBank data bases with the entire amino acid sequence revealed an interesting local homology with the carboxy terminus of all three prokaryotic DNA ligases present in the data bases (Fig. 2). The mRFC140 protein was most similar to the DNA ligase from Thermus thermophilus (23). The two proteins displayed over a stretch of 80 amino acids 50% identity and 66% similarity. A putative ATP binding site (43) was found around amino acid 640. This site as well as flanking sequences showed a remarkable homology to the putative ATP binding sites present in hRFC40, hRFC37, and hRFC36 (Fig. 3). The homology between these small subunits of RFC and our protein led to the assumption that our initially unknown DNA-binding protein might represent the large subunit of mouse RFC. This hypothesis was further supported by Western blot experiments described in detail below. The final proof for the identity of the isolated mouse cDNA clone came from a sequence comparison with the large subunit of human RFC. It revealed a high overall homology (83.2% identity and 89.5% similarity) between the two proteins (see Fig. 2 in reference 5). This result was further confirmed by a zooblot analysis (data not shown). An mRFC140 cDNA probe cross-hybridized under low-stringency conditions strongly with human, rat, cow, chicken, and Xenopus DNAs and weakly with Drosophila, Hydra, and Dictyostelium DNAs. cDNA clones encoding hRFC140 have been isolated very recently by using tryptic peptide sequence information derived from the purified protein (5). During the preparation of this report, a new nucleotide sequence from Drosophila DNA appeared in the EMBL data base (accession number L17340). The encoded protein displays 44% overall identity and 61% similarity to mRFC140. Although the sequence was submitted as that of a germ line transcription factor, we think that this gene most likely represents the Drosophila homolog of the largest subunit of RFC.

mRFC140 is ubiquitously expressed. To determine the size of the mRNA for mRFC140, Northern blots were performed

with poly(A)+ RNA. Filters were hybridized with a radiola-beled antisense RNA, and a typical autoradiogram is shown in Fig. 4A. Two transcripts with estimated lengths of 5.0 and 4.5 kb are detected. The longer one is at least three times more abundant than the shorter one. The same two transcripts have also been found in RNA isolated from brain, testis, and embryonic stem cells (data not shown). At the moment, it is unclear whether the two mRNAs arise by alternative splicing, alternative promoter usage, or alternative polyadenylation site usage (4). RNase protection experiments with total RNA from various mouse tissues revealed that mRFC140 is expressed ubiquitously (Fig. 4B). Expression was highest in strongly proliferative tissues like thymus, spleen, and testis. Contrary to
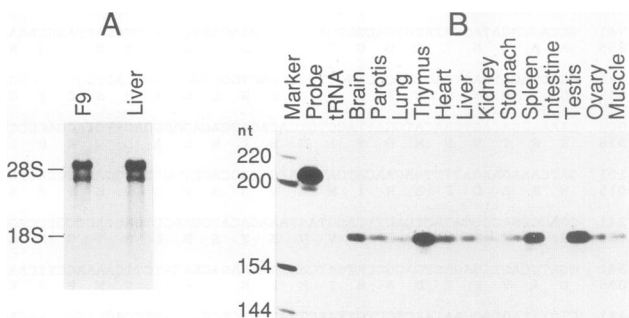

FIG. 4. Analysis of mRFC140 expression by Northern blot analysis and RNase protection experiments. (A) Poly(A)+ RNA from mouse F9 teratocarcinoma cells (1 μg) or mouse liver (5 μg) was analyzed by Northern blotting. The filter was hybridized with a radiolabeled antisense RNA transcribed in vitro from pmTCTTT8 (see Fig. 1A). The positions of 28S RNA (4,712 nt) and 18S RNA (1,869 nt) are indicated. (B) Total RNA (50 μg) from various mouse tissues was analyzed by RNase protection experiments. A 209-nt-long radiolabeled antisense RNA from the 3' end of pmTCTTT8 was used as a probe. A protected fragment of 175 bp was expected. On a longer-exposed autoradiogram, a faint signal can be detected in the lane with RNA from intestine. Yeast tRNA (50 μg) was used as a negative control.
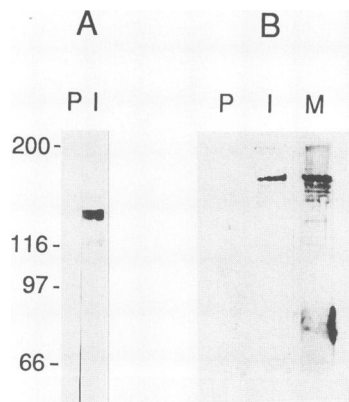
FIG. 5. Western blot analysis. Whole cell extracts from mouse NIH 3T3 cells were separated by SDS-polyacrylamide gel electrophoresis and transferred to nitrocellulose. After staining with Ponceau S, each lane was cut in half and incubated with rabbit preimmune serum (lane P) or anti-mRFC140 immune serum (lane I). Immune complexes were detected with an anti-rabbit IgG alkaline phosphatase conjugate. Size markers: rabbit myosin, 200 kDa; E. coli β-galactosidase, 116 kDa; rabbit phosphorylase b, 97.4 kDa; bovine serum albumin, 66.2 kDa. (B) Purified human RFC protein was separated by SDS-polyacrylamide gel electrophoresis and transferred to nitrocellulose. Lanes were incubated either with rabbit preimmune serum (lane P), anti-mRFC140 immune serum (lane I), or a monoclonal antibody raised against hRFC140 (lane M). Immune complexes were detected by enhanced chemiluminescence using species-specific secondary antibodies coupled to horseradish peroxidase. Lanes P and I were exposed for 3 min; lane M was exposed for 30 s.

this supposition, a relatively high level of expression was found in brain, whereas in intestine it was hardly detectable.

mRFC140 is a nuclear protein. To obtain protein for the immunization of rabbits, a cDNA fragment encoding 753 amino acids from the carboxy terminus of mRFC140 was cloned into the vector pRSET and expressed in E. coli. The protein was purified and injected into rabbits to raise polyclonal antibodies against mRFC140. These antibodies were subsequently used in Western blots to determine the size of the mRFC140 protein. The anti-mRFC140 antibodies recognize specifically in mouse whole cell extracts a protein with an apparent molecular mass of 130 kDa, as predicted from the cDNA (Fig. 5A). Interestingly, two closely spaced bands are visible on Western blots, which could well represent the same protein with differences in posttranslational modification (e.g., phosphorylation). To obtain additional evidence that the isolated mouse cDNA clone encodes the large subunit of RFC, Western blotting was performed with purified human RFC protein and anti-mRFC140 antibodies. As shown in Fig. 5B, the polyclonal antibodies raised against the mouse protein cross-react specifically with the large subunit of purified human RFC. To determine the intracellular localization of the mRFC140 protein, indirect immunofluorescence experiments using mouse B16 melanoma cells and affinity-purified antibodies were performed. As can be seen in Fig. 6, the mRFC140 protein is localized exclusively in the nucleus. The nuclear staining is in most cases diffuse and excludes the nucleoli. However, an association of the protein with discrete replication foci during certain phases of the cell cycle is not ruled out by these experiments (25).

Chromosomal localizations of the mouse and human RFC140 genes. The full-length mRFC140 cDNA was used as a probe for fluorescence in situ hybridization to mouse metaphase chromosome spreads. Hybridized probe was detected via FITC or rhodamine, and chromosomes were banded by DAPI (Fig. 7D to G). Several DAPI images were used to generate complete mouse karyotypes (not shown). Subsequent overlaying of the DAPI image with the image acquired from the probe signal revealed the localization of mRFC140 on chromosome 5C. Evaluation of 50 metaphases revealed signal
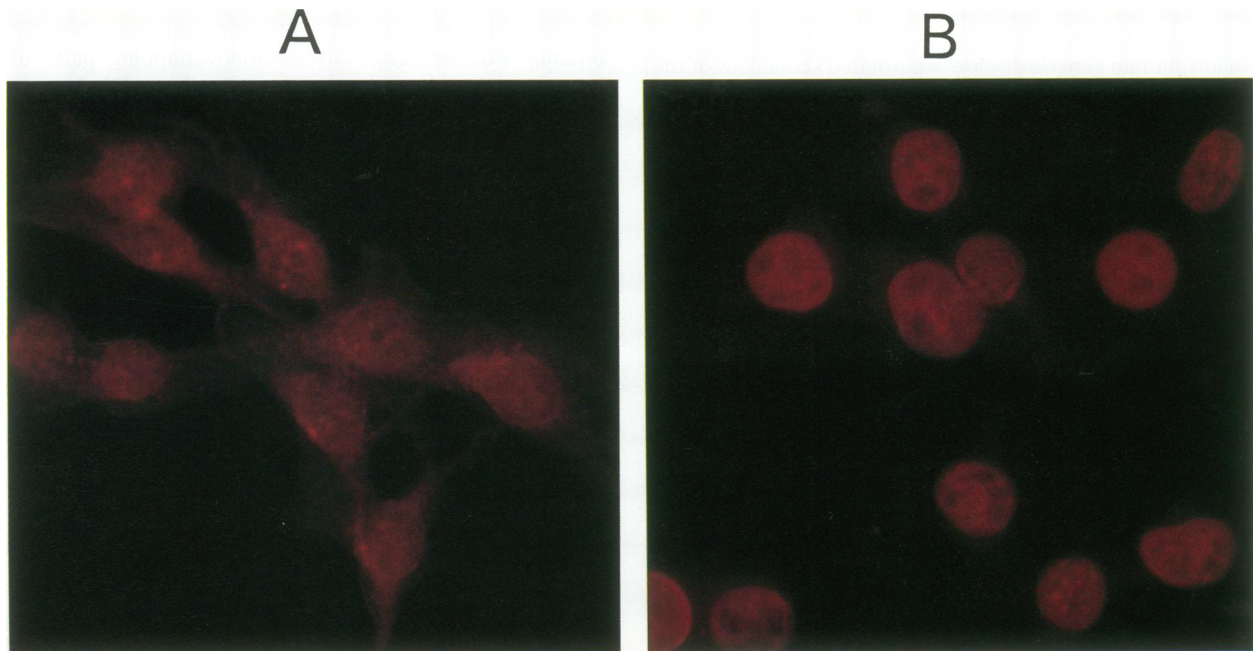
FIG. 6. Intracellular localization of mRFC140 protein. Mouse B16 melanoma cells were grown on coverslips, fixed with methanol and acetone, and incubated with either rabbit preimmune serum (A) or affinity-purified anti-mRFC140 antibodies (B). Immune complexes were detected by an anti-rabbit IgG Texas red conjugate. Magnification, ×250.
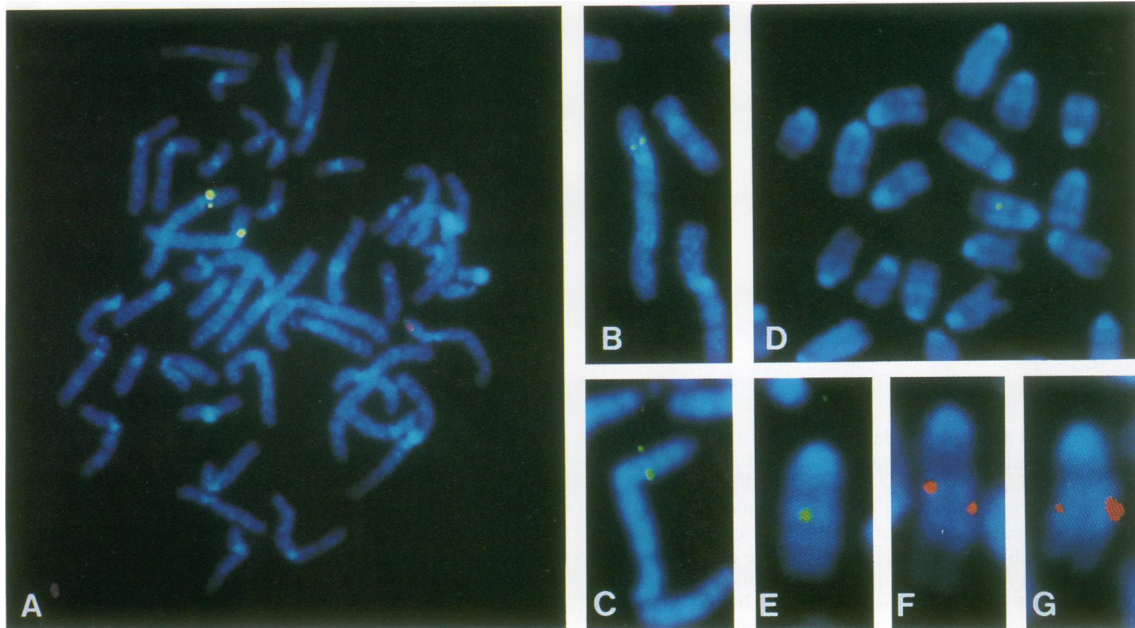
FIG. 7. Mapping of the RCF140 gene to human and mouse chromosomes. The biotin-labeled genomic clone λhgTCTTT6-3 was hybridized to normal human chromosomes and detected via FITC. Highly specific signals were found on the short arm of chromosome 4 (see complete metaphase in panel A) in the region 4p13-p14 (see single chromosome 4 homologs in panels B and C). (D to G) Hybridization of biotin (D and E)- and digoxigenin (F and G)-labeled mRFC140 cDNA to normal mouse chromosomes. Hybridized probes were detected by FITC-conjugated avidin and rhodamine-conjugated antidigoxigenin antibodies, respectively. Single (D and E) and doublet (F and G) signals were seen on chromosome 5C. (D) Section of a metaphase; (E to G) single chromosome 5 homologs. Chromosomes were banded with DAPI. Photographs were taken directly from the video screen.

doublets in 20% (Fig. 7F and G) and single signals in 40% (Fig. 7D and E) of the chromosomes 5 (a second set of experiments revealed 35% of the signals as doublets and about 20% as single fluorescent spots). Further fluorescent spots were not found consistently at any other chromosomal region and were therefore considered background.

To examine whether RFC140 gene mutations might play a role in any human genetic disease, we isolated sequences of the human gene to be used as probe for the chromosomal localization. A human liver cDNA library was screened by using the mouse cDNA as a probe, and four independent partial cDNAs for hRFC140 (termed phTCTTT15, -21, -24, and -25) were isolated. Two of these (phTCTTT24 and -25) were used to screen a human genomic library, and three independent recombinant phage clones were isolated. The identity of one of the genomic clones, termed λhgTCTTT6-3, was confirmed by sequence analysis showing that the region with DNA ligase homology is interrupted by an intron. Using λhgTCTTT6-3 as the probe, the human RFC140 gene could be mapped to the chromosome region 4p13-p14. From the 30 metaphases analyzed, 22 (approximately 70%) exhibited signals on both homologs of chromosome 4p13-p14 (Fig. 7A), whereas in 8 metaphases, only one homolog was labeled. Of the 52 specific hybridization signals on chromosome 4, 35 appeared as doublets (Fig. 7B and C). No additional signals were found at other chromosomal regions, and therefore the chromosomal location could be unambiguously assigned. So far, no clinical disorder for which RFC140 could play an apparent role has been localized to this region of chromosome 4. Since the human chromosome region 4p13-p14 is syntenic to mouse chromosome 5, our results add another example for the establishment of this syntenic group.

DISCUSSION

We have fortuitously cloned a cDNA for the large subunit of mouse RFC by virtue of its unspecific DNA binding activity. The fusion protein expressed by the original recombinant phage bound to five completely different pairs of oligonucleotides, which were used to test the sequence specificity of binding. Furthermore, various nonoverlapping parts of the protein were able to bind independently to DNA (data not shown). Since we could not detect any known DNA binding motif, mRFC140 protein may bind to DNA simply by its positive net charge at physiological pH, which is caused by the high content of basic amino acids more or less evenly dispersed throughout the molecule. Since the probes used for screening and analysis of DNA binding specificity always contained 4-nt-long overhanging 5′ ends, the protein could bind either to double-stranded DNA or to the partial single-stranded ends of the probe molecule. This second possibility is quite likely because purified RFC binds specifically to DNA of a primer-template junction containing a recessed 3′ end (24, 54). Previous experiments have indicated that the DNA binding activity of RFC might reside in the 140-kDa subunit (54), a notion corroborated by the results presented in this report.

The DNA-dependent ATPase activity was originally assigned to the 40-kDa subunit (54), and the cDNA for hRFC40 does indeed encode an ATP binding motif (6). By using E. coli-expressed hRFC40 protein, ATP binding but no ATPase activity was detected. Recently, cDNAs for the other RFC subunits have been reported (5, 7, 37; this report). A comparison of the protein sequences of all five RFC subunits leads to the unexpected result that all subunits except hRFC38 contain very similar ATP binding motifs (Fig. 3). Further experiments are required to determine whether each of these four subunits

can bind ATP and whether all of them show ATPase activity. Interestingly, the B motif of the ATP binding site found in these RFC subunits is extended to a DEAD box, which seems to be typical for a family of putative DNA and RNA helicases (13). Therefore, it might be worthwhile to test RFC140 for DNA or RNA helicase activity.

Another unexpected result was the striking homology between a region in the middle of the mRFC140 protein and the carboxy terminus of the three prokaryotic DNA ligases which have been cloned to date. Although the gene for DNA ligase from E. coli was isolated several years ago (19), the protein has not been further characterized. Therefore, no specific function can be assigned to the carboxy terminus of this enzyme. The prokaryotic DNA ligases differ from their eukaryotic counterparts mainly by the use of NAD^+ instead of ATP as a cofactor. A data base search revealed a weak homology between the carboxy-terminal one-third of E. coli DNA ligase and several NAD^+-dependent enzymes (50). On the basis of these homologies, the authors speculated that the carboxy terminus of E. coli DNA ligase contains an NAD^+ binding site. It will be interesting to determine whether RFC140 is able to bind NAD^+. In the same publication, a consensus sequence for the active site of DNA ligases (KY/ADGXR) was presented. At position 818, the sequence KADSQR is found in mRFC140. It fits well with the consensus sequence noted above because the substitution of glycine by serine represents a conservative amino acid exchange. It is tempting to speculate that either RFC140 alone or the entire RFC complex has DNA ligase activity. RFC is an essential part of the replisome and would be ideally located to ligate the Okazaki fragments generated during DNA replication of the lagging strand. However, all available data imply that DNA ligase I is the enzyme joining the Okazaki fragments during DNA replication (reference 29 and references therein).

Several findings support the assumption that pmTCTTTV represents a full-length mRFC140 cDNA clone. (i) We have isolated three different cDNA clones encoding the very 5' end which all start within a region of 25 bp. (ii) The GC content of the first 79 bp is considerably higher (65.8%) than for the rest of the sequence (45.8%). This feature is frequently found in the 5' untranslated regions of genes. (iii) The sequence around the ATG at position 80 fits well with the optimal translation start signal as determined by Kozak (22). (iv) On Northern blots, two mRNAs of approximately 5.0 and 4.5 kb are visible. The isolated cDNAs span a region of 4,679 bp without a poly(A) tail. (v) By Western blot analysis, the size of the protein was estimated to be approximately 130 kDa. The molecular size calculated from the cDNA was 126 kDa.

We wanted to know whether RFC140 might be involved in any human inherited disease for which the chromosomal locus has been identified. Therefore, we isolated human cDNA and genomic clones and determined the chromosomal localization of the gene. No mutation has been mapped to the RFC140 locus in mice or humans. Because many inherited diseases have not yet been mapped, the possibility cannot be excluded that a known syndrome is linked to RFC140. We would predict, however, that a complete loss of RFC140 function is most likely lethal for the organism.

In an attempt to estimate the number of mRFC140 molecules per cell, we prepared a whole cell extract from a dish with 10^7 cells and obtained a signal in a Western blot, using at least 1/100 of this extract. Taking into account the detection limit in our Western blot system (1 ng of a specific protein), the number of mRFC140 molecules per cell was calculated to be approximately 5×10^4. This number seems to be relatively high. It has been shown, however, that other proteins involved

in DNA replication are present in even higher numbers in the nucleus. For topoisomerase I, the estimated number was up to 10^6 molecules per nucleus, and for topoisomerase II, this value was still 10^5 (56). Our estimation is, of course, not very precise. Nevertheless, RFC140 represents a rather abundant protein within the nucleus. How many of these protein molecules are enzymatically active at the replication fork remains an open question.

The isolation of cDNA clones encoding the large subunit of RFC represents an important step toward a better understanding of the role of the RFC complex during DNA replication in eukaryotes. cDNA clones for all five subunits are now available and will considerably facilitate detailed studies on the role of each RFC subunit and the interplay between them.

REFERENCES

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1987. Current protocols in molecular biology. Greene/Wiley-Interscience, New York.
2. **Boshart, M., F. Weih, M. Nichols, and G. Schütz.** 1991. The tissue-specific extinguisher locus Tse-1 encodes a regulatory subunit of cAMP-dependent protein kinase. Cell **66:**849–859.
3. **Boshart, M., F. Weih, A. Schmidt, R. E. K. Fournier, and G. Schütz.** 1990. A cyclic AMP response element mediates repression of tyrosine aminotransferase gene transcription by the tissue-specific extinguisher locus Tse-1. Cell **61:**905–916.
4. **Breitbart, R. E., A. Andreadis, and B. Nadal-Ginard.** 1987. Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. Annu. Rev. Biochem. **56:**467–495.
5. **Bunz, F., R. Kobayashi, and B. Stillman.** 1993. Proc. Natl. Acad. Sci. USA **90:**11014–11018.
6. **Chen, M., Z.-Q. Pan, and J. Hurwitz.** 1992. Sequence and expression in Escherichia coli of the 40-kDa subunit of activator 1 (replication factor C) of HeLa cells. Proc. Natl. Acad. Sci. USA **89:**2516–2520.
7. **Chen, M., Z.-Q. Pan, and J. Hurwitz.** 1992. Studies of the cloned 37-kDa subunit of activator 1 (replication factor C) of HeLa cells. Proc. Natl. Acad. Sci. USA **89:**5211–5215.
8. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:**156–159.
9. **Church, G. M., and W. Gilbert.** 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81:**1991–1995.
10. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
11. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:**6–13.
12. **Fien, K., and B. Stillman.** 1992. Identification of replication factor C from Saccharomyces cerevisiae: a component of the leading-strand DNA replication complex. Mol. Cell. Biol. **12:**155–163.
13. **Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov.** 1989. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of

DNA and RNA genomes. Nucleic Acids Res. **17:**4713–4730.

14. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25:**263–269.

15. **Harlow, E., and D. Lane.** 1988. Antibodies—a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

16. **Hochuli, E.** 1990. Purification of recombinant proteins with metal chelate adsorbent, p. 87–98. *In* J. K. Setlow (ed.), Genetic engineering—principles and methods, vol. 12. Plenum Press, New York.

17. **Hurwitz, J., F. B. Dean, A. D. Kwong, and S.-H. Lee.** 1990. The in vitro replication of DNA containing the SV40 origin. J. Biol. Chem. **265:**18043–18046.

18. **Huynh, T. V., R. A. Young, and R. W. Davis.** 1985. Constructing and screening cDNA libraries in λgt10 and λgt11, p. 49–78. *In* D. M. Glover (ed.), DNA cloning—a practical approach, vol. I. IRL Press, Oxford.

19. **Ishino, Y., H. Shinagawa, K. Makino, S. Tsunasawa, F. Sakiyama, and A. Nakata.** 1986. Nucleotide sequence of the *lig* gene and primary structure of DNA ligase of *Escherichia coli.* Mol. Gen. Genet. **204:**1–7.

20. **Janknecht, R., G. D. Martynoff, J. Lou, R. A. Hipskind, A. Nordheim, and H. G. Stunnenberg.** 1991. Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus. Proc. Natl. Acad. Sci. USA **88:**8972–8976.

21. **Kornberg, A., and T. Baker.** 1992. DNA Replication. W. H. Freeman & Co., New York.

22. **Kozak, M.** 1991. Structural features in eukaryotic mRNAs that modulate the initiation of translation. J. Biol. Chem. **266:**19867–19870.

23. **Lauer, G., E. A. Rudd, D. L. McKay, A. Ally, D. Ally, and K. C. Backman.** 1991. Cloning, nucleotide sequence, and engineered expression of *Thermus thermophilus* DNA ligase, a homolog of *Escherichia coli* DNA ligase. J. Bacteriol. **173:**5047–5053.

24. **Lee, S.-H., A. D. Kwong, Z.-Q. Pan, and J. Hurwitz.** 1991. Studies on the activator 1 protein complex, an accessory factor for proliferating cell nuclear antigen-dependent DNA polymerase δ. J. Biol. Chem. **266:**594–602.

25. **Leonhardt, H., A. W. Page, H.-U. Weier, and T. H. Bestor.** 1992. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell **71:**865–873.

26. **Li, J. J., and T. J. Kelly.** 1984. Simian virus 40 DNA replication in vitro. Proc. Natl. Acad. Sci. USA **81:**6973–6977.

27. **Lichter, P., C.-J. Chang Tang, K. Call, G. Hermanson, G. A. Evans, D. Housman, and D. C. Ward.** 1990. High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. Science **247:**64–69.

28. **Lichter, P., and T. Cremer.** 1992. Chromosome analysis by non-isotopic in situ hybridization, p. 157–192. *In* D. E. Rooney and B. H. Czepulkowski (ed.), Human cytogenetics—a practical approach. IRL Press, Oxford.

29. **Lindahl, T., and D. E. Barnes.** 1992. Mammalian DNA ligases. Annu. Rev. Biochem. **61:**251–281.

30. **Linn, S.** 1991. How many pols does it take to replicate nuclear DNA? Cell **66:**185–187.

31. **Lu, Y., A. S. Zeft, and A. T. Riegel.** 1993. Cloning and expression of a novel DNA binding protein, PO-GA. Biochem. Biophys. Res. Commun. **193:**779–786.

32. **Luckow, B., and G. Schütz.** 1987. CAT constructions with multiple unique restriction sites for the functional analysis of eukaryotic promoters and regulatory elements. Nucleic Acids Res. **15:**5490.

33. **Luckow, B., and G. Schütz.** 1991. Linker scanning mutagenesis, p. 101–133. *In* M. McPherson (ed.), Directed mutagenesis—a practical approach. IRL Press, Oxford.

34. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridisation probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12:**7035–7056.

35. **Nagai, K., and H. C. Thøogersen.** 1987. Synthesis and sequence-specific proteolysis of hybrid proteins produced in *Escherichia coli.* Methods Enzymol. **153:**461–481.

36. **Nitsch, D., M. Boshart, and G. Schütz.** 1993. Activation of the tyrosine aminotransferase gene is dependent on synergy between liver-specific and hormone-responsive elements. Proc. Natl.

37. **O'Donnell, R. M., F. B. Dean, M. Chen, and J. Hurwitz.** 1993. Homology in accessory proteins of replicative polymerases—*E. coli* to humans. Nucleic Acids Res. **21:**1–3.

38. **Pabo, C. O., and R. T. Sauer.** 1992. Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem. **61:**1053–1095.

39. **Pan, Z.-Q., M. Chen, and J. Hurwitz.** 1993. The subunits of activator 1 (replication factor C) carry out multiple functions essential for proliferating-cell nuclear antigen-dependent DNA synthesis. Proc. Natl. Acad. Sci. USA **90:**6–10.

40. **Podust, V. N., A. Georgaki, B. Strack, and U. Hübscher.** 1992. Calf thymus RF-C as an essential component for DNA polymerase δ and ε holoenzymes function. Nucleic Acids Res. **20:**4159–4165.

41. **Robbins, J., S. M. Dilworth, R. A. Laskey, and C. Dingwall.** 1991. Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell **64:**615–623.

42. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

43. **Saraste, M., P. R. Sibbald, and A. Wittinghofer.** 1990. The P-loop—a common motif in ATP- and GTP-binding proteins. Trends Biochem. Sci. **15:**430–434.

44. **Shark, K. B., and T. Conway.** 1992. Cloning and molecular characterization of the DNA ligase gene (lig) from *Zymomonas mobilis.* FEMS Microbiol. Lett. **96:**19–26.

45. **Smith, D. E., and P. A. Fisher.** 1984. Identification, developmental regulation, and response to heat shock of two antigenically related forms of a major nuclear envelope protein in *Drosophila* embryos: application of an improved method for affinity purification of antibodies using polypeptides immobilized on nitrocellulose blots. J. Cell Biol. **99:**20–28.

46. **Stillman, B.** 1989. Initiation of eukaryotic DNA replication *in vitro.* Annu. Rev. Cell Biol. **5:**197–245.

47. **Studier, F. W., and B. A. Moffat.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189:**113–130.

48. **Tabor, S., and C. C. Richardson.** 1989. Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and *Escherichia coli* DNA polymerase I. Proc. Natl. Acad. Sci. USA **86:**4076–4080.

49. **Thömmes, P., and U. Hübscher.** 1990. Eukaryotic DNA replication. Eur. J. Biochem. **194:**699–712.

50. **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.

51. **Tsurimoto, T., T. Melendy, and B. Stillman.** 1990. Sequential initiation of lagging and leading strand synthesis by two different polymerase complexes at the SV40 DNA replication origin. Nature (London) **346:**534–539.

52. **Tsurimoto, T., and B. Stillman.** 1989. Purification of a cellular replication factor, RF-C, that is required for coordinated synthesis of leading and lagging strands during simian virus 40 DNA replication in vitro. Mol. Cell. Biol. **9:**609–619.

53. **Tsurimoto, T., and B. Stillman.** 1990. Functions of replication factor C and proliferating-cell nuclear antigen: functional similarity of DNA polymerase accessory proteins from human cells and bacteriophage T4. Proc. Natl. Acad. Sci. USA **87:**1023–1027.

54. **Tsurimoto, T., and B. Stillman.** 1991. Replication factors required for SV40 DNA replication in vitro. J. Biol. Chem. **266:**1950–1960.

55. **Vinson, C. R., K. L. LaMarco, P. F. Johnson, W. H. Landschulz, and S. L. McKnight.** 1988. In situ detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. Genes Dev. **2:**801–806.

56. **Vosberg, H.-P.** 1985. DNA Topoisomerases: enzymes that control DNA conformation. Curr. Top. Microbiol. Immunol. **114:**19–102.

57. **Weinberg, D. H., K. L. Collins, P. Simancek, A. Russo, M. S. Wold, D. M. Virshup, and T. J. Kelly.** 1990. Reconstitution of simian virus 40 DNA replication with purified proteins. Proc. Natl. Acad. Sci. USA **87:**8692–8696.

58. **Yoder, B. L., and P. M. J. Burgers.** 1991. *Saccharomyces cerevisiae* replication factor C. J. Biol. Chem. **266:**22689–22697.

Acad. Sci. USA **90:**5479–5483.