Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 11014–11018, December 1993
Biochemistry

cDNAs encoding the large subunit of human replication factor C

(simian virus 40/DNA replication/DNA binding/ATPase/DNA polymerase accessory protein)

FRED BUNZ*, RYUJI KOBAYASHI†, AND BRUCE STILLMAN†

†P.O. Box 100, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724; and *Graduate Program in Molecular and Cellular Biology, State University of New York at Stony Brook, NY 11794

Communicated by James D. Watson, August 18, 1993

ABSTRACT **Replication factor C (RFC) is a multisubunit, DNA polymerase accessory protein required for the coordinated synthesis of both DNA strands during simian virus 40 DNA replication *in vitro*. Previous studies have shown that RFC is a DNA-dependent ATPase that binds in a structure-specific manner to the 3′ end of a primer hybridized to a template DNA, an activity thought intrinsic to the 140-kDa component of this multisubunit complex. Here, the isolation and analysis of cDNAs encoding this subunit is described. Analysis of the full-length coding sequence revealed an open reading frame of 3.4 kb, encoding an 1148-amino acid protein with a predicted molecular mass of 130 kDa. A putative ATP-binding motif was observed that is similar to a motif in several of the smaller subunits of RFC and in functionally homologous replication factors of bacterial and viral origin. A "DEAD" box is also conserved among these proteins. The predicted protein shows significant identity with a DNA-binding protein of murine origin (B. Luckow, P. Lichter, and G. Schütz, personal communication). Regions of similarity were also seen between the amino acid sequences of the 140-kDa subunit of RFC, poly(ADP-ribose) polymerase, and bacterial DNA ligases—possibly representing a conserved structural feature of these proteins that bind similar DNA substrates.**

In all replication systems studied thus far, DNA is synthesized by a complex apparatus consisting of many protein components. Replication factor C (RFC) is a human-cell DNA replication factor composed of 140-, 40-, 38-, 37-, and 36-kDa subunits, which has been shown to be required for efficient simian virus 40 DNA replication *in vitro* (1–3). RFC (also known as activator 1; A1) has been shown to be essential in the coordination of DNA synthesis on both strands at a replication fork (1, 4). RFC functions in the switch from DNA polymerase α/primase to DNA polymerase δ during initiation of leading-strand DNA replication and, subsequently, as a component of the protein complex responsible for synthesis of the leading strands (4–8). The proliferating cell nuclear antigen (PCNA) cooperates with RFC in this capacity. Recent work has shown that the RFC/PCNA/polymerase δ complex functions in the synthesis of the lagging strand as well (S. Waga and B.S., unpublished work).

Characterization of the biochemical properties of RFC has shown that it possesses an intrinsic ATPase activity (8, 9). This ATPase is activated in the presence of primer/template DNA and is further stimulated by PCNA addition (8, 9). RFC also exhibits a structure-specific, primer/template DNA-binding activity (8, 9) that has been localized by photocrosslinking studies to the 140-kDa subunit p140 (6). The structure-specific DNA-binding activity of RFC is stimulated by ATP. Recently a relatively weak primer/template DNA-binding activity intrinsic to the 37-kDa subunit of RFC has also been reported (10).

In its biochemical properties, RFC is functionally analogous to the DNA-dependent ATPases of both *Escherichia coli* and bacteriophage T4, which also function as DNA polymerase accessory factors. In the DNA polymerase III holoenzyme complex from *E. coli*, a DNA-binding protein complex, the γ–δ complex, assembles the various components onto a primed template with the concomitant hydrolysis of ATP (11–13). Accessory proteins gp44 and gp62 from the bacteriophage T4 polymerase holoenzyme complex act similarly (14–17). Sequence analysis has revealed regions of similarity among the four cloned smaller subunits of human RFC, the *E. coli* DNA polymerase III γ protein, and the bacteriophage T4 gene product gp44 (18–20), which is a component of the DNA-dependent ATPase in that system (21).

Recent reports have described the molecular cloning and expression of several related subunits of human RFC in the 36- to 40-kDa size range (18–20). Here, the molecular cloning of cDNAs encoding the 140-kDa subunit is reported.‡ These data are consistent with the 140-kDa subunit of RFC functioning in the structure-specific recognition of the 3′ ends of primers during DNA synthesis.

MATERIALS AND METHODS

RFC Purification. RFC was purified as described by Tsurimoto and Stillman (5) from human 293 cells, with slight modifications.

Isolation of Tryptic Peptide Fragments and Protein Sequencing. Purified RFC (30 mg) was fractionated by SDS/PAGE and digested with trypsin as described (22). Peptides were purified by reverse-phase chromatography on a Hewlett–Packard model 1090 liquid chromatograph with a 2.1 × 250 mm Vydac C_{18} column (particle size 5 μm, 300 Å; The Separations Group). Peptides were eluted with a 50-min linear gradient of 10–35% acetonitrile in water/0.085–0.07% trifluoroacetic acid followed by a 10-min linear gradient from 35 to 50% acetonitrile. Peptide fractions were sequenced directly on an Applied Biosystems model 470 automated sequencer with an on-line 120A HPLC phenylthiohydantoin analyzer. Amino acid yields were generally in the range of 0.5–2 pmol.

Cloning of cDNAs. The peptide sequence IVAESLNN was used to design the degenerate 23-mer oligonucleotide 1407S (5′-ATWGTKGCWGARTCNCTVAACAA-3′), in which W is A or T, K is G or T, R is A or G, and V is A, C, or G. A random-primed λ ZAP II cDNA library containing HeLa cDNA inserts was amplified, and $\approx 5 \times 10^5$ plaques were transferred to nylon filters (Colony/Plaquescreen, DuPont/

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: RFC, replication factor C; PCNA, proliferating cell nuclear antigen; mAb, monoclonal antibody; ORF, open reading frame.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. L23320).

NEN). Two hundred picomoles of oligonucleotide 1407S was labeled with [γ-^{32}P]ATP (ICN) and polynucleotide kinase and incubated with the filters in Blotto/10 [0.05% heparin/1% SDS/0.5% nonfat dry milk/6% PEG 8000/5× standard saline phosphate/EDTA (SSPE; 1× SSPE is 0.15 M NaCl/10 mM phosphate, pH 7.4/1 mM EDTA)/sheared, single-stranded, salmon sperm DNA at 0.2 mg/ml/10% (vol/vol) formamide] at 50°C for 14 hr. Filters were washed at 3× standard saline citrate (SSC)/0.1% SDS at 45°C and autoradiographed (1× SSC is 150 mM sodium chloride/15 mM sodium citrate). Subsequent cDNA clones were identified by screening both random-primed and oligo-primed cDNA libraries in phage λgt10 (clones 23, 24, and 43) and an oligo(dT)-primed human cDNA library in λ ZAP (clone 39) with insert cDNAs labeled by the method of Feinberg and Vogelstein (23). The hybridization temperature was 60°C, and the filters were subsequently washed to 0.3× SSC/0.3% SDS at 65°C, as in ref. 24.

All phage inserts were subcloned into the vector pBluescriptSK(+) (Stratagene) and sequenced with a Sequenase 2.0 DNA-sequencing kit (United States Biochemical) (25).

Protein Sequence Analysis. The amino acid sequence predicted by the isolated cDNAs was compared with sequences in the GenBank (release 76.0) and European Molecular Biology Library (EMBL) data bases (release 34.0), using the BLAST network service at the National Center for Biotechnology Information (Bethesda, MD). Protein sequence motifs were detected with the aid of PROSITE 5.4 (IntelliGenetics). Sequence alignments were done with the BLAST network service and GENALIGN 5.4 (Intelligenetics).

Primer Extension and Amplification of 5′ End of mRNA. Rapid amplification of cDNA ends (RACE) was done essentially as described in Frohman *et al.* (26). First-strand cDNA was synthesized using primer 140F (5′-CTCCTCTGACTCTGAAT-3′) and 8 μg of total HeLa cell RNA as template. Synthesized DNAs were tailed with terminal transferase (Bethesda Research Laboratories) and dATP. Second-strand cDNA was synthesized using the adapter-T oligonucleotide (5′-GACTCGAGTCGACATCGATTTTTTTTTTTT-3′). PCR amplification was done with oligonucleotide 140G (5′-ATCCTCTTTACGGGAGC-3′), which anneals to the cDNA 5′ of primer 140F, and the adapter primer (5′-GACTCGAGTCGACATCG-3′), which is identical to the 5′ portion of adapter-T.

***In Vitro* Transcription/Translation.** The plasmid pBD-RFC was constructed by subcloning the *Bam*HI–*Bam*HI fragment of clone 39 (nt 520–3379) into the *Bam*HI site of the cell-free expression vector pBD7 (27), placing part of the open reading frame (ORF) of RFC under the control of the highly efficient translational initiation signal of black beetle virus mRNA1. Capped RNA was transcribed *in vitro* with T7 RNA polymerase (New England Biolabs) in the presence of 1 mM guanosine triphosphorylguanosine (GpppG; Pharmacia). One microgram of this material was translated in a nuclease-treated rabbit reticulocyte lysate (Promega).

Immunoprecipitation of Translated Gene Products. Two microliters of translated material, constituting 4% of the translation mixture, was mixed with 200 μl of Nonidet P-40 cell lysis buffer (50 mM Tris, pH 8.0/500 mM NaCl/0.2% Nonidet P-40) per immunoprecipitation. Labeled proteins were incubated with 100 μl of spent medium from the hybridoma cells on ice for 2 hr. Antibodies were collected by incubation with 100 μl of protein A-Sepharose (Pharmacia) added as a 25% slurry in lysis buffer for 45 min at 4°C. The beads were collected by centrifugation and washed two times with lysis buffer, once with 10 mM Tris, pH 8.0/150 mM NaCl/0.2% Tween 20 and twice more in lysis buffer. Bound proteins were eluted by boiling in 60 μl of Laemmli loading buffer for 5 min and subjected to electrophoresis on SDS/15% polyacrylamide gels.

RESULTS

Cloning of cDNAs. Tryptic peptides from the 140-kDa subunit of human RFC were isolated and sequenced. Three useful amino acid sequences were obtained. One of these, 140K7 [(A)IVAESLNNT(R)] was used to design a DNA probe that would be complementary to the p140 subunit coding sequence. Degeneracy was minimized by the incorporation of the most frequently used codons at each position, where such a bias exists (28). The resulting 23-mer oligonucleotide was 192-fold degenerate, containing eight possible mismatches to the complementary sequence and encoded the peptide sequence IVAESLN.

This oligonucleotide was used to probe a human cDNA library, and one positive clone, designated 140-1, was found to contain a perfect match with one of the DNAs in the degenerate pool of oligonucleotides used as a probe. In addition, sequences flanking the hybridizing site matched the codons for amino acids detected in the peptide sequence but not used in probe design. The peptide sequence and the predicted amino acid sequence differed at position 11. This residue was tentatively designated as an arginine from the peptide sequence data (data not shown), whereas the cDNA sequence predicts a serine at this position. This discrepancy is likely the result of the low yield obtained during that degradation cycle, which made the data interpretation difficult. Furthermore, the cDNA sequence encoded a lysine residue upstream from the first alanine, resulting in a tryptic site that would account for generation of the peptide that was isolated and sequenced.

This initial cDNA was used to isolate several overlapping cDNAs that were then sequenced (Fig. 1). Clone 140-43 encodes amino acids upstream of those of clone 140-39, including a methionine near the 5′ end; this codon is in a favorable context for the initiation of translation, as determined by Kozak (29). However, no in-frame stop codons were found 5′ to this sequence, so the presence of an upstream methionine codon remained possible.

To isolate a cDNA representing the extreme 5′ end of the RNA message, total RNA was used as a template for RACE (26), in which a specific cDNA was synthesized and then amplified by the PCR. Sequencing revealed that the amplified cDNA perfectly overlapped the 5′ end of clone 43 and extended farther 5′ by 150 bp. Analysis of this sequence revealed the presence of multiple stop codons upstream of the formerly identified 5′ methionine codon in all three reading frames. No other methionine codon was found in a favorable environment for translation initiation. This result suggests that the previously described methionine codon represents the start site of translation.

Sequence of cDNAs. Juxtaposition of the combined sequences resulted in a 4980-bp cDNA with a 3444-bp open

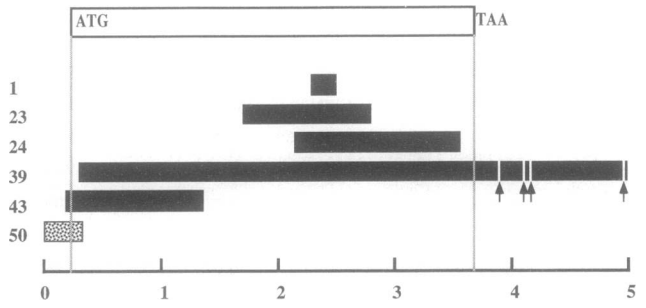

FIG. 1. RFC p140 cDNA clones. Black bars represent isolated cDNAs. Clones 1, 23, and 43 were from a random hexamer primed library; clones 24 and 39 were isolated from a library primed with oligo(dT); and clone 50 was isolated by the rapid analysis of cDNA end (RACE) method (26). The white box represents the ORF. The relative positions of the polyadenylylation signal AAUAAA in the 3′ untranslated region are shown by arrows.

reading frame that encodes a protein of 1148 aa (Fig. 2). Peptide sequences obtained from digested protein were found in the translated sequence of the ORF. In addition to the above mentioned peptide, close matches to two other peptide sequences, NLADDYLENK and IIDEEXLLNLI, were found in the translated sequence as NLADDSSENK and IIDEDGLLNLI, respectively. These two sequences were not used in the design of a probe for cloning because the molar yield obtained during the peptide sequencing was too low to be reliable. It is, therefore, not surprising that there are mismatches between the sequence of these peptides and that predicted by the cDNA clone.

The ORF is succeeded by a 1.3-kb untranslated region. Several polyadenylylation signal sequences (AAUAAA) were observed in this region of the cDNA (Fig. 1). Northern blot analysis of polyadenylylated RNA detected two distinct bands (data not shown), suggesting that these multiple mRNAs might differ in processing of their 3′ ends.

The predicted molecular mass of the translated protein was calculated at 130 kDa; this agrees reasonably with the observed migration of the protein on SDS/PAGE. p140 subunit is posttranslationally modified (F.B. and B.S., unpublished work), which might retard its migration somewhat. The sequence KKKTK (aa 1121–1125) is a close match to a known nuclear localization signal seen in large tumor antigen, although whether this sequence would be sufficient to target this protein to the nucleus is unclear (30).

Immunoprecipitation of Gene Product by Monoclonal Antibodies (mAbs) Directed Against RFC. To confirm that the cloned cDNA encodes the 140-kDa subunit of RFC, a portion of the ORF was expressed *in vitro* and incubated with three mAbs directed against p140 (F.B. and B.S., unpublished work). A restriction fragment encompassing ≈80% of the ORF was subcloned into a cell-free expression vector, transcribed, and translated *in vitro* using rabbit reticulocyte lysate. When fractionated by SDS/PAGE, a ladder of labeled translated products, leading up to the 130-kDa position, was produced (Fig. 3*A*). Protein was also translated from brome mosaic virus RNA. No protein was produced in the absence of input RNA. mAbs directed against the 140-kDa subunit of RFC were incubated with these samples, and immunoprecipitated proteins were analyzed by SDS/PAGE (Fig. 3*B*). RFC mAbs 6 and 19 bound to proteins translated by the lysate that had been programmed with the putative 140-kDa cDNA. pAb419, a mAb directed against simian virus 40 large tumor antigen (31) did not precipitate these proteins, and neither the RFC antibodies nor pAb419 specifically bound to proteins derived from brome mosaic virus RNA, suggesting that the proteins precipitated by the anti-RFC mAbs were specifically recognized. These results indicate that the translated gene product is antigenically similar to native RFC p140. It is, therefore, highly probable that the cDNA isolated represents the p140 message. mAb 11, which has been shown to recognize native p140 (F.B. and B.S., unpublished work), did not detect any protein. This result might indicate that the epitope for this mAb lies in the 20% of the coding sequence not included in the translation construct or that *in vitro* translation of mRNA does not generate the antibody epitope.

Regions of Sequence Similarity Between the Predicted Amino Acid Sequences of RFC p140 and Proteins from the Data Base. p140, the DNA-binding subunit of human RFC, was found to share 83% amino acid identity with a predicted protein from a murine cDNA isolated from a cDNA library by virtue of its DNA-binding properties (B. Luckow, P. Lichter, and G. Schütz, personal communication; Fig. 2). If conservative amino acid substitutions between these two proteins are taken into account, the overall similarity is 90% over the 1122 aa that their sequences overlap.

The RFC p140 also shared a region of similarity with several DNA ligases of prokaryotic origin, and, to a lesser extent, with eukaryotic poly(ADP-ribose) polymerases (Fig. 4*A*). Over a region 60 aa in length, RFC p140 is 55% identical and 65% conserved with DNA ligase from *Zymomonas mobilis*, the protein of this class that it most resembles. DNA ligases of eukaryotic origin do not share any appreciable similarity with these sequences. p140 also resembles both human (Fig. 4*A*) and mouse (data not shown) poly(ADP-ribose) polymerase over a smaller region. No apparent sequence similarities between these proteins have been found outside of the region shown.

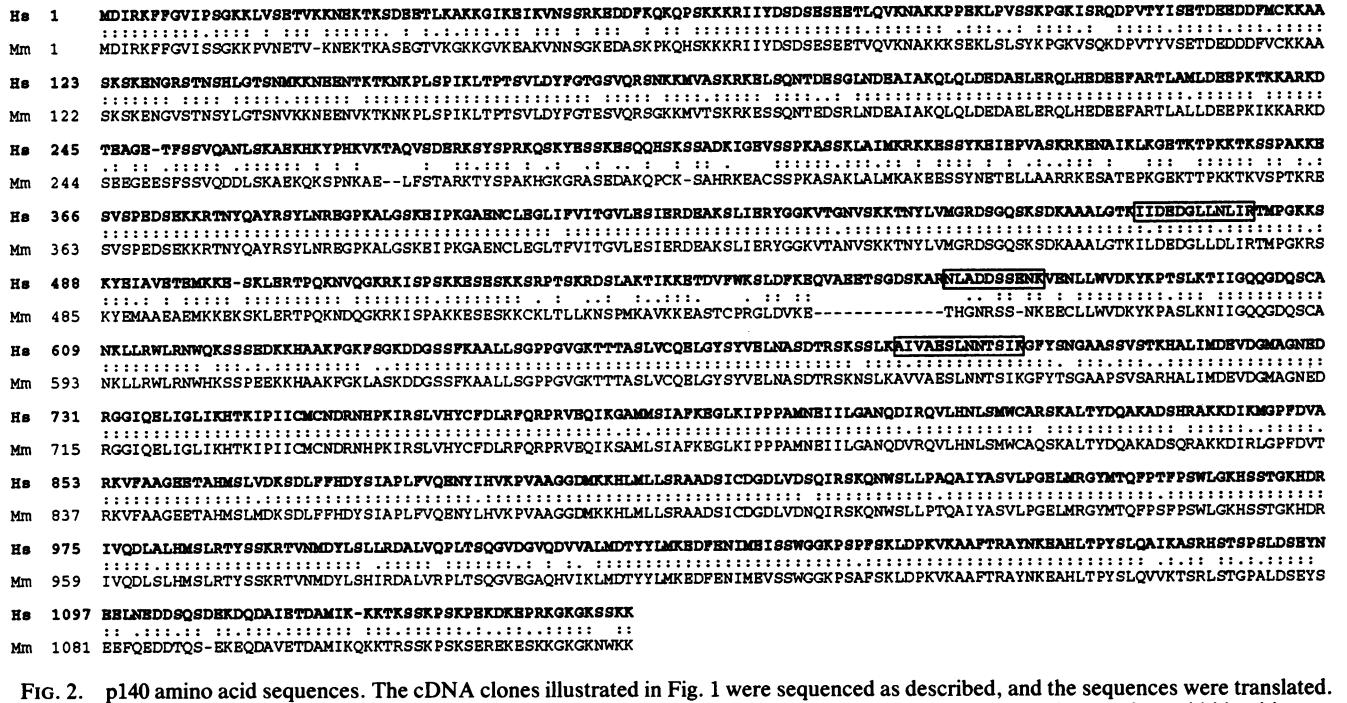

FIG. 2. p140 amino acid sequences. The cDNA clones illustrated in Fig. 1 were sequenced as described, and the sequences were translated. Human protein sequence (Hs; boldface) was aligned with that of mouse (Mm) by using the BLAST network service. Amino acid identities are indicated by a colon; conserved substitutions are indicated by a period. Human amino acid sequences corresponding to those of directly sequenced tryptic peptides are boxed.

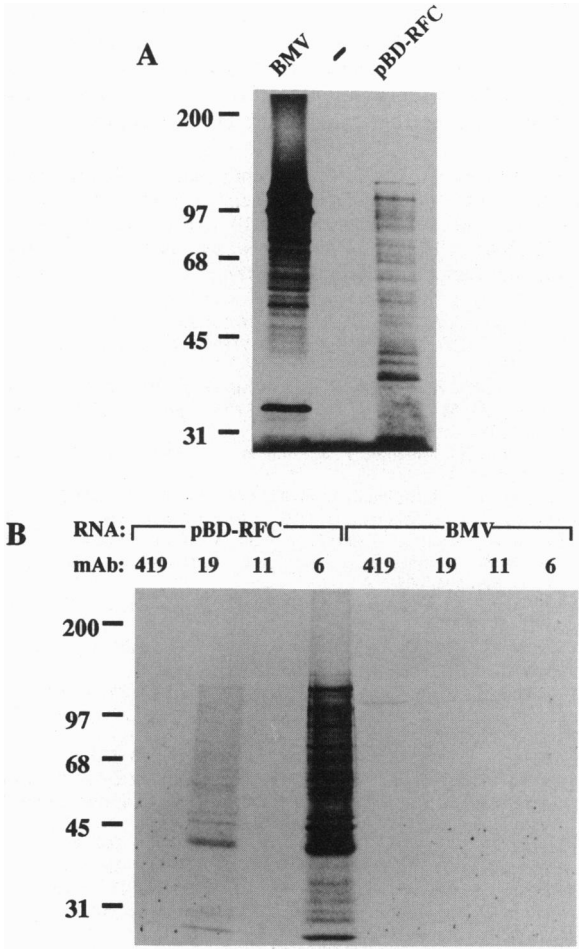

FIG. 3. Expression analysis of cDNA clone. RNA transcribed either from brome mosaic virus (BMV) or pBD-RFC was translated in a rabbit reticulocyte lysate with 40 μCi of [^{35}S]methionine. (*A*) Two microliters of these reactions and of a reaction containing no input RNA (−) were fractionated by SDS/PAGE. (*B*) Antibodies pAb419, directed against simian virus 40 large tumor antigen, and mAbs 19, 11, and 6, directed against native RFC p140, were incubated with the *in vitro* translated gene products. After immunoprecipitation, proteins bound to the antibodies were analyzed by SDS/PAGE. Positions of molecular mass markers (in kDa) are indicated.

The predicted amino acid sequence of p140 was also compared with those of the smaller subunits of human RFC and with the analogous γ–τ subunit of *E. coli* DNA polymerase III and gp44 of phage T4 (Fig. 4*B*), which have been shown to exhibit sequence similarity (18). An ATP-binding P-loop motif previously found conserved in p36, p37, p40, γ–τ, and gp44 (18), was also present in p140. C-terminal to the ATP-binding domain of these proteins is another highly conserved region that contains sequences resembling or identical to the DEAD motif. This region is a previously defined amino acid sequence motif that has been observed in a large, evolutionarily broad family of RNA helicases and RNA-dependent ATPases (for review, see ref. 33).

DISCUSSION

Two independent lines of evidence support the identity of the p140 subunit of human RFC with the gene product. (*i*) Amino acid sequences directly obtained from purified protein were found within the predicted amino acid sequence of the gene product. (*ii*) The translated gene product was specifically recognized by two RFC mAbs. We conclude that the cDNAs described here encode the 140-kDa subunit of RFC.

Significant homology was found between p140 and a recently characterized protein overproduced from a murine cDNA clone (B. Luckow, P. Lichter, and G. Schütz, personal communication), suggesting that the mouse protein, in fact, represents the murine homologue of RFC. This protein was originally detected as a DNA-binding protein expressed from a cDNA library probed with ligated oligonucleotides. Subsequently, this DNA-binding activity was found to be sequence independent. Because RFC is a structure-specific DNA-binding protein that binds to primer/template DNAs, the murine p140 subunit was probably detected by binding to nicked or recessed 3′ DNA ends present in the probe used to screen the cDNA expression library. Recently, a *Drosophila* DNA sequence encoding a protein with 42% identity and 59% similarity (using the GAP program, Genetics Computer Group, University of Wisconsin) appeared in the GenBank data base (accession no. L17340). The protein is listed as a female germ-line-specific transcription factor, although it could be the *Drosophila* homologue of the human RFC large subunit.

Homology was detected between p140 subunit and the smaller subunits of RFC. This homology was most apparent around a putative nucleoside triphosphate-binding P-loop motif that is also conserved in replication proteins from *E. coli* and bacteriophage T4 that are DNA-dependent ATPases (18–20). It is unclear which of the RFC subunits participate in ATP-binding and hydrolysis, but when ATP was crosslinked to purified RFC, binding to ATP was detected

A
E. coli ligase 604-VVLTGSLSQMSRDDAKARLVELGAKVAGSVSKKTDLVIAGEAAG-SKLAKAQELGIEVIDEAEM-665
Z. mobilis ligase 655-IVFTGSLQKITRDEAKRQAENLGAKVASSVSKKTNLVVAGEAAG-SKLSKAKELDISIIDEDRW-717
hRFC 412-FVITGVLESIERDEAKSLIERYGGKVTGNVSKKTNYLVMGRDSGQSKSDKAAALGTKIIDEDGL-475
hpADP-ribo. pol 394-KILTLGKLSRNKDEVKAMIEKLGGKLTGTANKASLCISTKKEVEKMNKKMEEVKEANIRVVSED-457

B
hRFC 36kDa 39-PHLLLY GPPGTGKTS TILACAKQLYKDKEFGSM.VLELNASDDRGIDIIRGPI-
hRFC 40kDa 70-PNIIIA GPPGTGKTT SILCLARALLGALKDA...MLELNASNDRGIDVVRNKI-
hRFC 37kDa 72-PNLLFY GPPGTGKTS TILAAARELFGPELFRLR.VLELNASDERGIQVVREKV-
hRFC 140kDa 645-KAALLS GPPGVGKTT TASLVCQELGYS.....YVELNASDTRSKSSLKAIV-
T4 gp44 43-HIILHS PSPGTGKTT VAKALCHDVNAD.....MMFVNGSDCKIDFVRGP..-
E. coli γ/τ 39-HAYLFS GTRGVGKTS IARLLAKGLNCETGITATPCGVCDNCREIEQGRFVDLI-

...LSFA.....TRTIFKK.....GFKLVILDEADAMTQDAQNALRRVIEKF-129 hRFC 36kDa
...KMFAQQKVTLPKG.....RHKIIILDEADSMTDGAQQALRRTMEIY-160 hRFC 40kDa
...KNFAQLTVSGSRSDGKPCP.....PFKIVILDEADSMTSAAQAALRRTMEKE-170 hRFC 37kDa
AESLNNTSIKGFYSNGAASSVS.....TKHALIMDEVDGMAGNEDRGGIQELIGL-740 hRFC 140kDa
..LTNFA.....SAASFDG.....RQKVIVIDEFDRSGLAESQRHLRSFMEA-126 T4 gp44
EIDAASRTKVEDTRDLLDNVQYAPARGRFKVYLIDEVHMLSRHSFNALLKTLEEP-146 *E. coli* γ/τ

FIG. 4. Alignments between regions of RFC p140 and structurally similar proteins, determined by the BLAST network server, GENALIGN, and manual inspection. (*A*) Amino acid sequences of *E. coli* and *Z. mobilis* DNA ligases, RFC p140, and human poly(ADP-ribose) polymerase (hpADP-ribo. pol). (*B*) Alignment of amino acid sequences of human RFC p36, p37, p40, and p140 subunits, with bacteriophage T4 gp44 and *E. coli* DNA polymerase III γ (18). The boxed sequence is the ATP-binding motif (32).

only on p40 (6). When overproduced in bacteria, this peptide also bound, but did not hydrolyze, ATP (19). The five-subunit RFC protein is a DNA- and PCNA-activated ATPase, but which subunit(s) has this activity remains to be determined.

A DEAD motif present in the smaller subunits of RFC (18) is within a region also conserved in p140 subunit. Although this motif appears in a number of putative RNA helicases, RFC is not known to exhibit such activity. Members of the DEAD-box family of proteins can function in the ordered assembly of multiprotein complexes onto RNA substrates with the concomitant hydrolysis of ATP, or GTP, during the processes of pre-mRNA splicing and translation (for review, see ref. 33). It remains to be determined whether a similar mechanism governs the assembly of RFC and PCNA upon the DNA template.

The apparent structural redundancy between the various RFC subunits is representative of a general feature of the DNA-dependent ATPases that function in DNA replication (18). The *E. coli* γ-τ complex is judged to contain two γ subunits and two structurally similar δ subunits (34), whereas the phage T4 analog is composed of one gp62 subunit and four gp44 molecules, the latter containing the ATP-binding domain (35). The functional consequence of this stoichiometry remains to be determined. Multiple versions of this domain might interface with PCNA, which is predicted to function as a trimer (36) containing six structural domains (37), or with other elements of the different polymerase assemblies as they simultaneously replicate the leading and the lagging strands. The class of DNA-dependent ATPases, which includes RFC, has been described as molecular matchmakers because these ATPases promote the assembly of other proteins onto DNA in a manner dependent on DNA structure and ATP hydrolysis (38).

Comparing the amino acid sequence of p140 subunit with those in the data bases revealed the presence of a conserved region that is shared by prokaryotic DNA ligases and, to a lesser extent, by eukaryotic poly(ADP-ribose) polymerases. RFC p140 and its mouse analog are nearly identical within this region: the only differences are the conserved substitutions. One common characteristic shared among these enzymes of nucleic acid metabolism is an ability to bind DNA with a recessed 3′ end or to bind gapped, duplex DNA. The region of alignment, however, does not fall within the domain of poly(ADP-ribose) polymerase that has been shown essential for binding nicked DNA (39). Although the significance of this observed region of similarity remains unknown, one possibility is that it is within a protein domain involved in the recognition of single-stranded DNA adjacent to 3′ ends, a region of template DNA shown to be protected from nuclease digestion by RFC (40). This region of similarity is outside the region of homology shared by p140 and the smaller RFC subunits, and no relationships were seen between the smaller subunits of RFC and the DNA ligases or poly(ADP-ribose) polymerase.

Prokaryotic DNA ligases and mammalian poly(ADP-ribose) polymerases use NAD as a cofactor. The NAD-binding site that has been mapped on these proteins is not within the observed region of similarity. Several sequences within p140 protein partly match to the sequences KLDGLA and KEDGSL, which are the NAD active sites in *E. coli* and bacteriophage T4 DNA ligases (41). Regions flanking these sequences in p140 do not show any apparent similarity with sequences found around known NAD-binding sites. RFC is known to use ATP as a cofactor for activity (9); it is unknown whether NAD can function in this capacity.

All cDNAs encoding the five human RFC subunits have now been isolated, which should facilitate the eventual overproduction of these proteins and the reconstitution of enzymatic activity (this report and refs. 18–20). Recently, a homologue of RFC has been purified from *Saccharomyces cerevisiae* (42, 43). The exploitation of the powerful methods of yeast genetics combined with biochemical analysis promises to shed considerable light on the mechanisms by which these proteins interact to catalyze DNA synthesis.

We thank Drs. B. Luckow, P. Lichter, and G. Schütz for communication of the murine cDNA sequence before publication. Random-primed and oligo(dT)-primed λgt10 libraries were the generous gift of Dr. Robert Tjian. This research was supported by a grant from the National Cancer Institute (CA13106).

1. Tsurimoto, T. & Stillman, B. (1989) *Mol. Cell. Biol.* **9**, 609–619.
2. Lee, S.-H., Eki, T. & Hurwitz, J. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7361–7365.
3. Weinberg, D. H., Collins, K. L., Simancek, P., Russo, A., Old, M. S., Virshup, D. M. & Kelly, T. J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8692–8696.
4. Tsurimoto, T., Melendy, T. & Stillman, B. (1990) *Nature (London)* **346**, 534–539.
5. Tsurimoto, T. & Stillman, B. (1989) *EMBO J.* **8**, 3883–3889.
6. Tsurimoto, T. & Stillman, B. (1991) *J. Biol. Chem.* **266**, 1961–1968.
7. Lee, S. & Hurwitz, J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 5672–5676.
8. Lee, S.-H., Kwong, A. D., Pan, Z.-Q. & Hurwitz, J. (1991) *J. Biol. Chem.* **266**, 594–602.
9. Tsurimoto, T. & Stillman, B. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1023–1027.
10. Pan, Z.-Q., Chen, M. & Hurwitz, J. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6–10.
11. Wickner, S. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 3511–3515.
12. Onrust, R., Stuckenberg, P. T. & O'Donnell, M. (1991) *J. Biol. Chem.* **266**, 21681–21686.
13. McHenry, C. S. (1988) *Annu. Rev. Biochem.* **5**, 519–550.
14. Cha, T.-A. & Alberts, B. M. (1988) in *Eukaryotic DNA Replication: Cancer Cells 6*, eds. Kelly, T. & Stillman, B. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 1–10.
15. Munn, M. M. & Alberts, B. M. (1991) *J. Biol. Chem.* **266**, 20034–20044.
16. Munn, M. M. & Alberts, B. M. (1991) *J. Biol. Chem.* **266**, 20024–20033.
17. Capson, T. L., Benkovic, S. J. & Nossal, N. G. (1991) *Cell* **65**, 249–258.
18. O'Donnell, M., Onrust, R., Dean, F. B., Chen, M. & Hurwitz, J. (1993) *Nucleic Acids Res.* **21**, 1–3.
19. Chen, M., Pan, Z.-Q. & Hurwitz, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 5211–5215.
20. Chen, M., Pan, Z.-Q. & Hurwitz, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2516–2520.
21. Mace, D. C. & Alberts, B. M. (1984) *J. Mol. Biol.* **177**, 279–293.
22. Brill, S. J. & Stillman, B. (1991) *Genes Dev.* **5**, 1589–1600.
23. Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267.
24. Ruppert, J. M., Kinzler, A. J., Wong, A. J., Bigner, S. H., Kav, F., Law, M. L., Seuanez, H. N., O'Brien, S. J. & Vogelstein, B. (1988) *Mol. Cell. Biol.* **8**, 3104–3113.
25. Henikoff, S. (1984) *Gene* **28**, 351–359.
26. Frohman, M. A., Dush, M. K. & Martin, G. R. (1988) *Proc. Natl. Acad. Sci. USA* **78**, 8998–9002.
27. Damahapatra, B., Rozhon, E. J. & Schwartz, J. (1987) *Nucleic Acids Res.* **15**, 3933.
28. Wada, K.-N., Wada, Y., Doi, H., Ishibashi, F., Gojobori, T. & Ikemura, T. (1991) *Nucleic Acids Res.* **19**, suppl. 1981–1986.
29. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8133.
30. Dingwall, C. & Laskey, R. A. (1991) *Trends Biochem. Sci.* **16**, 478–481.
31. Harlow, E., Crawford, L. V., Pim, D. C. & Williamson, N. M. (1981) *J. Virol.* **39**, 861–869.
32. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982) *EMBO J.* **1**, 945–951.
33. Wassarman, D. A. & Steitz, J. A. (1991) *Nature (London)* **349**, 463–464.
34. Maki, S. & Kornberg, A. (1988) *J. Biol. Chem.* **263**, 6561–6569.
35. Jarvis, T. C., Paul, L. S., Hockensmith, J. W. & von Hippel, P. H. (1989) *J. Biol. Chem.* **264**, 12717–12729.
36. Bauer, G. A. & Burgers, P. M. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7506–7510.
37. Kong, X.-P., Onrust, R., O'Donnell, M. & Kuriyan, J. (1992) *Cell* **69**, 425–437.
38. Sancar, S. & Hearst, J. E. (1993) *Science* **259**, 1415–1420.
39. Menissier-de Murcia, J., Molinete, M., Gradwohl, G., Simonin, F. & de Murcia, G. (1989) *J. Mol. Biol.* **210**, 229–233.
40. Tsurimoto, T. & Stillman, B. (1991) *J. Biol. Chem.* **266**, 1950–1960.
41. Tomkinson, A. E., Totty, N. F., Ginsburg, M. & Lindahl, T. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 400–404.
42. Fien, K. & Stillman, B. (1992) *Mol. Cell. Biol.* **12**, 155–163.
43. Yoder, B. L. & Burgers, P. M. J. (1991) *J. Biol. Chem.* **266**, 22689–22697.