THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 264, No. 13, Issue of May 5, pp. 7466-7472, 1989
Printed in U.S.A.
© 1989 by The American Society for Biochemistry and Molecular Biology, Inc.

Structure of the Human Gene for the Proliferating Cell Nuclear Antigen*

(Received for publication, October 7, 1988)

Salvatore Travali, De-Hui Ku, Maria Giulia Rizzo, Luigi Ottavio, Renato Baserga‡, and Bruno Calabretta

From the Department of Pathology and Fels Research Institute, Temple University Medical School Philadelphia, Pennsylvania 19140

The proliferating cell nuclear antigen (PCNA, cyclin) was originally defined as a nuclear protein whose appearance correlated with the proliferative state of the cell. It is now known to be a co-factor of DNA polymerase δ and to be necessary for DNA synthesis and cell cycle progression. cDNA clones of human PCNA have been isolated and, using one of these cDNA, we have now obtained from a λ phage library a clone containing the entire human PCNA gene and flanking sequences. The human PCNA gene is a unique copy gene and has 6 exons. It spans, from the cap site to the poly(A) signal 4961 base pairs. We have identified, in the 5′-flanking sequence, a region with promoter activity, a well as other structural elements common to other promoters. An interesting feature of the PCNA gene is the presence of extensive sequence similarities among introns and between introns and exons.

The proliferating cell nuclear antigen (PCNA)[1] was originally identified by immunofluorescence as a nuclear protein whose appearance correlated with the proliferative state of the cell (1–3). A cell cycle-dependent protein described by Bravo and called cyclin (for a review, see Ref. 4) was eventually shown to be identical with PCNA (5). PCNA is required for SV40 DNA replication *in vitro* (6), and it has been identified as the auxiliary protein for DNA polymerase δ (7, 8). The coordinated leading and lagging strand DNA synthesis is PCNA-dependent (9), and both DNA synthesis and cell cycle progression are inhibited when Balb/c 3T3 cells are exposed to antisense oligodeoxynucleotides to PCNA (10). Its association with DNA synthesis (11), and the fact that it is growth-regulated (see "Discussion"), suggest that PCNA is a key gene in the transition of cells from quiescence (or mitosis) to S phase. cDNA clones corresponding to the human (12, 13) and rat (14) PCNA have been isolated and described. We report here (1) the structure and complete sequence of the human PCNA gene and (2) the identification of a 5′-flanking sequence with promoter activity. We elected to report the entire sequence of the gene, including introns, because of the importance of the gene, but especially because of a striking feature, *i.e.* the extensive sequence similarities among introns and between introns and exons.

MATERIALS AND METHODS

Screening of a Genomic Library—A genomic library of human leukocyte DNA cloned in the EMBL3 vector was a kind gift of Dr. Carlo Croce (Wistar Institute). 500,000 recombinants were screened with a 5′-end probe (*Bam*HI-*Stu*I fragment of approximately 300 bp) of the human PCNA cDNA (13), according to the procedure described by Maniatis *et al.* (15). The DNAs of 5 positive individual clones were purified after tertiary screening. By restriction digestion analysis, using the restriction enzymes *Eco*RI, *Bam*HI, and *Hin*dIII, and hybridization with the full length PCNA cDNA insert, we identified two different clones representing the complete gene and a pseudogene of PCNA/cyclin.

Southern Blot Analysis of the PCNA Gene—Chromosomal DNA was purified from human peripheral blood leukocytes from different individuals, as described by Gross-Bellard *et al.* (16), with minor modifications. Twenty μg of DNA were digested with the restriction enzyme *Eco*RI and blotted by the method of Southern (17). The full length cDNA of human PCNA labeled by random priming (18) was hybridized to the digested genomic and phage DNAs by standard methods (15).

Mapping of Phage EMBL3-S2—One of the 5 clones selected, EMBL3-S2, was studied in detail. We used the "Phage Lambda Mapping Quick-kit" (Collaborative Research), as previously described (19). Determinations of fragment lengths identified the position of the restriction sites *Eco*RI, *Hin*dIII, and *Bam*HI within the human insert, approximately 15 kb in length.

Subcloning of the BamHI Fragments of EMBL3-S2 Human Insert—*Bam*HI digestion of EMBL3-S2 phage gave two fragments, 5.5 and 2.7 kb in size, hybridizing, respectively, to the 5′- and the 3′-ends of a PCNA cDNA clone (13). These *Bam*HI fragments were subcloned in the *Bam*HI site of the pGEM3 vector (Promega Biotec) by standard procedures. Large scale preparations of each subclone were obtained by CsCl-ethidium bromide equilibrium centrifugation technique. Restriction enzyme analysis was performed on each subcloned fragment by endonuclease cleavage selecting restriction sites present in the human cDNA of PCNA/cyclin and the multilinker of the pGEM3 vector.

Sequence Analysis—The 5.5-kb *Bam*HI fragment hybridizing to the 5′ end of PCNA cDNA was further subcloned in four smaller pieces (with 5′-3′ direction): *Bam*HI-*Pst*I, *Pst*I-*Hin*cII, *Hin*cII-*Eco*RI, and *Eco*RI-*Bam*HI (see Fig. 2). Each insert was sequenced using T7 and Sp6 primers with the Sequenase protocol (U. S. Biochemical).

Deletion Mutant Analysis of the 5′-Flanking Sequence—For deletion analysis of the PCNA promoter region we used the fragment *Bam*HI-*Nru*I derived from the 5.5-kb *Bam*HI fragment. The *Nru*I restriction site is contained in the first exon 100 bp upstream from the ATG starting codon. We constructed the following deletion mutants: *Bam*HI-*Nru*I, approximately 2.8 kb in length; *Eco*RI-*Nru*I, approximately 1.3 kb in length; *Eco*RV-*Nru*I, approximately 0.85 kb; *Hin*dIII-*Nru*I, approximately 0.625 kb; and *Pvu*II-*Nru*I, approximately 395 bp (see Fig. 6). These promoters were used to drive the human thymidine kinase cDNA (see below).

Transfection—Tk$^-$ts13 cells were transfected in suspension, as

* This work was supported by Grant CD214 from the American Cancer Society and GM 33694 from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J04718.

‡ To whom correspondence and reprint requests should be addressed.

[1] The abbreviations used are: PCNA, proliferating cell nuclear antigen; bp, base pairs; kb, kilobase pairs; TK, thymidine kinase.

This is an Open Access article under the CC BY license.

described in detail by Shen *et al.* (20). Selection of TK⁺ colonies was carried out by standard methods in hypoxanthine, aminopterin, thymidine medium (gHAT).

Primer Extension—A 16-mer antisense oligodeoxynucleotide was synthesized for primer extension. The oligomer 5′CTGAGACCTAGAAAGA 3′ extended from nucleotide +5 to nucleotide +50 of the human PCNA cDNA. The Oligo Sizing Ladder was purchased from American Bionetics, Inc., Emeryville, CA. Oligomer and markers were end-labeled, and extension was carried out as previously described (21, 22).

RESULTS

Isolation of the Genomic Clone Containing the Human PCNA Gene—We have previously reported (13) the isolation from an Okayama-Berg library (23) of a cDNA clone containing the coding sequence of the human PCNA gene. The PCNA/cDNA was 4l base pairs longer than the previous PCNA/cDNA described by Almendral *et al.* (12). Using a 0.3-kb *Bam*HI-*Stu*I fragment that is in the most 5′ region of the PCNA/cDNA, we screened a human genomic DNA library for clones carrying the PCNA gene. Several clones were isolated and one of them, EMBL3-S2, was shown to contain all the coding sequences of the PCNA gene. The Southern blot in Fig. 1 shows the hybridization of genomic DNA derived from three different individuals and two PCNA-containing phages to a full length cDNA of human PCNA. Three *Eco*RI bands were detected in genomic DNAs (*panel A*); two of them (3.7 and 2.7 kb) were found in phage EMBL3-S2 (*panel B, lane 1*), whereas the 1.5-kb band was found in phage EMBL3-HOL1 (*panel B, lane 2*). Since the 1.5-kb *Eco*RI fragment was found to hybridize with three different segments of a PCNA cDNA clone, a 5′- (*Bam*HI-*Stu*I), a middle region (*Eco*RV-*Eco*RI), and a 3′- (*Hpa*I-*Bam*HI) fragment, we suspected that the 1.5-kb band corresponded to a PCNA pseudogene. This has been confirmed by direct sequence analysis.[2]

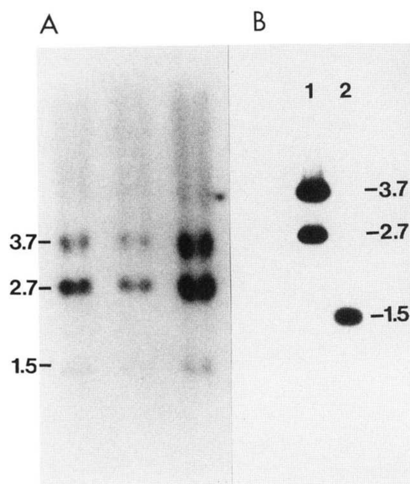

FIG. 1. **Southern blot analysis of human genomic DNA and of phages EMBL3-S2 and EMBL3-HOL1.** Human genomic DNA from three different individuals (*panel A*) and DNA from phages EMBL3-S2 and EMBL3-HOL1 (*panel B*) were digested with *Eco*RI and blotted as described (15, 17, 18). The blot was hybridized to a radioactive full length cDNA (insert only) of PCNA (13). Sizes are given in kilobases.

Fig. 2 shows the physical map of the human PCNA gene. The gene consists of 6 exons separated by five introns. The 1st exon (380 bp) contains 159 base pairs of 5′-untranslated region and 221 base pairs of coding sequence. The other exons are, respectively, 98, 68, 195, 124, and 416 bp. The longest intron is the 4th (1885 bp), and the shortest one is the last (86 bp). The distance of the *Bam*HI site in the 5′-flanking region is only approximate, since our sequencing (see below) stopped at the *Eco*RI site at −1267. It should also be noted that upstream of the *Bam*HI site in phage EMBL3-S2 there are 5 kb of additional human sequences, containing Alu repeats (not shown).

Sequence of the PCNA Gene—The entire sequence of the PCNA gene, including the introns and the proximate 5′- and 3′-flanking sequences, is shown in Fig. 3. The introns are indicated by lowercase letters. The CAP site was determined by primer extension and S1 nuclease assay (see below) and is numbered +1. The ATG initiation codon is located at +160, the TGA termination codon at +4656, and the polyadenylation signal AATAAA at nucleotide 4961; the sequence similarity with the PCNA/cDNA (13) ceases at nucleotide 4984.

Determination of the CAP Site—This was done both by primer extension (21, 22) and by S1 nuclease assay (15). We present here only the data obtained by primer extension (Fig. 4). The position of the band in respect to the Oligo Size Ladder indicates a length of 50 nucleotides.

The CAP site is therefore localized at the same nucleotide that began the PCNA/cDNA isolated by Jaskulski *et al.* (13), which therefore must be considered a full length cDNA. The S1 nuclease assay (not shown), actually gave multiple bands, including some faint ones longer than the main band located at the same nucleotide identified by primer extension. Multiple sites of initiation are not uncommon in genes lacking a perfect TATAA box (see below), especially when they are rich in GC boxes (24). The strong band in the S1 assay, the band obtained by primer extension, and the length of the cDNA all agree to indicate that the preferred initiation site is at the nucleotide designated in Fig. 3 as +1.

Sequence Similarities in the PCNA Gene—While sequencing the PCNA gene, we were struck by the recurrence of similar sequences. Computer analysis showed striking sequence similarities between introns and between introns and exons. Three of these sequence similarities (intron 1/intron 3; intron 1/intron 4; and introns 3 and 4) are illustrated in Fig. 5. Table I summarizes these sequence similarities, including those between introns and exons (the latter represented

[2] D.-H. Ku, S. Travali, M. G. Rizzo, L. Ottavio, R. Baserga, and B. Calabretta, manuscript in preparation.

FIG. 2. **Physical map of the human PCNA gene.** Partial map of the insert from phage EMBL3-S2. The exons are given as *boxes*.

```
                                                                   -1267
                                                                 gaattct
-1260
gctgaccaaggtattaaaagtaactaaagagaagtggtgtgaagaaagcaagagagaaacaacaaatcct
-1190
gtccatcctgtaacaattgaaaatttctggctgggcgtggtggctcaggcctgtaatcccagcactttga
-1120
gaggccgaggcaggtggatcacctgaggtcaggtgttcaagaccagcctggccaacatggtgaaaccccg
-1050
tctctactaaaaaaaataataataataatacaaaaattagccgggtgtggtggtaggcacctgtaatcc
-980
cagatactcgggaggctgaggcaggagactcacttgaacctgggaggcggaggttgcaatgagctgagat
-910
cgcgcgactgtactccagcctggatgacagagcaggactccatctcaaaaaggaaggcggggaaaagggg
-840
aaatattaaatgtgtacgctctttgactcagctgtattacttcaaggagttgatatcaccaaaattgcct
-770
aagtgctcaaaggtgtttgtagttaaacaacaggagattgataaattatgttatatacatgtgatgctat
-700
gttttaaagaggtactgatatgataaaaagatgtacgtggcataaattaaatgtacttattaagtactt
-630
ttccaagtgtttacggaatgagtgcattttgaaaaaaaaaagtgtattcgaactttaaaaaagcttt
-560
aaaagctttatacaataacgattgagtgattataagagctggcggggaatgttaagaggatgatagggа
-490
gctaagtttaacagaacaattcacctctttatcttgtgacacctacgagcgcatcaattctgtaattgaa
-420
aaataaagtgcatatttgcagcagctgtactctcttcaggctgcaaggaggcttttcctcccggtaggct
-350
tgatttgcatttcactttcactttcgtggctggaaactttctacccacgtagtgaggctagaggagccac
-280
ctaaagctggggcttgacgaagccgggaccgggacccgatctccacatatgcccggacttcttctgcggc
-210
cgggttcaggagtcaaagaggcggggagacctgcgcgacgctgccccgccctgcgcccgcttcctccaat
-140
gtatgctctagggggcgggcctcgcggggagcatggacacgattggccctaaagtcttccccgcaaggcc
-70
gtgggctggacagcgtggtgacgtcgcaacgcggcgcagggtgagagcgcgcgcttgcggacgcggcggc
+1                                                                   +70
ATTAAACGGTTGCAGGCGTAGAGAGTGGTCGTTGTCTTTCTAGGTCTCAGCCGGTCGTCGCGACGTTCGC
                                                                   +140
CCGCTCGCTCTGAGGCTCCTGAAGCCGAAACTAGCTAGACTTTCCTCCTTCCCGCCTGCCTGTAGCGGCG
                                                                   +210
TTGTTGCCACTCCGCCACCATGTTCGAGGCGCGCCTGGTCCAGGGCTCCATCCTCAAGAAGGTGTTGGAG
                                                                   +280
GCACTCAAGGACCTCATCAACGAGGCCTGCTGGGATATTAGCTCCAGCGGTGTAAACCTGCAGAGCATGG
                                                                   +350
ACTCGTCCCACGTCTCTTTGGTGCAGCTCACCCTGCGGTCTGAGGGCTTCGACACCTACCGCTGCGACCG
                                                                   +420
CAACCTGGCCATGGGCGTGAACCTCACCAGgtgagcctcgcgccccgggaagccgccccggcccgcctgc
                                                                   +490
acctccggctgtggcgagcgcttcgagcctagccctcattggctggcgtgggcatcccagagcttctcatt
                                                                   +560
ggcctgcacgcagtggtggggcccaagctgagatgagcggttacggaaaagcccgcgctggctgctgcgc
                                                                   +630
gaacctgctttttcgcgccaaagtcacaaagcgggtggtggcgggaaaatcaagggtttttccgcagtgc
                                                                   +700
caggaacactgttccagggactctttgctcactaaacctgttggccttgaatggacgctttagctgtggc
                                                                   +770
tttcttgtttctgagacggtctcggtctcggtgtgttgcccgggctggtctccaacttctgggctcaagc
                                                                   +840
gatcctcccggctcagtcgcgtcgactttaaatgctttataatgcccttgcgagaaatgtggcagcctgt
                                                                   +910
catcctacttagtggtaggagattgtttctatccagaaggacactgctggtggtattttagtataaata
                                                                   +980
ctgccagatgcgtccaaaacgtctgcattaataatggcatcctccagcagtccgtttaccctccaccagt
                                                                  +1050
tctgagacggcctgacgggtgagagtggtaaccccttctaaccgcgttcgaaatacagcccttcagcaga
                                                                  +1120
cggcgttgattttaaagcatgtgtctcctgtcttctagTATGTCCAAAATACTAAAATGCGCCGGCAATG
                                                                  +1190
AAGATATCATTACACTAAGGGCCGAAGATAACGCGGATACCTTGGCGCTAGTATTTGAAGCACCAAgtaa
                                                                  +1260
gtcgtaccttttaccgagtcacgaagctacaggaaaatcaaaactctgtgtgagtagaaactcaaaagc
                                                                  +1330
tatctgcgtttcttttggtaagACCAGGAGAAAGTTTCAGACTATGAAATGAAGTTGATGGATTTAGATG
                                                                  +1400
TTGAACAACTTGGAATTCCAgtgagtatcagtttctcattgtagagagtgctgtacacaggcacgatagt
                                                                  +1470
tatgtcatagaatgtttgtttatttttacagacagggtcttggctctgttgcccaggctggagtgcagta
                                                                  +1540
gtgccatatagctctctctaacctgggattcctgggctcaagcagtcctcttgccttagtctcctaagtg
                                                                  +1610
gctaggaaggactacgggcctgtcccaccacacctggctaatttttttcatttttgtgtgtgggacgtgg
                                                                  +1680
gggcagtctagccaggctggctggaactcctggcctcaagtgatcctcctccgtcaagatatgttaatat
                                                                  +1750
aatttaaagcctacttcataacaacttttctagaaatatatctactggtgcatgtttcaaagagatgatt
                                                                  +1820
ttagtatttggatagttgttcaccacaagtctaataatctccacaggttaaatttattgtttatgccagt
                                                                  +1890
tgtctatttgcattaacttccatgaactctttaaattgttctctagaatgcttgcttttattaatgagg
                                                                  +1960
ttttaaagctagcttgagagaaatttatccaggttaggttataaacaccaaaggagagaagaaatgtttg
                                                                  +2030
aatgttgaaaatgcctaatatattctcttgctttcttttagaaagtgattaggcctgcttgcgccatcat
                                                                  +2100
gatttctgtgccatactctaatgttctcttactttatccctggaggatgaggaggaggaggctcttgttc
                                                                  +2170
cctggatggtgcatttaatagccatttatttttttgagtggagtttgttaagaaattacgcaagtcatat
                                                                  +2240
tttaaagtaatcagaaaatatgattctgagttgtttaggtgttgccttttaagaaagtgagggtgccaaa
                                                                  +2310
tcattaaatttctaacaattaacttttggaaaattttgttcttaatagGAACAGGAGTACAGCTGTGTAG
                                                                  +2380
TAAAGATGCCTTCTGGTGAATTTGCACGTATATGCCGAGATCTCAGCCATATTGGAGATGCTGTTGTAAT
                                                                  +2450
TTCCTGTGCAAAAGACGGAGTGAAATTTTCTGCAAGTGGAGAACTTGGAAATGGAAACATTAAATTGTCA
                                                                  +2520
CAGACAAGTAATGTCGATAAAGAGGAGGAAGCTgtaagtagttttaagtaaaaagaaaatagtttgaag
                                                                  +2590
agaattataatactgcttattaggttaattgctaaaattaaaagtagacagaattggatcccaagtaatt
                                                                  +2660
tctgaaaattgagatactgttgaaatctgtgaatgatttataagtgtcatccaatttagaattatatttg
                                                                  +2730
caagaagggaatacaaattcagcacgtgtacataccacagcaacagtggtttatggatcaagtccacacc
                                                                  +2800
ggctcttaagggtaggattgggaagttaggcgtataacttagcttctggagatacttactctcttaccaa
                                                                  +2870
ataattgagcataggacagcagctcaatagaaggatatgttaggagtaaaagtctacctgtttggagcac
                                                                  +2940
ttatgtaatcctaatatagcttacatgttgtgggtccaattggtagcccattttaaaggtggagaagcag
                                                                  +3010
gctgagcaaccttaagtgacaatttagccaaagtcacaggctgtaggaatcaaaggttaaacaggaagga
                                                                  +3080
gactctcactaaggctagaaagcagactccatgcaactttgagagtacctagagagacccttatttaacc
                                                                  +3150
aaaatagaaagaacatagcaaaaccccatctctcataaaaatataaaattgaccgggtgatgagtggcg
                                                                  +3220
acacacctgtaaacccgactactggaatgcatgagatgggagaatgacttgaaccgaggaggcggaggtt
                                                                  +3290
gcagtgagcccagatcatgccactcccctccagcctgggtgacagagcaagattccatcttaaacaaaca
                                                                  +3360
aaaaaaactcgctaacctgggcataaattaaaactttgtaaatcaaggacaaaggtcctaaacctcataa
                                                                  +3430
cttgcattaggattaaatacggtagcattaaagagcttagcatatctgtgtgtggcatattataagctta
                                                                  +3500
caataaatactatatattgctctcttgtcccttgaatgggtagtcaacatttagtttaaataaaggtaaa
                                                                  +3570
attcagttgaaaggtttttttttaaattaataaagtctaggagctgattctttatctgtttcctgaatca
                                                                  +3640
catttccactcctgccaacctcgtttttttcttttgctgtttttctttgtttttgagacagggcttgctc
                                                                  +3710
tgtgccacccaggctggagtgcggtggtgcagtcggttcactacagcctcaaactccagggcttaagtga
                                                                  +3780
tcctcctgcctcagtttcccaagagccgggacacaggtgtgtgccaacacactagcctggtttccctaat
                                                                  +3850
ttcattttccccttgaccattacaactatttgttgaagaaattagatcatttattagtttcagagtttgg
                                                                  +3920
attttacctgattgcattcctgtgtatctaataacctctacctgtgtgtcctacagactggtagctatag
```

FIG. 3. **Sequence of the human PCNA gene.** The nucleotides are numbered from CAP site = +1. Introns are in *lowercase letters*.

```
                             .                .               .      +3990
cctggagccttgatatcagggtgttttgtttcggggtgagagagcaagaatatggtggtggtgtgtgtg
         .                .                .                         +4060
cctctagtaggaggcacagggtgtctggatgtgtttgcaatgttagcagctataagtcattgtctagatc
                             .                .               .      +4130
cattaagtcattaattagagtttgcagagctgaaattaatacgttttatcacttattggctgcttattag
         .                .                .                         +4200
aaaacttccataagaaaagcttcccattatataatttggttatctaaattatagctataccaaaagacaa
                             .                .               .      +4270
aggctagataatcgagtctttttgcatttatgtatcagtcttcaaaattttcatagcgtccctccaaagt
         .                .                .                         +4340
gaccaatacaagtgtttgtgggtttttataaatatataatgagctaatagattgcaactttcttgatgtt
                             .                .               .      +4410
tttcaatgatgaatcttttgttttgtagGTTACCATAGAGATGAATGAACCAGTTCAACTAACTTTTGCA
         .                .                .                         +4480
CTGAGGTACCTGAACTTCTTTACAAAAGCCACTCCACTCTCTTCAACGGTGACACTCAGTATGTCTGCAG
                             .                .               .      +4550
ATGTACCCCTTGgtaagataataaatttgaaccttgttttgatggtagtcatatgtgatacatactcctc
         .                .                .                         +4620
agtaattaaccatcttcctgtctttcagTTGTAGAGTATAAAATTGCGGATATGGGACACTTAAAATACT
                             .                .               .      +4690
ACTTGGCTCCCAAGATCGAGGATGAAGAAGGATCTTAGgcattcttaaaattcaagaaaataaaactaag
         .                .                .                         +4760
ctctttgagaactgcttctaagatgccagcatatactgaagtcttttctgtcaccaaatttgtacctcta
                             .                .               .      +4830
agtacatatgtagatattgttttctgtaaataacctatttttttctctattctctccaatttgtttaaa
         .                .                .                         +4900
gaataaagtccaaagtctgatctggtctagttaacctagaagtatttttgtctcttagaaatacttgtga
                             .                .               .      +4970
tttttataatacaaaagggtcttgactctaaatgcagttttaagaattgtttttgaatttaaataaagtt
         .                .                .                         +5040
acttgaatttcaaagatcacagggcagtgtcttcatttgaccaggactgttgaaagtatcctactgaatt
                             .         +5073
cccagctacagtcaccctttgttcaaactgttc
```

FIG. 3—*continued*

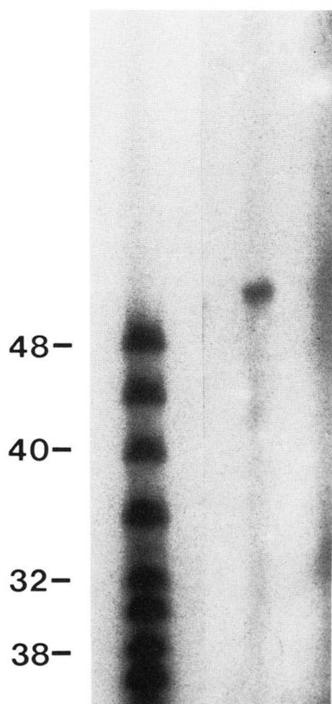

FIG. 4. **Determination of CAP site of the human PCNA gene.** Primer extension analysis was carried out as described under "Materials and Methods." Size markers are indicated. The band localizes at the nucleotide, designated as +1 in Fig. 3.

by the cDNA). There are clearly several regions of extensive sequence similarities, between various introns and between introns and exons. The occurrence of such extensive intragenic sequence similarities is not a common observation in

```
INTRON 1 - INTRON 3

733                                                 786
GTGTTGCCCGGGCTGGTCTCCAACTTCTGGGCTCAAGCGATCCTCCCGGCTCAG
:: : ::: :: :::: ::  :::: :::: :::::: ::::::::    :::
GTCTAGCCAGG-CTGG-CTGGAACTCCTGGCCTCAAGTGATCCTCCTCCGTCAA
1616                                               1667


INTRON 1 - INTRON 4

750                                 787
CTCCAACTTCTGGGCTCAAGCGATCCTCCCGGCTCAGT
::: :::: : ::::: ::: :::::::: : ::::::
CTCAAACTCCAGGGCTTAAGTGATCCTCCTGCCTCAGT
3688                               3725


INTRON 3 - INTRON 4

1415
TTTGTTTATTTTTACAGACAGGGTCTTGGCTCTGTTGCCCAGGCTGGAGTGCAGTAGTGCC
:::  ::  ::::  ::::::::  :    :: ::   :::::::::::::: :: ::: :
TTTTCTTTGTTTTTGAGACAGGGCTTGCTCTGTGCCACCCAGGCTGGAGTGCGGTGGTG-C
3611

ATATAGCTCTCTCTAACCTGGGATTCCTGGGCTCAAGCAGTCCTCTTGCCTTAGTCTCCTA
:    : :: ::  : :::   : ::: ::::: :::   ::::: ::::: ::: ::: :
AGTCGGTTCACTACAGCCTCAAACTCCAGGGCTTAAGTGATCCTCCTGCCTCAGTTTCCCA

                                                          1592
AGTGGCTAGGAAGGACTACGGGCCTGTCCCACCACACCTGGCTAATTTT-----TTTCATT
::     ::   :::: :: ::  ::: ::: :::::  : ::  :::      :::::::
AG-----AGCGGGGAC-ACAGGTGTGTGCCAACACACTAGCCTGGTTTCCCTAATTTCATT
                                                          3786
```

FIG. 5. **Sequence similarities in the human PCNA gene.** The following sequences are compared (the *top sequence* is mentioned first): intron 1 and intron 3, intron 1 and intron 4, and intron 3 and intron 4. *Dots* emphasize identities.

human genes, although Munn and Mues (25) have reported highly conserved repeats in the introns of heat-shock genes.

Structural Analysis of the Flanking Regions of the Human PCNA Gene—The salient structural features of the flanking regions of the PCNA gene can be summarized as follows. There is no conventional TATAA box in the 5′-flanking region. There is a similar sequence, TAAA at nucleotide −88, but it is not perfect and seems to be located a little too far away from the CAP site. There is a CCAAT sequence at nucleotide −145. The elements in the 5′-flanking region of the PCNA gene that have sequence similarities to well known consensus sequences in other promoters are summarized in Table II.

Functional Analysis of the PCNA Promoter—To delineate roughly which sequences in the 5′-flanking region may be important in the regulation of the human PCNA gene, we proceeded to a functional analysis of that region using hybrid genes in which fragments of various lengths of the 5′-flanking region of the PCNA gene were fused to the coding sequence of the human TK cDNA (26). The sequences cloned upstream of the human thymidine kinase coding sequence included the 5′-flanking sequence of PCNA from residue −2800 (*Bam*HI site, approximate) to nucleotide residue +60, which is in the 5′-untranslated region of the PCNA gene. Progressive deletions were then carried out using unique restriction sites of the 5′-flanking region. These deletions are summarized in Fig. 6, while Fig. 7 summarizes the activities of these constructs as tested in a standard assay. After transfection of exponentially growing TK^-ts13 cells (20) with the chimeric plasmids, colonies were selected in hypoxanthine aminopterin thymidine medium, the plates were fixed after 9 days, and the colonies were counted. Clearly, all promoters tested were active. The differences between the following promoters: *Bam*HI, *Eco*RI, *Eco*RV, and *Pvu*II, were not statistically

TABLE I
Sequence similarities in the human PCNA gene
In parentheses is the length of the introns or of the cDNA, in base pairs (first column). In the second column, are the cDNA sequences that were compared to intron sequences. The last 3 columns are self-explanatory. (For comparison among introns, no nucleotide numbers are given, but the sizes of the compared sequences can be gleaned from the "matched" and "unmatched" columns.)

			Nucleotides		
		Numbers of cDNA	Matched	Mismatched	Gaps
Intron 1 (708 bp)	PCNA cDNA (1300 bp)	170–310	71	69	2
Intron 2 (96 bp)	PCNA cDNA	1230–1270	26	18	1
Intron 3 (938 bp)	PCNA cDNA	725–910	99	75	8
Intron 4 (1885 bp)	PCNA cDNA	945–1275	169	118	14
Intron 5 (86 bp)	PCNA cDNA	1035–1090	40	13	4
Intron 1 (708 bp)	Intron 2 (96 bp)		28	7	2
Intron 1	Intron 3 (938 bp)		144	87	11
Intron 1	Intron 4 (1885 bp)		101	85	6
Intron 1	Intron 5 (86 bp)		18	9	0
Intron 2 (96 bp)	Intron 3 (938 bp)		34	17	2
Intron 2	Intron 4 (1885 bp)		40	24	2
Intron 2	Intron 5 (86 bp)		15	11	0
Intron 3 (938 bp)	Intron 4 (1885 bp)		288	188	21
Intron 3	Intron 5 (86 bp)		28	17	1
Intron 4 (1885 bp)	Intron 5 (86 bp)		28	13	1

TABLE II
Sequence similarities between the 5′-flanking sequence of the PCNA/cyclin gene and other promoter elements

GGCGGG	−122 to −127 N-*ras* gene (6/6)
GGCGGG	−160 to −165
	−186 to −191
	−515 to −520
ATTTGCAT	−341 to −348 Octamer (8/8)
ATTTGCAT	
$T_CC_{CC}A_NGCG$	−150 to −158 AP-2 binding site
C G C CGC	
CAANNTGT CT	−449 to −459 Glucocorticoid responsive element
CACAGTGTTCT	
ATTGG	−95 to −99 Inverted CCAAT
ATTGG	
CCAAT	−141 to −146 Direct CCAAT
CCAAT	

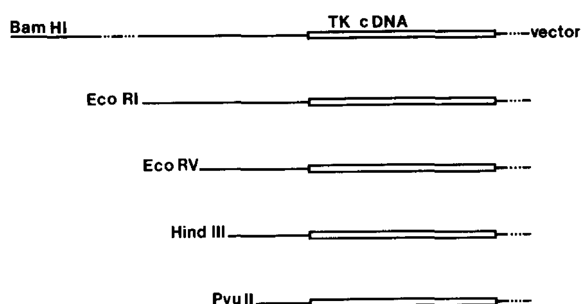

FIG. 6. **Deletion analysis of the promoter region of the PCNA gene.** The various fragments generated by the indicated restriction enzymes were cloned in the polylinkers of either pGEM 3 or Bluescript SK as described under "Materials and Methods."

significant. Only the *Hin*dIII promoter gave consistently fewer colonies than the other promoters.

DISCUSSION

The PCNA gene product (PCNA, cyclin, auxiliary factor of DNA polymerase δ) seems to occupy an important position in DNA replication and cell cycle progression (6–9, 13). The encoded protein product of the PCNA gene is probably present in the cell in a multimeric form (27), and its synthesis correlates with the proliferative state of the cell (3). Recent studies on the distribution of the PCNA protein during the cell cycle have revealed dramatic changes in its nuclear localization during the S phase (28, 29). PCNA has been identified not only by immunofluorescence staining with an appropriate antibody, but also by two-dimensional gel electrophoresis of labeled proteins (30, 31). In proliferative 3T3 cells, PCNA (cyclin) corresponds to 0.1% of the total radioactive proteins (4). The synthesis of PCNA is very low in quiescent cells and increases 6- to 7-fold after serum stimulation. It is induced by purified growth factors, such as platelet-derived growth factor and fibroblast growth factor (30). Epidermal growth factor, which is a weak mitogen in 3T3 cells, acts synergistically with insulin to induce both DNA replication in quiescent cells and synthesis of PCNA. Other growth factors, like insulin by itself, dexamethasone etc., which do not induce cell proliferation, do not stimulate PCNA synthesis. PCNA is induced in lymphocytes by interleukin-2 (32), and it is also induced by the E1A protein of adenovirus (33). On the basis of the cDNA and its sequence, the PCNA protein consists of 261 amino acids, M_r = 29,261, with a high content of acidic *versus* basic amino acids. PCNA is now recognized as a cofactor of DNA polymerase δ (6–8, 27, 31) which is involved in cellular DNA replication (34, 35). The protein has a DNA-binding domain and some homologies to the cAMP-binding protein (12), both of them encoded in the first exon.

The mRNA for PCNA, 1.3 kb, was shown by Almendral *et al.* (12) to be induced in quiescent 3T3 cells by stimulation with 10% fetal calf serum, becoming detectable 12 h after stimulation and reaching a maximum at about 16–18 h, which is also the peak of DNA synthesis. Jaskulski *et al.* (13) have investigated the effect of growth factors on the steady state levels of PCNA/mRNA in Balb/c 3T3 cells. PCNA/mRNA is already detectable in quiescent cells and begins to increase 4 h after stimulation with serum, reaching a maximum around 16–24 h. Platelet-poor plasma (which is essentially serum without platelet-derived growth factor) does not cause an increase in the steady state levels of PCNA/mRNA in Balb/c 3T3 cells. Platelet-derived growth factor by itself causes a clear increase in the levels of PCNA/mRNA. Epidermal growth increases the amount of PCNA/mRNA in quiescent Balb/c 3T3 cells at 4 h after stimulation, but the increase is not sustained, and, by 8 h the PCNA/mRNA is no longer detectable. Finally, the PCNA/mRNA, like TK mRNA, is not detectable when G_1-specific ts mutants of the cell cycle are

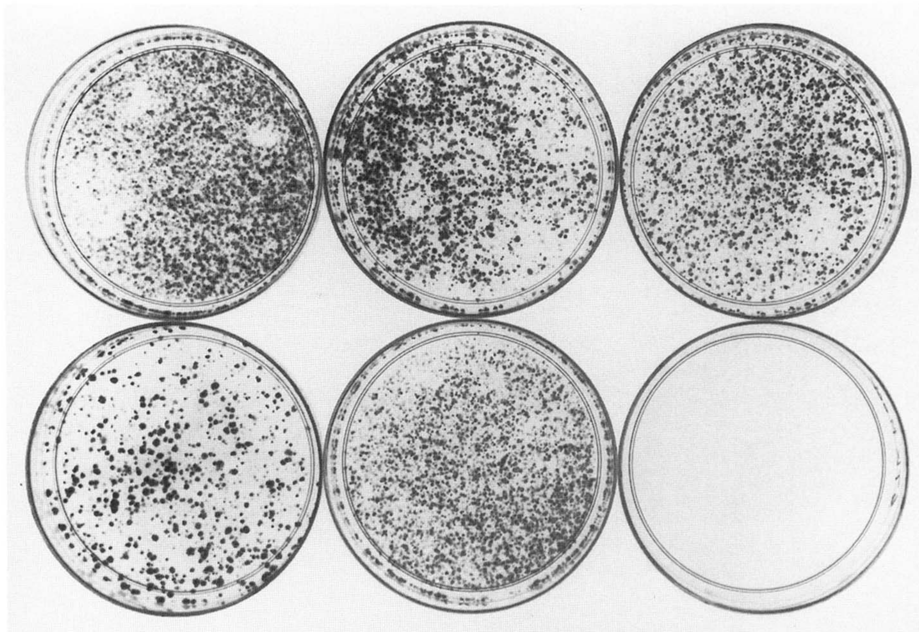

FIG. 7. **Promoter activity of the deletion mutants of the 5′-flanking sequence of the human PCNA gene.** The constructs are diagrammed in Fig. 6. TK⁻ts13 cells were transfected with the individual plasmids, and clones were selected in gHAT medium for 9 days. From *left* to *right* and from *top* to *bottom:* *Bam*HI promoter, *Eco*RI, *Eco*RV, *Hin*dIII, *Pvu*III, and mock.

stimulated with serum at the restrictive temperature, or when cells are exposed to concentrations of cycloheximide that cause only a moderate inhibition of protein synthesis (13).

We present here the full sequence of the human PCNA gene, including its proximate 5′- and 3′-flanking sequences. We have also identified the CAP site and the region in the 5′-flanking sequences that displays the function of a promoter. In this paper, we have concentrated our attention on the 5′-flanking sequence. Upstream of the CAP site are GC boxes that have shown to be binding sites for the transcription factor Sp1 (36, 37). GC boxes (GGCGGG or CCGCCC) are found in many other promoters. In some of these, as in the PCNA gene, the GC boxes are accompanied by the absence of a TATA box, notably in the Ha-*ras*-1 gene (38, 39), the N-*ras*-gene (40), the c-Ki-*ras* gene (24), the ribosomal protein promoters (41), and the human epidermal growth factor receptor gene (38).

The 5′-flanking region of the PCNA gene has promoter function, *i.e.* it can drive linked reporters in expression assays after transfection of the appropriate chimeric plasmids. Analysis of the deletion mutations given in this paper will serve as a basis to identify the specific sequences that make this gene growth regulated. For the moment, we can say that 395 bp of 5′-flanking sequence (*Pvu*II promoter) are sufficient for the full expression of a linked reporter. The activity of the *Hin*dIII promoter was always lower in this assay, suggesting the presence of negative regulatory elements between −557 and 396, as already reported in several other promoters (reviewed in Ref. 42). At −341, there is a perfect octamer motif ATTTGCAT which is found in disparate promoters such as those of immunoglobulin, histone H2B, U2 small nuclear RNA (43), herpes simplex virus thymidine kinase (44), and others. There are other recognizable motifs like a binding site for the transcription factor AP-2 (45), an inverted CCAAT sequence, and a glucocorticoid responsive element (46).

Another interesting feature of the PCNA gene is the sequence similarities between introns and between introns and exons. The meaning of these similarities escapes us at present, but it is a common assumption among geneticists that conserved sequences indicate some kind of functional significance.

In summary, we have presented the structure and functional analysis of the human PCNA gene which, because of its inducibility by growth factors and its role in DNA synthesis and cell cycle progression must be considered as one of the key genes in the control of cellular proliferation in animal cells.

REFERENCES

1. Miyashi, K., Fritzler, M. J., and Tan, E. M. (1978) *J. Immunol.* **121,** 2228–2234
2. Takasaki, J., Deng, J. S., and Tan, E. M. (1981) *J. Exp. Med.* **154,** 1899–1909
3. Takasaki, J., Fischwild, D., and Tan, E. M. (1984) *J. Exp. Med.* **159,** 981–992
4. Bravo, R. (1986) *Exp. Cell Res.* **163,** 287–293
5. Mathews, M. B., Bernstein, R. M., Franza, B. R., and Garrels, J. I. (1984) *Nature* **309,** 374–376
6. Prelich, G., Kostura, M., Marshak, D. R., Mathews, M. B., and Stillman, B. (1987) *Nature* **326,** 471–475
7. Prelich, G., Tan, C. K., Kostura, M., Mathews, M. B., So, A. G., Downey, K. M., and Stillman, B. (1987) *Nature* **326,** 517–520
8. Bravo, R., Frank, R., Blundell, P. A., and Macdonald-Bravo, H. (1987) *Nature* **326,** 515–517
9. Prelich, G., and Stillman, B. (1988) *Cell* **53,** 117–126
10. Jaskulski, D., DeRiel, J. K., Mercer, W. E., Calabretta, B., and Baserga, R. (1988) *Science* **240,** 1544–1546
11. Tan, C., Castillo, C., So, A. G., and Downey, K. M. (1986) *J. Biol. Chem.* **261,** 12310–12316
12. Almendral, J. M., Huebsch, D., Blundell, P. A., MacDonald-Bravo, H., and Bravo, R. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 1575–1579
13. Jaskulski, D., Gatti, G., Travali, S., Calabretta, B., and Baserga, R. (1988) *J. Biol. Chem.* **263,** 10175–10179
14. Matsumoto, K., Moriuchi, T., Koji, T., and Nakane, P. K. (1987) *EMBO J.* **6,** 637–642
15. Maniatis, T., Fritch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
16. Gross-Bellard, M., Oudet, P., and Chambon, P. A. (1973) *Eur. J. Biochem.* **36,** 32–38
17. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517
18. Feinberg, A. P., and Vogelstein, B. (1983) *Anal, Biochem.* **132,** 6–13
19. Ferrari, S., Calabretta, B., deRiel, J. K., Battini, R., Ghezzo, F., Lauret, E., Griffin, C., Emanuel, B. S., Gurrieri, F., and Baserga, R. (1987) *J. Biol. Chem.* **262,** 8325–8332
20. Shen, Y. M., Hirschhorn, R. R., Mercer, W. E., Surmacz, E.,

Tsutsui, Y., Soprano, K. J., and Baserga, R. (1982) *Mol. Cell. Biol.* **2,** 1145–1154
21. Sood, A. K., Pereira, D., and Weissman, S. M. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 616–620
22. Reeder, R. H., Roan, J. G., and Dunaway, M. (1983) *Cell* **35,** 449–456
23. Okayama, H., and Berg, P. (1983) *Mol. Cell. Biol.* **3,** 280–289
24. Hoffman, E. K., Trusko, S. P., Freeman, N., and George, D. L. (1987) *Mol. Cell. Biol.* **7,** 2592–2596
25. Munn, T. Z., and Mues, G. I. (1988) *Nature* **332,** 789
26. Bradshaw, H. D., and Deininger, P. L. (1984) *Mol Cell. Biol.* **4,** 2316–2320
27. Tan, C. K., Sullivan, K., Li, X., Tan, E. M., Downey, K. M., and So, A. G. (1987) *Nucleic Acids Res.* **15,** 9299–9308
28. Celis, J. E., and Celis, A. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 3262–3266
29. Bravo, R., and MacDonald-Bravo, H. (1985) *EMBO J.* **4,** 655–611
30. Bravo, R., and MacDonald-Bravo, H. (1984) *EMBO J.* **3,** 3177–3181
31. Celis, J. E., Madsen, P., Neilsen, S. V., Gesser, B., Nielsen, H. V., Petersen, O., Ratz, G., Lauridsen, J. B., and Celis, A. (1988) *Cancer Cells* **6,** 289–295
32. Moore, K., Sullivan, K., Tan, E. M., and Prystowsky, M. B. (1987) *J. Biol. Chem.* **262,** 8447–8450
33. Zerler, B., Roberts, R. J., Mathews, M. B., and Moran, E. (1987) *Mol. Cell. Biol.* **7,** 821–829
34. Hammond, R. A., Byrnes, J. J., and Miller, M. R. (1987) *Biochemistry* **26,** 6817–6824
35. So, A. G., and Downey, K. M. (1988) *Biochemistry* **27,** 4591–4595
36. Dynan, W. S., Saffer, J. D., Lee, W. S., and Tjian, R. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 4915–4919
37. McKnight, S., and Tjian, R. (1986) *Cell* **46,** 795–805
38. Ishii, S., Xu, Y. H., Stratton, R. R., Roe, B. A., Merlino, G. T., and Pastan, I. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 4920–4924
39. Ishii, S., Merlino, G. T., and Pastan, I. (1985) *Science* **230,** 1378–1381
40. Hall, A., and Brown, R. (1985) *Nucleic Acids Res.* **14,** 5255–5268
41. Dudov, K. P., and Perry, R. P. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 8545–8549
42. Rittling, S. R., and Baserga, R. (1987) *Anticancer Res.* **7,** 541–552
43. Parslow, T. G., Jones, S. D., Bond, B., and Yamamoto, K. R. (1987) *Science* **235,** 1498–1501
44. Sive, H. L., and Roeder, R. G. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 6382–6386
45. Mitchell, P. J., Wang, C., and Tjian, R. (1987) *Cell* **50,** 847–861
46. Chandler, V. L., Maler, B., and Yamamoto, K. R. (1983) *Cell* **33,** 489–499