
Amino acid sequence of the testosterone-regulated mouse kidney RP2 protein deduced from its complementary DNA sequence

Donna King¹, Y. Henry Sun^{2,3} and Jerry B. Lingrel*

Department of Microbiology and Molecular Genetics, and Program in Developmental Biology, University of Cincinnati College of Medicine, Cincinnati, OH 45267 and ³Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

Received 5 May 1986; Accepted 11 June 1986

ABSTRACT

The major forms of testosterone-regulated RP2 messenger RNA (also known as MAK mRNA and pMK908) in the mouse kidney were characterized by examining cDNA and genomic clones. Three sizes of RP2 mRNA are detected by Northern blot analysis and these were shown to result from polyadenylation at three distinct sites within the primary transcript of this single-copy gene. The complete RP2 mRNA sequence was obtained from overlapping cDNA clones, revealing an open reading frame of 357 amino acids that corresponds to a protein of 40,365 daltons. The detection of RP2 mRNA in all tissues examined to date suggests that the RP2 protein may function in a housekeeping role in all cells. This is supported by the finding of a high percentage of G + C residues at the 5' end of the gene, including a sequence homologous to the binding site of the transcription factor Sp1, which has been suggested to affect the regulation of other housekeeping genes that have been characterized. An examination of the amino acid sequence indicates that the RP2 protein is proline-rich and is composed of alternating alpha-helix and beta-sheet regions. RP2 is probably not integrated into a membrane structure in the cell as it does not appear to contain hydrophobic regions capable of spanning a membrane.

INTRODUCTION

RP2 messenger RNA was first identified as a testosterone-inducible gene product by the differential screening of a mouse kidney cDNA library using cDNA probes generated from uninduced and testosterone-induced kidney mRNA (1). The mRNA was originally termed pMK908 RNA after the cDNA clone that was used in the initial characterization. It was later given the more descriptive name MAK (mouse-androgen-kidney, ref. 2), whereas RP2 was used to designate a polymorphism in the gene (10). We use the term RP2 to avoid confusion in references to this gene and its mRNAs.

RP2 mRNA comprises approximately 0.1 to 0.2% of the polyadenylated RNA in mouse kidney. Stimulation by the hormone testosterone rapidly raises this level to nearly 1%, making RP2 a major product of the kidney (1,2). RP2 is one of only a few gene products known to be inducible in this tissue by testosterone. Others include β -glucuronidase (3), alcohol dehydrogenase (4), arginase (5), ornithine decarboxylase (6), KAP (7), and transferase II (8). Testosterone is also known to regulate RP2 mRNA

levels in the liver. Low amounts of RP2 mRNA have been detected in submaxillary gland, brain, heart, muscle, and testes tissues. Hormones produced or controlled by the pituitary gland have no effect on the regulation of RP2 mRNA levels in the kidney (2).

Two major sets of RP2 mRNAs are differentially expressed in the tissues that have been studied. Three RP2 species that are most abundant in kidney, liver, and submaxillary gland are 1350, 1450, and 1950 nt in length (2,9). These same three mRNAs are expressed at low levels in brain and muscle, but the more abundant RP2 mRNAs in these tissues are 1050, 1150, and 1650 nt long (2, D.K., unpublished observations). All of these different mRNAs are produced by the transcription of a single gene (2,10). Previous work has reported two different polyadenylation sites that are recognized to produce the 1450 and 1950 nt mRNAs in the kidney (9). In this paper we identify a third polyadenylation site, responsible for the generation of the 1350 nt species.

The hormonal regulation and the tissue-specific production of different size classes of mRNA from it make RP2 an interesting gene for study. We now present the complete sequence of the kidney mRNAs and the RP2 amino acid sequence as deduced from cDNA and genomic clones. This information allows preliminary conclusions to be drawn concerning the functional role of RP2 protein within the cell and provides essential groundwork for continued study of hormonal and tissue-specific mechanisms controlling gene expression.

MATERIALS AND METHODS

Clones.

cDNA clones were isolated from a library prepared from poly(A) cytoplasmic RNA from kidneys of DBA/LiHa mice induced with subcutaneous testosterone pellets (11). Screening of the cDNA library was described previously (2). Genomic clones were isolated from a cosmid library that was constructed from B10.D2-H-2^{dml} mouse liver DNA (12) by following the procedure of Steinmetz et al (13). The probe used in the genomic screening was the purified 1060 bp Bgl I fragment from the middle of cDNA clone pMAK-1 (see Figure 1). Sequencing of the clones was accomplished by the method of Maxam and Gilbert (14).

RNA Analysis.

Total cytoplasmic RNA was isolated from mouse kidneys as previously described (2). The probe for the S1 nuclease analysis of the 1350 and 1450 nt mRNA polyadenylation sites was the 412 bp Hinf I fragment that is shown in Figure 2. This fragment was 3' end-labeled with the Klenow fragment of DNA polymerase I and strand-separated (15). S1 nuclease analysis was performed by the method of Berk and

Sharp (16) as modified by Weaver and Weissmann (17). Total kidney RNA (10 μ g) was hybridized to excess probe at 50°C for 3.75 hrs. Nonhybridizing DNA probe was subsequently digested with 400 U of S1 nuclease (Sigma Chemical Co.) per ml at 18°C for 1 hour. The resulting protected end-labeled DNA fragments were electrophoresed on a 5% polyacrylamide -50% urea denaturing gel and autoradiographed.

The 5' end of kidney RP2 transcripts was identified by primer extension analysis. The 16-base long oligonucleotide 5'-CAGCTGCTCGAGCTGC-3', which was 5' end-labeled with [γ -³²P] ATP by polynucleotide kinase, was hybridized to 40 μ g of kidney cytoplasmic RNA from C57B1/6J mice at 50°C for 2 hrs. The RNA-DNA hybrids were ethanol precipitated, and the pellets were rinsed with cold 70% ethanol. Reverse transcription was performed with Moloney murine leukemia virus reverse transcriptase following the protocol recommended by the manufacturer (Bethesda Research Laboratories). Products of the reaction were electrophoresed on an 8% polyacrylamide-50% urea gel and autoradiographed.

A preparative-scale primer extension reaction was performed with 500 μ g of testosterone-induced C57B1/6J mouse total cytoplasmic kidney RNA. Hybridization occurred at 50°C over 40 hrs. Reverse transcription was performed with 200 units of the reverse transcriptase in a 100 μ l volume at 37°C for 1 hr. The resulting product was eluted from an 8% polyacrylamide-50% urea gel after it was identified by autoradiography. The DNA was then passed through silanized glass wool, phenol and chloroform extracted, ethanol precipitated, resuspended in distilled water, and subjected to Maxam and Gilbert sequencing reactions (14). The sequence of the first 61 nucleotides of the RP2 mRNA was determined in this manner.

Data Analysis.

The GenBank DNA sequence databank was searched for homologies to the RP2 cDNA sequence by Greg Wernke of the University of Cincinnati, using the Beckman Laboratories DNA analysis program. The RP2 amino acid sequence obtained from the cDNA was compared by Russell Doolittle to the contents of a protein databank maintained by himself at the University of California, San Diego. All other sequence analysis was performed with the International Biotechnologies, Inc. Pustell sequence analysis program.

RESULTS

Overlapping RP2 cDNA clones (Figure 1) were obtained from screenings of a DBA/LiHa testosterone-induced mouse kidney cDNA library (2,9). It is known that the sizes of RP2 mRNA in DBA/LiHa mouse kidney are 1350, 1450, and 2150 nt, while the sizes of RP2 mRNA in C57B1/6J mice are 1350, 1450, and 1950 nt. The difference in the sizes of the largest mRNA between these strains is due to the presence of a polymorphic B1 repetitive element in the 3' untranslated region in DBA mice (9). In

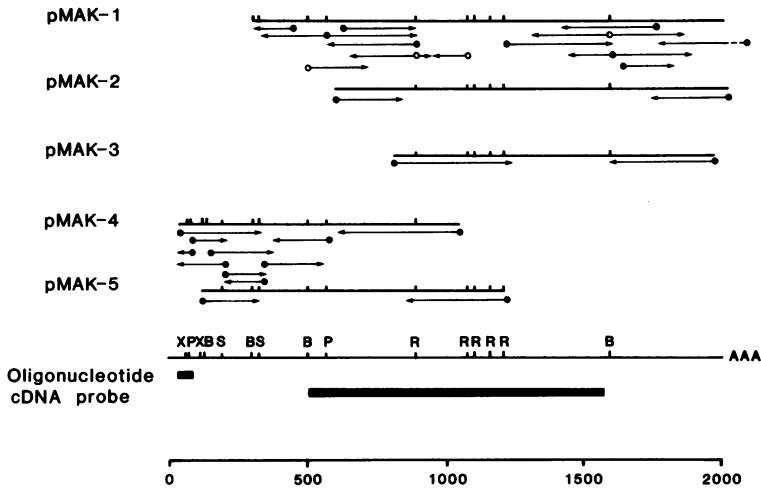


Figure 1

Restriction maps of overlapping RP2 cDNA clones from the DBA/LiHa library. Xho I, X; Sma I, S; Pvu II, P; Rsa I, R; Bgl I, B. The locations of an oligonucleotide used in primer-extension analysis and the 1060 bp Bgl I fragment used in the screen of the genomic library are also indicated. The sequencing strategy employed for each clone is shown below the maps. Closed circles indicate fragments sequenced from their 5' ends and open circles indicate those sequenced from their 3' ends. The dashed line represents flanking vector DNA that was also sequenced.

other respects the RP2 mRNAs from the two strains are the same. Taken together, these overlapping cDNA clones represent essentially the full length of the mRNAs in DBA mice. Their identical restriction maps reflect what was revealed at the nucleotide level; that these five cDNA clones represent mRNAs transcribed from the same gene.

The basis for the production of the 1450 and the 1950/2150 nt mRNAs from the single-copy RP2 gene was explained previously by the alternate use of two polyadenylation sites (9). This was determined by S1 nuclease mapping. Additional S1 nuclease mapping has now established that a third polyadenylation site is utilized to produce the 1350 nt mRNA. An S1 nuclease protection assay was performed with the 412 base long 3' end-labeled Hinf I fragment diagrammed in Figure 2. This fragment spans a region which includes the previously reported 1450 polyadenylation site. The band at 170 bases represents probe that was protected over this length by the 1350 nt mRNA. A band approximately 100 bases larger at position 265 corresponds to probe that was protected by the 1450 nt mRNA. Thus the two polyadenylation sites are separated from each other by an amount equivalent to the size differences observed on Northern blots. The same signals were generated from both DBA/2J and C57B1/6J kidney RNA, and stronger signals were seen when the RNA was prepared from

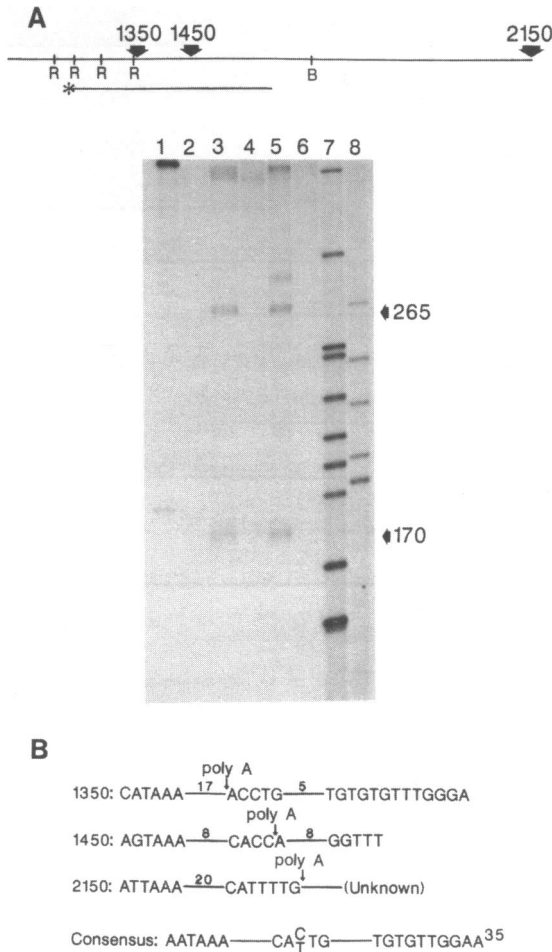


Figure 2

S1 nuclease analysis of multiple RP2 polyadenylation sites.

Part A. The 412 base 3' end-labeled Hinf I fragment probe spans two polyadenylation sites indicated on the map (arrows). This single-stranded probe was hybridized to equivalent amounts of testosterone-induced or uninduced total cytoplasmic kidney RNA from DBA/2J and C57B1/6J mice and subjected to S1 nuclease digestion. The resulting bands represent protection by the 1350 and 1450 nt RP2 mRNAs. The length of the poly (A) tails is assumed to be approximately 140 bases. Differences in signal intensities were also observed between testosterone-induced and uninduced female mice. This induction was previously determined to be 8-fold (2). Lanes: 1, probe alone; 2, yeast tRNA (negative control); 3, DBA/2J induced females; 4, DBA/2J uninduced females; 5, C57B1/6J induced females; 6, C57B1/6J uninduced females; 7 and 8, DNA size markers.

Part B. The three RP2 polyadenylation signals are compared to the consensus signal (18).

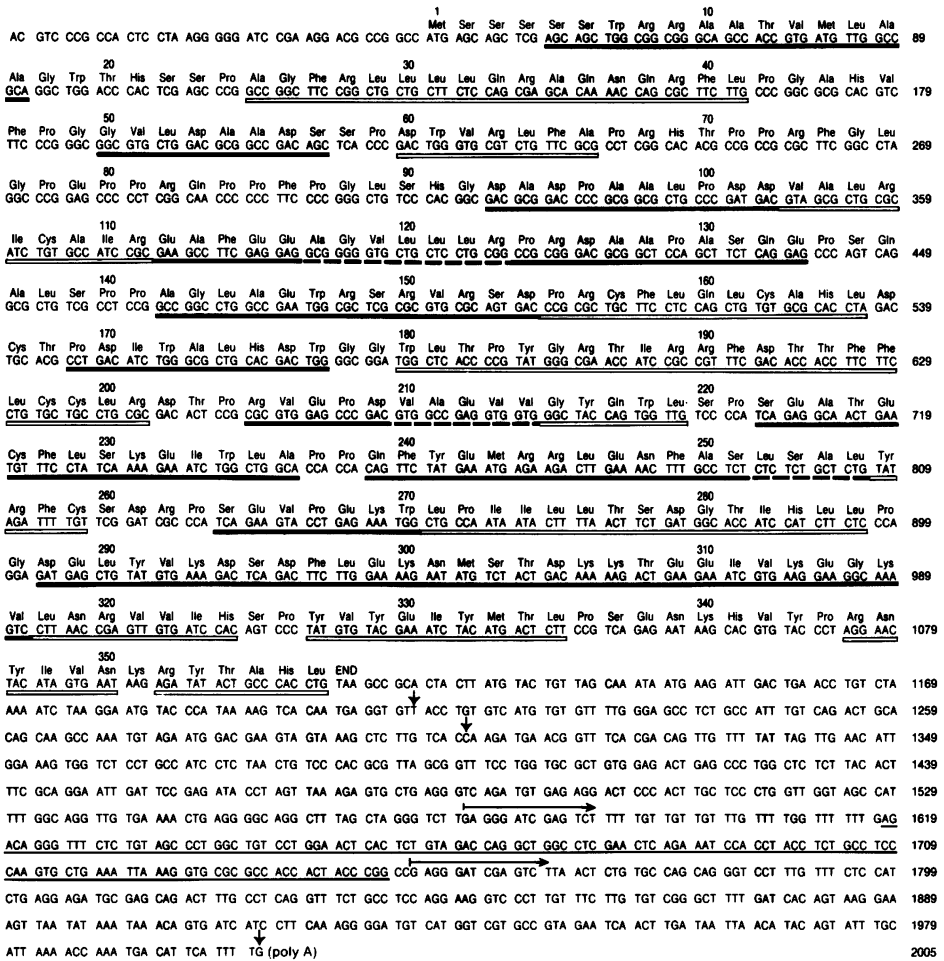


Figure 3

Complete sequence of DBA/LiHa mouse kidney RP2 mRNA and the amino acid sequence deduced from it. Regions calculated to likely form alpha-helix and beta-sheet secondary structures in the protein are indicated by shaded and open bars, respectively, below the sequence; three regions capable of forming either conformation are indicated by discontinuous bars. The three polyadenylation sites are indicated by vertical arrows. The B1 repetitive element is underlined and the 14 bp direct flanking repeats created by the insertion of the B1 element are overlined with arrows. C57B1/6J mice, representing the other form of the polymorphism, are precisely missing the B1 repeat, its 3' flanking region, and one of the 14 bp direct repeats. This has been sized as a 200 nt difference between strains in the largest RP2 mRNA. The 3' untranslated region from base 1184 has been published previously (9).

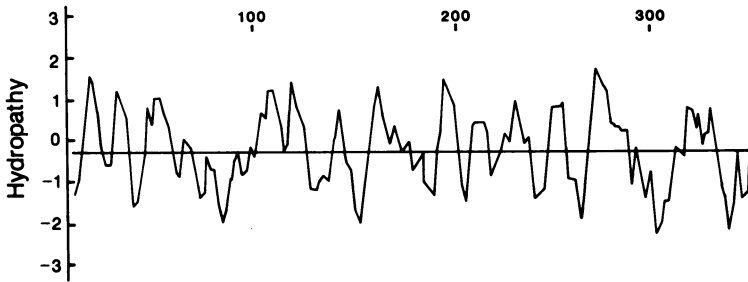


Figure 4

Hydropathy plot of RP2 protein sequence. Hydrophobic (plotted above the axis) and hydrophilic (below the axis) regions of the protein were plotted by computer over the length of the amino acid sequence. Each point represents the average polarity of nine amino acids spanning each position.

tetosterone-induced mice. A faint signal detected at position 285 in some lanes is not due to polyadenylation directed by any sequence known to function as a polyadenylation signal, nor does the sequence resemble an intron splice donor site.

Sequences surrounding the three RP2 polyadenylation sites are aligned with the polyadenylation consensus signal derived by Gil and Proudfoot (18) in part B of Figure 2. It is noted that none of the three sequences contains the very highly conserved portion of the consensus sequence, AATAAA (19). The three different variations of this sequence that are recognized as polyadenylation signals in the RP2 transcript have been reported to occur infrequently near the polyadenylation sites of other mRNAs, however (20,21).

The entire sequence of the largest RP2 mRNA from DBA mice is presented and translated in Figure 3. All three of the RP2 mRNAs detected in the kidney share the same amino acid coding region, differing only in the position of the polyadenylation site which is used. These are shown by vertical arrows at positions 1211, 1306, and 2005. The B1 repetitive element which creates the polymorphic size differences between mouse strains in the largest mRNA is underlined and is followed by the characteristic A-rich tract, represented here by T's because the element was inserted in the opposite transcriptional orientation with respect to RP2. The element is flanked by 14 bp directly repeated sequences (overlined arrows). A relatively short 5' untranslated region precedes two potential methionine initiation codons. Initiation at the first of these (as shown) is likely to be favored because the surrounding sequence more closely resembles the translation initiation consensus sequence described by Kozak (22). Initiation at the in-frame methionine codon at nucleic acid position 81 is the alternative possibility. This question can only be unequivocally resolved by the purification and amino acid sequence analysis of the RP2 protein. An open reading

frame of 357 codons follows the initial methionine, resulting in a polypeptide of 40,365 daltons. This agrees with the previous sizings of the *in vitro* synthesized polypeptide as 42,000 daltons (2) and 43,000 daltons (1). The amino acid sequence is rich in proline residues (9.8%) which undoubtedly have a significant effect on the secondary structure of the protein by introducing bends and distorting helical regions. Sequences that have a high probability of forming alpha-helix and beta-sheet structures (23) are indicated by shaded bars and open bars, respectively, in Figure 3. The regions which could form either structure are indicated by discontinuous bars. Those structures with a minimum length of 7 amino acids are shown. The N-terminus of the amino acid sequence was searched for similarities to signal sequences that direct the insertion of nascent polypeptides into membranes (24), but none could be identified. This suggests that the RP2 protein is not sequestered by the endoplasmic reticulum.

The amino acid sequence was also examined by hydropathy analysis (Figure 4). The polarity of the RP2 amino acid sequence was averaged over nine residues at a time and plotted over the full length of the sequence according to the algorithm of Kyte and Doolittle (25). None of the hydrophobic regions (indicated above the axis) are long enough or hydrophobic enough (a minimum of 19 residues with an average hydropathy greater than 1.6) to be capable of spanning a membrane. The lack of a transmembrane region and the lack of a signal sequence both indicate that the RP2 protein is probably not directly associated with membrane structures within the cell.

In order to determine whether the cDNA clones truly represent the entire length of RP2 mRNA, primer extension analysis was performed. This enabled us to determine the number and sequence of any missing bases at the 5' end. A 16 base long oligonucleotide was synthesized based on sequence at the 5' end of pMAK-4, the cDNA clone which extends farthest in that direction. The oligonucleotide was complementary to the coding strand starting at base 61 in Figure 3 and its position relative to the rest of the cDNA is indicated in Figure 1. The oligonucleotide was labeled with ^{32}P at its 5' end by T4 polynucleotide kinase and was hybridized to RNA from both testosterone-induced and uninduced C57B1/6J mouse kidneys. Reverse transcriptase was then used to extend the oligonucleotide primer to the mRNA cap site. When the reaction products were electrophoresed on a denaturing gel and autoradiographed, the distance from the end-labeled base of the oligonucleotide to the cap site could be measured. This data is shown in Figure 5. A strong band at 61 bases is generated from testosterone-induced kidney RNA whereas a much weaker band is seen at the same position when an equivalent amount of RNA from uninduced mouse kidney is used. Identical results were obtained using kidney RNA from DBA/2J mice (data not shown). Very faint bands in both lanes at approximately 200 bases do not reflect the induction and were not present in all repetitions of this experiment.

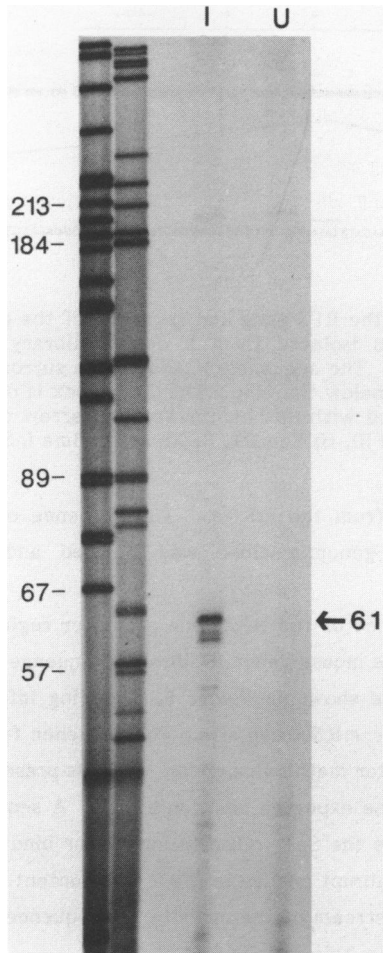


Figure 5

Analysis of the 5' end of RP2 mRNA by primer extension of an oligonucleotide. A 16 base long oligonucleotide complementary to the 5' end of the mRNA was synthesized based on sequence from the cDNA clone pMAK-4 (see figure 1). The oligonucleotide was labeled with ^{32}P at its own 5' end and hybridized to equivalent amounts of testosterone-induced and uninduced kidney RNA. The hybrids were then extended to the mRNA cap site by reverse transcriptase and the resulting product sized on a polyacrylamide gel. An autoradiograph is shown. The band at 61 bases indicates the site of initiation of transcription of all three major kidney RP2 mRNAs.

The primer extension result indicated that 28 bases at the 5' end of the mRNA were missing from our cDNA clones. A preparative-scale primer extension reaction was performed on testosterone-induced kidney RNA and the resulting product was subjected to Maxam and Gilbert sequencing. This allowed the determination of the

capable of spanning a membrane. Therefore, this testosterone-inducible gene product most likely exists in soluble form within cells of the kidney.

Searches were made of the available DNA and protein sequence databanks to determine whether an alignment could be made which would identify or further characterize the function of RP2. The results indicate that the sequence presented here is the first to be reported for this gene and its product. No significant homologies were found between the coding region of the RP2 mRNA and the contents of the GenBank DNA sequence databank. Similarly, the search of an amino acid sequence databank maintained by Russell Doolittle at the University of California, San Diego, revealed no alignment with any previously sequenced protein. The best match to RP2 that was observed showed some similarity (15 out of 55 amino acids) to the gag protein of a feline sarcoma virus (28). The determination of the amino acid sequence facilitates continued investigation of the role of RP2 within the kidney; moreover, polypeptides synthesized based on this information can now be used to generate antibodies, enabling the purification and further characterization of large quantities of the protein.

Several pieces of information suggest that RP2 protein may be functional in some form in all cells. RP2 mRNA has been detected in all the tissues that have been examined for its presence (kidney, liver, submaxillary gland, heart, brain, muscle, and testes, ref. 2), although the sizes of the mRNAs vary. This variation suggests that there are likely to be tissue-specific RNA processing mechanisms involved in addition to the polyadenylation sites that are utilized in the kidney. The isolation of RP2 genomic sequences will help to address this question in the future. For the present, it is known that the promoter region of this gene shares two characteristics with other genes that function in a ubiquitous, or housekeeping role. A region which resembles the binding site of the transcription factor Sp1 has been identified in the 5' flanking region, and 5' region of the mRNA is notably (68%) rich in G + C residues. Sequenced housekeeping genes with these features include mouse APRT (29), mouse HPRT (30), human DHFR (31), human triosephosphate isomerase (32), and hamster HMG CoA reductase (33).

ACKNOWLEDGEMENTS

The authors wish to acknowledge the assistance of Greg Wernke and Russell Doolittle with the computerized sequence analysis, Elizabeth Paine for help with the genomic screening, Steven Shapiro and Gary Shull for valuable discussion, and Mike Hughes for synthesis and purification of the oligonucleotide. Henry Sun is the recipient of an Earle C. Anthony Fellowship. This work was supported by American Cancer Society grant BC-480 and National Science Foundation Instrumentation Grant DMB-8414251.

¹Present address: Reproduction Laboratory, Building 200, Beltsville Agricultural Research Center, Beltsville, MD 20705, USA

²Present address: Department of Biology, Yale University, New Haven, CT 06511, USA

*To whom reprint requests should be addressed

REFERENCES

1. Berger, F.G., Gross, K.W. and Watson, G. (1981) *J. Biol. Chem.* 256, 7006-7013.
2. Snider, L.D., King, D. and Lingrel, J.B. (1985) *J. Biol. Chem.* 260, 9884-9893.
3. Swank, R.T., Paigen, K. and Ganschow, R.E. (1973) *J. Molec. Biol.*, 225-243.
4. Ohno, S., Stenius, C. and Christian, L.C. (1970) *Clin. Genet.* 1, 35-44.
5. Kochakian, C.D. (1945) *J. Biol. Chem.* 161, 115-125.
6. Bullock, L.P. (1983) *Endocrinology* 112, 1903-1909.
7. Toole, J.J., Hastie, N.D. and Held, W.A. (1979) *Cell* 17, 441-448.
8. Kochakian, C.D. and Mayumi, T. (1977) *Mol. Cell. Endo.* 6, 309-318.
9. King, D., Snider, L.D. and Lingrel, J.B. (1986) *Mol. Cell. Biol.* 6, 209-217.
10. Elliott, R.W. and Berger, F.G. (1983) *Proc. Natl. Acad. Sci.* 80, 501-504.
11. Palmer, R., Gallagher, P.M., Boyko, W.L. and Ganschow, R.E. (1983) *Proc. Natl. Acad. Sci.* 80, 7596-7600.
12. Sun, Y.H., Goodenow, R.S. and Hood, L. (1985) *J. Exp. Med.* 162, 1588-1602.
13. Steinmetz, M., Winoto, A., Minard, K. and Hood, L. (1982) *Cell* 28, 489-498.
14. Maxam, A.M. and Gilbert, W. (1980) *Meth. Enzymol.* 65, 499-560.
15. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual.* Cold Spring Harbor Laboratory.
16. Berk, A.J. and Sharp, P.A. (1977) *Cell* 12, 721-732.
17. Weaver, R.F. and Weissmann, C. (1979) *Nucl. Acids Res.* 7, 1175-1193.
18. Gil, A. and Proudfoot, N.J. (1984) *Nature* 312, 473-474.
19. Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature* 263, 211-214.
20. Whitelaw, E. and Proudfoot, N.J. (1983) *Nucl. Acids Res.* 11, 7717-7733.
21. Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) *Cell* 41, 349-359.
22. Kozak, M. (1984) *Nucl. Acids Res.* 12, 857-872.
23. Chou, P.Y. and Fasman, G.D. (1978) *Ann. Rev. Biochem.* 47, 251-276.
24. Watson, M.E.E. (1984) *Nucl. Acids Res.* 12, 5145-5164.
25. Kyte, J. and Doolittle, R.F. (1982) *J. Molec. Biol.* 157, 105-132.
26. Dynan, W.S. and Tjian, R. (1983) *Cell* 35, 79-87.
27. Blobel, G. and Dobberstein, B. (1975) *J. Cell. Biol.* 67, 835-851.
28. Hampe, A., Gobet, M., Sherr, C.J. and Galibert, F. (1984) *Proc. Natl. Acad. Sci.* 81, 85-89.
29. Dush, M.K., Sikela, J.M., Khan, S.A., Tischfield, J.A. and Stambrook, P.J. (1985) *Proc. Natl. Acad. Sci.* 82, 2731-2735.
30. Melton, D.W., Konecki, D.S., Brennand, J. and Caskey, C.T. (1984) *Proc. Natl. Acad. Sci.* 81, 2147-2151.
31. Yang, J.K., Masters, J.N. and Attardi, G. (1984) *J. Mol. Biol.* 176, 169-187.
32. Brown, J.R., Daar, I.O., Krug, J.R. and Maquat, L.E. (1985) *Mol. Cell. Biol.* 5, 1694-1706.
33. Reynolds, G.A., Basu, S.K., Osborne, T.F., Chin, D.J., Gil, G., Brown, M.S., Goldstein, J.L. and Luskey, K.L. (1984) *Cell* 38, 275-285.