Prostatic Steroid-binding Protein

ISOLATION AND CHARACTERIZATION OF C3 GENES*

(Received for publication, August 25, 1982)

Malcolm G. Parker, Roger White, Helen Hurst, Maurice Needham, and Rita Tilly

From the Imperial Cancer Research Fund, P. O. Box 123, Lincoln’s Inn Fields, London WC2A 3PX, England

Prostatic steroid-binding protein, whose expression is stimulated by androgens, consists of two subunits, one containing the polypeptides C1 and C3 and the other containing the polypeptides C2 and C3. We have isolated and sequenced cDNA clones specific for C3 mRNA and used them to isolate and characterize genomic clones for two C3 genes. Both genes are 3.2 kilobases with identical exon/intro arrangements, which is similar to the organization of the C1 and C2 genes, suggesting that they may have arisen by duplications of an ancestral gene. Finally, homologous human genes have not been detected.

Prostatic steroid binding protein is the principal secretory protein in rat prostatic fluid (1-6). It is a tissue-specific protein whose expression is regulated by androgenic steroids partly by stimulating rates of transcription and partly by effects on RNA stability (7-10). The protein is a tetramer, consisting of two subunits: one subunit containing the polypeptides C1 and C3 and the other subunit containing the polypeptides C2 and C3 (11).

Although we have previously reported the isolation of cDNA clones for C3 mRNA (12), the cDNA inserts obtained were smaller than the estimated C3 mRNA size. Therefore, we have now used an alternative procedure (13, 14) avoiding the use of S1 nuclease (15), which necessarily results in loss of DNA corresponding to the 5' end of the mRNA, to isolate full length clones of cDNA. One of these has been completely sequenced and used to isolate and characterize chromosomal DNA containing the C3 genes. In view of the composition of prostatic steroid-binding protein we wish to quantify C3 genes and investigate their organization so that ultimately we can study their interaction with androgen receptors and understand the mechanism by which steroid hormones regulate gene expression.

**EXPERIMENTAL PROCEDURES**

**Materials**—Animals and most chemicals were as previously described (7, 16). The following materials were generous gifts: reverse transcriptase (J. W. Beard, Life Science Inc., St. Petersburg, FL); and a partial EcoRI and a partial HaeIII rat DNA library in bacteriophage λ Charon 4A (Drs. T. Sargent, R. Wallace, L. Jagodzinski, and J. Bonner, Caltech, Pasadena, CA).

**cDNA Cloning and Screening**—The approach of Rougeon et al. (17) was used as modified by Land et al. (14) was used as previously described (18). Double-stranded cDNA molecules were inserted into the PstI site of pAT 153 (17) and colonies were first screened by in situ hybridization (18). Recombinant plasmids were isolated (19) and bound to diazobenzoyloxyethyl paper (20) to identify them by mRNA purification. First, they were hybridized with total prostatic poly(A)-containing RNA (6 μg), washed, and the bound RNA was eluted and translated in a cell-free system derived from wheat germ (12).

**DNA Sequencing**—DNA sequence analysis of pA34 was carried out using the method of Maxam and Gilbert (21) by sequencing in both directions from the BstEII, XbaI, and BglIII restriction enzyme sites (Fig. 1).

**Genomic DNA Cloning and Screening**—Sprague-Dawley rat DNA, obtained from the ventral prostate of a single animal, known to contain four C3 related EcoRI fragments, was digested to completion with EcoRI and ligated into the purified arms of bacteriophage λgt WESS. (22). A partial EcoRI and partial HaeIII rat DNA library constructed with liver DNA from a single Sprague-Dawley rat was also used (23). Screening was by the method of Benton and Davis (24). Initially, 32P-labeled total cDNA was used as the DNA probe because we wished to isolate all three prostatic binding protein genes, namely C1, C2, and C3. After the clones were purified, they were distinguished using specific 32P-labeled DNA plasmids labeled by nick-translation (25).

**Restriction Enzyme Mapping**—Rat liver and prostate DNAs were isolated using the method of Blin and Stafford (26) and, in cases where digestion with restriction enzymes proved difficult, were further purified through CsCl gradients. Rat DNA (20 μg) and recombinant phage DNA were digested with restriction enzymes and separated by electrophoresis on agarose gels. Transfer to nitrocellulose was as described by Southern (27). Hybridization was carried out with nick-translated 32P-labeled DNA probes and, in the case of cell DNA blots, dextran sulphate (28) was included in the hybridization buffers.

**Analysis of R-flogs in the Electron Microscope**—DNA samples (A11B and A61) were hybridized at 25 μg/ml with prostate mRNA at 50 μg/ml in 70% formamide, 0.4 M NaCl, and 0.1 M 1,4-piperazinediethanesulfonic acid buffer, pH 7.2. After incubating for 1 h at 47.5°C, the sample was diluted and spread on H2O as described by Wahl et al. (29). The nucleic acids were visualized for electron microscopy by shadowing with platinum.

**Analysis of Nuclear RNA on Agarose Gels**—Rat prostate nuclei were prepared by the citric acid method described by Busch (30) and nuclear RNA was isolated as described by Roop et al. (31). The RNA samples were made 5 mM in methylmercury hydroxide and separated by electrophoresis on agarose gels. Transfer of RNA was carried out with 5% serum albumin as binding agent. The DNA was transferred to diazobenzoyloxyethyl paper as previously described (12, 32). Hybridization with 32P-labeled DNA and autoradiography were then carried out (12).

**RESULTS**

**Identification and Characterization of C3 cDNA Clones**—Total prostate cDNA was cloned in the PstI site of the plasmid pAT 153, and tetracycline resistant colonies, which contained prostatic DNA sequences as shown by in situ hybridization, were selected for further study. Individual plasmids were identified by mRNA purification, and restriction enzyme analysis suggested that four contained C3 cDNA of approximately 600 nucleotides, which is similar to the size estimated for C3 mRNA (12). One clone, pA34, was selected for DNA sequencing (Fig. 1). From the DNA sequence, it was possible to predict a polypeptide sequence and this agreed completely with the sequence of the secreted C3 polypeptide (32). In addition, it is likely that the protein is translated with a so-called signal peptide of 18 amino acids with an AUG start codon at nucleotide position 55. Thus, C3 mRNA comprises a coding region of 285 nucleotides with a UAA termination
of restriction enzyme analysis and R-loop analysis. A restriction enzyme map of C3 cDNA in pA34 is also shown. PstI, RI represent exons and the following restriction enzymes were used: B

two C3 genes in X61 and XI 1B plus X5D were constructed on the basis

hgt
tail. 3' ["P"]pA34. Preliminary restriction enzyme analysis of the clones indicated that they covered similar regions of the rat poly(A) addition signal 13 nucleotides preceding the poly(A) exon. that most rats possess other C3 related sequences (Fig. 3). Structured our own nonamplified EcoRI library in bacteriophage Charon 4A (23). By screening 600,000 plaques of each DNA library we identified thirteen EcoRI clones and three HueIII clones which hybridized with ["P"]pA34. The positions of HindIII cut λ fragments are shown in track M and their sizes are shown alongside.

Characterization of C3 Genes—The organization of the C3 genes in representative clones Λ11B, Δ5D, and Λ61 and various derivative plasmid subclones has been investigated by restriction enzyme digestion and R-loop hybridization and by analysing R-loops in the electron microscope.

The cleavage sites of EcoRI, BamHI, BglII, BstEII, HindIII, XbaI, PstI, and MspI were mapped and the DNA fragments that contain coding sequence were identified by hybridizing blots with ["P"]-labeled pA34 which indicated the C3 genes consisted of three exons separated by two intervening sequences (Fig. 2). The orientation of the gene was obtained by using a 5' and 3' specific DNA probe. The cDNA clone pA34 was digested with PstI to yield 140 nucleotide and 450 nucleotide DNA fragments which represent the 5' and 3' ends of the mRNA respectively (Fig. 2). The exon/intron arrangement was confirmed with Λ61 by electron microscopy. Measurements of 28 R-loops (Fig. 4) indicated that the gene was some 8.72 ± 0.10 kb and 0.77 ± 0.08 kb.

In contrast, analysis of R-loops formed with Λ11B only partially confirmed the Southern blotting data and suggested that two exons of 0.15 and 0.20 kb were separated by an intervening sequence of 2.9 kb. However, a portion of mRNA appeared to form a “bridge” between the two exons which presumably is due to mismatching with the middle exon in Λ11B and results in failure to form an R-loop. It is noteworthy that although we detect this middle exon on blots, a comparison of an MspI blot of the two genomic clones shows that the C3 genes consisted of three exons separated by two intervening sequences (Fig. 2).

Isolation of C3 Genomic Clones—Initially, we isolated genomic clones from two amplified Sprague-Dawley rat DNA libraries which were constructed from a partial EcoRI digest and a partial HaeIII digest of liver DNA, cloned into the genomes of bacteriophage λ. The cDNA libraries which were constructed from a partial EcoRI digest and a partial HaeIII digest of liver DNA, cloned into the genomes of bacteriophage λ were partially confirmed the Southern blotting data and suggested that two exons of 0.15 and 0.20 kb were separated by an intervening sequence of 2.9 kb. However, a portion of mRNA appeared to form a “bridge” between the two exons which presumably is due to mismatching with the middle exon in Λ11B and results in failure to form an R-loop. It is noteworthy that although we detect this middle exon on blots, a comparison of an MspI blot of the two genomic clones shows that the C3 genes consisted of three exons separated by two intervening sequences (Fig. 2).

One possible explanation for failing to isolate clones containing such sequences was that they had been lost during amplification of the original libraries and therefore we have constructed our own nonamplified EcoRI library in bacteriophage λgt WES. Thus far, we have identified one additional 12.5 Kb clone (Λ61) which is distinct from those previously isolated.

1 The abbreviation used is: kb, kilobases.

Fig. 1. Nucleotide sequence of C3 cDNA (pA34) and deduced amino acid sequence. Nucleotide numbers are shown above the sequence and the putative cleavage point of the signal peptide is represented by I.

Fig. 2. Restriction enzyme maps of C3 genes. The maps of two C3 genes in Λ61 and Λ11B plus Δ5D were constructed on the basis of restriction enzyme analysis and R-loop analysis. A restriction enzyme map of C3 cDNA in pA34 is also shown. Solid blocks represent exons and the following restriction enzymes were used: B

= BamHI, Bgl = BglII, Bst = BstEII, H = HindIII, M = MspI, P = PstI, RI = EcoRI, X = XbaI. codon, a 5' noncoding region of at least 55 nucleotides, and a 3' noncoding region of 170 nucleotides with a ATTAAA poly(A) addition signal 13 nucleotides preceding the poly(A) tail.

Isolation of C3 Genomic Clones—Initially, we isolated genomic clones from two amplified Sprague-Dawley rat DNA libraries which were constructed from a partial EcoRI digest and a partial HaeIII digest of liver DNA, cloned into the genomes of bacteriophage λ. The cDNA libraries which were constructed from a partial EcoRI digest and a partial HaeIII digest of liver DNA, cloned into the genomes of bacteriophage λ were partially confirmed the Southern blotting data and suggested that two exons of 0.15 and 0.20 kb were separated by an intervening sequence of 2.9 kb. However, a portion of mRNA appeared to form a “bridge” between the two exons which presumably is due to mismatching with the middle exon in Λ11B and results in failure to form an R-loop. It is noteworthy that although we detect this middle exon on blots, a comparison of an MspI blot of the two genomic clones shows that the C3 genes consisted of three exons separated by two intervening sequences (Fig. 2).

One possible explanation for failing to isolate clones containing such sequences was that they had been lost during amplification of the original libraries and therefore we have constructed our own nonamplified EcoRI library in bacteriophage λgt WES. Thus far, we have identified one additional 12.5 Kb clone (Λ61) which is distinct from those previously isolated.

1 The abbreviation used is: kb, kilobases.
**Steroid-binding Protein Genes**

Fig. 4. Analysis of R-loops in the electron microscope. Electron micrographs (A, C, E, and G) and schematic representations (B, D, F, and H) show R-loops formed between prostate RNA and λ61 (A–D) and λ11B (E–H). The thick solid line represents double-stranded DNA, the thin line represents single-stranded DNA, and the dashed line represents RNA.

![Image](image-url)

**Fig. 5. Northern blot of C3 nuclear RNA.** [3H]pA34 was hybridized with prostatic nuclear RNA which had been covalently bound to diazobenzyloxymethyl paper after electrophoresis. RNA was from rats castrated for 6 days (track 1), castrated rats treated with testosterone for 1 h (track 2) or 24 h (track 3) and normal rats (track 4). The arrows indicate the position of 18S and 28S rRNA.

![Image](image-url)

**DISCUSSION**

We conclude from R-loop analysis and Southern blotting data that there are two C3 genes per haploid genome, both of which we have isolated from rat DNA libraries. However, in noninbred strains of rat, such as Sprague-Dawley, it is often difficult to decide whether a particular gene is represented by multiple copies or whether it exhibits polymorphism without doing breeding experiments. It is conceivable that both C3 genes are expressed in vivo to produce the two C3 polypeptide chains in prostatic steroid-binding protein. However, the R-loop analysis leads us to think that the 11B gene is not transcribed to a comparable level with the 61 gene, thus, the latter is probably responsible for the production of both C3 polypeptide chains. It should be noted though, that only the C3 polypeptide in the S subunit of prostatic binding protein has been sequenced (32) and its similarity with the chain in the F subunit is based solely on electrophoretic mobility (11).

The similarity in exon/intron arrangements in λ61 and λ11B and the DNA sequence homologies, at least in the 3' and 3' exons, suggest that the two genes have arisen from the duplication of an ancestral gene. Interestingly, the C1 and C2 genes also have similar exon/intron arrangements to one another and share considerable DNA sequence homologies which suggests that they also arose by gene duplication (16, 34). More remarkable is the fact that these two genes and the C3 genes are all 5.2 kb and contain three exons separated by introns of 1.7–1.8 kb at the 5' end and 0.8–0.9 kb at the 3' end. Obviously, it is conceivable that these two pairs of genes are themselves
derived by duplication of an ancestral gene but if this is the case, there has been considerable divergence between the C3 genes on the one hand, and the C1 and C2 genes on the other hand. Comparison of the DNA sequence of pA34 with the cDNA sequences of C1 and C2 (16) indicate several homologies. For example, in the first exon, nucleotides 58-79 in C1 share 75% homology with nucleotides 61-82 in C3 and in the second exon, nucleotides 210-235 in C2 share 76% homology with 229-251 in C3 but these homologies are not as extensive as those found between C1 and C2. It is also possible that all four genes are linked but we have no evidence for this inasmuch as there is no overlap between genomic clones.

Finally, it is surprising that we failed to isolate genomic clones which contained C3 homologous EcoRI bands in cell DNA of 10.5 and 11 kb since they should have been efficiently packaged in Agt WES. We believe these fragments contain alleles of the genes isolated and are constructing and screening other nonamplified libraries to investigate this. Also, we are sequencing λ61 and λ11B to compare the two C3 genes and confirm whether or not λ61 can encode C3.

Acknowledgments—We are extremely grateful to Drs. T. Sargent, R. B. Wallace, L. L. Jagodzinski, and J. Bonner for providing us with the rat DNA libraries and to Dr. C. Livingston for providing methylmercury hydroxide and M. Barier for typing the manuscript.

REFERENCES
1. Fang, S., and Liao, S. (1971) J. Biol. Chem. 246, 16-24
2. Heyns, W., and De Moor, P. (1977) Eur. J. Biochem. 78, 221-230
3. Parker, M. G., Scrace, G. T., and Mainwaring, W. I. P. (1978) Biochem. J. 170, 115-121
4. Forsgren, B., Bjork, P., Carlstrom, K., Gustafsson, J.-A., Pousette, A., and Hogberg, B. (1979) Proc. Natl. Acad. Sci. U. S. A. 76, 3149-3153
5. Lea, O. A., Petrusz, P., and French, F. S. (1979) J. Biol. Chem. 254, 6196-6202
6. Chen, C., Hirpakha, R. A., and Liao, S. (1979) J. Steroid Biochem. 11, 401-405
7. Parker, M. G., and Mainwaring, W. I. P. (1977) Cell 12, 401-407
8. Heyns, W., Peeters, B., and Mous, J. (1977) Biochem. Biophys. Res. Commun. 77, 1492-1499
9. Parker, M. G., and Scrace, G. T. (1979) Proc. Natl. Acad. Sci. U. S. A. 76, 1580-1584
10. Page, M. J., and Parker, M. G. (1982) Mol. Cell. Endocr. 27, 343-355
11. Heyns, W., Peeters, B., Mous, J., Rombouts, W., and De Moor, P. (1978) Eur. J. Biochem. 89, 181-186
12. Parker, M. G., White, R., and Williams, J. G. (1980) J. Biol. Chem. 255, 6996-7001
13. Rougeon, F., Kourilsky, P., and Mach, B. (1975) Nucleic Acids Res. 2, 2365-2378
14. Land, H., Grez, M., Hauser, H., Lindennmaier, W. and Schutz, G. (1981) Nucleic Acids Res. 9, 2251-2256
15. Maniatis, T., Kee, S. G., Efstradiatis, A., and Kafatos, F. C. (1976) Cell 6, 163-182
16. Parker, M., Needham, M., and White, R. (1982) Nature (Lond.) 298, 92-94
17. Twigg, A. J., and Sherratt, D. (1980) Nature (Lond.) 293, 216-218
18. Grunstein, M., and Henges, D. S. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 3961-3965
19. Birnboim, H. C., and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523
20. Alwine, J. C., Kemp, D. J., and Stark, G. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5350-5354
21. Maxam, A. H., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 560-564
22. Leder, P., Tiemeier, D., and Enquist, L. (1977) Science 196, 175-177
23. Sargent, T. D., Wu, J.-R., Sala-Trepat, J. M., Wallace, R. B., Reyes, A. A., and Bonner, J. (1979) Proc. Natl. Acad. Sci. U. S. A. 76, 3256-3260
24. Benton, W. D., and Davis, R. W. (1977) Science 196, 180-182
25. Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. (1977) J. Mol. Biol. 113, 237-251
26. Blin, N., and Stafford, D. W. (1976) Nucleic Acids Res. 3, 2303-2308
27. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517
28. Wahl, G. M., Stern, M., and Stark, E. R. (1979) Proc. Natl. Acad. Sci. U. S. A. 76, 3683-3687
29. Wahl, W., Wyler, T., Weber, R., and Ryfell, G. U. (1978) Eur. J. Biochem. 86, 225-234
30. Busch, H. (1967) Methods Enzymol. 12A, 434-439
31. Roop, D. R., Nordstrom, J. L., Tsai, S. Y., Tsai, M.-J., and O'Malley, B. W. (1978) Cell 15, 671-685
32. Peeters, B., Rombouts, W., Mous, J., and Heyns, W. (1983) Eur. J. Biochem. 115, 115-121
33. Bjork, P., Forsgren, B., Gustafsson, J.-A., Pousette, A., and Hogberg, B. (1982) Cancer Res. 42, 1935-1942
34. Parker, M., Needham, M., White, R., Hurst, H., and Page, M. (1982) Nucleic Acids Res. 10, 5121-5132