Volume 10 Number 17 1982 Nucleic Acids Research

Prostatic steroid binding protein: organisation of C1 and C2 genes

Malcolm Parker, Maurice Needham, Roger White, Helen Hurst and Martin Page

Imperial Cancer Research Fund, P.O. Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK

Received 29 July 1982; Accepted 17 August 1982

ABSTRACT

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 C2 and C3. We have characterised genomic clones containing the C1 and C2 genes by restriction enzyme analysis and DNA sequencing. Both genes are 3.2 Kb, have similar exon/intron arrangements and share considerable DNA sequence homologies in their coding regions, intervening sequences and 5' upstream DNA sequence which suggests that they have probably arisen from the duplication of an ancestral gene. The 5' termini of C1 and C2 mRNA have been mapped; the sequence TATAAA appears 30 nucleotides upstream but a CAAT-like sequence at -60 — -80 is absent. Finally, homologous human genes have not been detected.

INTRODUCTION

Prostatic steroid binding protein is the predominant protein synthesised and secreted by the rat ventral prostate (1,2) but, despite having steroid binding properties (3-6), its function *in vivo* is unknown. Nevertheless, since prostatic steroid binding protein expression is stimulated by androgenic steroids via increases in mRNA production it has been extremely useful in studying androgen action (1, 7-10).

The protein is a tetramer consisting of two subunits, one containing the polypeptides C1 and C3 and the other containing C2 and C3 (11). We have previously isolated cDNA clones for C1, C2 and C3 (9,12) and shown, in the case of C1 and C2, that the DNA sequence is in complete agreement with the amino acid sequence determined by Peeters *et al*. (13,14). It was striking that the cDNAs for C1 and C2 shared considerable DNA sequence homology and therefore it was likely that duplication of an ancestral gene followed by divergent evolution had occurred. In view of this and to investigate the influence of androgens on the expression of these genes we have isolated and characterised genomic DNA clones which contain the C1 and C2 genes.

0305-1048/82/1017-5121$2.00/0

MATERIALS AND METHODS

Materials

Animals and most chemicals were as previously described (7). Other reagents and suppliers were as follows:- T4 polynucleotide kinase (PL-Biochemicals Inc.); Eco RI, DNA polymerase and calf intestinal alkaline phosphatase (Boehringer); DNA ligase, restriction enzymes and M13 pentadecamer primer (New England BioLabs). The following materials were generous gifts: reverse transcriptase (J.W. Beard, Life Sciences Inc., U.S.A.); a partial Eco RI and a partial Hae III rat DNA library in bacteriophage λ Charon 4A (Drs. T. Sargent, B. Wallace, L. Jagodzinski and J. Bonner, Caltech, U.S.A.); E. coli JM 101 and M13 mp7 (J. Messing, University of Minnesota, U.S.A.).

Identification of genomic clones in rat DNA libraries

A partial Eco RI and a partial Hae III rat library was constructed with liver DNA from a single Sprague-Dawley rat (15) and screened using the method of Benton and Davis (16). Initially ^{32}P-labelled total cDNA was used as the DNA probe because we wished to isolate all three prostatic steroid binding protein genes, namely C1, C2 and C3. After the clones were purified they were distinguished using specific ^{32}P-cDNA plasmids labelled by nick-translation (17).

Restriction enzyme mapping

Rat liver and prostate DNAs were isolated using the method of Blin and Stafford (18) 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 (19). Hybridisation was carried out with nick-translated ^{32}P-labelled DNA probes (17) and, in the case of cell DNA blots, dextran sulphate (20) was included in the hybridisation buffers.

DNA sequencing

DNA sequencing was carried out using the method of Maxam and Gilbert (21) by 5' and 3' end-labelling DNA with [γ-^{32}P]-ATP and, in the case of the 5' end of C2, by subcloning into M13 mp7 (22) and sequencing by the methods of Sanger and colleagues (23-24). All sequences were analysed on thin acrylamide gels containing 7M urea (25).

RNA mapping by primer extension

Poly(A)-containing RNA and nuclear RNA were prepared from rat ventral prostate as previously described (7,10). Primer DNA fragments were

obtained for C1 and C2 from the cDNA clones pA13 and pB 44 (12). pA13 was digested with Bgl II, end-labelled with [γ-^{32}P]-ATP, digested with Hgi A1 and strand-separated on an acrylamide gel. A ^{32}P-labelled DNA fragment of 75 nucleotides was then eluted from the gel. The C2 primer was obtained by subcloning the 5' 115 nucleotide Dde I fragment from pB44 into the Hind III site of pAT153 as described by Donoghue and Hunter (26). After digestion with Hind III and end-labelling with [γ-^{32}P]-ATP, a 125 nucleotide ^{32}P-labelled DNA fragment was eluted from an acrylamide gel.

The single-stranded C1 DNA primer was incubated with RNA samples (0.01-1.0 µg) in 5 µl 0.12 M NaCl at 85°C for 5' and then transferred to 60°C for 1 hr whereas the double-stranded C2 DNA primer was incubated with RNA in 10 µl of 0.4 M NaCl, 1 mM EDTA, 40 mM PIPES pH 6.4 and 80% deionised formamide at 85°C for 5' and then transferred to 52°C overnight. After hybridisation, the primer DNAs were ethanol-precipitated and elongated in 10 µl buffer containing 50 mM Tris-HCl pH 8.3, 6 mM MgOAc, 6 mM dithiothreitol, 1.8 mM deoxynucleotide triphosphates and 5 units of reverse transcriptase. After 1 hr at 42°C the samples were denatured and run on a 6% acrylamide, 7 M urea sequencing gel (25).

RESULTS AND DISCUSSION

Isolation of cloned prostate C1 and C2 genes

To isolate C1 and C2 genomic clones we have screened two rat DNA libraries which were constructed from a partial Eco RI digest and a partial Hae III digest of Sprague-Dawley rat liver DNA cloned into the purified 'arms' of λ phage Charon 4A (15). By screening 600,000 clones of each DNA library, we identified one Eco RI clone (λ11A) and one Hae III clone (λ8D) which hybridised with ^{32}P-pA13 (C1) and one Eco RI clone (λ21B) which hybridised with ^{32}P-pB44 (C2).

Restriction enzyme maps were constructed using Eco RI, Bam HI, Hind III, Xba I, Msp I, Sac I and Pst I (Fig. 1) and these showed that λ11A and λ8D covered similar regions of the rat genome around the C1 gene while λ21B covered a region around the C2 gene. The DNA fragments which contained coding sequence were identified by hybridising blots with ^{32}P-labelled cDNA inserts derived from Pst I digestion of nick-translated pA13 and pB44 (Fig. 2). On the basis of these results we constructed Pst I and Bam HI subclones in pAT153 in order to map restriction enzyme sites in more detail and to analyse the gene by visualising R-loops in the electron microscope. For example, when a 2.8 Kb Pst I subclone that contained the

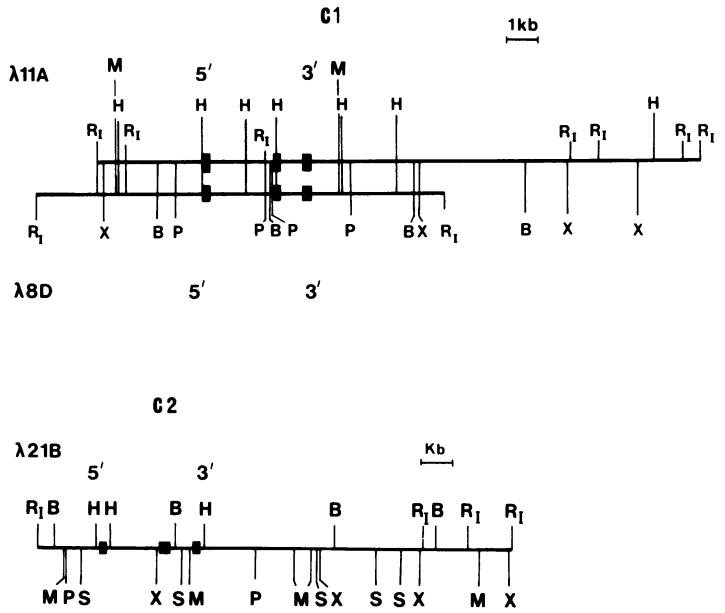

Fig. 1 Restriction enzyme maps of the C1 and C2 genes.
The maps were constructed on the basis of restriction enzyme analysis and blotting and R-loop analysis of λ11A and λ8D for C1 and λ21B for C2. Solid blocks represent exons. The following restriction enzymes were used: B, Bam HI; H, Hind III; M, Msp I; P, Pst I; RI, Eco RI; S, Sac I; X, Xba I.

3' portion of C1 was hybridised with prostate RNA, R-loops indicated that this DNA subclone contains two exons separated by an intervening sequence of 0.91 ± 0.05 Kb (unpublished data). The orientation of the genes was obtained by using 5' and 3' specific cDNA probes. For example, in the case of C1, pA13 was digested with Hind III and Pst I and the two DNA fragments which represent the 5' end and 3' end of the mRNA were separated by electrophoresis on an acrylamide gel. After nick-translation, the ^{32}P-labelled DNA fragments were hybridised with Hind III and Bam HI digests of the genomic clone. The 5' specific probe hybridised to Hind III fragments of 1.2 Kb and 0.8 Kb, the latter being reduced in size in a Hind III/Bam HI double digests whereas the 3' specific probe hybridised to a Hind III fragment of 2.0 Kb (Fig. 1). From these results we conclude that the C1 and C2 genes are both approximately 3.2 Kb and contain 3 exons separated by two intervening sequences of about 1.8 Kb and 0.9 Kb. There was no evidence of linkage between λ11A or λ8D and λ21B which indicates

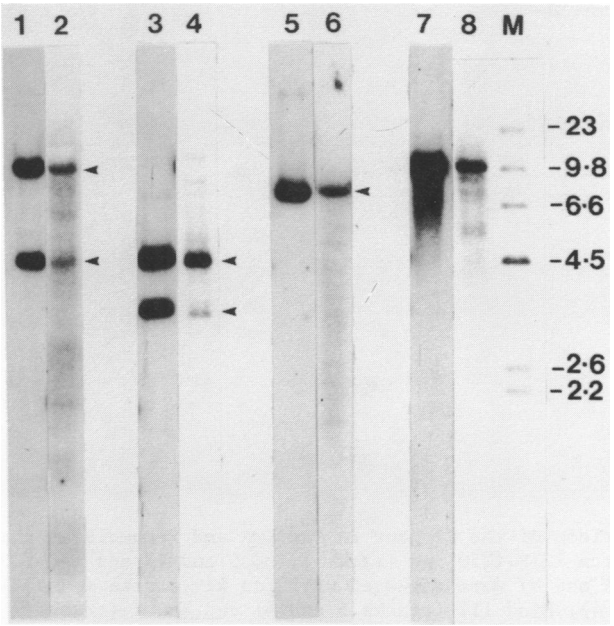

Fig. 2 Comparison of the C1 gene in rat DNA and recombinant phage DNA. DNA samples from λ11A (tracks 1, 3, 5 and 7) and rat (tracks 2, 4, 6 and 8) were digested with Eco RI (tracks 1 and 2), Bam HI (tracks 3 and 4), Msp I (tracks 5 and 6) and Xba I (tracks 7 and 8) and the blot was hybridised with ^{32}P-pA13. Track 9 shows the position of Hind III digested bacteriophage λ markers.

that the two genes are separated by a minimum of 7 Kb.

It was important to establish that the C1 and C2 genomic clones represented the genes in our colony of Sprague-Dawley rats and that genetic rearrangements had not taken place during propagation of the lambda-phage recombinants. This was verified by comparing Southern blots obtained for the cloned sequences with those of rat DNA. In the case of C1, Eco RI, Bam HI, Msp I and Xba I, gave hybridising fragments of exactly the same size in both λ11A and rat DNA (Fig. 2). From the relative intensities of the bands we conclude that the C1 gene is present in one copy per haploid genome. Finally, we have examined DNA from fifteen individual rats to investigate polymorphism of the C1 gene: Eco RI and Hind III (unpublished data) digests gave similar hybridising bands for each animal suggesting that the gene is homozygous, at least with respect to these restriction enzyme sites.

The situation for the C2 gene was slightly more complex (Fig. 3).

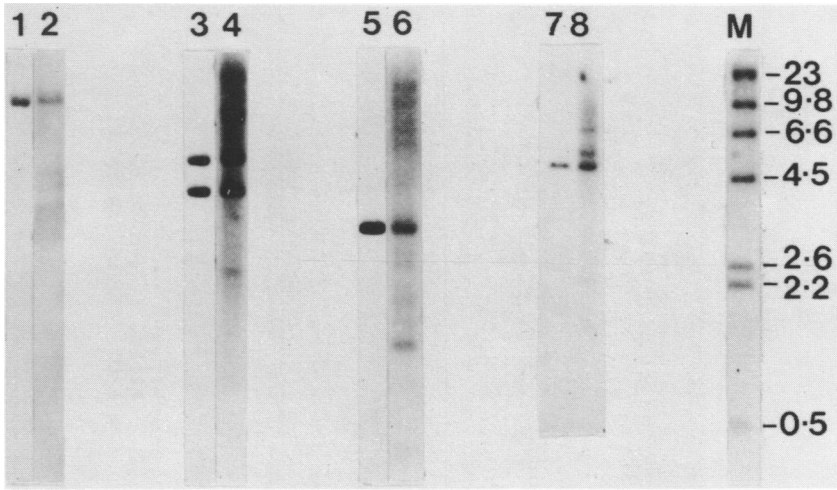

Fig. 3 Comparison of the C2 gene in rat DNA and recombinant phage DNA. DNA samples from λ21B (250 pg; tracks 1, 3, 5 and 7) and rat (20 μg; tracks 2, 4, 6 and 8) were digested with Eco RI (tracks 1 and 2), Bam HI (tracks 3 and 4), Hind III (tracks 5 and 6) and Xba I (tracks 7 and 8) and the blot hybridised with ^{32}P-pB44. Track 9 shows the position of Hind III digested bacteriophage λ markers.

Although several restriction enzyme digests of λ21B produced similar hybridising fragments to the predominant bands produced in rat DNA, certain enzymes such as Hind III and Xba I resulted in hybridising fragments in cell DNA which could not be accounted for in λ21B. These additional DNA fragments in rat DNA could represent, 1) part of the C2 gene; 2) polymorphism of the C2 gene; or 3) a second copy of the C2 gene. Since the DNA sequence of λ21B (see below) indicates that it contains both the 5' and 3' ends of the gene, we can eliminate possibility 1). Polymorphism of the C2 gene has been investigated by analysing DNA from individual rats. Hind III (Fig. 4) and Eco RI (unpublished data) digests gave similar hybridising bands for 15 animals tested and suggests that heterozygosity of the C2 gene cannot account for the 1.2 Kb Hind III fragment. Therefore we tentatively conclude that the additional hybridising DNA fragments represent a related C2 gene or pseudogene.

The DNA sequence of the C1 and C2 genes

We have determined the DNA sequence of parts of the C1 and C2 genes particularly around their 5' end to analyse in more detail the structural similarities between the two genes and because ultimately we wish to identify sequences involved in the regulation of their transcription. The

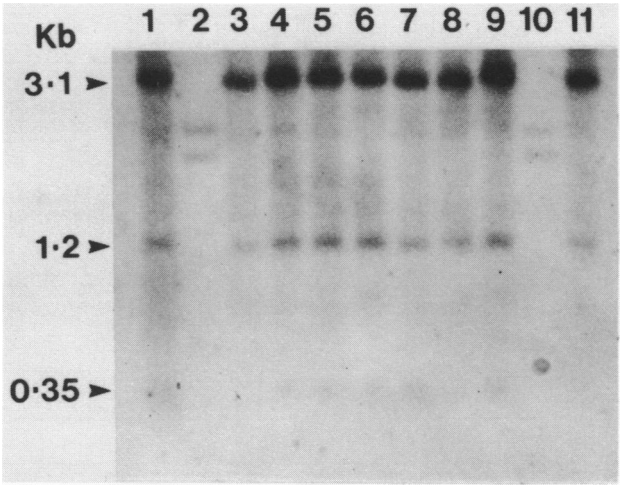

Fig. 4 Analysis of the C2 gene in individual rats.
DNA samples from 11 individual rats (tracks 1, 3-9, 11) were digested with Hind III and the blot was hybridised with ^{32}P-pB44. Tracks 2 and 10 contain markers.

method of Maxam and Gilbert (21) was used to sequence a 3.3 Kb Bam HI (p8D1) subclone of C1, which contains the 5' end of the gene and a 6.0 Kb Pst I subclone of C2, which contains the entire C2 gene (p21B1). p8D1 was labelled at the Hind III and Bcl I sites and p21B1 was labelled at the two Hind III sites flanking the first exon (Fig. 1). The DNA sequence containing the Hind III site at the 5' end of the C2 gene was also obtained by subcloning a Sau 3A DNA fragment into M13 mp7 and sequencing by the dideoxy methods of Sanger (23).

The DNA sequences (Fig. 5) showed that the first exon contained 104 nucleotides in p8D1, namely C1, and 93 nucleotides in p21B1, namely C2, and comprised the 5' non-coding regions and most of the coding region for the C1 and C2 signal peptides. However, the striking feature of the two sequences was the extent of homology, not only in the exons, which had been noted previously in the cDNA clones, but both 5' upstream regions and the first intervening sequence. The predominant homologous regions encompass, (i) the 5' upstream regions from nucleotides -230 to the cap site (see below) which share 76% homology, (ii) the first exon which share 85% homology and results in 60% amino acid sequence conservation of the C1 and C2 signal peptides and (iii) the intervening sequence from nucleotide 105-230 which share 76% homology. Therefore, we conclude that

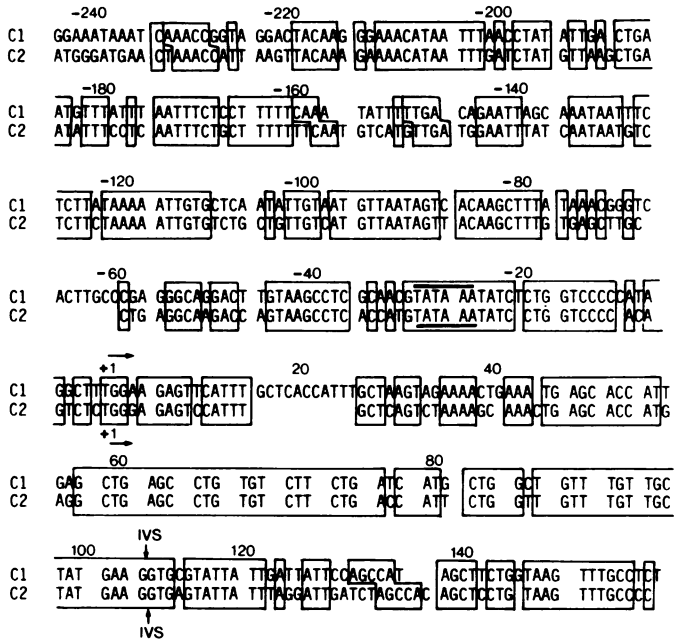

Fig. 5 Nucleotide sequence of the 5' end of the C1 and C2 genes. Nucleotide numbers are shown above the sequence; the position +1 is defined as the first nucleotide in C1 mRNA and sites of initiation of transcription are indicated with arrows. Gaps have been left to maximise sequence homology which are shown in boxes. The coding sequence is denoted by triplet codons; IVS denotes the first exon/intron boundary and the sequence TATAAA is underlined.

the genes coding for C1 and C2 have arisen by duplication of an ancestral gene and that the subsequent divergence has been slightly greater in the 5' upstream region and first intervening sequence than in the coding region.

The 5' end of C1 and C2 mRNA

The site of initiation of transcription of the C1 and C2 genes was determined by DNA primer extension. For C1, a 75 nucleotide Bgl II-Hgi AI DNA fragment was used as a primer and a single major stop band of about 120-121 nucleotides was obtained (Fig. 6A) irrespective of whether nuclear or cytoplasmic RNA was used. Thus there appears to be a single site of initiation of transcription and no RNA processing of the 5' end of the RNA occurs during export from the nucleus to the cytoplasm. By carrying out the primer extension reactions under chain termination conditions (Fig. 6B) it was possible to compare the sequence of the 5' end of the mRNA with that of the gene and show that transcription was initiated

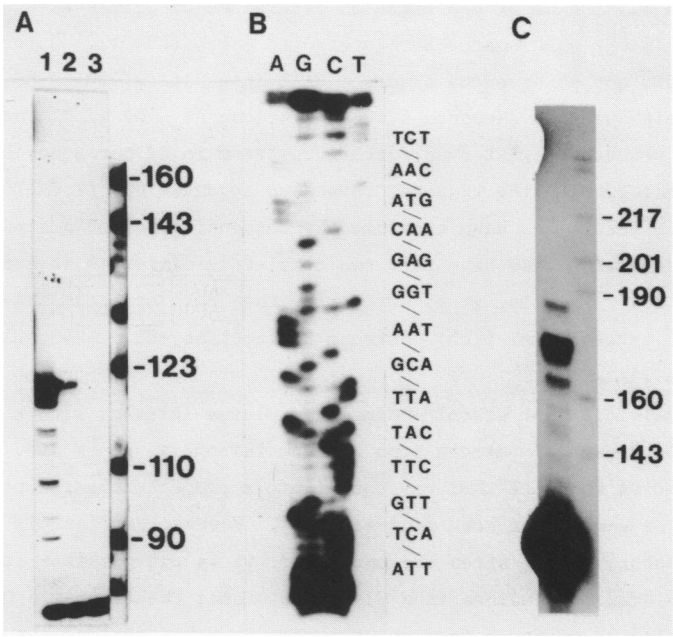

Fig. 6 Primer extension of C1 and C2 mRNA.
(A) shows the product of C1 primer extension using prostate RNA with the position of DNA markers alongside. Tracks 1, 2 and 3 represent reactions with 1 µg, 0.1 µg and 0.01 µg poly(A)-containing RNA respectively.
(B) shows the products of C1 primer extension under chain termination conditions with the DNA sequence corresponding to the coding strand alongside.
(C) shows the product of C2 primer extension using prostate RNA with the position of DNA markers alongside.

within the sequence GGCTTT*GGAA probably at the G marked with an asterisk.

For C2, a 125 nucleotide DNA fragment was used as a primer and two major stop bands of approximately 173-175 nucleotides and a minor stop band of approximately 185 nucleotides was obtained (Fig. 6C). Thus, the major site of initiation of transcription occurs within the sequence GTCTCT*G*GGA, probably at G residues marked with an asterisk; it is striking that this site corresponds precisely with the site of initiation of transcription of the C1 gene (Fig. 5). It is noteworthy that the cDNA clones pA13 and pB44 which were constructed using a method of Rougeon et al. (27) designed to contain the entire 5' end of C1 and C2 mRNA, were actually 6 nucleotides short and full-length respectively.

In both C1 and C2, the sequence TATAAA occurs 30 nucleotides upstream

from position +1, the major cap sites as determined by primer extension, which is consistent with most other eucaryotic genes (28-31). Surprisingly, however, we did not observe the sequence CAAT at position -70 — -80 in either gene although the sequence GCTCAATAT occurs at -105 in C1 and several CAAT sequences exist even further upstream in C2 but none contains extensive homology with the proposed consensus sequence GG $^{A}/_{T}$ CAATCT (29-31). This result has been confirmed by sequencing both DNA strands and, in the case of C2, we have used two completely different sequencing methods (21, 23). Finally, it is noteworthy that both upstream DNA sequences are extremely AT-rich: between nucleotides -80 — -240, C1 is 74% and C2 is 71% AT.

It is conceivable that steroid-hormone complexes interact with a homologous DNA sequence upstream from hormone responsive genes and, in fact, a consensus sequence that may constitute a progesterone-receptor complex binding region has been proposed (32). However, as far as androgen-receptor binding sites are concerned, it is not possible at this stage to draw conclusions from C1 and C2 alone; the extensive DNA sequence homologies upstream from the cap site probably reflect the limited divergence that has taken place since the gene duplication event and do not represent constraints on the sequence related to promoter function or hormonal regulation.

Identification of human prostatic steroid binding protein genes

An analogous steroid binding protein has been reported in human prostatic tissue (33) and therefore it was of interest to investigate whether there were homologus C1 and C2 genes in human DNA. However, analysis of DNA blots with [^{32}P]-pA13 and [^{32}P]-B44 both at high and low stringency (1 x SSC at 65°C) failed to reveal specific hybridising fragments in human DNA (unpublished data). It is noteworthy that cross-hybridisation between C1 genomic DNA (8D) and C2 cDNA (pB44) and between C2 genomic DNA (21B) and C1 cDNA (pA13) occurs at low stringency and therefore, since these two genes share approximately 75% homology, we presume that equivalent human genes, if they exist, must share less than 75% homology with the rat prostatic steroid binding protein genes.

ACKNOWLEDGEMENTS

We are extremely grateful to Drs. Sargent, Wallace, Jagodzinski and Bonner for providing us with the rat DNA libraries; R. Patient and J.G. Williams for many helpful discussions; Drs. B. Dalaey, W. Rombauts,

G. Volckaert, B. Peeters, J. Mous and W. Heyns, for communicating protein sequence data prior to publication; M. Barker for typing the manuscript.

REFERENCES

1. Heyns, W., Peeters, B. and Mous, J. (1977) Biochem. Biophys. Res. Commun. 77, 1492-1499.
2. Parker, M.G., Scrace, G.T. and Mainwaring, W.I.P. (1978) Biochem. J. 170, 115-121.
3. Fang, S. and Liao, S. (1971) J. Biol. Chem. 246, 16-24.
4. Heyns, W. and De Moor, P. (1977) Eur. J. Biochem. 78, 221-230.
5. Lea, O.A., Petrusz, P. and French, F.S. (1979) J. Biol. Chem. 254, 6196-6202.
6. Forsgren, B., Gustafsson, J.-A., Pousette, A. and Högberg, B. (1979) Cancer Res. 39, 5155-5164.
7. Parker, M.G. and Mainwaring, W.I.P. (1977) Cell 12, 401-407.
8. Parker, M.G. and Scrace, G.T. (1979) Proc. Natl. Acad. Sci. USA 76, 1580-1584.
9. Parker, M.G., White, R. and Williams, J.G. (1980) J. Biol. Chem. 255, 6996-7001.
10. Page, M.J. and Parker, M.G. (1982) Mol. & Cellular Endocrinol. in press.
11. Heyns, W., Peeters, B., Mous, J., Rombauts, W. and De Moor, P. (1978) Eur. J. Biochem. 89, 181-186.
12. Parker, M.G., Needham, M. and White, R. (1982) Nature 298, 92-94.
13. Peeters, B., Rombauts, W., Mous, J. and Heyns, W. (1981) Eur. J. Biochem. 115, 115-121.
14. Peeters, B., Heyns, W., Mous, J. and Rombauts, W. (1982) Eur. J. Biochem. 123, 55-62.
15. Sargent, T.D., Wu, J.-R., Sala-Trepat, J.M., Wallace, R.B., Reyes, A.A. and Bonner, J. (1979) Proc. Natl. Acad. Sci. USA 76, 3256-3260.
16. Benton, W.D. and Davis, R.W. (1977) Science 196, 180-182.
17. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
18. Blin, N. and Stafford, D.W. (1976) Nucleic Acids Res. 3, 2303-2308.
19. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
20. Wahl, G.M., Stern, M. and Stark, E.R. (1979) Proc. Natl. Acad. Sci. USA 76, 3683-3687.
21. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
22. Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucleic Acids Res. 9, 309-321.
23. Sanger, F. (1981) Bioscience Report 1, 3-18.
24. Schreier, P.H. and Cortese, R. (1979) J. Mol. Biol. 129, 169-177.
25. Sanger, F. and Coulson, A.R. (1979) FEBS Lett. 87, 107-110.
26. Donoghue, D.J. and Hunter, T. (1982) Nucleic Acids Res. 10, 2549-2564.
27. Rougeon, F., Kourilsky, P. and Mach, B. (1975) Nucleic Acids Res. 2, 2365-2378.
28. Goldberg, M.L. (1979) Ph.D. Thesis, Stanford University.
29. Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) Cell 21, 53-68.
30. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucleic Acids Res. 8, 127-142.
31. Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349-383.

32. Mulvihill, E.R., LePennec, J.-P. and Chambon, P. (1982) Cell 24, 621-632.
33. Bjork, P., Forsgren, B., Gustafsson, J.-A., Pousette, A. and Högberg, B. (1982) Cancer Res. 42, 1935-1942.