Volume 15 Number 4 1987 Nucleic Acids Research

Rat prostatic binding protein: the complete sequence of the C2 gene and its flanking regions

B.Delaey, L.Dirckx, J.-L.Decourt, F.Claessens, B.Peeters and W.Rombauts

Katholieke Universiteit Leuven, Fac. Geneeskunde, Afdeling Biochemie, Herestraat 49, B-3000 Leuven, Belgium

Received October 3, 1986; Revised December 31, 1986; Accepted January 21, 1987

ABSTRACT

The complete sequence (2879 bp) of the androgen-controlled rat prostatic binding protein C2 gene and 1023 bp of the 5'- and 2127 bp of the 3'-flanking regions have been determined. The gene contains three exons (93, 203 and 147 bp) and two introns (1630 and 806 bp). It is flanked by two homopurine-homopyrimidine stretches of 55 and 131 nucleotides respectively, located at positions -405 and 4151. These sequences are remarkably sensitive towards S1-nuclease, indicating an altered DNA conformation under superhelical stress. Several palindromes and dyad structures are observed in the 5'-upstream region of the gene and at position -457, an 80% homology to the consensus sequence of a glucocorticoid receptor binding site is found.

INTRODUCTION

Prostatic binding protein (PBP) is an androgen-controlled, steroid-binding protein in the rat ventral prostate. It constitutes the main protein fraction in this organ, where it is secreted into the seminal fluid (1-4). It consists of three different polypeptide chains, C1, C2 and C3, arranged in two non-identical dimeric subunits C1C3/C2C3 (5). The amino acid sequences of these three polypeptides have been determined by our group (6-8). Since PBP is under hormonal control, we believe it to be a useful model system for the study of androgen regulation.

cDNA clones for the three PBP mRNA's have been prepared both by us and other groups (9,10), and these have been used for the isolation of the corresponding genes from rat genomic libraries (11-15). To date, we have identified one C1-, two different C2- and three different C3-specific clones. Work is under way to establish whether the C2 and C3 sequences are truly different genes or whether they represent allelic differences of the same

gene. There is evidence for at least two non-allelic C3-genes (13,14). All these genes contain three exons with similar exon-intron arrangements and their coding sequences span a genomic DNA region of approximately 3 kb. In this report we present the sequence of the C2 gene and of the 5'- and 3'-flanking regions. In addition, we have studied the sensitivity of this gene to S1-nuclease, a test which is indicative for structural conformation changes in DNA under superhelical stress (42).

MATERIALS AND METHODS

Enzymes

Restriction enzymes were from BRL, Amersham or Boehringer-Mannheim. The Klenow fragment of E. coli DNA polymerase I was from P.L. Biochemicals. S1 nuclease, T4 DNA ligase and Exonuclease III were from BRL. Bal 31 nuclease was from Boehringer-Mannheim. Nucleotides and sequencing primers were from Boehringer-Mannheim or P.L. Biochemicals. [α^{32}P]-dNTP's (3000 Ci/mmol), [α^{32}P]-dATP (3000 Ci/ mmol) and [α^{35}S]-dATP (600 Ci/mmol) were from Amersham.

Strains

E. coli WB373 and M13 vector mWB2348 (16) were obtained through the courtesy of Dr. J. Vanderleyden (F.A. Janssenslaboratorium voor genetica, K.U.Leuven, Belgium).

Clones

Clone p21B1, a 6 kb PstI subclone of the C2 gene in pAT153, was obtained from Dr. M. Parker (Imperial Cancer Research Fund Laboratories, London). This clone contains the entire C2 gene and approximately 1 kb and 2 kb of 5'- and 3'- flanking sequences respectively. A 2.7 kb PstI-XbaI fragment containing the 5'- part of the gene was inserted into the M13 vectors mWB2348, mp18 and mp19, giving rise to subclones mWBC2PX9, mp18C2PX1 and mp19C2PX2 respectively. The same was done for the 3.3 kb PstI-XbaI fragment containing the 3'- part of the gene (clones mWBC2PX8, mp18C2PX2 and mp19C2PX3). mC2ΔH9 and mC2ΔC11 are deletion subclones of mWBC2PX9 (see below).

DNA sequencing and deletion strategy

Deletion clones were prepared according to different sequencing strategies. First, the "kilo-sequencing" procedure of Barnes

and Bevan was used (16). In short, 10 µg of supercoiled mWBC2PX9 DNA was treated with 150 ng DNase I in the presence of 500 µg/ml ethidium bromide to produce singly-nicked molecules. The DNA was then treated with 25 U Exonuclease III in a volume of 50 µl for 10 minutes at 25°C in order to widen the nick to a small gap. The gap was cut across by nuclease Bal 31 or S1 and the ends were filled in by the Klenow fragment of DNA polymerase I. XbaI linkers were added, and after removal of unligated linkers by Sepharose 4B chromatography, the DNA was cut with an excess of XbaI, thus generating the deletions by removing the insert fragment between the nick and the XbaI site on the vector. After recircularization of the deleted molecules, they were transformed into competent WB373. Another sequencing approach was to resect the isolated 6kb insert from p21B1 with Bal 31 nuclease. At different time points (1 to 15 minutes), aliquots were removed from the reaction mixture and phenol-extracted. After "polishing" the ends with Klenow polymerase, the molecules were cut with XbaI and ligated into mp18 cut with HincII and XbaI. Finally, some parts of the sequence were obtained by subcloning suitable restriction fragments into mp18 or mp19. For sizing of the insert lenght, individual plaques were grown in 2ml volumes of L-broth, 40 µl aliquots of culture supernatant were brought to 0.1% SDS and electrophoresed on 0.8 % agarose gels in Tris-borate buffer (pH 8.3). Sequencing was done according to Sanger et al (17). Sequence ladders for comparison with S1-treated DNA were prepared by 5'-end or 3'-end labeling of the appropriate restriction fragment, cutting off the label at one end followed by Maxam-Gilbert sequencing (18).

S1 nuclease hypersensitive site mapping

Digestions of supercoiled DNA with S1 nuclease were performed in a 100 µl volume containing 50 mM Na-acetate pH 4.5, 0.3 M NaCl, 4 mM ZnSO4, 3-10 µg DNA and 30-200 U of S1 nuclease. Incubation was at 37°C for 5 (fine mapping) or 60 minutes after which 15 µl 1 M Tris pH 8 and 3 µl 0.5 M EDTA were added. After phenol extraction and ethanol precipitation, the DNA was cut with different restriction enzymes and the fragments separated on agarose gels. For fine mapping of the hypersensitive sites at the nucleotide level, S1-treated mC2ΔH9 or mC2ΔC11 DNA was cut

with XbaI, 3'-end or 5'-end labelled and electrophoresed on 8% polyacrylamide-urea gels after removal of the label at one end by HaeIII cutting.

RESULTS AND DISCUSSION

A 6 kb PstI fragment of the C2 genomic clone p21B1, covering the entire gene and approximately 1 kb and 2 kb of the 5'- and 3'-sequences respectively, was used for the determination of the C2 sequence. Two XbaI-PstI fragments of this clone, containing the 5'-half (2.7 kb) and 3'-half (3.3 kb) of the gene respectively (fig. 1), were subcloned into M13 phages mWB2348,mp18 and mp19 (see materials and methods). Initially, all cloning was done in the M13 vector mWB2348, using E. coli WB373 as a host, because this vector allows a stable incorporation of very long inserts (16). However, due to the lack of suitable cloning sites in this vector, we later switched to the mp18 and mp19 vectors, grown in E. coli JM109. To facilitate dideoxy sequencing, we prepared several progressively deleted clones, using the "kilo-sequencing" procedure of Barnes and Bevan (16) or using Bal 31 nuclease to degrade insert termini. This enabled us to deduce all of the exon and 5'-upstream sequence and most of the intron sequence from both strands while the sequence of the 3'-flanking region was obtained from one strand.

The results show that the C2 gene insert in p21B1 contains 3 exons of 93, 203 and 147 nucleotides respectively, which are separated by two introns of 1630 and 806 nucleotides, and flanked by a 5'-upstream and a 3'-downstream region of 1023 and 2127 bp. Our data largely agree with the cDNA sequence and with that part of the genomic sequence (nucleotides -240 to 160) published by

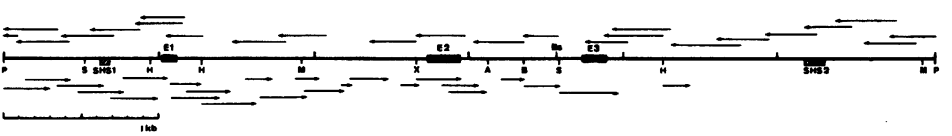
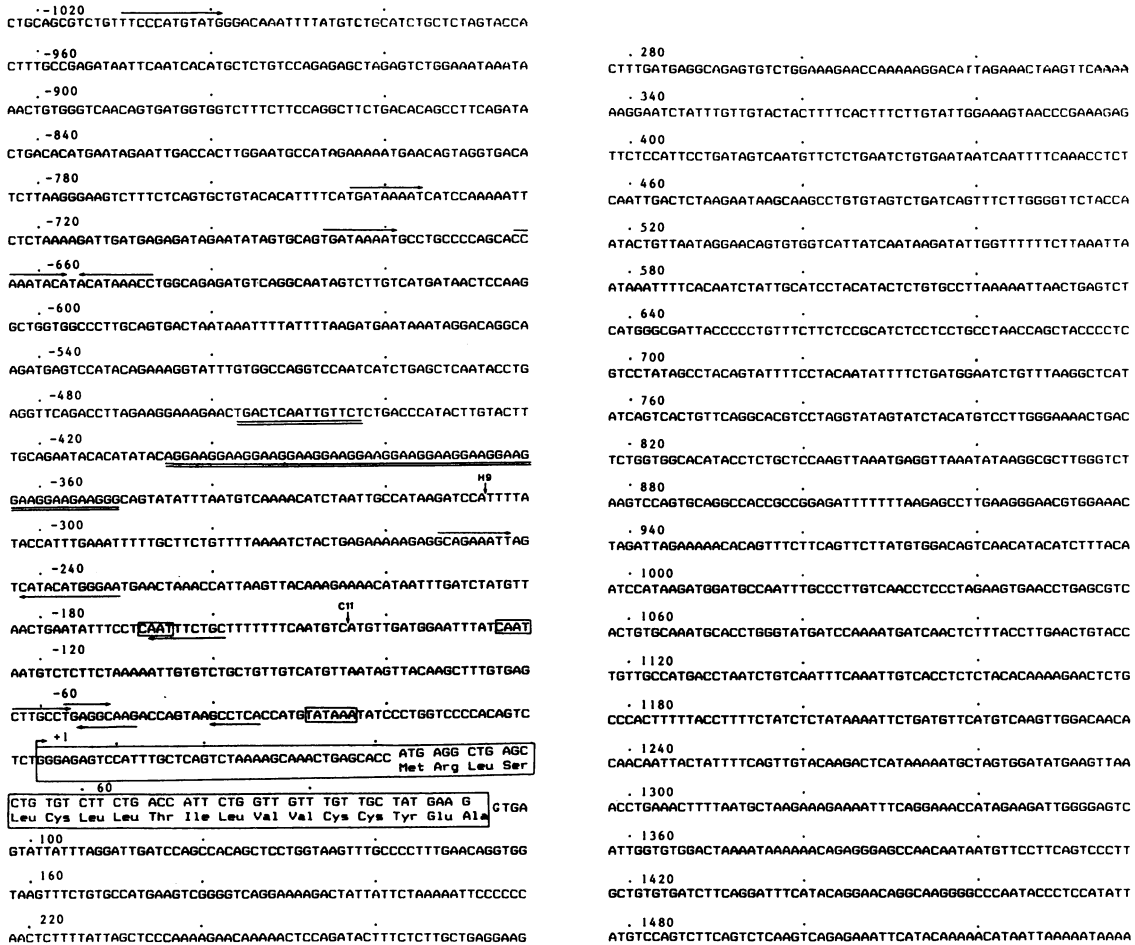


Fig. 1. Restriction map of the C2 gene with indication of S1-nuclease hypersensitive sites (SHS1 and 2). Restriction enzymes used are : AvaI (A), BamHI (B), BstEII (Bs), HindIII (H), MspI (M), PstI (P), SstI (S) and XbaI (X). Arrows indicate the direction and extent of nucleotide reading.


-1020
CTGCAGCGTCTGTTTCCCATGTATGGGACAAATTTTATGTCTGCATCTGCTCTAGTACCA
-960
CTTTGCCGAGATAATTCAATCACATGCTCTGTCCAGAGAGCTAGAGTCTGGAAATAAATA
-900
AACTGTGGGTCAACAGTGATGGTGGTCTTTCTTCCAGGCTTCTGACACAGCCTTCAGATA
-840
CTGACACATGAATAGAATTGACCACTTGGAATGCCATAGAAAAATGAACAGTAGGTGACA
-780
TCTTAAGGGAAGTCTTTCTCAGTGCTGTACACATTTTCATGATAAAATCATCCAAAAATT
-720
CTCTAAAAGATTGATGAGAGATAGAATATAGTGCAGTGATAAAATGCCTGCCCCAGCACC
-660
AAATACATACATAAACCTGGCAGAGATGTCAGGCAATAGTCTTGTCATGATAACTCCAAG
-600
GCTGGTGGCCCTTGCAGTGACTAATAAATTTTATTTTAAGATGAATAAATAGGACAGGCA
-540
AGATGAGTCCATACAGAAAGGTATTTGTGGCCAGGTCCAATCATCTGAGCTCAATACCTG
-480
AGGTTCAGACCTTAGAAGGAAAGAACTGACTCAATTGTTCTCTGACCCATACTTGTACTT
-420
TGCAGAATACACATATACAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAG
-360
GAAGGAAGAAGGGCAGTATATTTAATGTCAAAACATCTAATTGCCATAAGATCCATTTTA
-300
TACCATTTGAAATTTTTGCTTCTGTTTTAAAATCTACTGAGAAAAAGAGGCAGAAATTAG
-240
TCATACATGGGAATGAACTAAACCATTAAGTTACAAAGAAAACATAATTTGATCTATGTT
-180
AACTGAATATTTCCTCAATTTCTGCTTTTTTTCAATGTCATGTTGATGGAATTTATCAAT
-120
AATGTCTCTTCTAAAAATTGTGTCTGCTGTTGTCATGTTAATAGTTACAAGCTTTGTGAG
-60
CTTGCCTGAGGCAAGACCAGTAAGCCTCACCATGTATAAATATCCCTGGTCCCCACAGTC
+1
TCTGGGAGAGTCCATTTGCTCAGTCTAAAAGCAAACTGAGCACC ATG AGG CTG AGC
Met Arg Leu Ser
60
CTG TGT CTT CTG ACC ATT CTG GTT GTT TGT TGC TAT GAA G GTGA
Leu Cys Leu Leu Thr Ile Leu Val Val Cys Cys Tyr Glu Ala
100
GTATTATTTAGGATTGATCCAGCCACAGCTCCTGGTAAGTTTGCCCCTTTGAACAGGTGG
160
TAAGTTTCTGTGCCATGAAGTCGGGGTCAGGAAAAGACTATTATTCTAAAAATTCCCCCC
220
AACTCTTTTATTAGCTCCCAAAAGAACAAAAACTCCAGATACTTTCTCTTGCTGAGGAAG
280
CTTTGATGAGGCAGAGTGTCTGGAAAGAACCAAAAAGGACATTAGAAACTAAGTTCAAAA
340
AAGGAATCTATTTGTTGTACTACTTTTCACTTTCTTGTATTGGAAAGTAACCCGAAAGAG
400
TTCTCCATTCCTGATAGTCAATGTTCTCTGAATCTGTGAATAATCAATTTTCAAACCTCT
460
CAATTGACTCTAAGAATAAGCAAGCCTGTGTAGTCTGATCAGTTTCTTGGGGTTCTACCA
520
ATACTGTTAATAGGAACAGTGTGGTCATTATCAATAAGATATTGGTTTTTTCTTAAATTA
580
ATAAATTTTCACAATCTATTGCATCCTACATACTCTGTGCCTTAAAAATTAACTGAGTCT
640
CATGGGCGATTACCCCCTGTTTCTTCTCCGCATCTCCTCCTGCCTAACCAGCTACCCCTC
700
GTCCTATAGCCTACAGTATTTTCCTACAATATTTTCTGATGGAATCTGTTTAAGGCTCAT
760
ATCAGTCACTGTTCAGGCACGTCCTAGGTATAGTATCTACATGTCCTTGGGAAAACTGAC
820
TCTGGTGGCACATACCTCTGCTCCAAGTTAAATGAGGTTAAATATAAGGCGCTTGGGTCT
880
AAGTCCAGTGCAGGCCACCGCCGGAGATTTTTTTAAGAGCCTTGAAGGGAACGTGGAAAC
940
TAGATTAGAAAAACACAGTTTCTTCAGTTCTTATGTGGACAGTCAACATACATCTTTACA
1000
ATCCATAAGATGGATGCCAATTTGCCCTTGTCAACCTCCCTAGAAGTGAACCTGAGCGTC
1060
ACTGTGCAAATGCACCTGGGTATGATCCAAAATGATCAACTCTTTACCTTGAACTGTACC
1120
TGTTGCCATGACCTAATCTGTCAATTTCAAATTGTCACCTCTCTACACAAAAGAACTCTG
1180
CCCACTTTTTACCTTTTCTATCTCTATAAAATTCTGATGTTCATGTCAAGTTGGACAACA
1240
CAACAATTACTATTTTCAGTTGTACAAGACTCATAAAAATGCTAGTGGATATGAAGTTAA
1300
ACCTGAAACTTTTAATGCTAAGAAAGAAAATTTCAGGAAACCATAGAAGATTGGGGAGTC
1360
ATTGGTGTGGACTAAAATAAAAAACAGAGGGAGCCAACAATAATGTTCCTTCAGTCCCTT
1420
GCTGTGTGATCTTCAGGATTTCATACAGGAACAGGCAAGGGGCCCAATACCCTCCATATT
1480
ATGTCCAGTCTTCAGTCTCAAGTCAGAGAAATTCATACAAAAACATAATTAAAAATAAAA

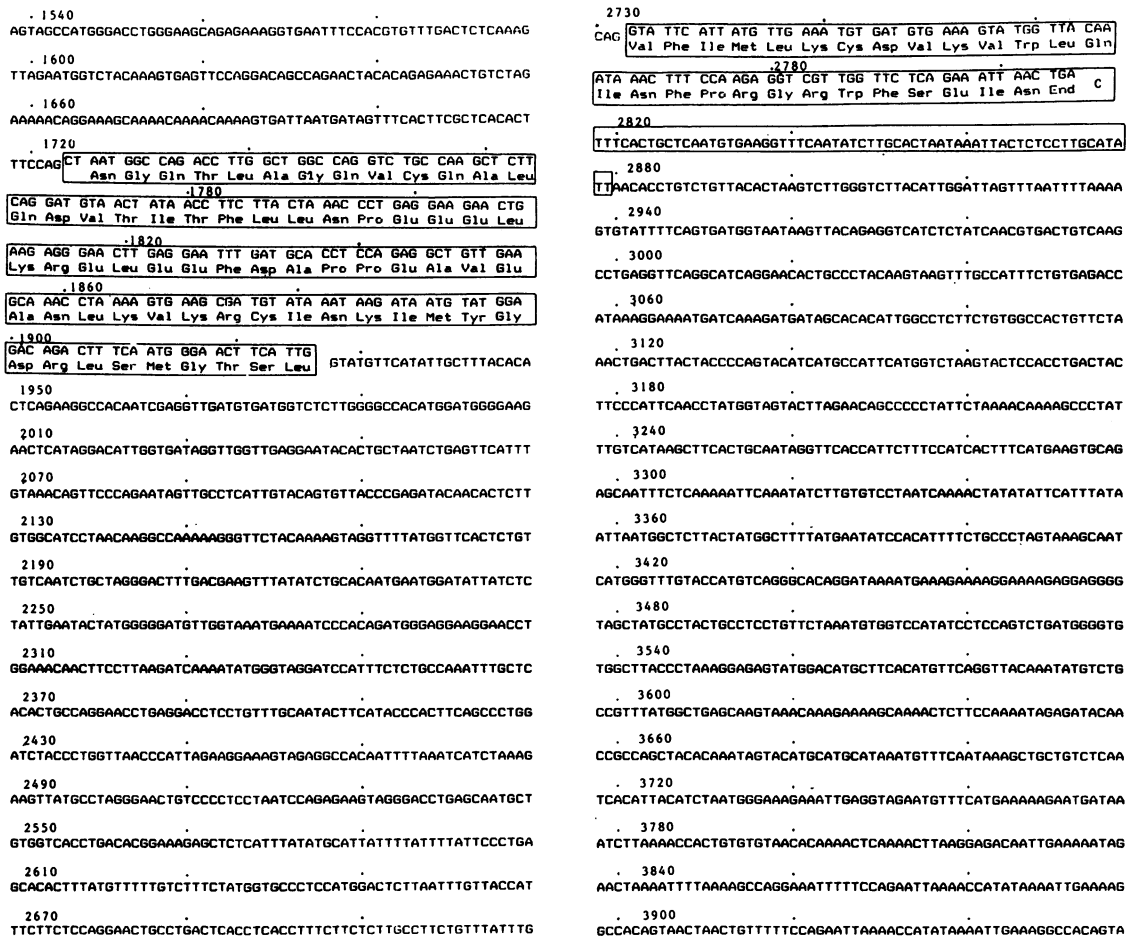

1540
AGTAGCCATGGGACCTGGGAAGCAGAGAAAGGTGAATTTCCACGTGTTTGACTCTCAAAG
1600
TTAGAATGGTCTACAAAGTGAGTTCCAGGACAGCCAGAACTACACAGAGAAACTGTCTAG
1660
AAAAACAGGAAAGCAAAACAAAACAAAAGTGATTAATGATAGTTTCACTTCGCTCACACT
1720
TTCCAG CT AAT GGC CAG ACC TTG GCT GGC CAG GTC TGC CAA GCT CTT
Asn Gly Gln Thr Leu Ala Gly Gln Val Cys Gln Ala Leu
1780
CAG GAT GTA ACT ATA ACC TTC TTA CTA AAC CCT GAG GAA GAA CTG
Gln Asp Val Thr Ile Thr Phe Leu Leu Asn Pro Glu Glu Glu Leu
1820
AAG AGG GAA CTT GAG GAA TTT GAT GCA CCT CCA GAG GCT GTT GAA
Lys Arg Glu Leu Glu Glu Phe Asp Ala Pro Pro Glu Ala Val Glu
1860
GCA AAC CTA AAA GTG AAG CGA TGT ATA AAT AAG ATA ATG TAT GGA
Ala Asn Leu Lys Val Lys Arg Cys Ile Asn Lys Ile Met Tyr Gly
1900
GAC AGA CTT TCA ATG GGA ACT TCA TTG
Asp Arg Leu Ser Met Gly Thr Ser Leu
GTATGTTCATATTGCTTTACACA
1950
CTCAGAAGGCCACAATCGAGGTTGATGTGATGGTCTCTTGGGGCCACATGGATGGGGAAG
2010
AACTCATAGGACATTGGTGATAGGTTGGTTGAGGAATACACTGCTAATCTGAGTTCATTT
2070
GTAAACAGTTCCCAGAATAGTTGCCTCATTGTACAGTGTTACCCGAGATACAACACTCTT
2130
GTGGCATCCTAACAAGCCAAAAAGGGTTCTACAAAAGTAGGTTTTATGGTTCACTCTGT
2190
TGTCAATCTGCTAGGGACTTTGACGAAGTTTATATCTGCACAATGAATGGATATTATCTC
2250
TATTGAATACTATGGGGGATGTTGGTAAATGAAAATCCCACAGATGGGAGGAAGGAACCT
2310
GGAAACAACTTCCTTAAGATCAAAATATGGGTAGGATCCATTTCTCTGCCAAATTTGCTC
2370
ACACTGCCAGGAACCTGAGGACCTCCTGTTTGCAATACTTCATACCCACTTCAGCCCTGG
2430
ATCTACCCTGGTTAACCCATTAGAAGGAAAGTAGAGGCCACAATTTTAAATCATCTAAAG
2490
AAGTTATGCCTAGGGAACTGTCCCCTCCTAATCCAGAGAAGTAGGGACCTGAGCAATGCT
2550
GTGGTCACCTGACACGGAAAGAGCTCTCATTTATATGCATTATTTTATTTTATTCCCTGA
2610
GCACACTTTATGTTTTTGTCTTTCTATGGTGCCCTCCATGGACTCTTAATTTGTTACCAT
2670
TTCTTCTCCAGGAACTGCCTGACTCACCTCACCTTTCTTCTCTTGCCTTCTGTTTATTTG
2730
CAG GTA TTC ATT ATG TTG AAA TGT GAT GTG AAA GTA TGG TTA CAA
Val Phe Ile Met Leu Lys Cys Asp Val Lys Val Trp Leu Gln
2780
ATA AAC TTT CCA AGA GGT CGT TGG TTC TCA GAA ATT AAC TGA C
Ile Asn Phe Pro Arg Gly Arg Trp Phe Ser Glu Ile Asn End
2820
TTTCACTGCTCAATGTGAAGGTTTCAATATCTTGCACTAATAAATTACTCTCCTTGCATA
2880
TTAACACCTGTCTGTTACACTAAGTCTTGGGTCTTACATTGGATTAGTTTAATTTTAAAA
2940
GTGTATTTTCAGTGATGGTAATAAGTTACAGAGGTCATCTCTATCAACGTGACTGTCAAG
3000
CCTGAGGTTCAGGCATCAGGAACACTGCCCTACAAGTAAGTTTGCCATTTCTGTGAGACC
3060
ATAAAGGAAAATGATCAAAGATGATAGCACACATTGGCCTCTTCTGTGGCCACTGTTCTA
3120
AACTGACTTACTACCCCAGTACATCATGCCATTCATGGTCTAAGTACTCCACCTGACTAC
3180
TTCCCATTCAACCTATGGTAGTACTTAGAACAGCCCCCTATTCTAAAACAAAAGCCCTAT
3240
TTGTCATAAGCTTCACTGCAATAGGTTCACCATTCTTTCCATCACTTTCATGAAGTGCAG
3300
AGCAATTTCTCAAAAATTCAAATATCTTGTGTCCTAATCAAAACTATATATTCATTTATA
3360
ATTAATGGCTCTTACTATGGCTTTTATGAATATCCACATTTTCTGCCCTAGTAAAGCAAT
3420
CATGGGTTTGTACCATGTCAGGGCACAGGATAAAATGAAAGAAAAGGAAAAGAGGAGGGG
3480
TAGCTATGCCTACTGCCTCCTGTTCTAAATGTGGTCCATATCCTCCAGTCTGATGGGGTG
3540
TGGCTTACCCTAAAGGAGAGTATGGACATGCTTCACATGTTCAGGTTACAAATATGTCTG
3600
CCGTTTATGGCTGAGCAAGTAAACAAAGAAAAGCAAAACTCTTCCAAAATAGAGATACAA
3660
CCGCCAGCTACACAAATAGTACATGCATGCATAAATGTTTCAATAAAGCTGCTGTCTCAA
3720
TCACATTACATCTAATGGGAAAGAAATTGAGGTAGAATGTTTCATGAAAAAGAATGATAA
3780
ATCTTAAAACCACTGTGTGTAACACAAAACTCAAAACTTAAGGAGACAATTGAAAAATAG
3840
AACTAAAATTTTAAAAGCCAGGAAATTTTTCCAGAATTAAAACCATATAAAATTGAAAAG
3900
GCCACAGTAACTAACTGTTTTTCCAGAATTAAAACCATATAAAATTGAAAGGCCACAGTA

Fig. 2. DNA sequence of the C2 gene. The exons, the TATA box and two CAAT motives are boxed. The homopurine-homopyrimidine tracts and the sequence resembling the glucocorticoid receptor binding site are underlined twice. Repeat and dyad structures in the upstream region are indicated by horizontal arrows and the ends of the mC2 C11 and mC2 H9 deletion clones by vertical arrows. Numbering is according to the putative transcription initiation start as determined by Parker et al. (11).

Table 1. Overview of the different repeat and dyad structures in the C2 gene promotor.

Dyad symmetries :		
-1010		-242
TTCCCATGTATG	and	CATACACATGGGAA
-254		-167
GCAGAAATT	and	AATTTCTGC
-63		-55
CTTGCCT	and	AGGCAAG
-57		-40
TGAGGC	and	GCCTCA
Direct repeats :		
TGATAAAAT at -744 and -687		
Inverted repeat :		
CCAAATACATACATAAACC at -665		

Parker et al. (43,11). However, in the 5'-flanking and first intron region 5 single-base discrepancies occur (positions -231, -182, -18, 117 and 132) while in the exon sequences several more differences with the data of these authors are found. These result in one "silent" codon change (pos. 1743), one amino acid insertion (pos. 1747-1749), one amino acid substitution (pos. 2740), one reading frame shift (pos. 2762) and two changes in the 3'-noncoding region (pos. 2846 and 2854). The amino acid sequence deduced from our C2 genomic sequence is in complete agreement with the protein sequence published by Peeters et al. (7). Except for the T's at positions 1924, 1925 and 1930, the two exon-intron junctions correspond with the consensus sequence ($^{C}_{A}$AG/GT$^{A}_{G}$AGT) of a splice donor site, while the two intron-exon junctions correspond with the consensus sequence of a splice acceptor site ($(^{T}_{C})_{n}$N$^{C}_{T}$AG/G), except for the C at position 1724 (19). A TATAAA sequence occurs at 29 bp upstream of the transcription start. Two CAAT motives occur at postions -126 and -168 but since they show little homology to the consensus CAAT box (20)

and because of their large distance from the transcription start, their functionality is doubtful.

An interesting feature in the sequence is the presence of several repeat and dyad sequences. From position -63 to -49, the 15 bp palindromic sequence CTTGCCTGAGGCAAG is found, partially overlapping a dyad structure at nucleotides -57 to -52 and -40 to -35 (fig.2). Since these sequences occur immediately upstream of the TATA box, they might be involved in transcriptional regulation. Other direct and inverted repeat structures further upstream of this sequence, are indicated in table 1. Similar palindromes have been detected at position -63 (18 bp) of the rat prolactin gene (21), at position -24 (12 bp, located after the TATA box) of the human thyroglobulin gene (22), in the Herpes virus thymidine kinase gene promotor (23) and in several Drosophila heat shock gene promotors (24). More interesting, the rat prostatic binding protein C3(1) gene (25), the rat seminal vesicle F and S genes (26-28) and the mouse renin 2 gene (29), all androgen-controlled genes, also show short inverted repeats in their promotor regions. However, since most genes do not contain such elements, their function, if any, remains to be established.

Also noteworthy is the presence of the sequence TGACTCAATTGTTCT at position -456, which shows 82% homology with the proposed consensus sequence ($\begin{smallmatrix}T\\C\end{smallmatrix}$GGTN$\begin{smallmatrix}A\\T\end{smallmatrix}CA\begin{smallmatrix}A&A\\C&T\end{smallmatrix}$NTGT$\begin{smallmatrix}T\\C\end{smallmatrix}$CT) for the binding site of the glucocorticoid receptor (30). Although there has not been shown any glucocorticoid effect on PBP expression to date, it has been suggested (31) that the regulatory elements of different steroid hormones could be either similar or at least share structural features. Indeed, similar sequences are found in the promotor regions of several other steroid-controlled genes, such as the chicken lysozyme gene, which is controlled by four classes of steroids : oestrogens, progestins, glucocorticoids and androgens and where the progesterone receptor was shown to bind to the same sites as in the MMTV promotor (31). Other homologies to this consensus sequence occur in the oestrogen receptor binding site of the vitellogenin promotor (32) and in the progesterone receptor binding site of the ovalbumin promotor (33). Furthermore, Parker et al (34) noted significant regions of homology between the consensus sequence of the progesterone receptor DNA

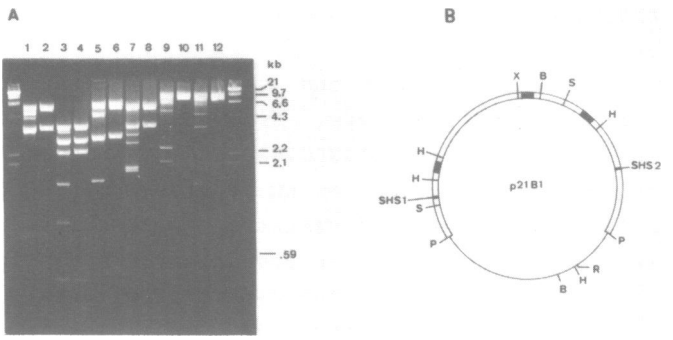


Fig. 3. A) S1 hypersensitive site mapping on plasmid p21B1, containing the entire C2 gene. Plasmid p21B1 was cut with PstI, HindIII, SstI, BamHI, XbaI or EcoRI either directly (lanes 2,4,6, 8,10,12 respectively) or after S1 treatment (lanes 1,3,5,7,9,11). The lanes at the sides contain HindIII size markers.
B) Restriction map of plasmid p21B1. B:BamHI, H:HindIII, P:PstI, R:EcoRI, S:SstI, X:XbaI. The C2 gene insert is indicated by a double line and the exons by black boxes.

binding site and parts of the PBP C3(1), C3(2), C2 and C1 genes. In this respect, the homology between the androgen -regulated PBP C1 and C2 polypeptides and the progesterone-controlled rabbit uteroglobin, as indicated by Baker (35), may be interesting. This author suggested a common ancestor existing for these proteins, which would imply that the control of the expression of these genes switched from one steroid hormone to another during evolution. Such a shift could be more readily explained if there were indeed a relationship between the control elements of different steroid-dependent genes.

One of the most striking features in the PBP C2 gene is the presence of two long homopurine-homopyrimidine stretches at positions -405 and 4151. The homopurine-homopyrimidine tract in the 5'-upstream region contains 12 repeats of the tetranucleotide GGAA, while the one in the 3'-downstream region is mainly composed of the GGAAA and GGGAA basic sequences. These sequences resemble the one found in the promotor region of the human thyroglobulin gene, where a 82.5% homology with the sequence $(AGGA)_{40}$ is found (22). Similar homopurine-homopyrimidine sequences have been found in a number of other genes including the mouse renin 1 and 2 genes (29), the rat casein gene (36), the rat somatostatin

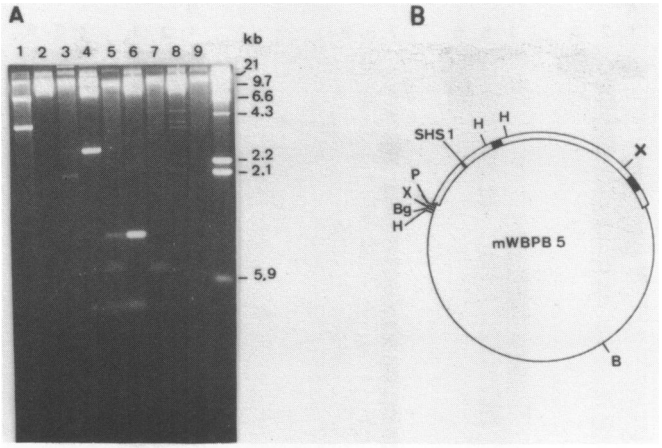


Fig. 4. A) S1 hypersensitive site mapping on M13 clone mWBPB5, containing the 5'-half of the C2 gene. Supercoiled phage DNA was cut with BamHI, XbaI, HindIII or BglII directly (lanes 2,4,6,8) or after S1-treatment (lanes 1,3,5,7), lane 9 contains supercoiled mWBPB5 DNA and lane 10 HindIII size markers.
B) Restriction map of mWBPB5. Bg:BglII, H:HindIII, P:PstI, X:XbaI. The insert (the 5'-part of the C2 gene) is indicated by a double line and the first and second exons are boxed.

gene (37) and the chicken ovotransferrin and ovalbumin X and pheasant ovotransferrin genes (38).

In fact, simple satellite-like DNA-sequences seem to be ubiquitous repetitive components of eucaryotic genomes (39). In the case of homopurine-homopyrimidine tracts, these sequences often show hypersensitivity towards single strand-specific nucleases (e.g. S1 or mung bean nuclease) (40,41,42) when under superhelical stress. When p21B1 (the entire C2 gene contained in pAT153) was treated with low amounts of S1 nuclease, followed by restriction analysis, two different S1-sensitive sites could be mapped: one in the 5'-flanking region (SHS1) and one in the 3'-flanking region (SHS2) of the gene (fig.3). The position of the S1-sensitive site in the 5'-upstream region was further confirmed by digestion of mWBPB5 (an M13 vector containing only the 5'-half of the gene) with the nuclease (fig.4). This analysis enabled us to map the site more precisely in the region of the homopurine-homopyrimidine tract at position -405. The position of the second hypersensitive site was located in the region of the homopurine-

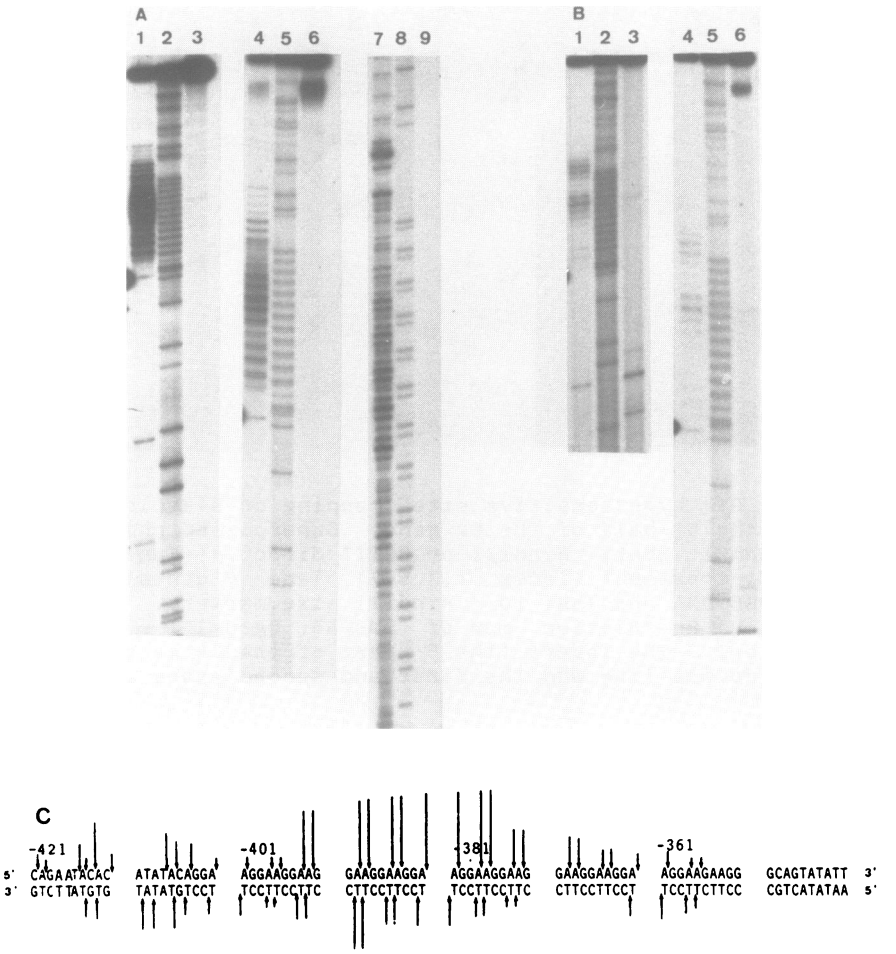


Fig. 5. A) High resolution mapping of nuclease S1 cutting sites on the upper (coding) strand of the C2 gene promotor.
lanes 1-6 : SHS mapping on mC2 C11.
lanes 1-3 : +S1, G-ladder and -S1, 2 hours electrophoresis
lanes 4-6 : +S1, G-ladder and -S1, 4 hours electrophoresis
lanes 7-9 : SHS mapping on mC2 H9 : +S1, G-ladder and -S1 respectively, 2 hours electrophoresis.

B) High resolution mapping of nuclease S1 cutting sites on the lower strand of the C2 gene promotor on mC2 C11.
lanes 1,4 : +S1, lanes 2,5 : C-ladder, lanes 3,6 : -S1.
Electrophoresis was for 2 (lanes 1-3) or 4 (lanes 4-6) hours.

C) Indication of S1 cutting sites on both strands of the homopurine-homopyrimidine region of the C2 gene. The length of the arrows corresponds with the extent of nuclease digestion.

homopyrimidine stretch at position 4151 to 4281. In the case of SH81, we have also used a fine-mapping procedure to localize the exact position of the S1-cuts in the DNA. Two M13-deletion clones, mC2 Δ C11 and mC2ΔH9, of which the 3' end is located respectively at 260 and 97 nucleotides downstream of the $(GGAA)_{12}$ repeat were S1-digested under conditions which produced few or no linear molecules, cut at the XbaI site and 3'- or 5'-labelled. After a second HaeIII digest to create fragments labelled uniquely at one end, the DNA was loaded on a 8% polyacrylamide-urea sequencing gel alongside a Maxam-Gilbert sequencing ladder of the same fragment (fig.5). This enabled us to localize the S1 nicks at the nucleotide level. From the results, it is clear that the S1-digestion pattern is slightly different on both strands of the DNA. On the top strand, the nicks are mainly found around the 5'-part of the $(GGAA)_{12}$-cluster and on a stretch of 13 alternating purine-pyrimidine nucleotides just in front of it (nucleotides -417 to -405). On the bottom strand, an additional (but less pronounced) S1-sensitive region is observed towards the 3'-end of the $(GGAA)_{12}$ cluster (nucleotides -362 to -357). This corresponds only partly with the results of Christophe et al. (22), reporting the generation of "staggered ends" after S1-digestion. Furthermore, in the S1-sensitive region of our clones, the nuclease seems to cut mainly after A- and T-residues, while in the alternating purine-pyrimidine region there seem to be no preferred nucleotides for nuclease attack.

The presence of an S1-sensitive site in the alternating purine-pyrimidine tract at position -416 to -404 is not surprising since this region can potentially adopt a Z-DNA conformation, and Z-DNA has been reported to be a preferential target for S1-nicking (37, 42).

The role, if any, of these S1-sensitive sites is still unclear. The presence of such sites in the promotor or upstream regions of many genes is however striking, and it is tempting to assume that they are indeed functional in gene expression or regulation. Possibly, S1-sensitive sites reflect the presence of conformational alterations in the DNA, which could act as general recognition signals for the nearby presence of a transcription-initiation or promotor region. These conformational changes

might facilitate or direct the binding of transcription or regulatory factors to more specific sequences on the promotor.

ACKNOWLEDGEMENTS

This work is supported by grants from N.F.W.O., F.G.W.O. (contract n° 3.0086.74), Nationale Bank and Onderzoeksfonds K.U. Leuven. We wish to thank Dr. M.Parker for supplying the C2 genomic clone, helpful suggestions and critical reading of the manuscript. Ms. A.M. Ickroth for technical assistance and Ms M. Coppens for typing the manuscript.

REFERENCES

1. Heyns, W. and De Moor, P. (1977) Eur. J. Biochem. 78, 221-230.
2. Heyns, W., Peeters, B. and Mous, J. (1977) Biochem. Biophys. Res. Commun. 77, 1492-1499.
3. Parker, M.G. and Scrace, T.G. (1978) Eur. J. Biochem. 85, 399-406.
4. Peeters, B.L., Mous, J.M., Rombauts, W.A. and Heyns, W.J. (1980) J. Biol. Chem. 255, 7017-7023.
5. Heyns, W., Peeters, B., Mous, J., Rombauts, W. and De Moor, P. (1978) Eur. J. Biohem. 89, 181-186.
6. Peeters, B., Heyns, W., Mous, J. and Rombauts, W. (1982) Eur. J. Biochem. 123, 55-62.
7. Peeters, B., Heyns, W., Mous, J. and Rombauts, W. (1983) Eur. J. Biochem. 132, 669-679.
8. Peeters, B., Rombauts, W., Mous, J. and Heyns, W. (1981) Eur. J. Biochem. 115, 115-121.
9. Delaey, B., Dirckx, L., Peeters, B., Volckaert, G., Mous, J., Heyns, W. and Rombauts, W. (1983) Eur. J. Biochem. 133, 645-649.
10. Parker, M.G., White, R. and Williams, J.G. (1980) J. Biol. Chem. 255, 6996-7001.
11. Parker, M., Needham, M., White, R., Hurst, H. and Page, M. (1982) Nucl. Acids Res. 10, 5121-5132.
12. Viskochil, D.H., Perry, S.T., Lea, D.A., Stafford, P.W., Wilson, E.M. and French, F.S. (1983) J. Biol. Chem. 258, 8861-8866.
13. Parker, M.G., White, R., Hurst, H., Needham, M. and Tilly, R. (1983) J. Biol. Chem. 258, 12-15.
14. Dirckx, L., Delaey, B., Peeters, B. and Rombauts, W. (1985) Arch. Int. Physiol. Biochim. 93, B29-B30.
15. Dirckx, L., Delaey, B., Claessens, F., Peeters, B. and Rombauts, W. (1985) Arch. Int. Physiol. Biochim. 93, B81.
16. Barnes, W.M. and Bevan, M. (1983) Nucl. Acids Res. 11, 349-368.
17. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
18. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
19. Mount, S.M. (1982) Nucl. Acids Res. 10, 459-472.
20. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142.
21. Maurer, R.A., Erwin, C.R. and Donelson, J.E. (1981) J. Biol. Chem. 256, 10524-10528.

22.Christophe, D., Cabrer, B., Bacolla, A., Targovnik, H., Pohl, V. and Vassart, G. (1985) Nucl. Acids Res. 14, 5127-5144.
23.McKnight, S.L. (1982) Cell 31, 355-365.
24.Pelham, H.R.B. (1982) Cell 30, 517-528.
25.Hurst, H.C. and Parker, M.G. (1983) EMBO J. 2, 769-774.
26.Harris, S.E., Mansson, P.6E., Tully, D.B. and Burkhart, B. (1983) Proc. Natl. Acad. Sci. USA 80, 6460-6464.
27.Kandala, J.C., Kistler, M.K. and Kistler, W.S. (1985) Biochem. Biophys. Res. Commun. 126, 948-952.
28.Williams, L., McDonald, C. and Higgins, S. (1985) Nucl. Acids Res. 13, 659-672.
29.Panthier, J.-J., Dreyfus, M., Tronik-Le Rouse, D. and Rougeon, F. (1984) Proc. Natl. Acad. Sci. USA 81, 5489-5493.
30.Karin, M., Haslinger, A., Holtgreve, H., Richards, R.I., Krauter, P., Westphal, H.M. and Beato, M. (1984) Nature 308, 513-519.
31.Van der Ahe, D., Janich, S., Scheidereit, C., Renkdwitz, R., Schütz, G. and Beato, M. (1985) Nature 313, 706-709.
32.Jost, J.P., Seldran, M. and Geiser, M. (1984) Proc. Natl. Acad. Sci. USA 81, 429-433.
33.Dean, D., Knoll, B.J., Riser, M.E. and O'Malley, B.W. (1983) Nature 305, 551-554.
34.Parker, M., Hurst, H. and Page, M. (1984) J. Steroid Biochem. 20, 67-71.
35.Baker, M. (1983) Biochem. Biophys. Res. Commun. 114, 325-330.
36.Jones, W.K., Yu-Lee, L.-Y., Clift, S.M., Brown, T.L. and Rosen, J.M. (1985) J. Biol. Chem. 260, 7042-7050.
37.Hayes, T.E. and Dixon, J.E. (1985) J. Biol. Chem. 260, 8145 -8156.
38.Maroteaux, L., Heilig, R., Dupret, D. and Mandel, J.L. (1983) Nucl. Acids Res. 11, 1227-1243.
39.Tautz, D. and Renz, M. (1984) Nucl. Acids Res. 12, 4127-4138.
40.Schon, E., Evans, T., Welsh, J. and Efstratiadis, A. (1983) Cell 35, 837-848.
41.Pulleyblank, D.E., Haniford, D.B. and Morgan, A.R. (1985) Cell 42, 271-280.
42.Evans, T., Schon, E., Grazyna, G.-M., Patterson, J. and Efstratiadis, A. (1984) Nucl. Acids Res. 12, 8043-8058.
43.Parker, M.G., Needham, M. and White, R. (1982) Nature 298,92-94