Volume 11 Number 8 1983 Nucleic Acids Research

The uteroglobin gene region: hormonal regulation, repetitive elements and complete nucleotide sequence of the gene

G.Suske, M.Wenz, A.C.B.Cato and M.Beato+

Institut für Physiologische Chemie, Philipps Universität, 3550 Marburg, FRG

Received 1 March 1983; Accepted 22 March 1983

ABSTRACT

Differential uteroglobin induction represents an appropriate model for the molecular analysis of the mechanism by which steroid hormones control gene expression in mammals. We have analyzed the structure and hormonal regulation of a 35 Kb region of genomic DNA in which the uteroglobin gene is located. The complete sequence of 3,700 nucleotides including the uteroglobin gene and its flanking regions has been determined, and the limits of the gene established by S1 nuclease mapping. Several regions containing repeated sequences were mapped by blot hybridization, one of which is located within the large intron in the uteroglobin gene. Analysis of the RNAs extracted from endometrium, lung and liver, after treatment with estrogen and/or progesterone shows that within the 35 Kb region, the uteroglobin gene is the only DNA segment whose transcription into stable RNA is induced by progesterone.

INTRODUCTION

Uteroglobin synthesis is subject to differential hormone regulation in various tissues of the rabbit (for a review see Ref.1). In the endometrium, progesterone and to a lesser extent estradiol, enhance the transcription of the uteroglobin gene (2,3). In the oviduct, estradiol is the only inducing steroid and in the lung, uteroglobin is synthesized at high rates and subject to glucocorticoid control. In the liver, this gene is not expressed, although receptors for the different hormones have been found in this tissue.

According to the generally accepted model for steroid hormone action, the hormone influences gene transcription by virtue of its interaction with a specific receptor protein. The steroid-receptor complex, after translocation into the cell nucleus and binding to chromatin, modulates specific gene expression. An important question that should be answered before starting an investigation on the interaction between steroid-receptor complex and the inducible gene, is whether other genes in the same region of DNA are subjected to similar hormonal control. Certainly the answer to this question will determine the strategy to be followed in later experiments, i.e. the region of DNA that should be considered in binding experiments with the hormone

receptor.

The uteroglobin gene has been recently characterized (4). It is a single copy gene composed of three exons and two introns. In this paper we describe a structural and functional analysis of a 35 Kb region of the rabbit genome, that contains the uteroglobin gene. The limits of the gene have been established by S1 mapping, and the nucleotide sequence of a 3.7 Kb region including the complete gene and flanking sequences has been determined. The organisation of repeated sequence elements within the 35 Kb region is described. The results of blotting experiments with RNA extracted from different tissues after treatment with estradiol and progesterone show that within the 35 Kb region the only progesterone inducible gene is the uteroglobin gene.

MATERIALS AND METHODS

The isolation of recombinant phages containing the uteroglobin gene from a rabbit gene library has been reported (4). For this study two recombinant phages λUG 9.2 and λUG 9.3 covering 35 Kb of rabbit genomic DNA were used (Fig.7). The insert of phage λUG 9.2 contains the complete uteroglobin gene plus 15 Kb of DNA upstream of the 5' end and 700 bp downstream of the 3' end. Phage λUG 9.3 contains an insert that starts in the first intron of the uteroglobin gene and extends 17 Kb downstream of the 3' end.

DNA sequence analysis

For the analysis of the nucleotide sequence several restriction fragments containing uteroglobin gene sequences were subcloned into the plasmid pBR322. All restriction sites used for end-labelling of the DNA fragments, were also sequenced starting from alternative sites. The procedure for labelling 5'ends and 3'ends were as described (4), and the Maxam and Gilbert procedure was used throughout (5). 90% of the sense strand and 80% of the antisense strand were sequenced.

S1 mapping

The 5' end and 3' end of the uteroglobin gene were determined by the procedure of Berk and Sharp (6) as modified by Weaver and Weissmann (7). Partially purified preuteroglobin mRNA (8) or poly(A)$^+$RNA extracted from various tissues (9,10) were hybridized to [^{32}P]end-labelled DNA (specific activity 250,000cpm/pmol) and subjected to treatment with S1 nuclease. The conditions of hybridization and nuclease digestion are given in the legend to the corresponding figures. The digestion products were analyzed by electrophoresis through 8% or 20% polyacrylamide sequencing gels followed by autoradiography (5).

Repeated sequences

Two methods were employed to detect the presence of repeated sequences in the uteroglobin gene region. In the first, recombinant phages or plasmids containing subcloned fragments of the gene were digested with restriction enzymes and, after electrophoresis in agarose gels, were transfered to either nitrocellulose (11) or DBM-paper (12). In some cases the transfer was bidirectional (13).

The filters were then hybridized in 4xSSC at 67°C overnight (4) to rabbit liver DNA (14), labelled by nick-translation to a specific activity of 2-8x10^7 cpm/µg (15). After extensive washing the filters were dried and autoradiographed with intensifying screens. The second method, contact hybridization, was as described (16,17) with the exception that the restriction fragments were labelled at the 5' ends with [γ^{32}P]ATP and T4-polynucleotide kinase (5). The gel was then contact hybridized to a DBM-filter loaded with rabbit liver DNA. Washing and autoradiography were as described (16).

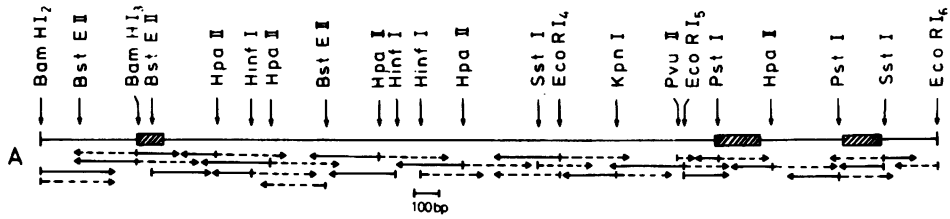
Transcription mapping

To detect transcripts of the uteroglobin gene, poly(A^+) RNA was extracted from different tissues (9,10), partially digested with alkali and labelled with [γ^{32}P]ATP and T4 polynucleotide kinase to specific activity of 1-5x10^7 cpm/µg (18). The labelled RNA was hybridized to the same DBM-filters used to detect repeated sequences (see above). Hybridization was at 42°C overnight in the presence of 50% formamide, 5xSSC and 10 µg/ml poly $(A)^+$RNA. The filters were extensively washed, and autoradiographied as mentioned above.

RESULTS

The restriction map and the complete nucleotide sequence of the uteroglobin gene and its flanking sequences is shown in Figure 1. These data essentially confirm the previously published sequence of the coding region of the uteroglobin gene (4) and further supplement it by completing the sequence analysis of the first long intron and the flanking sequences.

The location of the 5' end of the gene was established by the S1 nuclease mapping procedure (6,7) using partially purified uteroglobin mRNA isolated from endometrial polysomes of pseudopregnant rabbits treated with estradiol and progesterone (8). The DNA probe was obtained from the plasmid pUG 5.0, that contains the 5 Kb EcoRI $_{3-4}$ fragment cloned into pBR 322 (Fig.7C). The plasmid was digested with BstEII, labelled at the 5' end with [γ^{32}P] ATP, and the 337 bp BstEII fragment was digested with HinfI (Fig. 2). After hybridization and treatment with S1 nuclease the resistant products were



B

```
1 GATCCTGGGTGCTCCTATCCTGTCACAGCCACTTGGGCCTCAGCTTCTCCTGGGGTGACAGGAGACACACTTGCTTCTGTTGGTCTGGGCTGCCCAAATG
101 CCTTGGTTCAAAGAGAAGACGCAGGTGGCTAGGTCACCATGCCCTCGGGGGCAGGCACCCCTTACCACACCCCCGCACAAGACTCCGGGCTCTCCCTCCC
201 CCCGAGGGACCCATTGTGTGAGTTCAGTTTCAATAGGGATGGAAACTGGATTGAGAAAAGGGAATATTTACTTATCCCACCAAGTCAATGCCCAAGTAAA
301 TAATGCAGTCAAGTAAGTGGAGCCCAGGCCCTGCCCTCTTCTATCTGGGCACTGCCCGGAGAATACAAAAAGCACCTGACGGCCGTCCCCCTCAAGATC
401 ACCGGATCCAGAGCCAGCCCAGAGCCTTCCCATTCTGCCACCATGAAGCTCGCCATCACCCTCGCCCTGGTCACCCTGGCTCTCCTCTGCAGCCCTGGTG
                                    MetLysLeuAlaIleThrLeuAlaLeuValThrLeuAlaLeuCysSerProA
501 AGTGCCCAGCGTCCCTTTCCCCCAGGCAACCCTCTCCGCCACCCACTTCTGCTCAGAGGGAGAGGTGAGCTGCCTGTGCCCTCCCTGCAGCTGCTGGCAG
601 GGCCTGAGAACGTGGACTTTCGGAAAATGGGGTCTGGTGAGCGACCCCACGTTGAGAACCTGGGGAGGGCCCTGCGGCTGTTAGGAAACTAGGAAGGATG
701 TCCTCTTCCACATCCAACAACCTCCGCGTTCCGGGGGACAGGACAGGCTAGAGGGAGGGAGGAGGCCCCTGCGGGTGAACAGTGCAACAGCAAGTGGGAG
801 AGGAATACAGTCTCCCAAAGCCCAGACTTGTGTAGACCCCAAGATGCCCAAGGACCACTGGGGCCTCCGCCCGACTCTGGCAGCCACAGGGCCTGAAAGG
901 GGAGCTGCAGAGACGGACATCCCCGAAAGAAACCCAAGCCAAGGCACAGCTTCCGGGTCTGGGCCTAAAGGACAAGTCCAGAAAAAGGATCTTCCCTTTG
1001 GGACAACGCCAGCCCAGAGTGCCGTTTTCTAAACAATTCTGTTCTCTGCAAAAAGGGCTACAAGTCTCCAAGTGGCTGCTGCTGTTGCTAGGAGGAAAGG
1101 AATGGGAAAGGAGAAAGCGGGGAGGAAGGAGGGGAGAAGAGGGGAGGGCTCACATCTCCCTCCGAGCGCCCTGCCATTTGGGAAGGAGGTTACCTAACT
1201 CGGCTTTTCACCCTGCATCAGCCTCTTCCACTCCCATCTGAAAATGCTGTTGCTGTTGTTATTGTTGTTGTTTTCATCAGCAAACCCCAACGGAAAGTTT
1301 CACACAACTCCTAGATGTGAGTTTGACGACCTTGGGGGATTGGAAAAACCAAACCGAGAGGGAGTGGGAACAACTGTTTGGCAGAGAGTGTTACTTGTTT
1401 CTCCGGTGGGAAATGACATCATTTGGGGCTGAAATGTATGAATTAGAAGATCTGGGTACTGAGTATAGAGGTGGTGACTCAGGAGGACAGGACTGAACCA
1501 CCCCCCGCCTCTTTTTTAAAAAACAGATTGTATTTATTTGAGAGGTAGAGTTACAGAGAGAGGGAGAGAGAGAGAGAATCTTCCATCCACTGGTTCACTC
1601 CCCAGTTGGCCACAGTGGCAGAGCTGGGCCGATCCAAAGCCAGGAGCTTTTTCTGGGTTTCCCATGCAGGTATAGGGGCCCAAACACTTGGGCCGTTTTC
1701 CACTGCTTTCCTAGGCCAAAAGCAGAGAGCTGGATCGGAAGAGGAGCTGCCGGACATGAGCTGACATCCATATGGGATGCTGGTGCTGCAGGTGGAGGCT
1801 TAGCCTACAATGCCACAGCACCAGCCCTAGACTGAAACCCTTTAAGGGGTGGATCAGGGACACAACACCCATTCATATACTTTTGTGTGCACACGTGTAA
1901 TAATGTCTGCTCTTAACCCTTTGACATACACCTCTTTTCTTTCTTACAACCTCATGAGGAGGGGACATTGTACCCATTTTACAGATGATAAAAACTGAGC
2001 TATTAGCACTTTGTAAGATCAAGTAACTCCTGAGGGCATAGTAAGTTGCAAGGTGGGGATCTGAGCTCTGGAGAGGGTGTGGTTTGGAATGGCCACCCTG
2101 AGTAGGGCCTCCAGTCTGAGAGCAGACATTGAGTTGGCACATGAAGCAGAGAATTCAAGGACAGGGATTTGAGACTTCATGGGAAGATGGAGGCAAATAG
2201 AGGTGGGGCGTGGTGAGTCCCCCCCACAACCTGGGGGCAGGGGCCTGAGAAACAGTTATAGGGACAGTGGCACAAGTGGTTGGTCTGTGTTGAGCTGAGC
2301 AAACCACAAGGCCCACCACTCACATGTGCACGCCTGTCCCGTGCCAGGTGGGGAGCCCCATGCTGGGCTGGTACCAGCCTCAGCCTGGCTTTGCCACATC
2401 CCCTCAGTGACACGTGGCTCCAACCACCATGGACCACAGAAGACAGGCAAGGTCCTGGTAGGAACCAAAGCTCTCAGTGGCCTCTTCCCAGCCCCACCCT
2501 GCATTAAAAATCAGCCCAGCAGCTTCTATTATCAATACTCCACCCAGAACATGAGCATAAAAAACAAATGGCACCCTGCCCCACGCCAGCCCAGGAAGTC
2601 CAGGGAAGGCTCAGCCAACCACAAAACACCAGCTGTAACCTAGAGAGAATTCACCCAGCGCTTAGAGCACCAGCAGAGCTGATTTCCCCTCCCCGTCC
2701 ATGAAAAGCTGTCTAGTCTCAAACTGACTGACGAGCTTGGCTGTGTGCCCCTGGTTCTTGGGCTTCTCCTGCTCTGTTCTCAGCATCTGCAGGCATCTGC
                                                    CaSerAlaGlyIleCys
2801 CCGAGATTTGCCCACGTCATTGAAAACCTCCTCCTGGGCACGCCCTCCAGTTACGAGACATCCTTGAAGGAATTTGAACCTGATGACACCATGAAAGATG
     ProArgPheAlaHisValIleGluAsnLeuLeuLeuGlyThrProSerSerTyrGluThrSerLeuLysGluPheGluAspAspThrMetLysAspA
2901 CAGGGATGCAGATGAAGAAGGTCTTGGACTCCCTGCCCCAGACGACCAGAGAGAACATCATGAAGCTCACGGTAACCAGTGCCTATTGCCCACGCCTATA
     laGlyMetGlnMetLysLysValLeuAspSerLeuProGlnThrThrArgGluAsnIleMetLysLeuThr
3001 AGTGGCTTCCGGAGCCCCTCAGCTGTGGTCTCATCCATGCTAAGACCCTTTCCTAACCCAGAGAAAAGACTTGCAATACCAACATCCCCGTAGTACCCA
3101 TGTCCTCACGGGCAAGCCACCTCTCTAGGCATTTCCCCTTACCCATTTCTAAAATGGCTTTTTGAGAAGGAAAAAAAATGTCCTTATAATCATCATTCTT
3201 CCCAGAGTTCATTGGACATCTCCACTGACATTGACGTTTGTGTCTCCGCATCTCTGCCTCACTGCAGAGGTCACGGGGCTTTTGTTTCTCTTTTTCAGGA
                                                                                                    Gl
3301 AAAAATAGTGAAGAGCCCACTGTGTATGTAGGATGGAGGAATCCGAGGTCCTGCGGACTGAGAAGCCGAAGATTTCCACCTGCTGAAGCCCCTGCTGCTG
     uLysIleValLysSerProLeuCysMet
3401 CCCCTGGCCCCTTGGGTCCCCCACCCACCCAACCCAGCCAGCCTTTGCTTTCAATAAACTGCAAGCAGATCACATCCGTCGTCCTGAGCTCTTATTTACC
3501 TGCTTGAGGGAGAGAGGCGTGCACCTGCCAGGAAGGGTAGCCAGGACTCCCAGGAAGGCGGCGGAGCCTGCCCTTGGACCCCACGAGGAAACTCTGACGC
3601 CTTCCAAAGTCCTTTCCTCTAGGTTTCTGCCATTGGCCACCCAAAAAGCTGAGCCCTCTCCCCTTACATTCTACCACCTACGTCTCAAACTCCACCTGTG
3701 GGAGAATTC
```

analysed on a 8% acrylamide/7M urea sequencing gel (5). The autoradiogram shows a prominent group of bands located about 47 nucleotides upstream of the translation initiation codon (Fig. 2). The assignment of the initiation of transcription to the adenine at position 396 (Fig.1) was based on a comparison with the corresponding DNA sequence ladder, taking into consideration the different electrophoretic mobility of fragments generated by S1 nuclease treatment (19). This position, that we will tentativelly call "cap" site, coincides with a previously published 5' end of uteroglobin RNA, determined by sequencing and primer extension with a cloned cDNA fragment (20). When poly(A)$^+$RNA from endometrium or lung of rabbits treated with estradiol was used instead of partially purified uteroglobin mRNA from pseudopregnant rabbits, fragments of similar lengths were protected against digestion with S1 nuclease (Fig. 2, lanes 2,3).

A longer exposure of the autoradiogram results in the detection of weak bands located about 9, 18, 30, 50, 70 and 100 nucleotides upstream of the main "cap" site (Fig. 2A). With the exception of the band at position -30 all other bands persist after longer time of digestion with S1 nuclease (Fig. 2A, lane 2'). The bands generated by initiation upstream of the "cap" site are particularly evident when the appropriate single stranded $\underline{\text{BamHI}}_{2\text{-}3}$ restriction fragment is used as labelled probe (Fig. 3). Under these conditions the hybrid formed between the main population of mRNA molecules and the labelled DNA is only 9 nucleotides long and, therefore, unstable, and not protected against digestion with S1 nuclease. The other bands, representing minor start points can be thus clearly detected (Fig.3, lane1,2). The same pattern of protected bands upstream of the main "cap" site is observed when poly(A)$^+$ RNAs from endometrium or lung from estradiol treated rabbits were used for the S1 nuclease mapping (Fig.2A, lanes 2,3). Thus it is evident that these minor start sites are not restricted to the transcription of the

Fig.1 Restriction map and nucleotide sequence of the uteroglobin gene and its flanking sequences.
A. Restriction map of the 3.7 Kb region of rabbit DNA containing the uteroglobin gene. Only sites relevant to the sequences analysis or S1 nuclease mapping are shown. The boxes represent the exons of the uteroglobin gene. The sequencing strategy is indicated underneath of the restriction map. The solid lines refer to the antisense strand.
B. Nucleotide sequence of the region shown in A. The main "cap" site and the polyadenylation site are indicated by arrows. The ATA-box and the polyadenylation signals were underlined. The region of the gene expressed into protein is shown by the corresponding amino acid sequence.

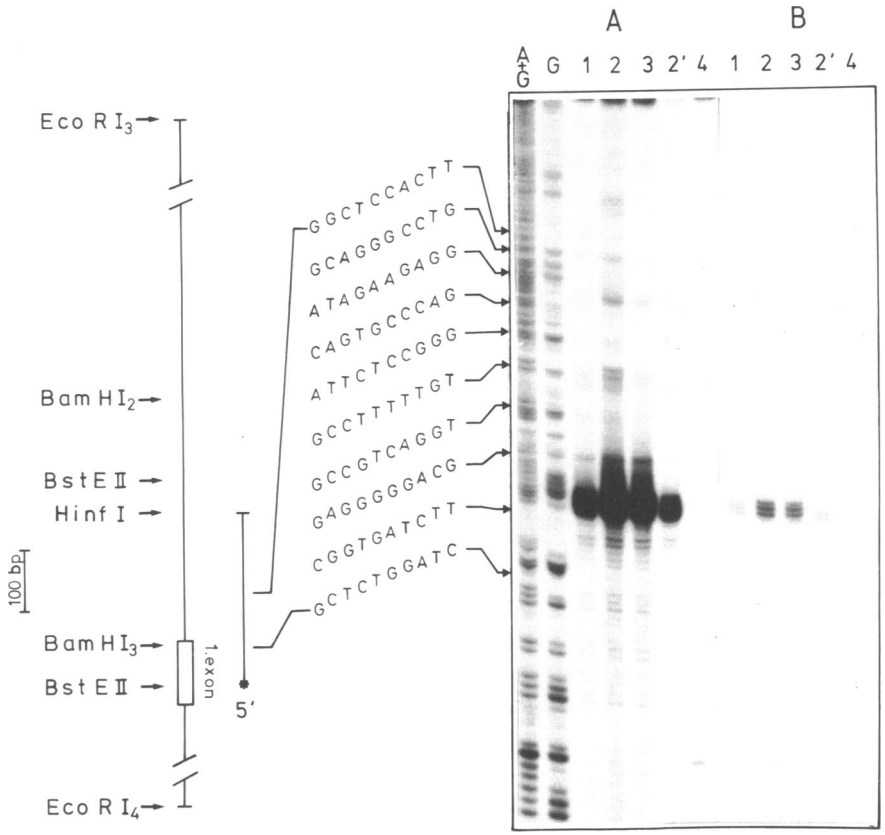


Fig.2 Mapping of the 5' end of the uteroglobin mRNA, by S1 nuclease digestion
The diagram on the left shows the restriction map of the DNA region around the first exon of the uteroglobin gene, and the BstEII-HinfI fragment used as radioactive probe. On the right are shown the autoradiograms of an 8% acryl amide sequencing gel exposed for either 24 hr (B) or 4 days(A). Two sequencing ladders, A+G and G, are shown together with the corresponding nucleotide sequence. Hybridization was performed at 43°C for 16 hr in 80% formamide, 0.4M NaCl, 40mM pipes, pH=6.5, 5mM EDTA. Treatment with S1 nuclease (250 U/ml) was performed at 30°C for 2 hrs. The following RNAs were used for hybridization : Lane 1, partially purified uteroglobin mRNA from endometrium of pseudopregnant rabbits, treated with estradiol and progesterone (0.3µg), Lane 2, total poly(A)$^+$RNA from endometrium of estradiol treated rabbits (3µg). Lane 3, total poly(A)$^+$RNA from lung of estradiol treated rabbits (3µg). Lane 2', the same as Lane 2, but treated with S1 nuclease for 5 hrs instead of 2 hrs. Lane 4, control assay without RNA.

gene in endometrium nor are they subject to differential hormonal control. Alternative initiation, therefore, does not appear to be physiologically relevant for the regulated expression of the uteroglobin gene.

The polyadenylation site of uteroglobin mRNA was also determined using the

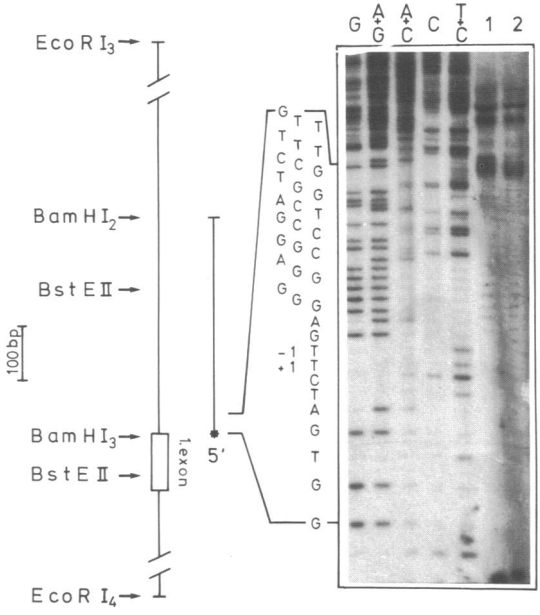


Fig.3 S1 nuclease mapping of alternative initiation site on the uteroglobin gene. The diagram on the left shows the single stranded BamHI fragment used as radioactive probe. On the right is shown the autoradiogram of a 20% acrylamide sequencing gel, containing 5 sequence ladders. Partially purified uteroglobin mRNA (0.3µg) was used for hybridization and the s1 nuclease treatment (250 U/ml) was carried out at 30°C for either 1 hour (lane 1) or 5 hours (lane 2).

S1 nuclease mapping technique (Fig.4). The 3' end of the uteroglobin mRNA was mapped to the cytidine residue at position 3076 starting from the "cap" site (Fig.1). This position is located 19 nucleotides downstream of the sequence 5'-AATAAA-3', that is generally accepted as a polyadenylation signal for eucaryotic mRNAs (21). Therefore, the 3' untranslated region of uteroglobin mRNA comprises 143 nucleotides in addition to the polyadenylyl residues. In conformity with the results obtained at the 5' end of the gene the position of the polyadenylation site at the 3' end is the same in endometrium and lung and is not influenced by the hormonal treatment of the animals (Fig.4, lanes 1,2,3). Overexposure of the autoradiograms did not result in the detection of any additional bands other than those already described.

Repetitive elements

Repeated sequences are ubiquitously interspersed through the eukaryotic genome (22), and have been proposed to be involved in the regulation of gene expression (23). For this reason we have studied the distribution of repetitive sequence elements in the 35 Kb region containing the uteroglobin gene. Two different procedures were employed. In the first method DNA from the 35 Kb region was cleaved with restriction endonucleases, and after electrophoretic separation the fragments were transferred to nitrocellulose or DBM-filters (11,12) and hybridized with nick-translated total genomic rabbit

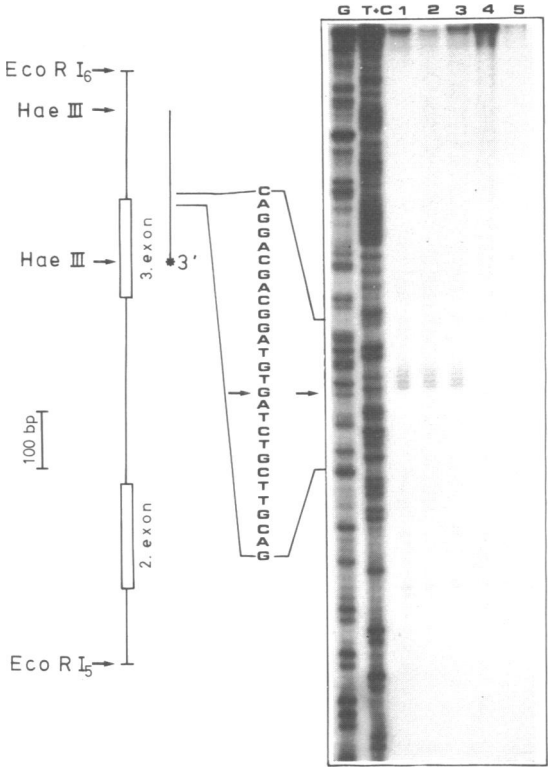


Fig.4 Mapping of the 3'end of uteroglobin mRNA by S1 nuclease digestion : the diagram on the left shows the restriction map around the third exon of the uteroglobin gene and the single stranded HaeIII fragment used as radioactive probe. On the right is shown the autoradiogram of a 8% acrylamide sequencing gel with the sequence ladders for G and T+C. After hybridization with various RNAs, the hybrids were treated with S1 nuclease (250 U/ml) at 30°C for 2 hrs. The following RNAs were used: Lane 1, partially purified uteroglobin mRNA from endometrium of pseudopregnant rabbits (0.3µg). Lane 2, total poly(A)$^+$RNA from endometrium of estradiol treated rabbits (3µg). Lane 3, total poly(A)$^\pm$ RNA from lung of estradiol treated rabbits (3µg). Lane 4, control without added RNA,not treated with S1 nuclease.Lane 5, control without added RNA, treated with S1 nuclease.

DNA. This method includes a partial alkaline hydrolysis of the DNA and therefore detects preferentially repeated sequences located in large restriction fragments. In the second method, the restriction fragments of the 35 Kb region, were labelled at the 5' end with [γ^{32}P] ATP, and after electrophoresis were contact hybridized to total rabbit DNA bound to DBM-filters (16, 17). This method detects preferentially repetitive elements located in small restriction fragments. The results of a set of experiments with λUG 9.2 and λ UG 9.3 digested with EcoRI and BamHI are shown in Fig. 5 and summarized in Fig. 7. Repeated sequence elements were found scattered throughout the uteroglobin gene region and are present even within the large intron of the gene, as shown by the positive signal of the 1.7 Kb BamHI-EcoRI-fragment (Fig. 5B). The exact position of this repeated sequence was located within a 340 bp BglII-PstI$_5$fragment by analysis of the subcloned BamHI-EcoRI 1.7 Kb fragment of the first intron (Fig. 5D,E). This region contains a cluster of 24 purines with the sequence 5'-GAGAGA-3' repeated 3 times (Fig. 1b, nucleotides 1555 to 1578). In addition, a weaker signal was detected in the

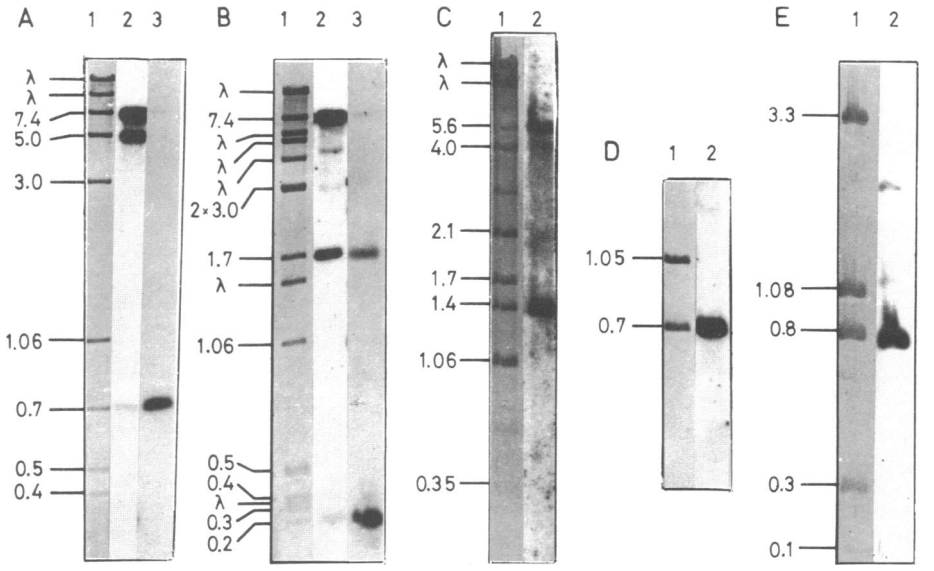


Fig.5 Repetitive sequence elements in the uteroglobin gene region.

Rabbit liver DNA was hybridized as described in Materials and Methods to the following blots :
A. λ UG 9.2 digested with EcoRI
B. λ UG 9.2 digested with EcoRI and BamHI
C. λ UG 9.3 digested with EcoRI
D. Insert of pUG 1.7 (BamHI-EcoRI) digested with BglII
E. pUG 1.7 digested with Pst I
The size of the fragments is indicated in kilobases.
λ represents fragments originating from the phage vector.
Lane 1 : ethidium bromide staining of the restriction fragments.
Lane 2 : autoradiogram of the blot hybridization with nick-translated rabbit liver DNA.
Lane 3 : autoradiogram of the contact hybridization with labelled restriction fragments. Longer exposures of these autoradiograms (not shown) allowed the detection of signals corresponding to the larger fragments seen in Lanes 2, A and B.

adjacent $PstI_5$-$EcoRI_4$ fragment (Fig. 7A). The significance of these repeated sequences is uncertain. Hybridization of the blots shown in Figure 5 with a labelled plasmid containing human Alu-family sequences, pBLUR 8 (24), yielded negative results (data not shown). Thus it does not appear to be any homology between sequences in the 35 Kb region and the human Alu-family of repetitive sequences.

Hybridization with labelled RNA

The occurrence of tissue specific or hormone dependent transcripts from the uteroglobin gene region was investigated by hybridization of Southern

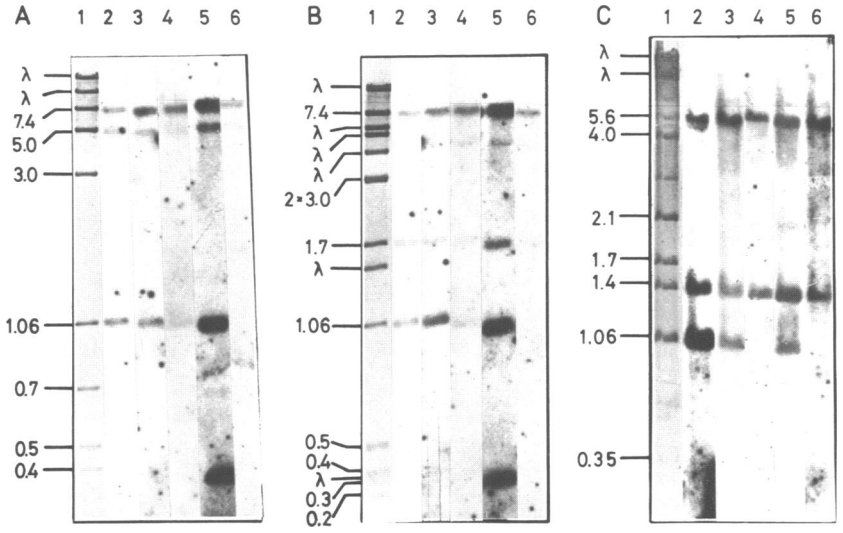


Fig.6 Blot hybridization with labelled RNA. Blots identical to those shown in Fig.5 A,B and C were hybridized to labelled poly(A)$^+$RNA from the following sources : Lane 2, endometrium from rabbits treated with estradiol and progesterone (8), Lane 3, endometrium of rabbits treated with estradiol (8), Lane 4, endometrium of control estrous rabbits, Lane 5, lung of rabbit treated with estradiol, Lane 6, liver of rabbits treated with estradiol, Lane 1, ethidium bromide staining of the restriction fragments. Symbols as in Fig.5.

blots similar to those shown in Fig. 5 to labelled poly(A)$^+$RNA extracted from various tissues after different hormonal treatments (Fig.6). All restriction fragments that showed a clearly positive signal in the repeated sequence analysis were also positive in the transcription mapping study (Fig. 7). The 5.6 Kb $EcoRI_{7-8}$ fragment showed a stronger hybridization signal with endometrial RNA from estradiol treated animals than with RNA from other sources (Fig. 6C). All the other fragments containing repeated sequences, showed hybridization signals of similar intensity independently of the source of labelled RNA. The hybridization signal observed with the 5.0 Kb $EcoRI_{3-4}$ fragment (Fig.6A) was mapped to the 1.7 Kb $BamHI_3$-$EcoRI_4$ fragment containing the long intron of the uteroglobin gene (Fig.7B). This signal was mainly due to hybridization of the RNA to the 0.34 Kb BglII-PstI fragment containing the repeated sequence elements (Fig.7A), but faint signals corresponding to other fragments of the large intron were also detected with labelled RNA from lung or endometrium of rabbits treated with hormones (data not shown).

In addition to these signals we detected hybridization of labelled RNA to the 1.0 Kb $EcoRI_{5-6}$ fragment containing most of the expressed uteroglobin

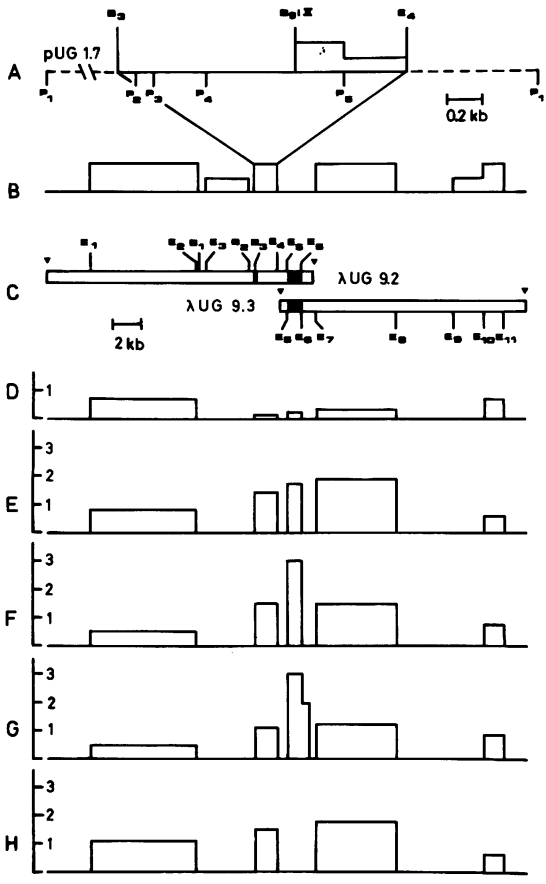


Fig.7: Summary of the blot hybridization data. The restriction maps of λUG 9.2 and λUG 9.3 are shown in C. The black boxes indicate the exons of the uteroglobin gene. The fragments containing repeated sequence elements are shown in B. The intensity of the signal was quantitated by scanning the autoradiograms. An expanded map of the 1.7 Kb BamHI-EcoRI fragment subcloned in pB322 is shown in A. The location and intensity of the repeated sequences is indicated.

The results of hybridization with labelled RNA from endometrium of control estrous rabbits is shown in D. The ordinate represents the relative intensity of the individual bands in the autoradiogram. The results obtained with other RNAs are shown in lines E to H. These ordinates represent the relative intensity of the individual bands divided by the relative intensity of the corresponding band in the control plot (D). Labelled poly(A)$^+$RNA from the following sources are shown : E, endometrium of estradiol treated rabbits, F,endometrium of rabbits treated with estradiol and progesterone, G, lung of estradiol treated rabbits, H, liver of estradiol treated ones.

mRNA sequences (Fig.6). The relative intensity of this signal was low with endometrial RNA from control estrous rabbits and increased 8-fold after treatment with estradiol (Fig.7E). A similar increase was found for the signals corresponding to the flanking 1.7 Kb $BamHI_3$-$EcoRI_4$ and to the 5.0 Kb $EcoRI_{7-8}$ fragments. After the administration of progesterone, however, only the signal corresponding to the uteroglobin gene increases (15 fold over the control, Fig.7F) whereas the relative intensity of the other signals remain similar. With poly(A)$^+$RNA from liver, a tissue that does not express the uteroglobin gene (1), no signal can be detected corresponding to the 1 Kb $EcoRI_{5/6}$ fragment (Fig. 7H). Thus, within the 35 Kb region only the transcription of uteroglobin gene responds to progesterone administration in

a way that leads to an increased level of stable transcripts in endometrial cells. A high concentration of stable transcripts hybridizing to the expressed uteroglobin gene region was also detected with lung RNA from estradiol treated rabbits (Fig.7G). This is not surprising as the uteroglobin gene is expressed at high rates in the lung, independently of progesterone administration (25). With labelled lung RNA an additional signal is detected corresponding to a 0.4 Kb EcoRI fragment immediately downstream of the uteroglobin gene extending from $EcoRI_6$ to the end of the DNA insert in λUG 9.2 (Fig. 7G). The significance of this signal remains unclear, as we could not find indication of alternative polyadenylation sites with lung RNA.

DISCUSSION

The nucleotide sequence of the uteroglobin gene reported here confirms published data, and serves to correct minor errors of a previous report that only included one thousand nucleotides (4). Salient features of this nucleotide sequence including the mosaic structure, exon-intron boundaries and palindrome structure of the 5' flanking sequences have been discussed previously (4).

A computer program (26, M. Kröger, personal communication) was used in a search for sequence homologies between the uteroglobin gene and other steroid hormone inducible genes. A 19 nucleotides consensus sequence reported to be characteristic of progesterone inducible genes (27), was found in several positions in the large intron of uteroglobin, but also in the introns and 5' flanking sequences of the rabbit ß-globin gene (28,29). Similar results were obtained in a search for homologies to another 23 nucleotides consensus sequence that has been reported to be present in several glucocorticoid responsive genes (30). Recently a 10 nucleotides sequence, 5'-TCAGTTCTGA-3' has been found around 100 base pairs upstream of the glucocorticoid inducible gene for tryptophan oxygenase (31), that is homologous to sequences in the same position of the promoter for MMTV (32). No homologous sequence was found in the putative promoter region of the uteroglobin gene, but the sequence, 5'-TGAGCTGTTA-3', is found 14 base pairs downstream of the polyadenylation site in the uteroglobin gene (Fig.1). This sequence is identical to that found in the LTR of MMTV by Fassel et al. (33). Other sequences exhibiting 90% homology are found in the large intron of uteroglobin, whereas in the rabbit ß-globin gene the maximal homology to this sequence was less than 80%. The significance of these homologies remains to be established in receptor binding studies (34).

Besides the reading frames used for uteroglobin expression there are several open reading frames, whose physiological meaning remains unknown. Two of these are located in the first intron and are preceded by hypothetical regulatory sequences. Between nucleotides 844 and 1029 in the sense strand there is an open reading frame for 62 amino acids preceded by a GAATA-box (at position 802) similar to that found at the 5' end of the gene. In the opposite strand there is another open reading frame for 82 amino acids (positions 2504 to 2258) preceded by a "TAATA"-box located 27 nucleotides upstream. Since none of these open reading frames gave a positive signal in the transcription mapping experiments, they do not appear to be efficiently transcribed into stable RNA products.

The main initiation site for uteroglobin gene transcription could be mapped by the S1 nuclease technique to an adenine located 47 nucleotides upstream of the ATG initiation triplet. Similar results were obtained with poly(A)$^+$RNA from endometrium and lung of either pseudopregnant or estrogen treated rabbits, thus demonstrating that the site of initiation is not subject to tissue specific or hormonal regulation. The significance of the weak signals observed upstream of the main "cap" site remains questionable. That they are not trivial artefacts is evident from controls with tRNA or poly(A)$^+$RNA from cells that do not synthesize uteroglobin, where no signals were detected. Some of these bands map to positions (-70 and -100) that are preceded by short nucleotide sequences resembling the cannonical TATA-box. At -99 there is a 5'-TAAATAA-3' sequence and at -134 there is a 5'-GAATAT-3' sequence very similar to the one preceding the main initiation site. Further experiments with primer extension and eventually *in vitro* transcription are needed to elucidate the relevance of these point bands.

Repeated sequences, as detected by two independent methods, were distributed throughout the uteroglobin gene region, and within the first long intron of the gene. A comparison of the sequence of the *BglII-EcoRI*$_4$ fragment that yields the main signal in the repeated sequence analysis, with published *Alu*-family sequences (for a review see Ref. 35) resulted in the detection of a sequence, 5'-CCAGCCCTAG-3', (bp 1822-1831 in Fig.1), homologous to the consensus sequence 5'-CCAGCCTGG-3', that is highly conserved in human, monkey, mouse and chinese hamster (35). Nevertheless, no positive hybridization signal was detected when a human *Alu*-family clone was used to probe the 35 Kb region.

As judged by transcriptional mapping the poly(A)$^+$RNA from different tissues contain sequences that hybridize to the DNA fragments where repeated sequences

are located. Whether this RNA is directly transcribed from the repeated sequences or originates in other regions of the rabbit genome, can not be decided with this technique. Whatever the origin of this RNA, we found only small differences in the signal intensity after different hormonal treatments. The only region of the DNA that yields an enhanced signal with RNA from progesterone treated rabbits, is the expressed uteroglobin gene region. Thus, in the 35 Kb region no other transcriptional unit as the uteroglobin gene is subject to similar progesterone control. We do not know whether progesterone induction of the uteroglobin gene occurs through a direct interaction of the hormone receptor with the uteroglobin promoter, as proposed for glucocorticoid induction of the transcription of mouse mammary tumor provirus (34). The results presented here, however, encourage us to search for DNA sequences around the uteroglobin gene that could be recognized by the progesterone receptor.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (SFB 103,B1) and by Fonds der Chemischen Industrie.

+To whom correspondence should be sent.

REFERENCES

1. Beato, M., Arnemann, J., Menne, C., Müller, H., Suske, G. and Wenz, M. (1983) in "Regulation of Gene Expression by Hormones" (ed. K.W. McKerns) Plenum Publ.Co, New York, in press.
2. Müller, H. and Beato, M. (1980), Eur. J. Biochem. 112, 235-241
3. Kumar, N.M., Chandra, T., Woo, S.L.C. and Bullock, D.W. (1982) Endocrinol. 111, 1115-1120
4. Menne, C., Suske, G., Arnemann, J., Wenz, M., Cato, A.C.B. and Beato, M. (1982) Proc. Natl. Acad. Sci. USA 79, 4853-4857
5. Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560
6. Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732
7. Weaver, R.F. and Weissmann, C. (1979) Nucleic Acid Res. 7, 1175-1193
8. Arnemann, J., Heins, B. and Beato, M. (1977), Nucleic Acids Res. 4, 4023-4036
9. Auffray, C. and Rougeon, F (1980), Eur. J. Biochem. 107, 303-314
10. LeMeur, M., Glanville, N., Mandel, J.L., Gerlinger, P., Palmiter, R., and Chambon, P., (1981) Cell 23, 561-571
11. Southern, E.M. (1975), J. Mol. Biol. 98, 503-517
12. Wahl, G.M., Stern, M. and Stark, G.R. (1979) Proc. Natl. Acad. Sci. USA 76, 3683-3687
13. Smith, G.E., Summer, M.D. (1980) Anal. Biochem. 109, 123-129
14. Blin, N. and Stafford, D.W. (1976) Nucleic Acids Res. 3, 2303-2308
15. Rigby, P.W.I., Dieckmann, M., Rhoades, C. and Berg, P. (1977), J. Mol. Biol. 113, 237-251

16. Zasloff, M. and Santos, T. (1980) Proc. Natl. Acad. Sci. USA 77, 5668-5672
17. Wood, W.I., Nickol, J. and Felsenfeld, G. (1981) J. Biol. Chem. 256, 1502-1505
18. Maizels, N. (1976) Cell 9, 431-438
19. Sollner-Webb, B. and Reeder, R.H. (1979) Cell 18, 485-499
20. Chandra, T., Bullock, D.W. and Woo, S.L.C. (1981) DNA 1, 19-26
21. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature (London) 263, 211-214
22. Britten, R.J. and Khone, D.E. (1968) Science 161, 529-540
23. Britten, R.J. and Davidson, E.H. (1969) Science 165, 349-357
24. Jelinek, W.R., Toomey, T.P., Leinwand, L., Duncan, C.H., Biro, P.A., choudary, P.V., Weissmann, S.M., Rubin, C.M., Houck, L.M., Deininger,P.L., and Schmid, C.W. (1980) Proc. Natl. Acad. Sci. USA 77, 1398-1402
25. Savouret, J.F., Loosfelt, H., Atger, M. and Milgrom, E. (1980) J. Biol. Chem. 255, 4131-4136
26. Kröger, M. and Kröger-Block, A. (1982) Nucleic Acids Res. 10, 229-236
27. Mulvihill, E.R., Le Pennec, J.P. and Chambon, P. (1982) Cell 24, 621-632
28. Van Ooyen, A., Van den Berg, J., Mantei, N. and Weissmann, C. (1979) Science 206, 337-344
29. Dierks, P., Van Ooyen, A., Mantei, N. and Weissmann, C. (1981) Proc. Natl. Acad. Sci. USA. 78, 1411-1415
30. Cochet, M., Chang, A.C.Y. and Cohen, S.N. (1982) 297, 335-339
31. Schmidt, W., Scherer, G., Danesch, U., Zentgraf, H., Matthias, P., Strange, C.M., Röwekamp, W. and Schütz, G., (1982) EMBO J. 1, 1287-1293
32. Donnehower, L.A., Huang, A.L. and Hager, G.L.(1981) J. Virol. 37. 226-238
33. Fassel, N., Pearson, K., Buetti, E. and Diggelman,H. (1982) EMBO J. 1,3-8
34. Geisse, S., Scheidereit, K., Westphal, H.M., Hynes, N.E., Groner, B. and Beato, M. (1982) EMBO J. 1, 1613-1619
35. Schmid, C.W. and Jelinek, W.R. (1982) Science 216, 1065-1070