The study of the regulation of uteroglobin gene in the rabbit endometrium constitutes a model for analyzing the mechanism of action of progesterone in mammals. The gene has been cloned into λ phage and sequenced. Comparison of the sequence of the gene with the amino acid sequence of preuteroglobin and the three-dimensional structure of uteroglobin established by crystal x-ray diffraction showed that the 3 exons correspond to different functional domains of the protein and that at least one of the splice junctions does not map at the surface of the protein.

S1 mapping allowed us to define the RNA polymerase initiation site. No difference was observed when analyzing premessengers from the endometrium, where the gene is controlled by progesterone and estradiol, and from lung where the gene is constitutively expressed and not controlled by these hormones. In addition, S1 mapping revealed the existence of several minor transcription initiation sites. In the 5’ flanking region between positions -33 and -24 there is the sequence AAATAAAAAA which may correspond to a Goldberg-Hogness box. Two other A- and T-rich sequences were found further upstream from the gene, one of these preceding by about 30 nucleotides a minor start of transcription. No obvious feature, possibly related to steroid regulation, was observed in the nucleotide sequence.

A fragment of the gene containing the “promoter” region (from nucleotide +10 to nucleotide -394) was preferentially retained on nitrocellulose filters after incubation with purified rabbit uterine receptor. A competitive binding assay was used to compare the affinity for the receptor of various DNA fragments. Labeled “promoter” region DNA was incubated with receptor and various concentrations of nonlabeled competing DNA, and the nitrocellulose-bound radioactivity was measured. This method showed the existence of several high affinity binding sites in the 5’ part of the gene and in adjacent regions. However, no high affinity binding sites were observed in the 3’ part of the gene. Also, within the “promoter” region there were at least two high affinity binding sites for the receptor.

Morphological changes in the rabbit endometrium are the basis of several classical tests used to characterize and assay steroids having progestational activity (1, 2). These changes consist mainly in the development of glands and the enhancement of the endometrial secretion. It has been shown that the major protein secreted into the uterine lumen is uteroglobin (3) also called blastokinin (4). Progesterone induction of uteroglobin is mainly due to increased levels of uteroglobin messenger RNA (5, 6). Recently it has been shown that progesterone increases transcription of the gene (7). The cDNA (8) and genomic fragments (9-11) coding for uteroglobin have been cloned. The rabbit progesterone receptor has been purified and polyclonal (12) and monoclonal1 antibodies obtained. It has thus become possible to use this system to study the interaction of a purified mammalian progesterone receptor with a cloned gene which it controls. A prerequisite to such studies is knowledge of the exact structure of the uteroglobin gene and of the adjacent regions. We report here the sequence of the gene (with the exception of the central part of the largest intron). The 3’ and 5’ flanking regions, with a special emphasis on the latter, have also been sequenced. The site of initiation of transcription has been determined by S1 mapping (13) with RNAs prepared from endometrium and lung, i.e. two tissues where the uteroglobin gene is expressed but where it is under completely different hormonal control (14). We also show that the rabbit progesterone receptor recognizes preferentially several regions in the 5’ part of the gene and in the 5’ flanking region. During the completion of these studies the intron-exon composition of the uteroglobin gene was reported by Mennen et al. (11).

**MATERIALS AND METHODS**

**Clones**—The isolation of plasmids containing uteroglobin cDNA and of phages λUG1 and λUG2 containing rabbit genomic fragments encompassing the uteroglobin gene has previously been described (8, 9). DNA from phage λUG1 was digested with EcoRI and the 7 fragments subcloned in plasmid pBR325. The BamHI fragment starting at nucleotide -394 and finishing at nucleotide +10 was subcloned in plasmid pBR322. The experiments were performed in compliance with the French National guidelines for containment.

**DNA Sequence Determinations**—The method of Maxam and Gilbert (15) was used. Restriction fragments were end labeled using T4 polynucleotide kinase (P-L Biochemicals) and [γ-32P]ATP, after dephosphorylation with calf intestine alkaline phosphatase (Boehringer Mannheim). Fragments labeled at one end were generated by digestion with an appropriate restriction enzyme or by separation of strands and 96.3% of the nucleotides were determined from at least two restriction sites. In most cases both strands were analyzed and all restriction sites were sequenced through.

**S1 Mapping**—The protocol of Weaver and Weissmann (13) was used.
used. Total poly(A) RNA (nuclear and cytoplasmic) was prepared (5) from endometria and lung of 5-day pregnant rabbits. A Diel fragment extending from nucleotide -355 to nucleotide +158 was end labeled with T<sub>4</sub> polynucleotide kinase. It was cleaved with EcoRI and the labeled fragment extending from nucleotide +108 to nucleotide -183 was purified by polyacrylamide gel electrophoresis. This fragment was eluted from the gel and aliquots of 50,000 cpm were hybridized with 25 μg of poly(A) RNA during 16 h at 55 °C as described (13). Nuclease S1 digestion (100 units) was performed for 30 min at 45 °C and the nuclease-resistant fragments were separated by electrophoresis on 6% polyacrylamide gels (13).

**Receptor Purification**—It was performed as previously described (12) except that the preparation was chromatographed on a 0.7-ml DNA-cellulose column. After washing the column, the receptor was eluted in 0.8 ml of 5 mM sodium phosphate, 0.5 M NaCl, pH 8.3, buffer. Specific activity of the receptor preparation averaged 5000 pmol of bound hormone/mg of protein.

**Receptor Binding to DNA**—This binding was studied as described for the Escherichia coli CAMP binding protein attachment to specific regions of the lac promoter DNA (16). DNA fragments were labeled at 5' termini by T<sub>4</sub> polynucleotide kinase using [γ-<sup>32</sup>P]ATP (specific activity of 1000–3000 Ci/mmol) (average specific activity of the labeled 404-base pair BamHII fragment, 7 × 10<sup>6</sup> cpm/μg). Incubations contained 2×P[γ<sup>-</sup>]DNA (1000–2000 cpm/50 μl), nuclease-free bovine serum albumin, the radioactivity bound to the filters was measured by nitrocellulose filters (B.A. 85, Schleicher & Schuell, Dassel, Germany). After washing with 100 μl of the same Tris buffer containing no bovine serum albumin, the radioactivity bound to the filters was determined.

**RESULTS**

**Sequence of the Uteroglobin Gene and of the Flanking Regions, Exons and Introns**—Comparison with the cDNA sequence (17) shows that the gene contains 3 exons. The first one is 103 nucleotides long (for start of the gene see S1 (1) are underlined). The acceptor consensus sequence (17) shows that the gene contains 3 exons. The first exon is 103 nucleotides long (for start of the gene see S1 (1)) and start with GT and finish with AG. More extensive consensus sequences have been proposed (19). The donor consensus sequence AG/CT matches in 8 of 9 nucleotides the corresponding sequence in exon 1/intron 1 (CTG/GTGAGT) and in only 6 of 9 nucleotides in exon 2/intron 2 (ACG/GTAACC). The acceptor consensus sequence T<sub>A</sub>C<sub>S</sub><sup>1</sup>N<sup>3</sup>T<sub>A</sub>G/G is closely followed by the corresponding regions of both intron-exon borders. At the 3' end of the gene is found the sequence AATAAA preceding by 19 nucleotides the site of addition of poly(A). The intron-exon composition of the uteroglobin gene observed in these experiments is identical with that reported by Mennen et al. (11).

**Regions Flanking the Gene**—We have sequenced 597 nucleotides upstream from the transcription start since this region may be important in the hormonal control of the gene.

3 A synthetic progesterin (17,21-dimethyl-19-norpregna-4,9-diene-3,20-dione).
The sequence AATACAAAAA which might be a somewhat unusual form of a Goldberg-Hogness box (20) is found between positions -33 and -24. At positions -81 to -85 there is the sequence CAAGT which might be the equivalent of a CAT box (20).

Two other sequences resembling the Goldberg-Hogness consensus sequence are found further upstream from the gene, TAAATAAT (−98 to −91) and AATATTTA (−132 to −125). S1 mapping experiments showed that at least the first of these sequences may be used as a minor transcription start site.

The sequence of the uteroglobin gene has been studied in a recent publication (11). Differences with this sequence are shown in Table I.

S1 Mapping of the Start of Transcription—Defining the nucleotide where transcription starts was especially interesting since steroid receptors may regulate the transcription of genes. It is thus important to precisely localize sequences recognized by the receptors in relation to the starting of transcription. Moreover, the uteroglobin gene is expressed in genes. It is thus important to precisely localize sequences that transcription usually initiates at an A which is preceded by a pyrimidine, actually a C in most cases (21). It is thus probable that transcription starts at the A labeled +1 in the Table I. RNA molecules which are capped at the 5′ end can often protect an additional 1–5 nucleotides of the DNA probe from digestion by S1 nuclease (22). This position of the start of transcription differs by a single nucleotide from the beginning of messenger as determined by cDNA sequencing (17). Some minor larger bands were also observed on the gel. They did not disappear with further S1 digestion (data not shown). It is thus possible that they correspond to minor transcription starts. One of the bands lies at nucleotide −71. An A is found at position −69 which again corresponds to the proposed consensus sequence (21). This start of transcription is preceded by Goldberg-Hogness-type sequence (TAAATAAT) at about 30 nucleotides upstream (positions −98 to −91). A minor smaller band was also observed at position +81; it may correspond to an allelic variation of the messenger sequence. Finally it was observed that the major start of transcription was identical in endometrium and lung.

Progestosterone Receptor Interaction with the Uteroglobin Gene—We hypothesized that the receptor could specifically interact within the 5′ control region adjacent to the uteroglobin gene. Thus the BamHI fragment extending from position +10 to −394 was isolated, 5′ end labeled, and aliquots incubated with various concentrations of receptor. After filtration and washing, the radioactivity bound to nitrocellulose filters was determined. The binding of the receptor to this gene fragment was compared to that of a similar fragment of nonspecific DNA (404 base pairs long pBR322). As shown in Fig. 2 there was a preferential retention of the BamHI fragment when compared to the nonspecific control segment of DNA.

The specificity in terms of receptor binding was studied in two types of experiments. To eliminate a preferential binding of the BamHI fragment independent of the nature of the interacting protein, and experiment was performed in which nonspecific proteins (hepatic cytosol) were used in place of receptor. As shown in Fig. 3 there was no preferential binding to the BamHI fragment. Moreover, it is known that receptor binding to DNA is specifically inhibited by pyridoxal phosphate (23). At very low concentrations (0.4 mM) pyridoxal phosphate almost completely prevented the retention on nitrocellulose filters of "promoter" DNA after incubation with the receptor preparation (Fig. 4).

To analyze the distribution of high affinity binding sites for the progesterone receptor in the region of the uteroglobin gene we used a competitive binding assay. Unlabeled gene fragments were added to receptor and 32P-labeled BamHI fragment of 404 base pairs which contains the promoter. The ability of these fragments to displace bound [32P]DNA was determined. The uteroglobin gene was cut into 3 fragments by EcoRI: fragment a of ~5000 base pairs contained about 3200 base pairs 5′ from the gene, the first exon, and ~1700 base pairs of the first intron; fragment b containing ~490 base pairs of the first intron; and fragment c (1064 base pairs) which contained the 3′ part of the gene, the end of the first intron, all of the second intron as well as the second the third exons, and 236 base pairs downstream from the 3′ terminus of the gene. As may be seen in Fig. 5A only fragment a exhibited a high affinity for the receptor; however, the affinity was somewhat lower than that of the 404-base pair BamHI.
FIG. 2. Progesterone-receptor binding to the 404-base pair BamHI fragment situated in the "promoter" region of the uteroglobin gene (a). The 404-base pair HpaII fragment prepared from pBR322 was used for comparison (b). Both DNA fragments were labeled at their 5' termini. Aliquots (1600 cpm) were incubated with varying concentrations of receptor and filtered through nitrocellulose filters (see under "Materials and Methods"). Bound radioactivity was measured.

FIG. 3. Nonspecific protein binding to the 404-base pair BamHI fragment situated in the "promoter" region of the uteroglobin gene (a). The 404-base pair HpaII fragment prepared from pBR322 was used for comparison (b). Experimental details are as in Fig. 2 except that rabbit liver cytosol was used instead of receptor. Progesterone receptor concentration is very low in liver cytosol (~0.01 pmol/mg of protein).

FIG. 4. Pyridoxal phosphate inhibition of the binding of progesterone receptor to the "promoter" region of the uteroglobin gene. Labeled 404-base pair BamHI fragment was incubated with progesterone receptor (4.7 µg/ml) in 16 mM barbital 1.2 mM EDTA, 50 mM NaCl, pH 8.3, buffer containing nuclease-free bovine serum albumin (50 µg/ml) and 10^{-7} M R5020. Various concentrations of pyridoxal phosphate were added to aliquots. Bound radioactivity was measured.

FIG. 5. Affinity of various parts of the uteroglobin gene for the progesterone receptor. Labeled 404-base pair BamHI fragment (1600–2500 cpm in the various experiments) was incubated with progesterone receptor (2.6 µg/ml) in the presence of various concentrations of nonlabeled competing DNA. Bound radioactivity was measured. The DNA fragments which were used are shown at the bottom of the figure. Also represented is the structure of the uteroglobin gene (B, exons). A, a, b, and c are EcoRI fragments of about 5, 0.49, and 1 kilobases, respectively, encompassing the uteroglobin gene; d is the 404-base pair BamHI fragment (nucleotides −394 to +10). e is BamHI-digested calf thymus DNA. Experimental conditions are as in Fig. 2. B, a, b, and c are fragments of ~2800 base pairs, 404 base pairs, and ~1800 base pairs, respectively, obtained by BamHI digestion of the 5-kilobase EcoRI fragment.

FIG. 6. Presence of several high affinity sites for progesterone receptor in the "promoter" region of the uteroglobin gene. Labeled 404-base pair BamHI "promoter" fragment was incubated with progesterone receptor (8 µg/ml) in presence of various concentrations of unlabeled competing DNA. Bound radioactivity was measured. The DNA fragments which were used are shown at the bottom of the figure (B, the first exon). a, fragment extending from nucleotide −394 to nucleotide −251; b, fragment extending from nucleotide −250 to nucleotide −194; c, fragment extending from nucleotide −193 to nucleotide +10; d, fragment extending from nucleotide −394 to nucleotide +10. "promoter" fragment (d). Nonspecific DNA (e = calf thymus) was a very poor inhibitor. The same experiment was performed with the same result after S1 digestion of DNA fragments (not shown). Thus single stranded regions generated.
Uteroglobin Gene and Progesterone Receptor

by restriction enzyme digestion were not responsible for differences in the affinity for the receptor of various DNA fragments.

To further analyze the distribution of the high affinity sites, the 5000-base pair EcoRI fragment was digested with BamHI yielding 3 DNA fragments. The 404-base pair (+10 to −394) “promoter” fragment had the highest affinity. The −2800-base pair fragment (upstream from position −394) and the −1800-base pair segment (downstream from position +10) also exhibited some affinity for the receptor (Fig. 5B).

We also determined if the 404-base pair BamHI “promoter” fragment contained one or several regions responsible for high affinity binding of receptor. After digestion with AvaI 3 fragments were obtained; one of 144 base pairs (−394 to −251), one of 57 base pairs (−250 to −194), and one of 203 base pairs (−193 to +10). As shown in Fig. 6 the 144- and 203-base pair fragments had both a high affinity for the receptor (comparable to that of the 404-base pair BamHI fragment), whereas the central 57-base pair fragment had a markedly lower affinity. To confirm this result the 404-base pair BamHI fragment was digested with HindII yielding two fragments of 181 base pairs and 223 base pairs, respectively. Both were labeled and incubated with various concentrations of receptor. The two fragments bound the receptor with similar affinities (not shown).

DISCUSSION

The uteroglobin gene contains 2 introns. The first one begins at the end of the region coding for uteroglobin signal peptide. Similar situations have been observed for many proteins (24, 25). The second intron interrupts the coding sequence at nucleotides corresponding to threonine 60. The three-dimensional structure of uteroglobin has been established by x-ray diffraction of crystals at a resolution of 2.2 Å (26). The protein is made of two identical subunits which together encompass the hydrophobic pocket constituting the steroid binding site and are held together by noncovalent bonds and 2 disulfide bridges. The amino acids preceding threonine 60 constitute the major part of the protein and are sufficient to encompass the steroid binding site, while the amino acids following threonine 60 are mainly involved in the tight binding of both subunits. As in many other proteins, the introns separate functional domains of the protein. However, this intron-exon boundary does not follow the rule proposed by Crail et al. (27) where splice junctions map at protein surfaces. Threonine 60 is one of the few hydrophilic residues which are buried inside the uteroglobin molecule (26).

The TATA consensus sequence preceding by about 30 nucleotides the start of transcription is known to be an important feature of the recognition of the eukaryotic genes by RNA polymerase II (20). In the uteroglobin gene the sequence AATACAAAAA is found between positions −33 to −24. This sequence is somewhat unusual since the transversion from A-T to C-G in both eukaryotic and prokaryotic Goldberg-Hogness or Pribnow boxes has been considered as a mutation decreasing the efficiency of promoters (28, 29).

Two other A- and T-rich sequences resembling more the usual type of TATA box are found further upstream from the gene. However, only one of them may be used in the recognition of a minor transcription initiation site. These observations illustrate the difficulties in defining exactly the DNA sequences recognized by RNA polymerase II.

S1 mapping experiments showed that the nucleotides at position +1 are identical in endometrial and pulmonary uteroglobin messengers. Thus the major site of transcription initiation of a unique gene (9), which can be regulated by completely different hormonal mechanisms in two different tissues, is identical in the two tissues. Studies of DNA modification or chromatin structure might shed some light on the differential nature of uteroglobin gene regulation. However, minor sites of initiation have been reported for other genes (30). Their significance remains unclear.

During the composition of this study a paper was published (11) describing a sequence analysis of the uteroglobin gene. The major structural features which are described in this work are analogous to our observations but there are many minor differences (see Table I).

The method which has been used to study progesterone receptor interaction with the uteroglobin gene was similar to that used for the study of the binding of E. coli chloramphenicol (cAMP binding) protein to the lac operon (16). Similar concentrations of protein were necessary to observe filter retention of DNA suggesting a similar affinity for “promoter” DNA. The existence of several high affinity binding sites for receptors resembles the results obtained with other systems (egg white proteins (31,32) and mouse mammary tumor virus (33)). On the other hand Govindan et al. (34) observed a single interaction between glucocorticoid receptors and mouse mammary tumor virus DNA.

Our preparation of receptor was about 50% homogeneous and was thus of similar purity to that used on some of the studies on mouse mammary tumor virus (33). However, we cannot completely eliminate the possibility that the DNA binding we observe was due to some contaminating protein. The inhibitory effect of pyridoxal phosphate (23) is, however, evidence in favor of the DNA binding being due to receptor. Further evidence in favor of this interpretation is the fact that when using receptor preparations of lower specific activity (i.e. less purified) we observed a diminished preferential binding to specific DNA. Since purification yielded steroid-receptor complexes and not free receptor it has been impossible to study the effect of the hormone on receptor binding.

The high affinity DNA binding sites observed in this study may be related to the biological activity of steroid-receptor complexes but further experiments will be necessary to establish this point. However, if this hypothesis is true then the existence of several binding sites for receptor in steroid-regulated genes suggests that receptors act by changing the structural or functional properties of the chromatin on a large portion of the gene.

Acknowledgments—We thank Dr. J. C. Mercier for help in the sequence experiments and N. Malpoint for typing the manuscript.

REFERENCES

1. Clauber, C. (1933) Die Weiblichen Sexual Hormone, pp. 88–100, J. Springer, Berlin
2. McGinty, D. A., Anderson, L. P., and McCullough, N. B. (1939) Endocrinology 24, 829–832
3. Beier, H. M. (1969) Biochem. Biophys. Acta 160, 289–291
4. Krishnan, R. S., and Daniel J. C., Jr. (1967) Science 158, 490–492
5. Atger, M., and Milgrim, E. (1977) J. Biol. Chem. 252, 5412–5418
6. Loosef, H., Fridmanovsky, F., Savouret, J.-F., Atger, M., and Milgrim, E. (1981) J. Biol. Chem. 256, 3465–3470
7. Kumar, N. M., Chandra, T., Woo, S. L. C., and Bullock, D. W. (1982) Endocrinology 111, 1115–1120
8. Atger, M., Perricaudet, M., Tiollais, P., and Milgrim, E. (1980) Biochem. Biophys. Res. Commun. 93, 1082–1088
9. Atger, M., Atger, P., Tiollais, P., and Milgrim, E. (1981) J. Biol. Chem. 256, 5970–5972
10. Snead, R., Day, L., Chandra, T., Mace, M., Jr., Bullock, D. W., and Woo, S. L. C. (1981) J. Biol. Chem. 256, 11911–11916
11. Menne, C., Suske, G., Arnesmann, J., Wenz, M., Cato, A. C. B., and Beato, M. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 4853–4857
Uteroglobin Gene and Progesterone Receptor

12. Logeat, F., Vu Hai, M. T., and Milgrom, E. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78**, 1426–1430
13. Weaver, R. F., and Weissmann, C. (1979) *Nucleic Acids Res.* **7**, 1175–1193
14. Savouret, J.-F., Loosef, H., Atger, M., and Milgrom, E. (1980) *J. Biol. Chem.* **255**, 4131–4136
15. Maxam, A., and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 560–564
16. Majors, J. (1975) *Nature (Lond.)* **256**, 672–674
17. Chandra, T., Bullock, D. W., and Woo, S. L. C. (1981) *DNA* **1**, 19–26
18. Breathnach, R., Benoist, C., O’Hare, K., Gannon, F., and Chambon, P. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75**, 4853–4857
19. Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472
20. Benoist, C., and Chambon, P. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77**, 3865–3869
21. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Keding, L., and Chambon, P. (1980) *Science* **209**, 1406–1411
22. Lai, E. C., Roop, D. R., Twai, M. J., Woo, S. L. C., and O’Malley, B. W. (1982) *Nucleic Acids Res.* **10**, 5553–5557
23. DiSorbo, D. M., and Litwack, G. (1982) in *Biochemical Actions of Hormones* (Litwack, G., ed) Vol. IX, pp. 205–219, Academic Press, New York
24. Gilbert, W. (1978) *Nature (Lond.)* **271**, 501
25. Balke, C. C. (1978) *Nature (Lond.)* **273**, 267
26. Mormon, J. P., Fridlansky, F., Bally, R., and Milgrom, E. (1980) *J. Mol. Biol.* **137**, 415–429
27. Caik, C. S., Sprang, S., Fletterick, R., and Rutter, W. J. (1982) *Nature (Lond.)* **299**, 180–182
28. Poncz, M., Ballantine, M., Solowieczyk, D., Barak, L, Schwartz, E., and Surrey, S. (1982) *J. Biol. Chem.* **257**, 5994–5996
29. Siebenlist, U., Simpson, R. B., and Gilbert, W. (1980) *Cell* **20**, 269–281
30. Grez, M., Land, H., Giesecke, K., and Schütz, G. (1981) *Cell* **25**, 743–752
31. Mulvihill, E. R., LePennec, J. P., and Chambon, P. (1982) *Cell* **24**, 621–632
32. Compton, J. G., Schrader, W. T., and O’Malley, B. W. (1982) *Biochem. Biophys. Res. Commun.* **105**, 96–104
33. Payvar, F., Wrang, O., Carlstedt-Duke, J., Ökret, S., Gustafsson, J.-Å., and Yamamoto, K. R. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78**, 6628–6632
34. Govindan, M. V., Spiess, E., and Majors, J. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **78**, 5157–5161
The rabbit uteroglobin gene. Structure and interaction with the progesterone receptor.
A Bailly, M Atger, P Atger, M A Cerbon, M Alizon, M T Vu Haï, F Logeat and E Milgrom

J. Biol. Chem. 1983, 258:10384-10389.

Access the most updated version of this article at http://www.jbc.org/content/258/17/10384

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/258/17/10384.full.html#ref-list-1