Volume 12 Number 22 1984 Nucleic Acids Research

Sequence homologies in the region preceding the transcription initiation site of the liver estrogen-responsive vitellogenin and apo-VLDLII genes

Philippe Walker, Jacques-Edouard Germond, Marianne Brown-Luedi, Françoise Givel and Walter Wahli

Institut de Biologie animale, Université de Lausanne, Bâtiment de Biologie, CH-1015 Lausanne, Switzerland

Received 6 September 1984; Revised and Accepted 24 October 1984

ABSTRACT

In the liver of oviparous vertebrates vitellogenin gene expression is controlled by estrogen. The nucleotide sequence of the 5' flanking region of the Xenopus laevis vitellogenin genes A1, A2, B1 and B2 has been determined. These sequences have been compared to each other and to the equivalent region of the chicken vitellogenin II and apo-VLDLII genes which are also expressed in the liver in response to estrogen. The homology between the 5' flanking region of the Xenopus genes B1 and B2 is higher than between the corresponding regions of the other closely related genes A1 and A2. Four short blocks of sequence homology which are present at equivalent positions in the vitellogenin genes of both Xenopus laevis and chicken are characterized. A short sequence with two-fold rotational symmetry (GGTCANNNTGACC) was found at similar positions upstream of the five vitellogenin genes and is also present in two copies close to the 5' end of the chicken apo-VLDLII gene. The possible functional significance of this sequence, common to liver estrogen-responsive genes, is discussed.

INTRODUCTION

Hormonal regulation of the vitellogenin genes, which encode yolk proteins in the liver of oviparous vertebrates, is controlled to an important extent at the level of transcription (1, 2). Results obtained with other steroid-regulated genes such as those coding for ovalbumin, metallothionein, MMTV and lysozyme (3 - 8) demonstrate that DNA sequences close to the initiation site of transcription are involved in gene regulation. It is thought that these sequences function by interacting directly or indirectly with the hormone receptor complex. In vitro binding experiments involving these DNAs, as well as chicken vitellogenin DNA, have indeed revealed sites of preferential binding of steroid receptors (4, 8, 9 - 15). Consensus nucleotide sequences for the sites which interact with glucocorticoid and progesterone receptors have been deduced by comparing the binding elements (4, 8, 9,

10). DNA fragments containing these binding sites have been shown to control gene expression in DNA transfer experiments (3-8). Unfortunately, no satisfactory expression system controlled by estrogens is yet available. However, comparison of the 5' end region of estrogen controlled genes might reveal the presence of sequence elements involved in the expression of these genes. In this paper, we therefore analyze the 5' end region of the four Xenopus laevis vitellogenin genes (1, 16, 17) and compare it to the equivalent region of the chicken vitellogenin II gene (18-20) and of the chicken very low density apolipoprotein II (apo-VLDLII) gene which is also controlled by estrogen in the liver (21).

MATERIALS AND METHODS

DNA sequence analysis

The DNA sequencing strategy used is described by Germond et al. in the accompanying paper. The nucleotide sequences were analyzed with computer programs prepared and kindly provided by Carolyn Tolstoshev (Laboratoire de Génétique Moléculaire des Eucaryotes, Strasbourg).

Preparation of Xenopus nuclear RNA and nuclease S1 mapping

Nuclear RNA was prepared as described from the liver of animals which were treated with 1 mg 17 β-estradiol for 3 days (17). Nuclease S1 mapping was also described previously (22).

RESULTS

Nucleotide sequences of the 5' flanking region of the Xenopus laevis vitellogenin genes

Figure 1 indicates the portions of the four Xenopus laevis vitellogenin genes and of the chicken vitellogenin gene which have been compared, i.e. the region upstream of the transcription initiation site. Figure 2 gives the nucleotide sequence of the A1, B1 and B2 DNAs determined using the dideoxy- and associated methods (23-26). Sequence data for the Xenopus gene A2 are taken partly from Walker et al. (22) and was extended further upstream. These four sequences span the region from positions -1020 for the genes A1 and A2, -938 and -770 for the genes B1 and B2, respectively, to the translation initiation codon. The chicken sequences used in this study are

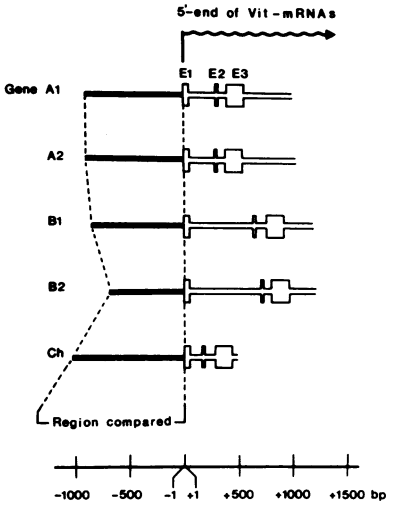

Figure 1 Structural organization of the 5' end region of five vitellogenin genes. The diagram indicates the region of the four Xenopus genes A1, A2, B1, B2, and of the chicken gene (ch) which were either sequenced in this work (A1, part of A2, B1 and B2) or for which sequence data were available (chicken gene and part of A2 gene, see ref. 21, 26, 27). Boxes E1, E2 and E3 represent exons. The regions compared in this paper are indicated by the heavy lines between the dashes.

taken from recently published work (22, 27, 28), and are therefore not listed in Figure 2. The nucleotide sequences downstream of the transcription initiation site of these five vitellogenin genes have been compared in the accompanying paper by Germond et al.

Transcription initiation sites of the Xenopus genes A1, B1 and B2

The position of the putative 5' end of the vitellogenin genes A1, B1 and B2 has been determined earlier by electron microscopy (R loop analysis) and by in vitro transcription (16, 17, 29). We now characterize the transcription initiation site of these genes by nuclease S1 mapping. The DNA fragments used as probes were 5' end-labelled at restriction sites within the first intron of the genes. These regions show enough sequence divergence between the two A as well as between the two B genes to render unspecific hybrids (A2 RNA x A1 probe, B2 RNA x B1 probe or B1 RNA x B2 probe) sensitive to S1 nuclease. The probes were hybridized either to liver nuclear RNA from an estradiol treated animal or to liver nuclear RNA from an unstimulated male not producing vitellogenin. The samples were then treated with nuclease S1 and the protected DNA fragments analyzed by gel electrophoresis (Fig.3). Protected fragments were only found with nuclear RNA from hepatocytes stimulated by estrogen. Their length indicate that the transcription initiation site of the Xenopus genes A1, B1 and B2 occupies the homologous position to the one defined earlier for the Xenopus gene A2

gene A1

```
-1020      -1010      -1000      -990       -980       -970
TCGTACACTT GGGGTGCACC CTCAAGTGAT TGTAACCCCA GGCTGGTGCA GTTTTCTGCT
-960       -950       -940       -930       -920       -910
GATGGGAGCA CCGGCCCAGG GTTTCAGGTA AGTCAGTACA ATCACTTGGG GGTGACTAAC
-900       -890       -880       -870       -860       -850
ATTTGGCACT GGCCCGGGGT TATACCAGTG AGCACTGCTA TTTAAGAAAG CTATTTAGTT
-840       -830       -820       -810       -800       -790
AAACTGAGTG TACAACGCCG TTCCTTTTCC TTTGGTGTGG AGATTCAGAT TACAGAAAGA
-780       -770       -760       -750       -740       -730
CCCCTTATCC AGAAAGCCCC AAGTCCCAAA CATTCTGGAT AACAGTTCCC ATTCCTGTCT
-720       -710       -700       -690       -680       -670
ATCATAAATG TCAGCAAACA TCAGAAGCAC ATTCTCTTCC ATGCAGTGAT AGTACTGAGT
-660       -650       -640       -630       -620       -610
GCCACACTGA GCACTGTCAT TTATCTGAAC CCTTTGTTTG CAATGCATAT AACATGTGGG
-600       -590       -580       -570       -560       -550
AAGAAAGGAT TCTGACTGCA GTGGAATGAA TCATAAATTA CTCTCATTAT TGCTATATAC
-540       -530       -520       -510       -500       -490
AATTATACAA GGAAATGATAA ATTAGCCCCT AAATATTGGG TGCACCATAT CTTGGGTATT
-480       -470       -460       -450       -440       -430
CTATCTAGTA CCACCAGAGG GATATCAATG CAAAAGGGTA TCAAAACAGA TTTATAACTT
-420       -410       -400       -390       -380       -370
AATAAAAAGT TATTTATTGT GTGCTCTTTG ATTTAAAAGT TACACGCAGT CACTTTGACT
-360       -350       -340       -330       -320       -310
TGATCTGCTT ACCAAGTTTG GCAAGGGTCA CAATGACCTC AACAAGGTTG ATTTAAATTG
-300       -290       -280       -270       -260       -250
CAACCTGGCC AAATTGATTA GAAAATAAAC TACAAAATAA AGCCAAAAAA AGTACAACTA
-240       -230       -220       -210       -200       -190
TAATATTGTC AGTAATGTTC ATATTTGCAT AAGTCTAAAC ATATATATTG CTAGCACATA
-180       -170       -160       -150       -140       -130
CCTGTCCATA CAATAAAAAT AATTTACAAA TAATAATGGT GATTTTTTAC CTGCATGTAA
-120       -110       -100       -90        -80        -70
CCTTAATGTG GTGATGTAAC TATTTGCATA ACCTGATAAC AATTTCTGAA TGAAGCATCC
-60        -50        -40        -30        -20        -10        -1
AAAAAGTCAC GTGTTACAAA TTTTCCTGTA ATAAATATGG CAAGTTGTGC ATGGTTCAGT
+1
ATTCGCCATC ACCATG
```

gene A2

```
-1020      -1010      -1000      -990       -980       -970
AGTTTTCTGG ATAACGGATT TCCGGATATT GGATCCCTAC CTGTATTTCC CCCTGTGATT
-960       -950       -940       -930       -920       -910
TTGTGGGTTT TCTCCAGGAA GTCTGGTTTC CCTGACACAC TCCAAAAGTA TACAGACAGG
-900       -890       -880       -870       -860       -850
TGACATGGTT CCTTATAAGA ATGACAAAAA CGTGTGTGTG AATGTGATAA GGAGCTTAGA
-840       -830       -820       -810       -800       -790
TTGTAAGCTT CGCCGAGACA AGAACTGATG TAAATGATAT ATAATATCTG TAAAGCACTG
-780       -770       -760       -750       -740       -730
CAGAATGTGT TAGGCCTGTA ATAAAGATTG ATACTAAACA AGATGCTTTG CTCAGTATTC
-720       -710       -700       -690       -680       -670
TTAACAACTG TGTCTGTCAC ATTCAAAAGT ATTAGTACCC CAACCCATTT TTATATTATA
-660       -650       -640       -630       -620       -610
TACCTACTGC CCAGAGTGCC CTCTGAACCT GAAAGTCTCT TACTATGTAG GCATGTTCTA
-600       -590       -580       -570       -560       -550
TCTGGGAAAA TAAATAGGGT TATCTTTTTT ATTTGAACAT TTTATATAGC TTTATTATAG
-540       -530       -520       -510       -500       -490
CTATTATTTA TATCTTAAAG ATTAAAAAAG CTGTCAATTC TCTTATCTTA TGTATTTTAA
-480       -470       -460       -450       -440       -430
ATTCATTTTA GAGAATGAAA ATAAGTGTAT TCATGTTGAA ATAAAAATCA AACTTTTTTT
-420       -410       -400       -390       -380       -370
GGACATCAAA ATGTGTTTTT TACTGGGACT TTTGTTCATA AAATAGTTTT CTGCATAGCC
-360       -350       -340       -330       -320       -310
TCTTTGATTA GATTAACTGT CCAAAGTCAG GTCACAGTGA CCTTGATCAAA GTTAAATGTAA
-300       -290       -280       -270       -260       -250
CCTCAACCTG GACAAAGCTA CTAGTTTTGA AATCAGAAAT TAAATTAATT AATAATAATT
-240       -230       -220       -210       -200       -190
TTATTAGACA TATTACATTT TTTTTTACAT TTTATTGGTT AAAATTCATT GTCATATAAA
-180       -170       -160       -150       -140       -130
TAAAGCAAAA ATTCACATCA AATATAAATC TGTTTATGCT GTTTGATGCA GAATAGATAT
-120       -110       -100       -90        -80        -70
GGTGAGGTAA TTGTTTACAC AACCTGATAA CAGTACTGGA ATGAGCCAAA ATCACACATC
-60        -50        -40        -30        -20        -10        -1
TAAAGATCAC GTGTTACAGA TTTTCCTGCA ATAAATATGG CAGGCTTTTC TGGGTTCAGT
+1
GTTCACCATC ACCATG
```

gene B1

```
                                -930       -920       -910
                     GATCAGCA   CAGCTCCCAC TGACTTCTAT GGGACCTTGA
-900       -890      -880       -870       -860       -850
TTGCTATTAT GTGGAGCAGT TTTGTGGTTT TATCACTTAA TAAATCTCTA CATTTTTAGA
-840       -830       -820       -810       -800       -790
GTGTTGTAGA ATTTTTTTTT ATAAAAAAAA AAGGATTTGT AAAAAAAAAG GAAATCCAAA
-780       -770       -760       -750       -740       -730
TGTTAGTAAA TAGGTTCCTA ACAGAGGTGT GAACAGCCAT TATGATTTTG AAATCATCTG
-720       -710       -700       -690       -680       -670
ATGTAAGTTA AAGTGCTGTT GCTTTTGTTC ACAGTATCTT TCAGCCATGT ATTAACTGTT
-660       -650       -640       -630       -620       -610
GCAGTAATAA ACAGCAGTGG GCCTTGTTCT GTTACTGTTT TCTTAAATCC TATTACGATT
-600       -590       -580       -570       -560       -550
AGAGAGTCTG ATATTTATTC AGAAAGATCAA ATTCATCAGC TGGAGGGCCA CAGTGAACAT
-540       -530       -520       -510       -500       -490
GCAAATAATA TAGCAGATAT GCTCCACCAT ATCAAAAGAA TATATAATAA ATATAGGGGT
-480       -470       -460       -450       -440       -430
GCAAGTTCCA TTAGGCCATA ATGCTCTTAG GCAGTAGACA ATGAATTAGT TAAACTAAAC
-420       -410       -400       -390       -380       -370
TTGGACCGAT CATCATGGAG AAGACCCAAA ATGCTGACTG GCACGTATCT AGTATTCTTA
-360       -350       -340       -330       -320       -310
TCGATCAGAT TAGAGCTGAT GCCTCCAGTC ACTGTGACCC AACCCAAGTT ATCATGACCT
-300       -290       -280       -270       -260       -250
CTTAGTTGGC TATGGTCTAT GGGTAATTAT TAACCATTAT TTAATTCAAA GAAAAATATA
-240       -230       -220       -210       -200       -190
TAATATTAAG CATTAGTTTA AAATAAAATG TTATATTATT TGATAAAGTG ACCAAGGGTA
-180       -170       -160       -150       -140       -130
TATGTTTTGT TGTTTAATTT TGATTTTTTT TTTCTGCAAC AGCTATGAGC ATTGTGCAAA
-120       -110       -100       -90        -80        -70
CATATTTGGG TGGATGCCTA TAATTGGGTA AAGGTTTATT ATTTACCAGA CCCAACCACT
-60        -50        -40        -30        -20        -10        -1
TCTGTGTGCA CATGCGCCAG ATCTTTCTGC TATAAATACA CTACAGTATC TCAAACTGAC
+1
ATTCGCCATC ACCATG
```

gene B2

```
           -770       -760       -750       -740       -730
           AATTCAAAAT ATCATTCTGA TGTGTTTCAG GAATAACCTA CACCTAGGGG
-720       -710       -700       -690       -680       -670
CACATTCACG TTTCGTTTCG AAGGATTTAA TTGTTCAATC GAATGATTTT TCCATCGATC
-660       -650       -640       -630       -620       -610
GCTCGATCAA ACAAATTGCA GTAAATCCTT AGCTTCGAT ATTCAAAGTC GAAGGATTTA
-600       -590       -580       -570       -560       -550
ACTTTGATGG TTGAATATCG AGGGTTAATT AACCCTCGAT ATTCGACCCA TAGTAAATGT
-540       -530       -520       -510       -500       -490
GCCCCCTAAT GTAGCAGATT TTCTTCTCTA CATAATGAAT ATATAATGAA TTTAGGGGTG
-480       -470       -460       -450       -440       -430
CAAGTTTCAT TACGACATAA TACTCCCAGG CAGTAGACAA TGAAGTAGTT AAATGAAACT
-420       -410       -400       -390       -380       -370
TGCACCAGTC ATCATGGAGA AGACCCAAAC TGCTAACTGG CAAGTATCTG GTATTCTTAT
-360       -350       -340       -330       -320       -310
CAATCAGATT AGAGCTGATG CCTAAAGTCA CTTTGACCCA ACCCAAGTTA TCATGACCTC
-300       -290       -280       -270       -260       -250
TTAGTTGGCT ATGGTCTACA GGTAATTATT AACCATTATT TAATTCAAAG AAAAATATGT
-240       -230       -220       -210       -200       -190
AGTGTTAAGT AATTATTTAA AATAACGTTT TATATTATTT GATAAAGTGA CGAAGGGTAC
-180       -170       -160       -150       -140       -130
ATTTTTTGTT GTTTAATTAT ATATATTTTT TGCAACAGGT ATCAACCAAT AGCATTGCAA
-120       -110       -100       -90        -80        -70
GCATATTTGG GTGGTTAATT TGGGCTGGGT AAAGATTAAT TATGTACCAG ACAAGTCACT
-60        -50        -40        -30        -20        -10        -1
TCTGTGTGCA CATGCGCCAG ATCTTTCCGC TATAAATACG CTAACGTTTC TCAAACTGAC
+1
ATTCACCATC ACGATG
```

(Fig. 2 and ref.22) and for the chicken gene (22, 27, 28).

Sequence homologies within the 5' flanking region of the Xenopus and chicken vitellogenin genes

The 5' flanking regions of the Xenopus and chicken vitellogenin genes were compared in order to identify conserved sequence elements that might participate in the hormonal regulation of these genes. Figure 4 shows the comparison of the closely related Xenopus gene pairs A1-A2 and B1-B2. The A1 x A2 gene comparison reveals two blocks with 73% sequence homology. The first directly flanks the 5' of the genes from position -1 to about -130 and the second from position -260 to about -340 . Analysis of the B1-B2 gene pair reveals a strong continuous homology of 88% from position -1 to about -535 where homology abruptly stops. Thus the 5' flanking regions in the B genes are more homologous to each other than they are in the A genes.

The comparative analysis by Walker et al. (22) between the Xenopus gene A2 and the chicken vitellogenin II gene revealed several blocks of sequence homology within the 500 nucleotides preceding the capping site of these genes. We have extended the comparison to the other Xenopus vitellogenin genes. Four sequence blocks were found for which homology appears in all analyzed vitellogenin genes (Blocks -1, -2, -3, -4; Fig. 5 ; Table 1).

Block-1 covers the 30 nucleotides upstream of the TATA box which is preceded by a stretch of seven pyrimidines and three purines (Fig. 5, Table 1). Between the TATA box and the transcription initiation site the sequences of the five genes are less related.

In Block-2, the chicken gene is closer to the Xenopus A genes than to the Xenopus B genes. Homology between chicken and Xenopus A sequences is even better than between Xenopus A and B genes (Fig. 5; Table 1). Thus homology between the chicken gene and the Xenopus B genes is rather low in this block which is relatively A+T rich (63% in A2). In Block-3, another A+T rich region (82% in A2), homology is also best between the Xenopus A genes and the chicken gene (Fig. 5 and Table 1).

Figure 2 Nucleotide sequences of the 5' flanking region of the four Xenopus vitellogenin genes. The display starts in 5' at position -1020 for the genes A1 and A2, -938 and -770 for the genes B1 and B2, respectively. All listings end at the ATG translation initiation codon found 13 nucleotides downstream of the transcription initiation site (+1).

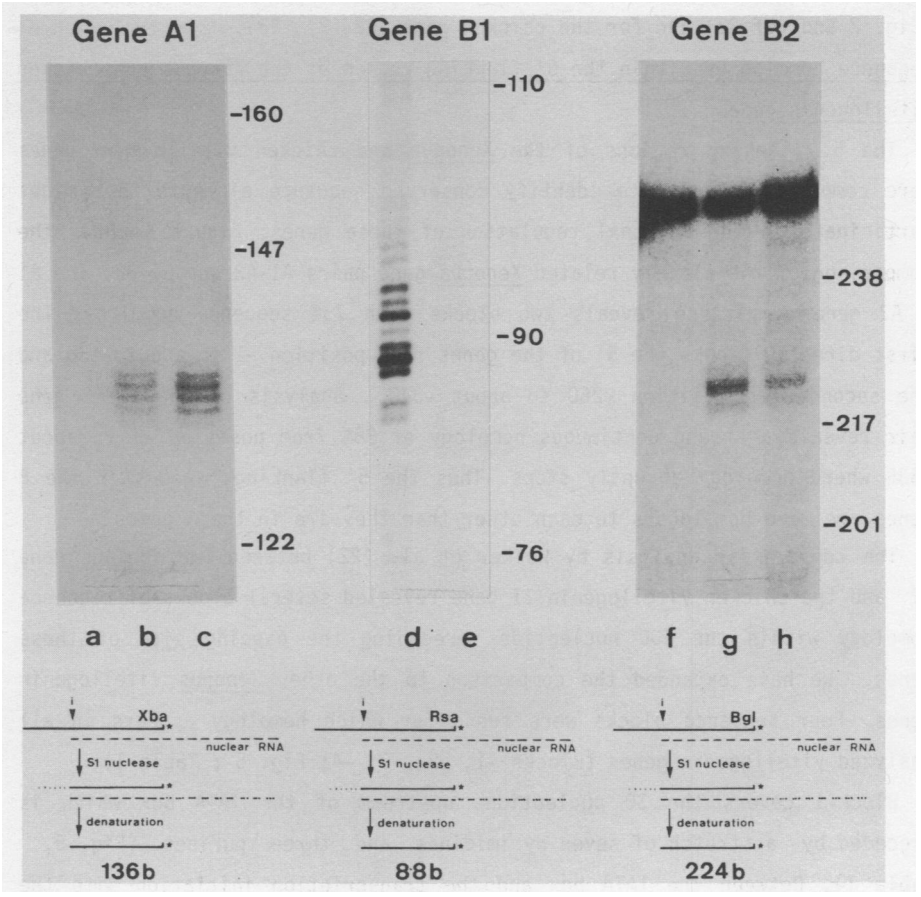

Figure 3 Nuclease S1 mapping of the transcription initiation site of the Xenopus vitellogenin genes A1, B1 and B2. The probe A1 was labelled at a Xba I site located 136 bp downstream of the putative transcription inititiation site, the B1 probe at a Rsa I site 88 bp downstream of the initiation site and the B2 probe at a Bgl II site 224 bp downstream of this site.The Xba I probe (gene A1) was hybridized to 6 µg of nuclear RNA from the liver of a control male (lane a), to 2 µg or 4 µg of liver nuclear RNA from an estrogen treated female (lanes b and c, respectively). The Rsa I probe (gene B1) was hybridized to 4 µg of liver nuclear RNA from an estrogen treated female (lane d) or to 6 µg of liver nuclear RNA from a control male (lane e). We cannot yet decide if the minor bands in lane d represent additional weak starts or alternatively are the result of S1 digestion artefacts. The Bgl II probe (gene B2) was hybridized to 6 µg of liver nuclear RNA from a control male (lane f), to 4 µg and 2 µg of liver nuclear RNA from a female treated with estrogen (lanes g and h, respectively). The strong band of 260 bp in lanes f, g, h represents renatured probe. The positions of length markers are indicated and given in bases. The schemes give the design of the experiment and the expected results.

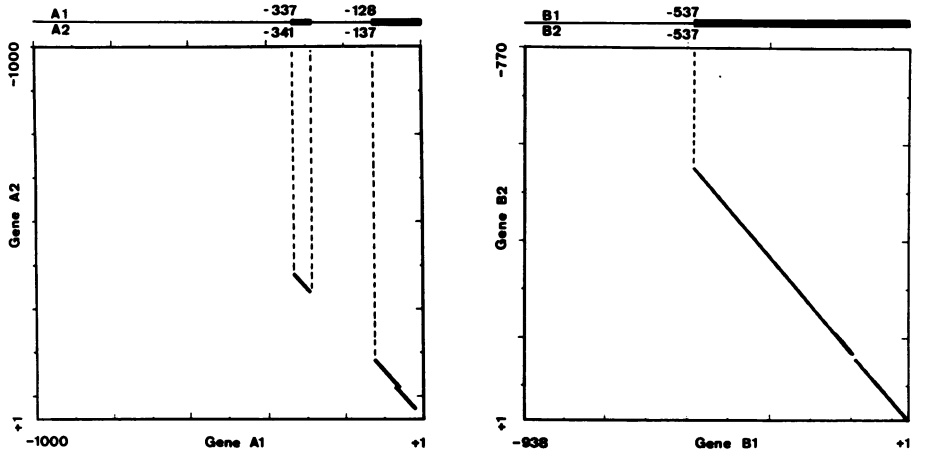

Figure 4 Comparison between the 5' flanking regions of the Xenopus genes A1 and A2 or of the genes B1 and B2. To prepare the diagram, a two dimensional matrix program was used with the following criteria : minimal block length of 40 nucleotides with more than 65% homology. A scheme of the detected regions of homology with their 5' position is drawn above the diagram.

Comparative analysis of the Xenopus gene A2 and of the chicken gene revealed a block of sequence homology with a palindromic structure around position -324 and -343 in the A2 and the chicken gene, respectively (22). A homologous block is found at the equivalent position of the other Xenopus vitellogenin genes (Fig. 5). The core of this block, located at a remarkably conserved position, is the region showing the highest level of homology in all five genes (Table 1). Closer analysis of this region revealed the short palindromic consensus sequence 5' -GGTCANNNTGACC- 3'. Direct or inverted repetition of subsets of this consensus gives a palindromic structure to the entire Block-4.

Distribution of the -GGTCANNNTGACC- element within the 5' flanking region of genes regulated by estrogen in the liver

As mentioned above, the analysis of the palindromic Block-4 revealed a well conserved core represented by a short symmetrical nucleotide sequence, -GGTCANNNTGACC-. The presence of this element was tested for in all the available Xenopus and chicken vitellogenin sequences (11.6 kb), chicken apo-VLDLII sequences (-475 to 265; ref.21) chicken albumin DNA (-266 to 305; ref.21) and the chicken egg white protein gene sequences, namely ovalbumin

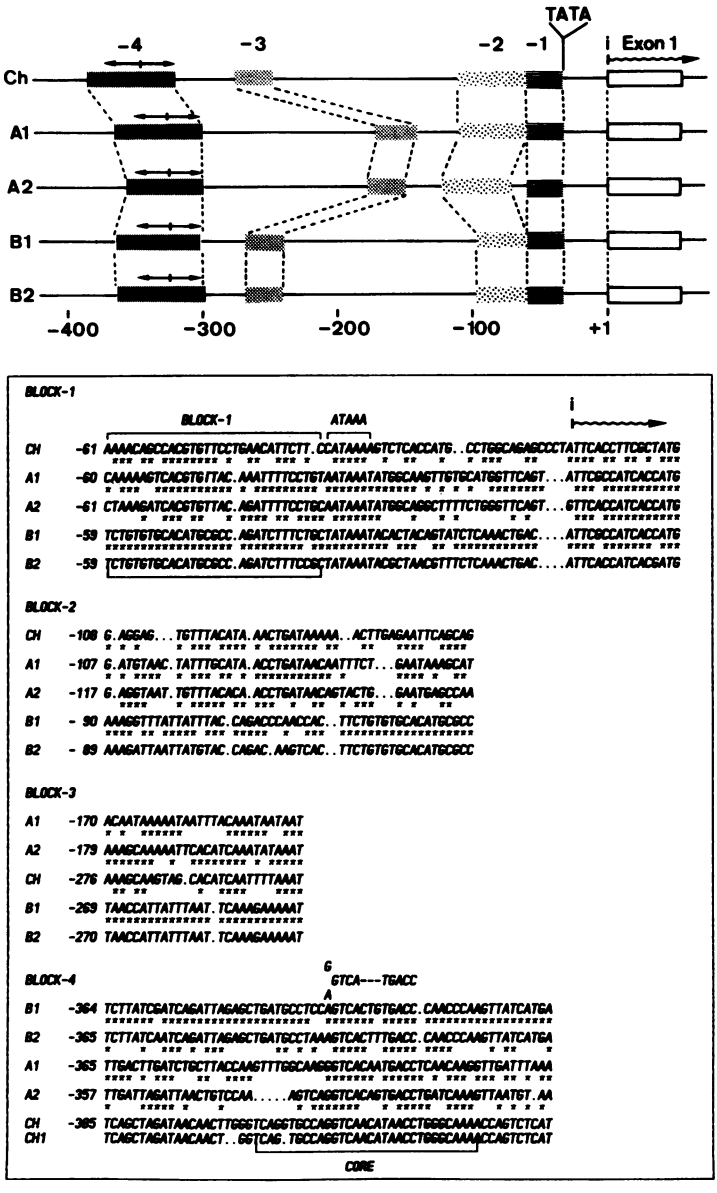

Figure 5 Blocks of sequence homology common to the 5' flanking region of the Xenopus and chicken vitellogenin genes. The scheme indicates the size and position of the homology Blocks -1, -2, -3 and -4. The palindromic sequence in Block-4 is represented by arrows. The positions of the TATA box and the transcription initiation site (i) are indicated. The nucleotide sequence within the four blocks of homology is given with reference to the

different genes and to the position of the 5' nucleotide of the sequences shown. The consensus sequence of a highly conserved element of 13 nucleotides is presented on top of Block-4. The core region indicated below Block-4 contains the sequences whose percentages of homology are given in Table 1 (Block-4, values in parentheses). Ch and Ch 1 refer to the sequences of the same chicken vitellogenin II gene taken from ref. 26 and 21, respectively.

(-300 to 7564; ref.30, 31), ovalbumin-like X and Y (-1613 to 624 and 1 to 6641; ref.32, 33), lysozyme (-440 to 280; ref.34), ovomucoid (-437 to 133; ref.35), and conalbumin (-266 to 200; ref.36). This analysis is limited by the available sequence data and consequently does not strictly include equivalent portions of the genes tested. With one exception (X gene sequences; position -305) the element was only found in the 5' flanking DNA of the vitellogenin genes and of the apo-VLDLII gene, also estrogen-responsive in the liver. Figure 6A gives the positions at which the element appears, up to three times in each vitellogenin DNAs and twice in apo-VLDLII DNA. The nucleotide sequence of each of these elements, the consensus derived from them with the nucleotide frequencies in percent at each position is given in Figure 6B. The element, which itself is G+C rich, is generally embedded in A+T rich regions.

Distribution of putative steroid receptor binding sites within the vitellogenin sequences

Recently, consensus sequences derived from elements present in all glucocorticoid or progesterone receptor binding sites have been proposed (8-10). Although estrogen is the only class of steroid hormones known to regulate the expression of vitellogenin genes (37), we screened the DNA flanking them for the presence of these consensus sequences. The hexanucleotide TGTTCT that is present in all glucocorticoid binding sites was found in the 5' flanking sequence of the gene A2 at position -607, of the gene B1 at position -637 and of the chicken vitellogenin gene at positions -888 and -1065 or at positions -604 and -910 in the inverted orientation. However, sequences common to the progesterone receptor binding sites were not found.

It has been suggested that a short region of homology at an equivalent position in chicken lysozyme, ovalbumin, conalbumin, ovomucoid and apo-VLDLII DNA might be involved in the estrogen-regulated expression of

Table 1

Percent homology between the best conserved regions (Blocks -1, -2, -3 and -4) within the 5' flanking sequences of Xenopus and chicken vitellogenin genes.

	Block -1	Block -2	Block -3	Block -4
A1 x A2	83	70	61	68 (80)*
x B1	53	54	68	60 (73)
x B2	47	48	68	56 (70)
x Ch	67	67	46	45 (56)
A2 x B1	63	56	64	50 (73)
x B2	60	54	64	50 (67)
x Ch	57	69	78	56 (67)
B1 x B2	97	87	100	93 (93)
x Ch	53	46	53	59 (63)
B2 x Ch	50	41	53	51 (72)

* Values in parentheses represent the percent homology between the different genes in the core region of 30 nucleotides in Block-4 (see Fig. 6).

these genes (8). Owing to the lack of an estrogen-responsive experimental system no functional analysis has yet been performed with this element (8). Although some similarities can be found with the vitellogenin genes in the corresponding region we nevertheless judge them to be too weak to join the consensus.

Chicken oviduct estrogen receptor was shown to bind preferentially to a region flanking the chicken vitellogenin gene and containing the estrogen-dependent hypomethylated site (15). Based on these results, combined with those obtained from DNAse I protection experiments, the sequence -GCGTGACCGGAGCTGAAAGAACAC- has been proposed as a possible estrogen receptor binding site (15). Interestingly, the eight 5' proximal nucleotides of this sequence are part of the -625 copy of the symmetrical element -GGTCANNNTGACC- described above (Fig. 6). For the rest of the proposed

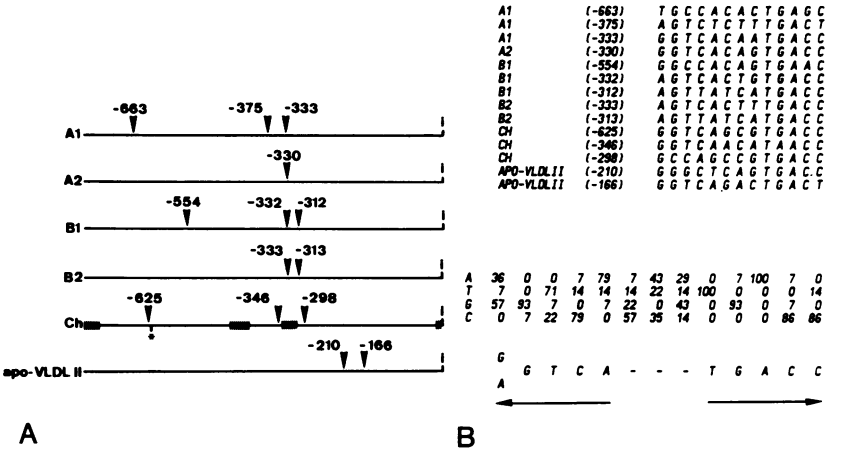

Figure 6 Distribution and characteristics of the GGTCANNNTGACC element. (A) Map of the positions (arrow heads) the element occupies in the 5' flanking region of six estrogen-responsive genes in the liver i.e. four Xenopus vitellogenin genes (A1, A2, B1, B2), the chicken vitellogenin II gene, and the chicken apo-VLDLII gene. In the scheme representing the region flanking the chicken vitellogenin gene the positions of nuclease hypersensitive sites (IIIIIIII , ref. 26) and of the estrogen-dependent hypomethylated site (*, ref. 20, 39, 40) are indicated. (B) Nucleotide sequence of each copy of the element with its position in brackets. Also shown are the frequency of each of the four nucleotides at each position of the element and the consensus sequence derived from this data.

binding sequence no homology was found either in the Xenopus vitellogenin genes or in the chicken apo-VLDLII gene.

DISCUSSION

All five vitellogenin genes analyzed, whose expression is regulated by estrogen, have a homologous 5' end organization (see also accompanying paper), and the initiation of transcription takes place at corresponding positions. Comparative analysis of their 5' flanking region reveals and highlights several structural features deserving special attention.

Different degrees of homology between the flanking sequences of the Xenopus closely related vitellogenin gene pairs

Our analysis shows that among Xenopus vitellogenin genes, the region upstream of the site of transcription initiation is better conserved in the

B1-B2 gene pair than in the A1-A2 pair. Similar observations have been made in the accompanying paper for intron sequences. This contrasts with the quite similar sequence divergence found between A1-A2 and B1-B2 homologous exons. According to the current hypothesis, A1-A2 and B1-B2 gene pairs arose simultaneously as the result of a whole genome duplication in Xenopus laevis about 30 Million years ago (16,38). From our data it is clear that the 5' flanking region and the introns of the A genes have diverged more rapidly than their B gene equivalents. There is no obvious explanation for this observation. Though it is known that the A1 and B1 genes are closely linked, information as to the organization of the whole gene family is still limited. Its future elucidation should aid interpretation of our results and possibly challenge our current hypothesis concerning the evolution of the vitellogenin multigene family in Xenopus laevis (16).

The B1-B2 flanking regions are very homologous (88%) up to position -535 where they diverge very abruptly. However, our electron microscopic analysis indicates that some homology appears again further upstream where no nucleotide sequence data is yet available (17). These results suggest that some sequence rearrangements may have taken place in this region without destroying the hormonal control elements of the B genes.

Homologies between Xenopus laevis vitellogenin genes and chicken vitellogenin and apo-VLDLII genes

Recently, we described the presence of several blocks of homology between the chicken vitellogenin gene and the Xenopus laevis gene A2 (22). We now discuss the results of an analysis extended to all available sequences upstream of the 5' end of the other Xenopus vitellogenin genes. Of the four blocks found in the region flanking the initiation site, two are relatively A+T rich (Blocks-2 and -3) and the degree of homology is quite variable according to the pairs of genes compared. Since the results of functional tests are not yet available, we shall not speculate as to the relevance of these regions for gene regulation. In contrast, Blocks -1 and -4 have a lower A+T content and present interesting characteristics. Block-1 is located immediately upstream of the TATA box. In the chicken lysozyme gene the equivalent position has been described as a strong glucocorticoid receptor binding site (8). By analogy, this region might be involved in

similar functions in estrogen-regulated gene expression, assuming that the mode of action of the different classes of steroids possess common features.

Block-4 contains a well conserved core showing a twofold rotational symmetry from which the consensus sequence 5' -GGTCANNNTGACC- 3' was derived. The molecular analysis of DNA control elements of gene regulation in procaryotes has shown that symmetrical nucleotide sequences play a central role in the recognition of important sites on the DNA by regulatory factors e.g. activators or repressors (39). Recognition processes of regulatory sites of specific genes might be similar in eucaryotes. Indeed, recently Borgmeyer et al. (40) described an eucaryotic nuclear protein recognizing a symmetrical sequence element on double-stranded chicken DNA. It is worth noting that though different, the consensus element -TGGCANNNTGCCA- deduced by Borgmeyer et al. is similar to the one -GGTCANNNTGACC- we found located, most of the time in a few copies, close to the transcription initiation sites of the vitellogenin and apo-VLDLII genes.

Possible functional significance

In the 5' flanking region of the chicken vitellogenin gene, the three copies of the -GGTCANNNTGACC- element are located between, rather than within, nuclease hypersensitive sites (20, 26). Such sites have not been found in the Xenopus gene A2 which has been analyzed (41) and so a comparative observation is not possible. Another interesting feature is the proximity on the chicken gene of the symmetrical element at position -625 and of the MspI estrogen-dependent hypomethylated site (20, 42, 43). Indeed the two 3'-proximal cytosines of the element are part of the MspI site. Furthermore, this same -625 element is at the 5'-edge of a putative estrogen receptor binding site (15). This, together with the more general finding that the element flanks genes which are estrogen-responsive in the liver, suggest that it might be involved in liver-specific expression of estrogen-controlled genes. If this symmetrical sequence is a protein binding site it might represent the target of the estrogen receptor complex. However, the egg white protein genes, which are estrogen-responsive in the oviduct do not contain the element. This suggests a model in which the hepatocyte estrogen receptor is different from the oviduct estrogen receptor, each one recognizing a tissue specific binding site, the possible basis of the

differential effect of estrogen on gene expression in the two cell types. Alternatively, the element could represent the binding site for a factor responsible for the "determination" of liver estrogen- responsive genes (44). In this case there is no absolute need for different estrogen receptors in hepatocytes and oviduct cells. The estrogen receptor complex would promote binding of RNA polymerase in the A+T rich vicinity of the "open determination site" by directing it to its entry site. Yet another possibility is that the symmetrical sequence described interacts with an intermediary protein transducing receptor binding into transcriptional activation (45). A clearer understanding of the complex mechanism of gene activation by estrogens will require a step-by-step reconstitution of the process using several different and independent means.

ACKNOWLEDGEMENTS

We would like to thank Carolyn Tolstoshev and Mauro Rezzonico for the gift and adaptation of computer programs, David Cooper, Susan Gerber-Huber, Bob Hipskind and Riccardo Wittek for comments on the manuscript, Hannelore Pagel and Nelly Buchi for typing it. This work was supported by the Etat de Vaud and by the Swiss National Science Fondation.

REFERENCES

1. Wahli, W., Dawid, I.B., Ryffel, G.U. and Weber, R. (1981), Science 212, 298-304.
2. Brock, M.L. and Shapiro, D.J. (1983), J. Biol. Chem. 258, 5449-5455.
3. Dean, D.C., Knoll, B.J., Riser, M.E. and O'Malley,B.W. (1983), Nature 305, 551-554.
4. Karin, M., Haslinger, A., Holtgreve, H., Richards, R.I., Krauter, P., Westphal, H.M. and Beato, M. (1984), Nature 308, 513-519.
5. Buetti, E. and Diggelmann, H. (1983), EMBO J. 2, 1423-1429.
6. Hynes, N., van Ooyen, A.J.J., Kennedy, N., Herrlich, P., Ponta, H. and Groner, B. (1983), Proc. Natl. Acad. Sci. USA, 80, 3637-3641.
7. Majors, J. and Varmus, H. (1983), Proc. Natl. Acad. Sci. USA, 80, 5866-5870.
8. Renkawitz, R., Schuetz, G., von der Ahe, D. and Beato, M. (1984), Cell 37, 503-510.
9. Mulvihill, E.P., LePennec, J.-P. and Chambon, P. (1982). Cell 28, 621-632.
10. Compton, J.G., Schrader, W.T. and O'Malley, B.W. (1983), Proc. Natl. Acad. Sci. USA 80, 16-20.
11. Payvar, F., Wrange, O., Carlstedt-Duke, J., Okret, S., Gustafson, J.-A., and Yamamoto, K.R. (1981), Proc. Natl. Acad. Sci. USA, 78, 6628-6632.

12. Givindan, M.V., Spiess, E. and Majors, J. (1982), Proc. Natl. Acad. Sci. USA, 79, 5157-5161.
13. Pfahl, M. (1982), Cell 31, 475-482.
14. Geisse, S., Scheidereit, C., Westphal, H.M., Hynes, N.E., Groner, B. and Beato, M. (1982), EMBO J. 1, 1613-1619.
15. Jost, J.-P., Seldran, M. and Geiser, M. (1984), Proc. Natl. Acad. Sci. USA 81, 429-433.
16. Wahli, W., Germond, J.E., ten Heggeler, B. and May, F.E.B. (1982), Proc. Natl. Acad. Sci. USA, 79, 6832-6836.
17. Germond, J.-E., ten Heggeler, B., Schubiger, J.-L., Walker, P., Westley, B. and Wahli, W. (1983), Nucleic Acids Res., 11, 2979-2997.
18. Arnberg, A.C., Meijlink, F.C.P.W., Mulder, J., van Bruggen, E.F.J., Gruber M., and AB, G. (1981), Nucleic Acids Res. 9, 3271-3286.
19. Wilks, A., Cato, A.C.B., Cozens, P.J., Mattaj, I. and Jost, J.-P. (1981), Gene 16, 249-259.
20. Burch, J.E.B. and Weintraub, H. (1983), Cell 33, 65-76.
21. Haché, R.J.G., Wiskocil, R., Vasa, M., Roy, R.N., Lau, P.C.K., and Deeley, R. (1983), J. Biol. Chem. 258, 4556-4564.
22. Walker, P., Brown-Luedi, M., Germond, J.-E., Wahli, W., Meijlink, F.C.P.W., van het Schip, A.D., Roelink, H., Gruber, M. and AB, G. (1983). EMBO J. 2, 2271-2279.
23. Messing, J., Crea, R. and Seeburg, P.H. (1981), Nucleic Acids Res. 9, 309-321.
24. Sanger, F., Nicklen, S. and Coulson, A.R. (1977), Proc. Natl. Acad. Sci. USA 74, 5463-5467.
25. Barnes, W.M. and Bevan M. (1983), Nucleic Acids Res. 11, 349-368.
26. Hong, G.F. (1982), J. Mol. Biol. 158, 539-549.
27. Burch, J.B.E. (1984), Nucleic Acids Res. 12, 1117-1148.
28. Geiser, M., Mattaj, I.W., Wilks, A.F., Seldran, M. and Jost, J.-P. (1983), J. Biol. Chem. 258, 9024-9030.
29. Wahli, W., Dawid, I.B., Wyler, T., Weber, R. and Ryffel, G.U. (1980), Cell 20, 107-117.
30. Woo, S.L.C., Beattie, W.G., Catterall, J.F., Dugaiczyk, A., Staden, R., Brownlee, G.G. and O'Malley, B.W. (1981), Biochemistry 20, 6437-6446.
31. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980), Nucleic Acids Res. 8, 127-142.
32. Heilig, R., Muraskowsky, R. and Mandel, J.L. (1982), J. Mol. Biol. 156, 1-19.
33. Heilig, R., Muraskowsky, R., Kloepfer, C., Mandel, J.L. (1982), Nucleic Acids Res. 10, 4363-4382.
34. Grez, M., Land H., Giesecke, K., Schuetz, G., Jung, A. and Sippel, A.E. (1981), Cell 25, 743-757.
35. Gerlinger, P., Krust, A., LeMeur, M., Perrin, F., Cochet, M., Gannon, F., Dupret, D. and Chambon, P. (1982), J. Mol. Biol. 162, 345-364.
36. Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F. and Chambon, P. (1979), Nature 282, 567-574.
37. Wallace, R.A. (1978) in Vertebrate Ovary; Jones, R.E. (ed.), Plenum Publishing, New-York, pp.469-502.
38. Bisbee, C.A., Baker, M.A., and Wilson, A.C., Hadzi-Azimi, J. and Fischberg, M. (1977), Science 195, 785-787.

39. Takeda, Y., Ohlendorf, D.H., Anderson, W.F., and Matthews, B.W. (1983), Science 221, 1020-1026.
40. Borgmeyer, U., Nowock, J. and Sippel, A.E. (1984), Nucleic Acids Res. 12, 4295-4311.
41. Gerber-Huber, S., May, F.E.B., Westley, B.R., Felber, B., Hosbach, H.A., Andres, A.C. and Ryffel, G.U. (1983), Cell 33, 43-51.
42. Wilks, A.F., Cozens, P.J., Mattaj, I.W. and Jost, J.P. (1982), Proc. Natl. Acad. Sci. USA 79, 4252-4255.
43. Meijlink, F.C.P.W., Philipsen, J.N.J., Gruber, M. and AB, G. (1983), Nucleic Acids Res. 11, 1361-1373.
44. Brown, D.D. (1984), Cell 37, 359-365.
45. Palmiter, R.D., Mulvihill, E.R., Shepherd, J.H. and McKnight, G.S. (1981), J. Biol. Chem. 256, 7910-7916.