Chromatin studies reveal that an ERE is located far upstream of a vitellogenin gene and that a distal tissue-specific hypersensitive site is conserved for two coordinately regulated vitellogenin genes

John B.E.Burch* and Andrew H.Fischer[†]

Institute for Cancer Research, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111, USA

Received April 10, 1990; Revised and Accepted June 8, 1990

ABSTRACT

Estrogen induces the expression of three vitellogenin genes in chicken hepatocytes. To survey the vitellogenin III (VTGIII) gene region for possible distal regulatory sequences, we identified tissue-specific hypersensitive (HS) sites within a 45 kb chromatin region spanning this gene. Five constitutive HS sites were found to mark the VTGIII gene region in hormone-naïve hepatocytes. Strikingly, the constitutive HS site located 5.5 kb upstream of the VTGIII gene and a previously identified HS site located within the coordinately regulated VTGII gene mapped to nearly identical copies of a 72 bp sequence. Moreover, it would appear that there has been evolutionary pressure to retain specifically this 72 bp of VTGII-like sequence near the VTGIII gene subsequent to the VTGIII and VTGII genes becoming unlinked ~16 Myr ago. Two additional sets of HS sites were induced in the VTGIII gene region in response to estrogen. One set mapped immediately upstream of the gene in the vicinity of what we show to be a functional estrogen response element (ERE). The other induced HS site mapped 7.5 kb upstream of the gene. This far-upstream region was sequenced and was found to contain two imperfect ERE consensus sequences spaced 88 bp apart. In transient expression assays neither of these individual imperfect ERE sequences was functional, but a fragment spanning both sequences behaved as a strong ERE. In contrast to this synergism between imperfect ERE sequences, the presence of an NF-1 binding site 23 bp away from the more distal imperfect ERE sequence was not sufficient to render the latter a functional ERE in our assays.

INTRODUCTION

The chicken genome contains three vitellogenin genes (VTGI, VTGII, VTGIII) that encode the major yolk proteins (1,2). These

genes are coordinately expressed in the livers of egg-laying hens in response to the elevated estrogen levels that also serve to orchestrate both the terminal differentiation of the oviduct as well as the expression of the egg-white protein genes in that tissue. Whereas precocious expression of the egg-white protein genes by exogenous estrogen is restricted to female birds, precocious expression of the vitellogenin genes can be effected in birds of either sex as early as day 9 of embryonic development (3,4). The control of vitellogenin gene expression is very tightly regulated in avian hepatocytes, with mRNA levels for individual vitellogenin genes increasing by as much as five orders of magnitude in response to a single injection of estrogen (4). These vitellogenin genes, as well as the coordinately regulated apoVLDLII gene, thus represent attractive models for studying steroid-dependent tissue-specific gene expression (5).

The induction of vitellogenin gene transcription occurs as a direct response to estrogen (6,7). Thus the vitellogenin genes must reside in 'potentially active' chromatin domains within hormone-naïve hepatocytes. This suggests that control elements that operate at the level of chromatin domains probably contribute to the regulation of these genes prior to the addition of estrogen. Since it has generally been difficult to gain insights into tissue-specific chromatin restructuring events that occur independent of gene transcription, the vitellogenin genes provide valuable models for addressing this level of regulation. With this in mind, several years ago we initiated a chromatin structural analysis of the VTGII gene region as it exists within hormone-naïve hepatocytes (8). Among other things, we looked for tissue-specific HS sites since virtually all control regions that have been described to date are found to be marked by such sites (9,10). Aside from a set of HS sites that mapped to a CR1 middle-repetitive element (11), the only constitutive tissue-specific HS site found within a 30 kb gene region spanning this gene was one that mapped within intron 17. We reasoned that if this HS site indeed marked an important control region we might expect to find a comparable HS site in the vicinity of the evolutionarily related VTGIII gene that we recently cloned (12). We addressed this point in the

* To whom correspondence should be addressed

[†] Present address: Department of Pathology, Thomas Jefferson University Hospital, 11th and Walnut Streets, Philadelphia, PA 19107, USA

present study. Strikingly, a HS site located 5.5 kb upstream of the VTGIII gene was found to be homologous to the HS site located within the 17th intron of the VTGII gene. The 72 bp sequences that underlie these two HS sites are shown to be nearly identical. This finding provides interesting insights into the evolution of this coordinately regulated multigene family.

The sequences immediately upstream of two chicken vitellogenin genes (VTGII and VTGIII), as well as four frog vitellogenin genes, have been determined (12,13,14). In each case consensus sequences (13) for estrogen response elements (EREs) have been identified and many of these have been shown to represent functional EREs in transient expression experiments (15,16,17). In addition, we have shown that estrogen induces a set of nuclease hypersensitive (HS) sites in the vicinity of the two functional EREs that flank the chicken VTGII gene (8,15). Estrogen also induces a set of HS sites downstream of the chicken VTGII gene (11) near an ERE consensus sequence that happens to lie within the remnant promoter of a vitellogenin pseudogene (12). Given this correlation between EREs and estrogen-dependent HS sites, we anticipated that HS sites would also be induced near the consensus ERE sequence that we previously identified immediately upstream of the chicken VTGIII gene (12). We show that this is indeed the case and that a fragment containing this consensus sequence functions as an ERE in transient expression assays. More importantly, we also found an estrogen-dependent HS site 7.5 kb upstream of this gene. This unexpected result led us to discover a complex functional ERE in this far-upstream region that is the most distal ERE presently identified for any gene (18–21).

MATERIALS AND METHODS

Nuclease hypersensitive sites (HS sites) were mapped for the VTGIII gene region as described previously for the VTGII gene region (8) except that the Southern blots were transferred to Nytran membranes and hybridized using 2×10^6 CPM/ml of SP6 or T7 derived RNA probes (22). DNaseI digestions were done at 15°C (to minimize the activity of endogenous nucleases) for 10 min. using between 0.1 and 20.0 μ g/ml of enzyme. Hybridizations were done in 50% formamide (23) at 49°C for 18 hrs. The filters were washed for 2×30 min. at room temperature in 1st blot wash ($2 \times$ SSC, 1% SDS) and 2×30 min. at 65°C in 2nd blot wash ($0.1 \times$ SSC, 1% SDS). Filters were exposed to film overnight at -70°C using two intensifying screens.

DNA sequences were determined from double stranded templates (24) using synthetic oligonucleotide primers and the Sequenase kit (U.S. Biochemical).

Subfragments of the parental HS-I₁ containing 1.2 kb far-upstream region of VTGIII (the HindIII×EcoRI fragment corresponding to probe 3 in Figure 1) were synthesized using polymerase chain reactions (PCR). Pairs of synthetic oligonucleotides that contained HindIII and BamHI ends, respectively, were used in conjunction with reagents from the Perkin-Elmer-Cetus kit. The PCR reactions were programmed for 30 cycles (1 min. at 95°C, 1 min. at 50°C, 1 min. at 72°C) followed by 4 min. at 72°C. The PCR fragments were isolated, digested with HindIII and BamHI and cloned into the pTK.CAT test vector (15) using standard techniques (25). This test vector contains the $-105/+52$ TK promoter cloned upstream of the CAT gene (26). Each of the resultant clones was sequenced to verify that the PCR amplification had not introduced mutations. The oligonucleotides used to prepare the various plasmids were

as follows: 5' GCGAAGCTTGCACAGGTTTTGGCAGA 3' and 5' GCGGGATCCATGTTGACTCCTCCTG 3' (SE.TKCAT); 5' GCGAAGCTTGCACAGGTTTTGGCAGA 3' and 5' GCGGGATCCTTTTTGAAGACTGAAG 3' (SM.TKCAT); GCGAAGCTTCAGTCTTCAAAAAG 3' and 5' GCGGGATCCATGTTGACTCCTCCTG 3' (ME.TKCAT); 5' GCGAAGCTTGCTGCTCTCTATCATCT 3' and 5' GCGGGATCCATGTTGACTCCTCCTG 3' (AE.TKCAT). HS.TKCAT was constructed by digesting the plasmid HE.TKCAT with SphI and religating the vector. The plasmid denoted LL.TKCAT in Figure 4 was prepared by cloning a PCR-amplified fragment (corresponding to sequences between -1433 and -330 upstream of the VTGIII gene) into the HindIII site immediately upstream of the TK promoter in the test vector TK.CAT. The oligonucleotides used to direct this PCR synthesis were 5' GCGAAGCTTAAACCGATGAAGGCTTC 3' and 5' GCGAAGCTTGGTGCCAGGTTATCCTG 3'.

Cotransfections with 3μ g each of the test plasmid and an estrogen receptor expression vector (pCEO; 27) were carried out using the method of calcium phosphate precipitation (28). The recipient cells were primary chick embryo fibroblast (CEF) cells. As with our previous analysis of the EREs located immediately upstream of the VTGII gene (15) moxestrol, which is a synthetic estrogen (29), was used at a concentration of 10^{-6} M and was found to give results comparable to those obtained with 17 β -estradiol. CAT assays were performed on $10-50 \mu$ g of whole cell extracts using a routine protocol (30) except that the radioactivity on the TLC plates was quantified using an Ambis Radioanalytic Imaging System instead of film.

RESULTS

Constitutive and Estrogen-induced Liver-specific Hypersensitive Sites in the VTGIII Gene Region

Our first step in surveying the VTGIII gene region for potential control regions was to search for nuclease hypersensitive (HS) sites. The VTGIII gene region was analyzed essentially as previously described for the VTGII gene region (8). Preparations of liver nuclei from hormone-naïve or estrogen-treated chicks were digested with varying amounts of DNaseI and then the DNA was analyzed on Southern blots. Nuclei from blood cells were also analyzed to address whether the HS sites seen with liver nuclei were tissue-specific. This control was important since erythrocytes were the major contaminating cell type in our liver preparations.

We used thirteen probing protocols to screen for HS sites over more than 45 kb of DNA containing the VTGIII gene (Figure 1). All of the HS sites that we observed can be seen in the six representative sets of Southern blots shown in Figure 2: the seven additional protocols depicted in Figure 1 yielded consistent results (data not shown). It is clear from this analysis (i) that five sites within the VTGIII gene region were marked by tissue-specific constitutive HS sites in hormone-naïve birds (labelled HS-C_{1,2,3,4,5}) and (ii) that estrogen treatment induced the formation of two additional HS sites (labelled HS-I_{1,2}).

A Constitutive Tissue-specific Nuclease Hypersensitive (HS) Site has been Conserved in the Context of Two Coordinately Regulated Vitellogenin Genes

Since the VTGIII gene is one of three coordinately regulated chicken vitellogenin genes, one might expect that at least a subset of the HS-C sites identified in the present study would map to, and hence highlight, regulatory sequences that are also conserved

Fig. 1 Strategy for mapping nuclease hypersensitive sites (HS sites) over the VTGIII gene region. The solid arrow at the top shows the position of the VTGIII gene within the 45 kb that we cloned and restriction mapped (12). The vertical arrows on the top line mark the positions of HS sites (see Figure 2). HS sites labelled 'C' refer to constitutive (but tissue-specific) HS sites that are evident before the gene is transcribed whereas those labelled 'I' refer to induced sites which are observed only when the gene is being transcribed in response to estrogen. A total of thirteen fragments were used for probing Southern blots of DNase I digested nuclear DNA. The probes used are indicated by boxes and the genomic fragments they revealed are represented by horizontal arrows. The restriction enzymes that define the ends of these DNA fragments are designated as follows: B=BamHI, H=HindIII, K=KpnI, R=EcoRI and X=XbaI. The six Southern blots shown in Figure 2 are indicated by the dark boxes which are numbered to match the panels in Figure 2. Southern blots that either yielded redundant information or failed to reveal HS sites (data not shown) are indicated by light boxes. The two HS sites that are analyzed in greater detail in subsequent figures are HS-I₁ and HS-C₃.

in the context of other vitellogenin gene regions. We and others had previously shown that the VTGII gene contains an HS-C site within the 17th intron as well as a cluster of HS-C sites that map to a CR1 repetitive element located immediately downstream of the gene (8,11). The VTGIII gene region, in contrast, is both devoid of CR1 sequences (31) and does not display an HS-C site in the region of intron 17 (Figure 1). The possibility that the HS-C₃ site located 5.5 kb upstream of the VTGIII gene might nevertheless be related to the HS-C site located in intron 17 of VTGII was suggested by the results of a cross-hybridization analysis that we previously reported for the VTGII and VTGIII gene regions (12; see Discussion). With this in mind, a 0.2 kb region corresponding to our best estimate for the location of the HS-C₃ site was sequenced. As shown in Figure 3A, this sequence revealed a 72 bp segment that was nearly identical to the sequence that underlies the HS-C site within intron 17 of VTGII. Based on evolutionary considerations presented in the Discussion section and summarized in Figure 8, we infer that this conserved sequence serves a role in the coordinate regulation of these genes.

An Estrogen-induced HS Site Located Immediately Upstream of the VTGIII Gene Maps in the Vicinity of a Functional Estrogen Response Element

As noted above, the region immediately upstream of the VTGIII cap site was devoid of chromatin HS sites in hormone-naïve birds but became restructured when the gene was activated in

hepatocytes in response to estrogen (Figure 2, panel 4). We previously reported a similar observation for the VTGII gene (8) and showed (15) that the restructured region contained two functional estrogen response elements (EREs). Since the VTGIII gene is flanked by an imperfect ERE consensus sequence identical to the more proximal ERE of VTGII (12), it seemed likely that the doublet HS-I site induced immediately upstream of the VTGIII gene would similarly map near a functional ERE. To test this prediction, we cloned a fragment from this flanking region of VTGIII into the pTK.CAT test vector and performed transient expression assays. In this and all subsequent experiments, test plasmids were cotransfected with an estrogen receptor expression vector to compensate for the lack of estrogen receptors in the recipient chick embryo fibroblast (CEF) cells. As shown in Figure 4, it is clear that this VTGIII sequence confers estrogen-dependent expression to the otherwise insensitive TK promoter (compare TKCAT and LL.TKCAT). Thus the HS-I₂ site maps in the vicinity of a functional ERE located immediately upstream of the VTGIII gene.

An Estrogen-induced HS Site Reveals a Functional Estrogen Response Element Far Upstream of the VTGIII Gene

As shown in panel 2 of Figure 2, faint HS sites could be seen in the HS-I₁ region in hormone-naïve hepatocytes and these sites were dramatically augmented in response to estrogen. To address whether an ERE might reside in this region ~7.5 kb upstream of the VTGIII gene, a 1.2 kb fragment spanning the HS-I₁ site

Fig. 2 Constitutive and induced nuclease HS sites over the VTGIII gene region. Liver nuclei from hormone-naïve (labelled '-') or estrogenized (labelled '+') chicks were digested with various amounts of DNaseI before the DNA was isolated and digested with restriction enzymes for Southern blot analysis. Nuclei from RBC (labelled 'R') were similarly analyzed. The probing protocol for each autoradiogram is depicted in Figure 1. HS sites present in the hormone-naïve liver nuclei are indicated as C₁-C₅ whereas the two regions which become hypersensitive in response to estrogen are indicated as I₁ and I₂. The locations of these sites are indicated in Figure 1. The location of the distal ERE identified by transient expression assays (see Figure 4) is indicated in Panel 3. Note that some samples were apparently overdigested with DNaseI: these in no way detract from the conclusions being drawn. Note also that the HS sites labelled HS-I₂, HS-C₁, HS-C₅ are each doublets whereas the HS site labelled HS-I₁ is an ~200 bp 'window' of hypersensitivity.

(corresponding to probe 3 of Figure 1) was cloned into the pTK.CAT test vector and the resultant plasmids were analyzed in transient expression assays. As shown by the constructs HE.TKCAT and EH.TKCAT in Figure 4, placement of this fragment in either orientation upstream of the TK promoter yielded CAT expression that was strongly induced in response to estrogen, indicating that a functional ERE resides within this distant region upstream of the VTGIII gene.

The sequence of this ERE-containing fragment was determined (Figure 5) and searched for consensus (13,32) ERE sequences (GGTCANNNTGACC). Although no perfect matches to this

consensus were found, we did find two imperfect ERE consensus sequences spaced 88 bp apart; each of these matched the consensus in 11 out of 13 positions (see Figure 5). To determine whether either or both of these sequences contributed to the ERE activity observed in Figure 3, subfragments from this region were cloned into the pTK.CAT test vector and analyzed in transient expression assays (Figure 6). As expected, the HS.TKCAT plasmid (which contains ~1.0 kb of sequence from the left (HindIII) end of the parent fragment but lacks both of the imperfect ERE consensus sequences) was devoid of ERE activity. Conversely, the SE.TKCAT plasmid (which contains the

Fig. 3 DNA sequence comparison for the HS-C sites that are conserved in the VTGIII and VTGII gene regions. The 72 bp sequences that are conserved in association with the VTGIII HS-C₃ site (Panel 3 of Figure 2) and the VTGII HS-C(A₁) site (8) are shown in Panel A. The sequence of the far-upstream region of VTGIII containing the HS-C₃ site is arbitrarily numbered from 1–120 in Panel A: sequences from the VTGII gene are numbered relative to the transcriptional start site (45). In Panel B, sequences from an overlapping portion of the far-upstream region of VTGIII (61–180) are compared with intron 31 sequences of VTGII. Previous studies from our laboratory showed that a VTGII pseudogene (ΨVTGII) lies immediately upstream of the VTGIII gene and that the major fragment of ΨVTGII is homologous to the 3' end of VTGII beginning within intron 31 (12). Thus the VTGIII HS-C₃ site is revealed to be a small second fragment of ΨVTGII which is flanked by deletions of more than 7 kb on either side (see Figure 8).

remaining ~0.2 kb of sequence inclusive of both imperfect ERE sequences) displayed ERE activity. Unexpectedly, this fragment also effected a significant increase in basal expression from the TK promoter, suggesting that it contains binding sites for other positive-acting factors in addition to estrogen receptors (see below). The fact that this effect on basal activity was suppressed when the entire 1.2 kb fragment was assayed further suggests that negative elements may reside upstream of the ERE. Another curious aspect of these results is that, as indicated in the cartoon at the top of Figure 6, the HS-I₁ site that is induced by estrogen actually maps 0.8 kb upstream of the ERE. Thus it would appear that this may be a rather complex distal control region, with the ERE activity being only one part.

To determine whether either or both of the distal imperfect ERE sequences could function independently, the 0.2 kb of sequence from the right end of the 1.2 kb parental fragment was divided into two fragments that were individually cloned into the pTK.CAT test vector to yield the plasmids SM.TKCAT and ME.TKCAT. As indicated in Figure 6, these plasmids were each essentially devoid of estrogen-dependent expression and showed basal expression on par with the parental pTK.CAT plasmid. We infer that the activity of this far-upstream ERE is dependent on synergism between the two imperfect ERE sequences.

We also searched the sequence of the 1.2 kb distal region for other known factor binding sites and, as highlighted in Figure 5, found a perfect consensus binding site for the ubiquitous

Fig. 4 Fragments containing the HS-I₁ and HS-I₂ sites each confer estrogen-responsiveness to a heterologous promoter. A 1.2 kb HindIII×EcoRI fragment containing the HS-I₁ site was cloned in both orientations (using HindIII linkers) immediately upstream of the TK promoter in the pTK.CAT test vector to yield the plasmids HE.TKCAT and EH.TKCAT, respectively. Similarly, a 1.0 kb fragment containing the HS-I₂ site was cloned immediately upstream of the TK promoter in the TK.CAT test vector to yield LL.TKCAT (see Materials and Methods). These three derivative plasmids, as well as the parental pTK.CAT vector, were each cotransfected into chick embryo fibroblast (CEF) cells along with an estrogen receptor expression vector (pCEO; 27) and transient expression assays were performed after culturing the cells in either the absence (-E) or presence of estrogen (+E) for 48 hrs. as described in Materials and Methods. The CAT activities were average values obtained from two independent experiments and represent percent conversion of substrate.

transcription factor NF-1 (33,34). As expected, *in vitro* footprinting experiments confirmed that this sequence specifically binds NF-1 (data not shown). Despite the fact that this NF-1 binding site is only 23 bp away from the more distal imperfect ERE consensus sequence, a fragment containing this NF-1/ERE pair failed to effect estrogen-dependent expression through the TK promoter (see construct SM.TKCAT of Figure 6). This result was surprising considering that imperfect glucocorticoid response elements which do not function alone can nevertheless be rendered functional when placed near binding sites for other transcription factors, including NF-1 (35,36). To test whether the NF-1 binding site might nevertheless contribute to the ERE activity seen with the pair of imperfect ERE palindromes in construct SE.TKCAT (see Figure 6), a test plasmid was constructed (AE.TKCAT) that was identical to SE.TKCAT except for a 23 bp deletion which removed the NF-1 binding site from the left end of the insert. As shown in Figure 7, transient expression assays using these sister plasmids revealed that both basal and estrogen-dependent expression were reduced by a factor of two when the NF-1 binding site was deleted. Thus, despite the close spacing of their respective binding sites, NF-1 and estrogen receptors appear to act independently in this context.

DISCUSSION

We have presented evidence for two different types of HS sites that are each located far upstream of the chicken VTGIII gene. These regions were detected on the basis of their being marked,

Fig. 5 The sequence of the 1.2 kb fragment containing the far-upstream HS-I₁ site reveals two imperfect estrogen response element (ERE) consensus sequences as well as a consensus NF-1 binding site. The 1.2 kb HindIII×EcoRI fragment containing the HS-I₁ site was sequenced on both strands as described in Materials and Methods. Consensus ERE and NF-1 binding sites, respectively, are indicated above the related sequences found within this 1.2 kb fragment. A few restriction sites are indicated (see Figures 6 and 7 below).

respectively, by a constitutive tissue-specific HS site at -5.5 kb (HS-C₃) and an estrogen-dependent tissue-specific HS site at -7.5 kb (HS-I₁). The significance of these two findings is discussed below.

The Conserved Constitutive Tissue-specific Hypersensitive Site

We infer that the 72 bp sequence which underlies the constitutive HS site located 5.5 kb upstream of the VTGIII gene very likely marks an important control region for this gene. This inference is based largely on the finding that a nearly identical HS site has also been conserved in the context of the coordinately regulated VTGII gene, albeit in the latter case the HS site lies within rather than upstream of the gene. We have recently deduced that these two genes were subject to an unusual evolutionary history (12). When one considers this fact, it becomes obvious how the two copies of this conserved 72 bp sequence (and associated constitutive HS sites) came to be located in such different positions relative to the VTGII and VTGIII promoters (see below). Moreover, this unusual evolutionary history may account for why this HS site was not simply conserved within the 17th introns of both the VTGII and VTGIII genes as might have been expected if it is indeed an important control element for both of these genes.

As summarized in Figure 8A, the VTGII and VTGIII genes arose from the tandem duplication of an ancestral vitellogenin

Fig. 6 The two imperfect ERE consensus sequences identified in Figure 5 behave cooperatively to comprise a functional ERE. Various subfragments of the far-upstream parental 1.2 kb HindIII×EcoRI fragment were individually cloned into the pTK.CAT test vector. The resultant plasmids were each cotransfected along with an estrogen receptor expression vector into CEF cells and assayed for transient CAT expression (bottom panel) after being cultured for 48 hrs. in either the absence (light boxes) or presence (black boxes) of estrogen. The percent CAT conversions represent average values obtained from two independent experiments for each construct. The cartoon at the top of the figure shows the location of the HS-I₁ site observed in vivo (indicated by a hatched box) as well as the locations of the imperfect ERE consensus sequences (indicated by open boxes). The letter designations H, S, M and E refer to the positions of the HindIII, SphI, MboII and EcoRI sites, respectively (see Figure 5).

gene (12). This must have occurred very early in avian evolution, as evidenced by the fact that comparable introns for the VTGII and VTGIII genes have diverged to the point of being completely unrelated at the sequence level. One can infer that this tandem arrangement of VTGII and VTGIII genes must have persisted during the vast majority of their long evolutionary history. However this tight linkage was broken by a second duplication event which occurred recently in evolutionary time (approximately 16 Myr ago). As a consequence, the VTGII and VTGIII genes in the contemporary chicken genome are each flanked by partially deleted pseudogene copies of each other.

One can imagine that such a biphasic evolutionary history would have novel consequences for critical control elements, such as enhancers and locus activation regions (37,38,39) that are capable of exerting their effects over appreciable distances. Whereas a single control element of this type might be competent to act on multiple genes within a single chromatin domain, in order for dispersed genes to retain their coordinate regulation it would obviously be necessary for each gene to take with it copies of all of the requisite control elements. It is tempting to speculate that this may be precisely what happened in the case of the 72 bp sequence that underlies the conserved HS site that

Fig. 7 Deletion of the NF-1 Binding site results in a 2-fold decrease in basal and estrogen-dependent expression. Test plasmids containing various portions of the 1.2 kb HindIII $\times$ EcoRI far-upstream fragment were cotransfected into CEF cells and assayed for CAT activity after culturing for 48 hrs. in either the absence (stippled boxes) or presence (black boxes) of estrogen. The percent CAT conversion for each represents the average of two independent experiments. The deletion of the NF-1 site from SE.TKCAT to yield AE.TKCAT results in a 2-fold decrease in basal and estrogen-dependent expression as indicated. The letter designations H, S, A and E refer to the positions of the HindIII, SphI, AluI and EcoRI sites, respectively (see Figure 5).

we have identified. Whereas intragenic sequences would have been copied in the original duplication event which gave rise to the VTGII and VTGIII genes, the enormous evolutionary time during which these genes remained tandemly linked would have provided ample opportunity for one of the two copies of this 72 bp sequence to diverge to the point of it no longer being functional. Since the 72 bp sequence that presumably remained conserved during this time lies within the VTGII gene, the recent duplication event would necessarily have retained a copy of this HS site in association with the VTGII promoter. In contrast, for the dispersed VTGIII gene to retain a copy of this HS site, the duplicated region would have to have included not only the VTGIII gene but also the HS site from the neighboring VTGII gene. This is indeed observed. Moreover, the notion that there may have been selective pressure to retain specifically this HS site in association with the VTGIII promoter would appear to be supported by the striking observation that the underlying 72 bp sequence is highlighted by deletions of more than 7 kb of VTGII-related sequences on either side. The possibility that all of this is simply coincidental seems exceedingly unlikely.

Elucidating the precise role of this 72 bp sequence will require further studies. However, examination of the HS site sequence upstream of the VTGIII gene reveals two features worth noting. First, the deletion breakpoint at the 3' side of this 72 bp sequence has created a consensus binding site (MTTRCNMA; nucleotides 112–120 in Figure 5) for the transcription factor

C/EBP (40). This factor is expressed in hepatocytes (41) and binds within control regions of several liver genes (42,43) including the promoter region of VTGII (manuscript in preparation). Secondly, this deletion resulted in the close juxtaposition of two sequences (nucleotides 60–69 and 167–158 in Figure 3) which fit the consensus GTGGTGANGG. These sequences bind a factor(s) present in liver nuclear extracts (unpublished results) and are similar to sequences (GTGGTGATGT and ATGGTGAGGT) that have been inferred to be important promoter elements for the frog vitellogenin A1 and A2 genes, respectively (44).

The Distal Estrogen-dependent Hypersensitive Site

The functional ERE that we discovered 6.7 kb upstream of the VTGIII gene was shown to contain two imperfect ERE consensus sequences separated by 88 bp. Whereas neither of these sequences was functional in the absence of the other, the two displayed synergistic effects and together constituted a strong ERE that was on par with the pair of EREs located immediately upstream of the VTGII gene (data not shown). We also showed that the perfect NF-1 binding site that lies only 23 bp upstream of the more distal imperfect ERE was not sufficient to render this solo imperfect ERE functional. In contrast, studies have shown that the inactivity of an imperfect glucocorticoid response element can be rescued by placing an NF-1 site 29 bp away (35). Although the details of the two sets of experiments are slightly different, our results suggest that estrogen receptors and glucocorticoid receptors may differ in their ability to synergize with other transcription factors such as NF-1 that bind nearby.

The ERE that we identified 6.7 kb upstream of the VTGIII gene is unusual in several respects. Most notably, no other EREs have been found to lie even as far as 2 kb away from an estrogen-responsive promoter (13,18–21). This raises the possibility that EREs may be able to exert effects over much larger distances than previously appreciated. Alternatively, rather than acting on the VTGIII promoter located 6.7 kb away, it is possible that this ERE is used to regulate a different type of control element located closer to this ERE. An obvious possibility is that the ERE regulates a control element that maps 0.8 kb further upstream, since a strong HS site is induced in this region in response to estradiol.

Although we do not know what type of control element, if any, the distal estrogen-dependent HS site might correspond to, several possibilities can likely be ruled out. For example, this HS site probably does not highlight another estrogen-responsive promoter since probes from this region failed to detect transcripts from either strand when laying hen RNA was analyzed on Northern blots (data not shown). A related possibility that we entertained was that the HS-I₁ site might map to a remnant vitellogenin pseudogene promoter, as previously shown for the HS sites that are induced downstream of the VTGII gene (12). However, a comparison of the sequence underlying this HS site with published vitellogenin gene and promoter sequences failed to reveal any regions of homology (data not shown). On the other hand, several of our transient expression experiments suggest that a negative element may reside in the general vicinity of this HS site. Whereas sequences overlapping the ERE were found to augment basal transcription from the TK promoter, this effect was suppressed by the addition of sequences spanning the HS site (Figures 6 and 7). These results indicate that the ERE is only one component of what appears to be a rather complex distal control region for this gene.

Fig. 8 A comparison of HS sites over the coordinately regulated VTGII and VTGIII genes. Panel A depicts the evolutionary relationship between the VTGII and VTGIII genes and pseudogenes (Ψ) inferred from a previous study (12) and modified to incorporate the data presented in Figure 3. Panel B shows the positions of HS sites and EREs in the VTGII (8,15) and VTGIII (this study) gene regions. In Panel B the genes and pseudogenes are indicated by the thick lines (with arrowheads at the respective 3' ends), the flanking regions are indicated by medium thickness lines and the deletions within the pseudogenes are indicated by thin lines. The locations of the tissue-specific HS sites (which, in some cases are actually clusters of HS sites) are marked by solid (fuzzy) ovals: a 'C' over an oval indicates that it is a constitutive HS site whereas those labelled with an 'I' are induced in response to estrogen. The two HS sites that are the focus of the present study (HS-I₁ and HS-C₃) are located upstream of the VTGIII gene and are highlighted in boldface. As indicated, a VTGIII pseudogene (ΨVTGIII) is located downstream of the VTGII gene and a VTGII pseudogene (ΨVTGII) is located upstream of the VTGIII gene (12). The portions of the two gene regions that are vertically aligned show approximately 96% homology at the DNA sequence level (12; unpublished results). Note that the HS-C(A₁) site previously identified within the 17th intron of the VTGII gene (Figure 3). Note also that all of the HS-I sites map in the vicinity of EREs, the approximate locations of which are indicated by open diamonds. See text for details.

ACKNOWLEDGMENTS

We thank Leonard Cohen, Kathleen Conklin, John Pehrson, Robert Perry and Alfred Zweidler for helpful comments on the manuscript, Dorene Davis and Susan Costello for excellent technical assistance, Pierre Chambon for providing the chicken estrogen receptor expression plasmid (pCEO), Gunter Schutz for providing the pTK.CAT test vector, Judy Herman and Anne Carson of Secretarial Services at the Institute for Cancer Research for help with preparing the manuscript. We also thank Bob Nathan and Glen Miller of the Oligonucleotide Synthesis Facility at the Institute for Cancer Research for preparing the oligonucleotides used in this study. This work was supported by NIH Grant # 35535 (awarded to JB) and NIH Grant CA-06927, a Pathology Resident Fellowship from the University of Pennsylvania Medical School (awarded to AF) and an appropriation from the Commonwealth of Pennsylvania.

REFERENCES

- Wang, S., Smith, D.E. and Williams, D.L. (1983) *Biochemistry* 22, 6206–6212.
- Evans, M.I., O'Malley, P.J., Krust, A. and Burch, J.B.E. (1987) *Proc. Natl. Acad. Sci. USA* 84, 8493–8497.
- Elbrecht, A., Lazier, C.B., Protter, A.A. and Williams, D.L. (1984) *Science* 225, 639–641.
- Evans, M.I., Silva, R. and Burch, J.B.E. (1988) *Genes & Dev.* 2, 116–124.
- Wahli, W. (1988) *Trends in Genetics* 4, 227–232.
- Wilks, A., Seldran, M. and Jost, J.-P. (1984) *Nucleic Acids Res.* 12, 1163–1177.
- Jost, J.-P., Moncharmont, B., Jiricny, J., Saluz, H. and Hertner, T. (1986) *Proc. Natl. Acad. Sci. USA* 83, 43–47.
- Burch, J.B.E. and Weintraub, H. (1983) *Cell* 33, 65–76.
- Elgin, S.C.R. (1988) *J. Biol. Chem.* 263, 9259–9262.
- Gross, D.S. and Garrard, W.T. (1988) *Ann. Rev. Biochem.* 57, 159–197.
- van het Schip, F., Samallo, J., Meijlink, F., Gruber, M. and Ab, G. (1987) *Nucleic Acids Res.* 15, 4193–4202.
- Silva, R., Fischer, A.H. and Burch, J.B.E. (1989) *Mol. Cell. Biol.* 9, 3557–3562.
- Walker, P., Germond, J.-E., Brown-Leudi, M., Givel, F. and Wahli, W. (1984) *Nucleic Acids Res.* 12, 8611–8626.
- Burch, J.B.E. (1984) *Nucleic Acids Res.* 12, 1117–1135.
- Burch, J.B.E., Evans, M.I., Friedman, T.M. and O'Malley, P.J. (1988) *Mol. Cell. Biol.* 8, 1123–1131.
- Klein-Hitpass, L., Schorpp, M., Wagner, U. and Ryffel, G.U. (1986) *Cell* 46, 1053–1061.
- Seiler-Tuyns, A., Walker, P., Martinez, E., Merillat, A.M., Givel, G. and Wahli, W. (1986) *Nucleic Acids Res.* 14, 8755–8770.
- Berry, M., Nunez, A.-M. and Chambon, P. (1989) *Proc. Natl. Acad. Sci. USA* 86, 1218–1222.
- Maurer, R.A. and Notides, A.C. (1987) *Mol. Cell. Biol.* 7, 4247–4254.
- Tora, L., Gaub, M.-P., Mader, S., Dierich, A., Bellard, M. and Chambon, P. (1988) *EMBO J.* 7, 3771–3778.
- Waterman, M., Adler, S., Nelson, G., Greene, G.L., Evans, R.M. and Rosenfeld, M.G. (1988) *Mol. Endocrin.* 2, 14–21.
- Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) *Nucleic Acids Res.* 12, 7035–7056.
- Wahl, G.M., Stern, M. and Stark, G.R. (1979) *Proc. Natl. Acad. Sci. USA* 76, 3683–3687.
- Chen, E.Y. and Seeburg, P.H. (1985) *DNA* 4, 165–170.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor University Press, Cold Spring Harbor.
- Luckow, B. and Schutz, G. (1987) *Nucl. Acids Res.* 15, 5490.
- Krust, A., Green, S., Argos, P., Kumar, V., Walter, P., Bornet, J.-M. and Chambon, P. (1986) *EMBO J.* 5, 891–897.
- Graham, F.L. and van der Eb, A.J. (1973) *Virology* 52, 456–467.
- Raynaud, J.P., Bouton, M.M., Gallet-Bourquin, D., Philibert, D., Tournemine, C. and Azadian-Boulanger, G. (1973) *Mol. Pharmacol.* 9, 520–533.
- Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982) *Mol. Cell. Biol.* 2, 1044–1051.
- Silva, R. and Burch, J.B.E. (1989) *Mol. Cell. Biol.* 9, 3563–3566.
- Klein-Hitpass, L., Ryffel, G.U., Heitlinger, E. and Cato, A.C.B. (1988) *Nucleic Acids Res.* 16, 647–663.
- Nowock, J., Borgmeyer, U., Puschel, A.W., Rupp, R.A.W. and Sipple, A.E. (1985) *Nucleic Acids Res.* 13, 2045–2061.
- Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J. and Tjian, R. (1987) *Cell* 48, 79–89.
- Schüle, R., Muller, M., Kaltschmidt, C. and Renkawitz, R. (1988) *Science* 242, 1418–1420.
- Strähle, U., Schmid, W. and Schütz, G. (1988) *EMBO J.* 7, 3389–3395.
- Forrester, W.C., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G. and Groudine, M. (1987) *Nucleic Acids Res.* 15, 10159–10177.
- Greaves, D.R., Wilson, F.D., Lang, G. and Kioussis, D. (1989) *Cell* 56, 979–986.
- Grosfeld, F., Vanassen, G.B., Greaves, D.R. and Kollias, G. (1988) *Cell* 51, 975–985.
- Ryden, T.A. and Beemon, K. (1989) *Mol. Cell. Biol.* 9, 1155–1164.
- Johnson, P.F., Landschulz, W.H., Graves, B.J. and McKnight, S.L. (1987) *Genes & Dev.* 1, 133–146.
- Johnson, P.F. (1990) *Cell Growth & Differ.* 1, 47–52.
- McKnight, S.L., Lane, M.D. and Gluecksohn-Waelsch, S. (1989) *Genes & Dev.* 3, 2021–2024.
- Dobbeling, U., Ross, K., Klein-Hitpass, L., Morley, C., Wagner, U. and Ryffel, G.U. (1988) *EMBO J.* 7, 2495–2501.
- van het Schip, F., Samallo, J., Broos, J., Ophuis, J., Mojet, M., Gruber, M. and Ab, G. (1987) *J. Mol. Biol.* 196, 245–260.