The EMBO Journal Vol.2 No.12 pp.2271 – 2279, 1983

Sequence homologies within the 5′ end region of the estrogen-controlled vitellogenin gene in *Xenopus* and chicken

Philippe Walker, Marianne Brown-Luedi, Jacques-Edouard Germond, Walter Wahli*, Frits C.P.W.Meijlink[1], Alfred D.van het Schip[1], Henk Roelink[1], Max Gruber[1] and Geert AB[1]

Institut de Biologie animale, Université de Lausanne, Bâtiment de Biologie, CH-1015 Lausanne, Switzerland, and [1]Biochemisch Laboratorium, University of Groningen, 9747 AG Groningen, The Netherlands

Communicated by Max Gruber
Received on 12 August 1983; revised on 23 September 1983

In oviparous vertebrates vitellogenin, the precursor of the major yolk proteins, is synthesized in the liver of mature females under the control of estrogen. We have established the organization and primary structure of the 5′ end region of the *Xenopus laevis* vitellogenin A2 gene and of the major chicken vitellogenin gene. The first three homologous exons have exactly the same length in both species, namely 53, 21 and 152 nucleotides, and present an overall sequence homology of 60%. In both species, the 5′-non-coding region of the vitellogenin mRNA measures only 13 nucleotides, nine of which are conserved. In contrast, the corresponding introns of the *Xenopus* and the chicken vitellogenin gene show no significant sequence homology. Within the 500 nucleotides preceding the 5′ end of the genes, at least six blocks with sequence homologies of >70% were detected. It remains to be demonstrated which of these conserved sequences, if any, are involved in the hormone-regulated expression of the vitellogenin genes.

Key words: vitellogenin genes/hormonal regulation/reverse transcription/nucleotide sequence comparison

Introduction

Vitellogenin is the precursor of the major yolk proteins in oviparous vertebrates (Wallace, 1978). It is synthesized under estrogen control in the liver of mature females. Normally, males do not produce vitellogenin, but they can be induced by estrogen injection. This estrogen-controlled gene expression is mainly regulated at the transcriptional level (Wahli *et al.*, 1981; Brock and Shapiro, 1983); in addition, preferential stabilization of vitellogenin mRNA has also been observed (Wiskocil *et al.*, 1980).

The region around the start site of transcription is probably a major regulation site of hormone-controlled gene expression. This hypothesis is based on results obtained with different experimental systems, such as proviral mouse mammary tumor virus DNA (Buetti and Diggelmann, 1981; Hynes *et al.*, 1981), $\alpha_{2\mu}$ globulin genes (Kurtz, 1981), human and rat growth hormone genes (Robins *et al.*, 1982; Doehmer *et al.*, 1982), the chicken lysozyme gene (Renkawitz *et al.*, 1982) and the rat prostatic C3 gene (Page and Parker, 1983). DNA sequences that influence accuracy and rate of transcription initiation by polymerase II have also been found in the 5′ end flanking region of many genes the expression of which is not hormone dependent (Grosschedl and Birnstiel, 1980; Benoist and Chambon, 1981; Fromm and Berg, 1982; Struhl, 1981; Dierks *et al.*, 1981a, 1981b; McKnight *et al.*, 1981; McKnight and Kingsbury, 1982; Grosveld *et al.*, 1982a, 1982b; Mellon *et al.*, 1981; Pelham, 1982; Dierks *et al.*, 1983).

By comparing the 5′ end region of different genes whose expression is under similar control, one might find common structural features which play a role in important control mechanisms of gene expression. As a step towards such analyses, the estrogen-controlled vitellogenin genes of *Xenopus laevis* (A1, A2, B1 and B2) and of chicken have been isolated and their molecular organization has been determined (Wahli and Dawid, 1979; Wahli *et al.*, 1980, 1982; Germond *et al.*, 1983; Arnberg *et al.*, 1981; Wilks *et al.*, 1981). We now present a comparison of the 5′ end region of a *X. laevis* and of a chicken vitellogenin gene and discuss observed homologies and differences.

Results

Mapping of the transcription initiation site of the A2 gene of X. laevis

The position of the putative 5′ end of the vitellogenin A2 gene has been determined by electron microscopic analysis of hybrids between vitellogenin mRNA and cloned A2 gene sequences (Wahli *et al.*, 1980). To map the position of the transcription initiation site more precisely, an *Eco*RI DNA fragment of 2.3 kb containing the putative 5′ end was subcloned in pBR322 (Gerber-Huber *et al.*, 1981), isolated and tested for transcription in a HeLa whole-cell extract (Manley *et al.*, 1980). The run-off transcripts from the entire or truncated DNA fragment were analysed by gel electrophoresis and the position of the *in vitro* initiation site of the A2 gene was determined (Figure 1A, lanes a and b). When the entire chimeric plasmid, that is the *Xenopus* 2.3-kb *Eco*RI fragment joined to the pBR322 plasmid, was digested and tested for transcription in the cell-free extract, three additional initiation sites were found (Figure 1, lane d), which all map in pBR322 sequences (lane e). Others have reported similar observations concerning transcription initiation events within pBR322 sequences (Sassone-Corsi *et al.*, 1981). Fine mapping of the 5′ end of *in vitro* transcripts was performed by S1 nuclease analysis (Figure 1B). A 5′ end-labelled DNA fragment containing the putative initiation site was used as probe and hybridized with the *in vitro* transcripts. After S1 nuclease digestion a protected fragment of 135 bases was detected. Based on this result, *in vitro* transcription starts at the T or G within the sequence TTCAGTGTTCAC (Figures 1B and 7).

Next, we tested if this initiation site corresponds to the one used *in vivo*. The DNA probe 5′ end-labelled at the *Sma*I site located 450 bp downstream from the *in vitro* initiation site was hybridized to either purified vitellogenin mRNA (Figure 2A, lane a) or to nuclear RNA from hepatocytes producing vitellogenin (lane b). Nuclear RNA from hepatocytes which do not produce vitellogenin or tRNA were included as controls (Figure 2A, lanes c and d). The RNA-DNA hybrids were treated with S1 nuclease and the size of the protected DNA fragments were measured by gel electrophoresis. As expected, the 450-bp digestion-resistant hybrid was only found with

*To whom reprint requests should be sent.

© IRL Press Limited, Oxford, England.

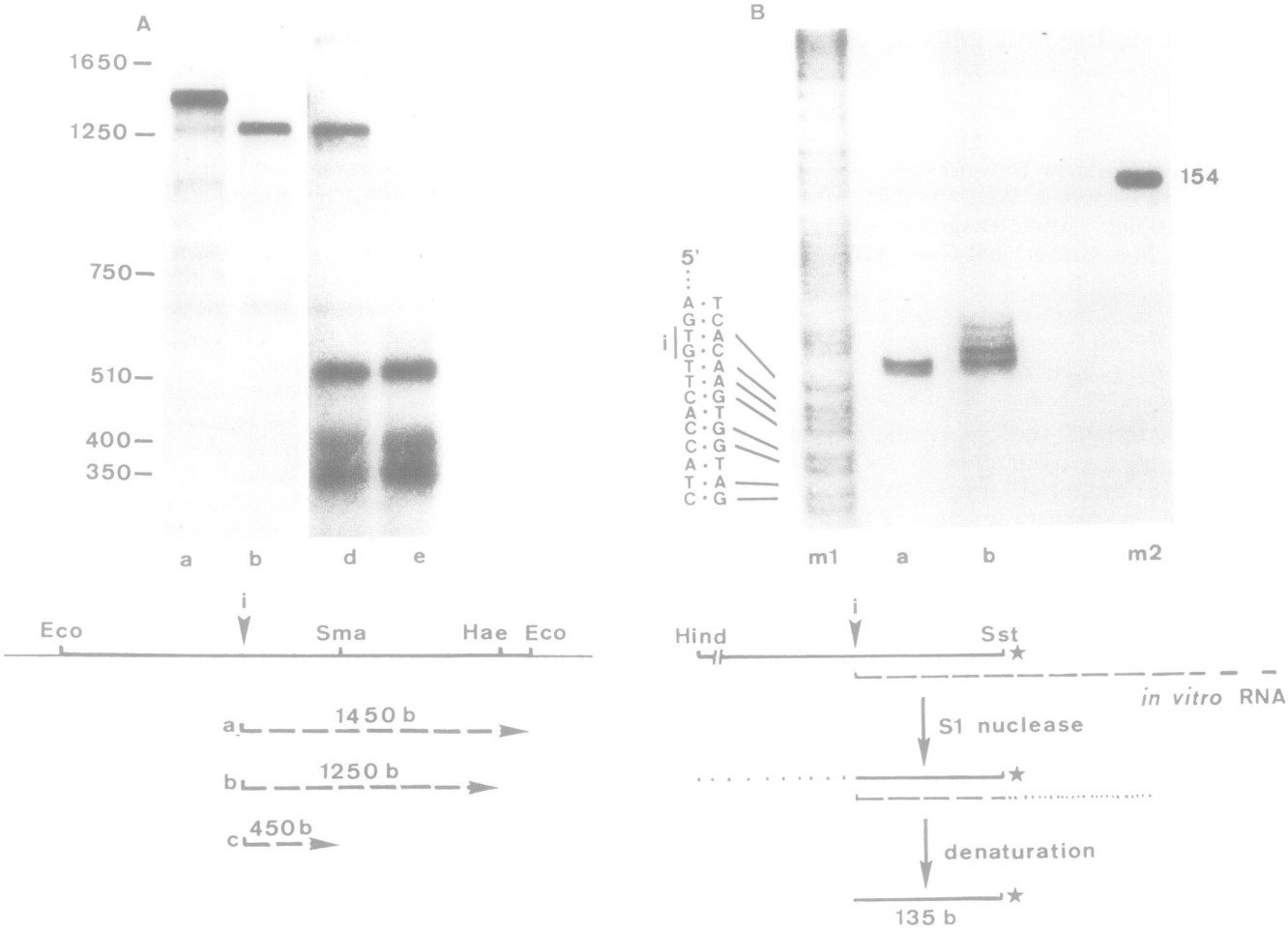

Fig. 1. Mapping of the *in vitro* transcription initiation site of the *Xenopus* vitellogenin A2 gene. **(A)** Electrophoretic analysis of *in vitro* transcription products. Transcription of the isolated 2.3-kb *Eco*RI fragment containing the putative 5′ end of the gene **(a)**; same fragment truncated in 3′ with *Hae*III **(b)** or *Sma*I (data not shown, see **c** in scheme); same fragment in pBR322, hybrid plasmid digested with *Hae*III **(d)**; pBR322 digested with *Hae*III **(e)**. The size markers are given in bases. The scheme represents the run-off transcripts and gives the position of the initiation site (i) of *in vitro* transcription. (b = bases.) **(B)** Nuclease S1 mapping of the *in vitro* transcription initiation site. The hybridization probe was a 970-bp *Hind*III-*Sst*I fragment 5′ end-labelled at the *Sst*I site. *In vitro* synthesized RNA was mixed with 5 μg tRNA, treated with DNase to eliminate the template and hybridized to the probe. The DNA-RNA hybrids were treated with 15 U **(a)** or 1 U **(b)** of S1 nuclease. The end-labelled probe was sequenced according to Maxam and Gilbert (1980) and the A+G track is shown **(m1)**. The position of a size marker of 154 bases is given in **m2**. The scheme represents the different steps of the nuclease S1 experiment and gives the position of the transcription initiation site (i). (b = bases.) In positioning the initiation site on the A+G track a correction of 1.5 bases has been made because of the elimination of the terminal nucleotide during sequencing (Moss and Birnstiel, 1979; Sollner-Webb and Reeder, 1979).

nuclear RNA from estrogen-stimulated hepatocytes (Figure 2A, lane b, arrow). Hybrids between the same probe and purified vitellogenin mRNA yielded S1-resistant fragments of ~85 bp and 100 bp (Figure 2A, lane a). While the shorter band corresponds to the distance between the 5′ end of exon 3 and the *Sma*I site within this exon (84 bp, arrow in lane a) the resistant hybrids of ~100 bp are not explained and might represent partially S1 nuclease-digested material. As seen in lane b of Figure 2A, these two bands can also be detected with nuclear RNA. They probably originate from transcripts which have already undergone processing.

To map the 5′ end of *in vivo* transcripts, the same experimental design as shown in Figure 1B was used for nuclear RNA from estrogen-stimulated hepatocytes (Figure 2B, lanes b, c and d) or from control hepatocytes (lane a). The results show that the initiation sites used *in vivo* and *in vitro* coincide precisely.

Sequence of the 5′ end of the Xenopus vitellogenin A2 mRNA

The primary sequence of the 5′ end region of the vitellogenin A2 mRNA was determined using the dideoxynucleotide method of RNA sequencing. A DNA primer of 44 bp isolated from the putative third exon of the A2 gene was denatured and hybridized to purified vitellogenin mRNA and elongated with reverse transcriptase in four different reactions each containing a different chain terminator (Figure 3). The sequence for this part of the A2 mRNA agrees with the corresponding gene sequence (see below). Comparison of these two sequences accurately defines the structural organization of this gene region: exon 1 has a length of 53-54 bp, intron 1 of 213 bp, exon 2 of 21 bp, intron 2 of 80 bp and exon 3 of 152 bp (Table I and Figure 7). The 3′ end of exon 3 was tentatively determined earlier by electron

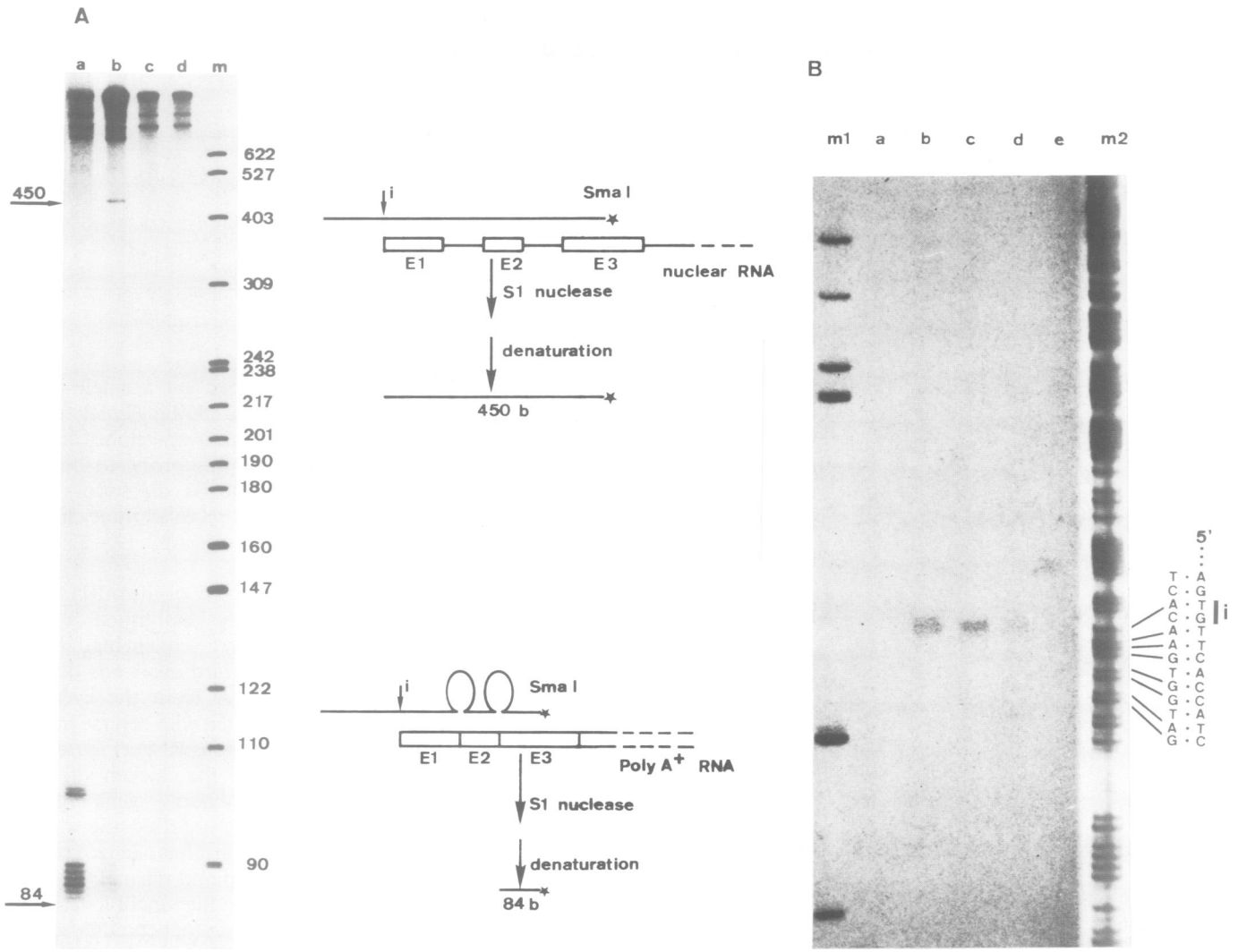

Fig. 2. Nuclease S1 mapping of the *in vivo* transcription initiation site of the *Xenopus* vitellogenin A2 gene. **(A)** A 5′ end-labelled probe (*Sma*I site, see scheme) containing the putative 5′ end of the A2 gene was hybridized to purified vitellogenin mRNA **(a)**, to nuclear RNA from the liver of an estrogen treated female **(b)** or of a control male **(c)** and to tRNA **(d)**. The DNA-RNA hybrids were treated with S1 nuclease. Size markers are given in bases **(m)**. The scheme gives the design of the experiment and the expected results. **(B)** A 970-bp *Hind*III-*Sst*I DNA fragment, 5′ end-labelled at the *Sst*I site (see scheme of Figure 1B) was hybridized to nuclear RNA from the liver of a control male **(a)**, of an estrogen-treated female **(b,c,d)** or to tRNA **(e)**. The DNA-RNA hybrids were treated with S1 nuclease for 30 min: 280 Units in **(a)**, **(c)** and **(e)** and 25 Units in **(b)**. In **(d)** S1 nuclease was added in three 95 Unit aliquots at times 0 min, 30 min and 90 min; the total digestion time was 120 min. The end-labelled probe was sequenced (Maxam and Gilbert, 1980), the G+A track is shown **(m2)** and the initiation site (i) is indicated. Size markers are given in bases **(m1)**. The scheme of Figure 1B applies to this experiment with the difference that in this case, *in vivo* synthesized nuclear RNA was used.

microscopy of RNA-DNA hybrids (Wahli *et al.*, 1980). The precise 3′ boundary of this exon is indicated by the presence of a sequence typical for splice donor sites (Mount, 1982). Furthermore, this putative donor site is identical to the two first sites over seven nucleotides, namely GGTAAGT. The position of the transcription initiation site determined by S1 nuclease experiments was confirmed by the size of the longest fragments synthesized in this primer extension experiment.

Mapping and sequencing of the 5′ end of the chicken vitellogenin gene

The position of the putative 5′ end of the chicken vitellogenin gene has been determined earlier by electron microscopy and Southern blot analysis of cloned gene sequences (Arnberg *et al.*, 1981). To map the position of the 5′ end more precisely, a 3.6-kb *Bam*HI fragment containing the putative transcription initiation site was subcloned in pBR322. Electron microscopic analysis of hybrids between the 3.6-kb *Bam*HI fragment and vitellogenin mRNA showed that this fragment is made up of 1.7 kb of flanking DNA, the first three exons and introns, and a short stretch ($\leq$50 nucleotides) of exon 4 (Arnberg, in Meijlink, 1983). Based on these data, a DNA restriction fragment that was expected to contain the 5′ end region of the gene was isolated and sequenced. Inspection of this sequence for possible splice junctions (Mount, 1982) yielded an exon-intron structure which is in good agreement with the data obtained by electron microscopy.

The 5′ end of exon 1 was mapped by nuclease S1 analysis using a probe end-labelled at the *Sau*3AI site located within the putative first exon. The labelled probe was hybridized to polysomal RNA from the liver of an estradiol-treated rooster (AB *et al.*, 1976). A cluster of nuclease S1-resistant fragments of 23–26 nucleotides was found which maps the 5′ end of exon 1 at the Ts and the A in the sequence CCCTATTC (Figures 4 and 7). From this experiment we cannot conclude

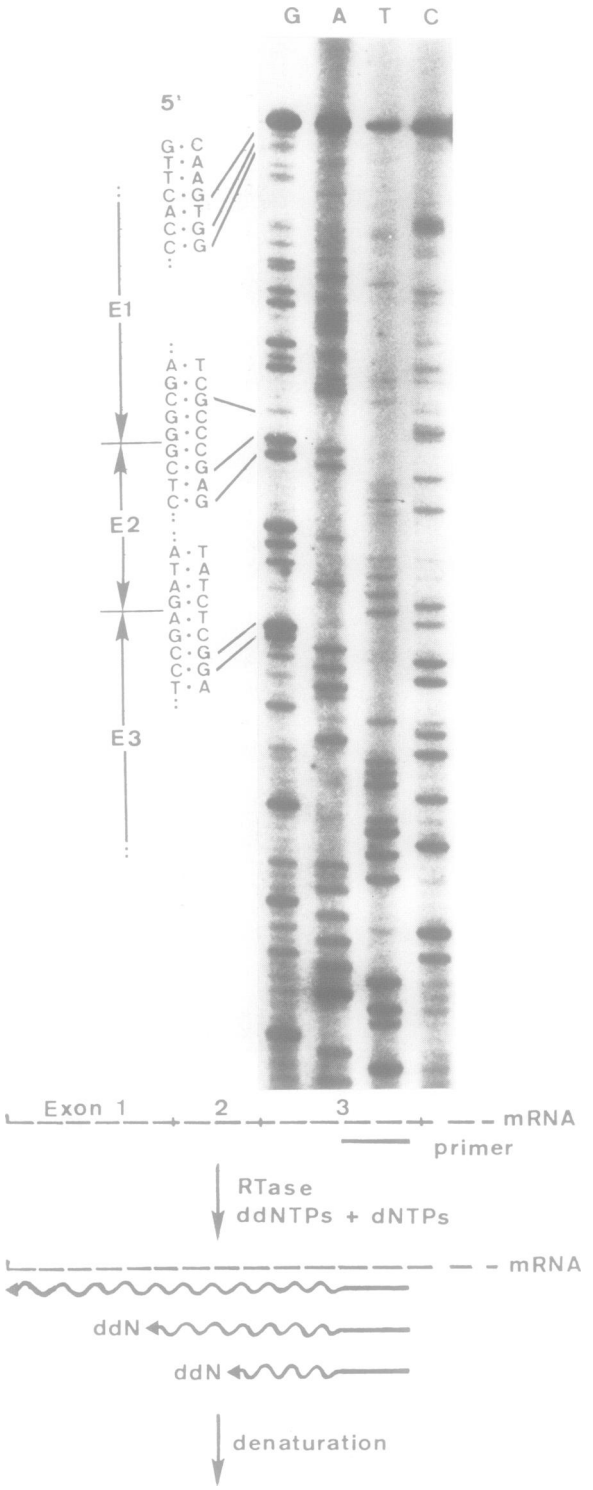

Fig. 3. Determination of the 5′ end sequence of the *Xenopus* vitellogenin A2 mRNA using the dideoxynucleotide method adapted to RNA sequencing. Reverse transcriptase elongation of a 44 base primer originating from the putative third exon hybridized to purified vitellogenin mRNA is shown. Four reactions were carried out each containing a different dideoxynucleotide: ddGTP (G), ddATP (A), ddTTP (T) and ddCTP (C). The exon-intron boundaries are indicated, based on the comparison with the DNA sequence of the 5′ end region of the A2 gene.

whether this heterogeneity reflects mRNAs with different 5′ ends or is due to a nuclease S1 digestion artifact (see below).

The size of the first three exons was determined by using a uniformly labelled probe synthesized on a single-stranded DNA template. For this purpose, a 550-bp *Eco*RI fragment that spans the exon 1 to exon 3 region was cloned into M13 mp9 RF DNA. Hybridization of this probe to polysomal RNA was followed by a nuclease S1 digestion. Electrophoresis of the protected fragments revealed three clusters of bands of the expected sizes: 45 – 54 bp, 19 – 27 bp and ~150 bp (Figure 5).

Table I. Structural organization of the 5′ end of the *Xenopus* and the chicken vitellogenin gene

	No. of nucleotides		Homology (%)
	Xenopus A2	Chicken	
Exon 1	53 (40)[a]	53 (40)[a]	66
Exon 2	21	21	48
Exon 3	152	152	60
Intron 1	213	115	–[b]
Intron 2	80	100	–[b]

[a]In parentheses: coding part of exon 1.
[b]No significant sequence homology detected.

A more precise location of the 5′ end of the mRNA and of the splice junctions was definitely established using the dideoxynucleotide method of RNA sequencing. A 90-bp primer isolated from the putative third exon was denatured and hybridized to polysomal RNA from the liver of an estradiol-treated rooster and elongated with reverse transcriptase as described above for the *Xenopus* RNA (Figure 6). The nucleotide sequence of this part of the chicken vitellogenin mRNA agrees with the corresponding gene sequence. Together these sequences accurately define the structural organization of this gene region: exon 1 has a length of 53 bp, intron 1 of 115 bp, exon 2 of 21 bp and intron 2 of 100 bp. Since the primer used is very close to the junction between intron 2 and exon 3, the first nucleotide that could be read was in exon 2. However, the length of the extended DNA chain allowed us to determine the exact location of the boundaries of intron 2. The size of exon 3, 152 bp, was tentatively deduced by attributing its 3′ end to a specific splice-junction donor sequence. It is in agreement with earlier data obtained by electron microscopy (Arnberg *et al.*, 1981).

Comparison of the fine structure of the Xenopus 5′ end gene region with its homologous chicken gene region

The nucleotide sequences of DNA fragments containing the *Xenopus* and chicken 5′ end vitellogenin gene region were determined (Sanger *et al.*, 1977; Maxam and Gilbert, 1980) and are compared in Figure 7. This comparison covers ~1 kb of homologous sequences representing ~500 bp of 5′-flanking region, the first three exons and the first two introns. In both *Xenopus* and chicken, the three homologous exons have exactly the same size (Table I).

The sequence homology between the two exon 1 is considerable (66%). In both species this exon is composed of 13 bp of non-coding sequences and of a 40-bp coding block, assuming that the first AUG after the initiation site of transcription serves as translation initiation codon. This first block encodes a very apolar peptide sequence (Figure 7 and Table II). Therefore it may represent part or all of the signal peptide of the vitellogenin protein. Because four out of the five amino acid substitutions observed are conservative the matching of the amino acids with identical properties is very good in this part of the protein (92%). Tables I and II demon-

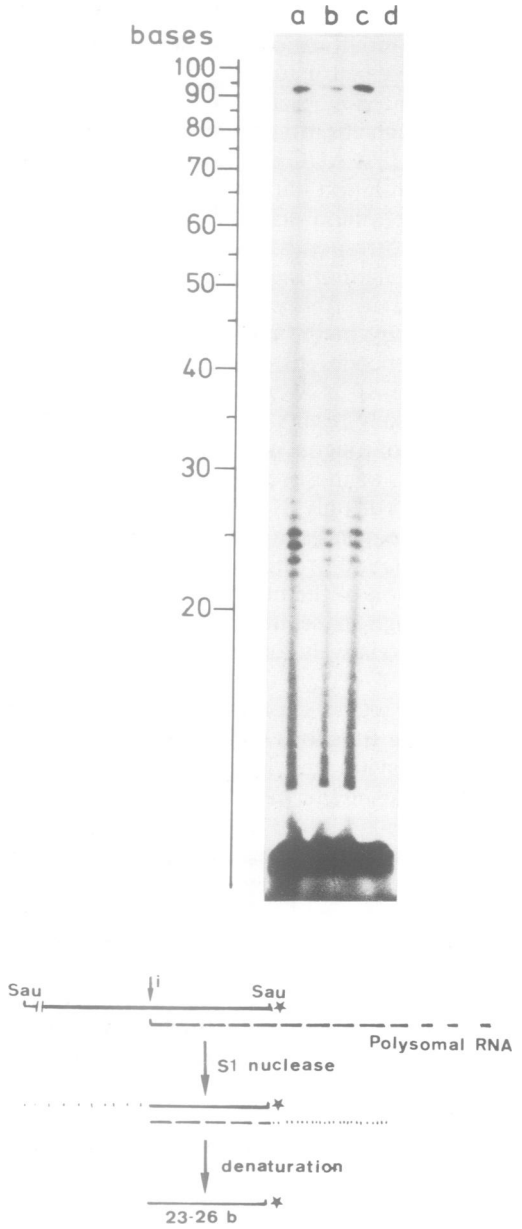

Fig. 4. Nuclease S1 mapping of the 5′ end of chicken vitellogenin mRNA. A 5′ end-labelled probe (*Sau*3AI) containing the putative 5′ end of the gene was hybridized to polysomal RNA from the liver of an estradiol-treated rooster **(a,b,c)** or to total cellular RNA from the liver of a control rooster **(d)**. The DNA-RNA hybrids were treated with 400 U **(a)**, 200 U **(b,d)** and 100 U **(c)** S1 nuclease at 37°C for 30 min. The scheme gives the design of the experiment and represents the obtained results.

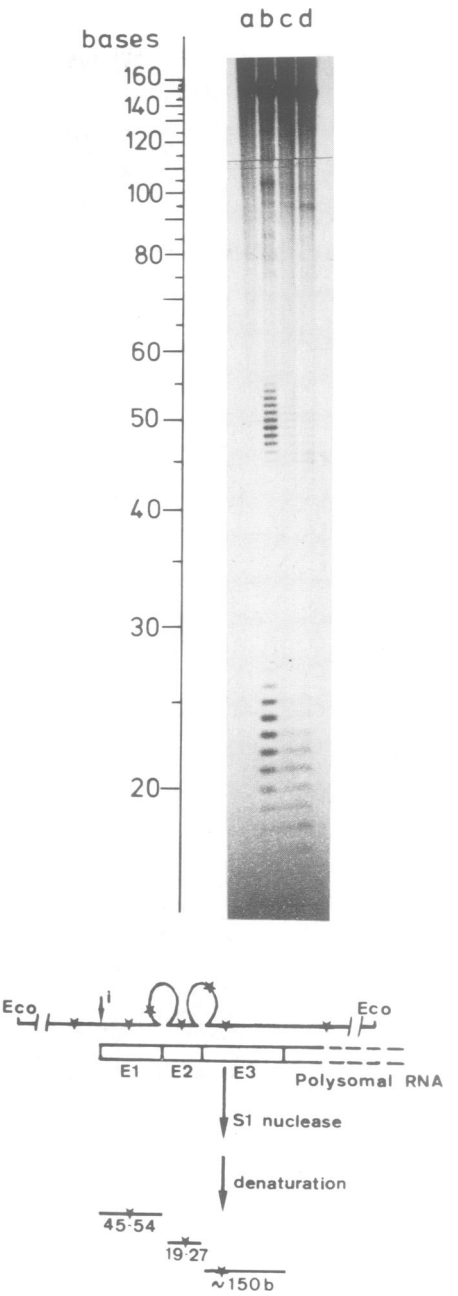

Fig. 5. Nuclease S1 mapping of exons 1, 2 and 3 of the chicken vitellogenin gene. A labelled probe was synthesized on a single-stranded M13 template containing a genomic 550 nucleotide *Eco*RI fragment overlapping exons 1 to 3 and hybridized to different RNAs: total cellular RNA from control rooster liver **(a)**, polysomal RNA from the liver of an estradiol-treated rooster **(b, c** and **d)**. Nuclease S1 amounts were 400 U in **(b)**, 200 U in **(a,c)** and 100 U in **(d)**. The scheme gives the design of the experiment and represents the obtained results.

strate that exon 2 and exon 3 are less conserved than exon 1 and that the part of the protein encoded by these two exons is more polar. Nonetheless, the overall nucleotide homology is 60% in the first three exons. While the amino acid homology is 43%, the amino acid side chain similarity amounts to 73%.

In contrast to the exons, the homologous introns in the 5′ end gene region have different lengths. Except for the splice junctions no significant sequence homology is observed. The 5′ end flanking regions are presented in Figure 7 after optimizing the alignment of the *Xenopus* and chicken sequences. Several blocks with >70% homology can be observed. The most proximal block to the transcription initiation site is the ATA homology (TATA box) common to almost all eucaryotic protein-coding genes. Up to position −110 there are three other blocks and, finally, two additional homology regions are observed further upstream. The block located around position −324 in *Xenopus* and −343 in chicken is especially interesting by its palindromic structure (Figure 7). The sequences between the blocks of homology show no more similarity than that found between random sequences.

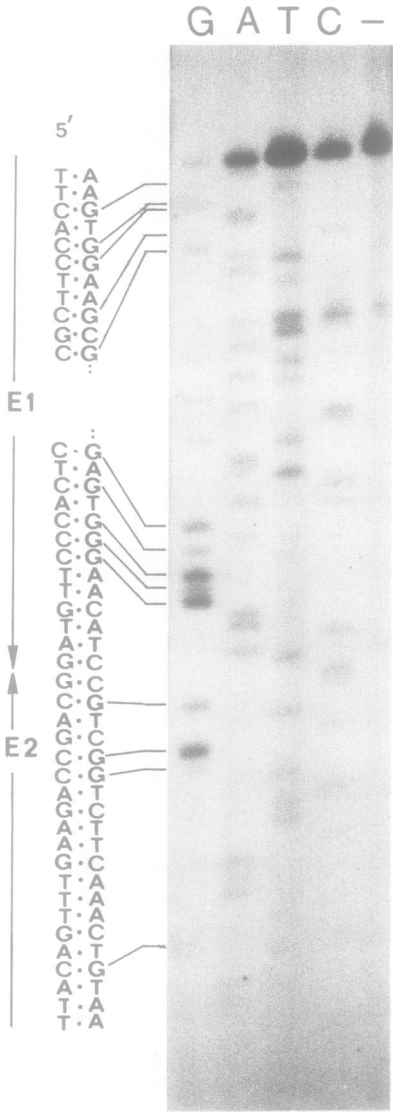

Fig. 6. Determination of the 5′ end of chicken vitellogenin mRNA using the dideoxy method adapted to RNA sequencing. Reverse transcriptase elongation of a 90 base primer (*Bst*NI-*Hha*I) originating from the putative third exon hybridized to polysomal RNA is shown. Four reactions were carried out each containing a different dideoxynucleotide: ddGTP (G), ddATP (A), ddTTP (T) and ddCTP (C). A fifth reaction was without dideoxynucleotides (–). The scheme of Figure 3 applies to this experiment since it was performed with a similar design (see text).

Discussion

Here we compare the 5′ end regions of vitellogenin genes from *Xenopus* and chicken. The comparison is especially worthwhile because these genes are estrogen-regulated in both amphibians and birds. The two gene sequences are very similar immediately downstream of the transcription initiation site. If the first AUG is the translation initiation codon, the untranslated part of both *Xenopus* and chicken mRNA is only 13 nucleotides long, nine of which are conserved. Compared to most eucaryotic mRNAs it is relatively short. This region can base-pair with the 3′ end of 18S rRNA (Hagenbuechle *et al.*, 1978) which may explain the sequence conservation observed between *Xenopus* and chicken.

The amino acid sequence of vitellogenin is not yet available. Due to the heterogeneity of the protein resulting from multiple genes in both species (Wahli *et al.*, 1979; Wang and Williams, 1980; Wiley and Wallace, 1981) and heavy post-translational modifications (Wallace, 1978) a detailed analysis of the protein structure will be difficult. At the level of the DNA sequence we observe an open reading frame of 71 amino acids through the first three exons in both *Xenopus* and chicken if the first AUG downstream of the mRNA capping site is the translation initiation codon. It appears likely, therefore, that the amino acid sequence deduced from the DNA sequence corresponds indeed to that of authentic vitellogenin. The length of the first three homologous coding blocks is identical in *Xenopus* and chicken. The highest nucleotide homology in the coding regions is found in the first exon which encodes 13 amino acids, 11 of them with non-polar side chains in *Xenopus* and 10 in chicken. This homology most likely results from the pressure to maintain the characteristics of a signal peptide which has been found at the amino terminus of most secreted proteins. It is striking to find a second exon of only 21 bp surrounded by 293 bp and 215 bp of intron sequences in *Xenopus* and chicken, respectively. The contribution of this short region may be essential for the conformation of the protein since it is present in both species, even though its sequence is less conserved. Length conservation, but relatively low sequence homology are also observed in exon 3 except for a few short blocks. Since vitellogenin can be regarded as storage protein, sequences not involved in specific functions such as uptake by the oocytes, cleavage into phosvitin and lipovitellin, formation of yolk platelets or granules might have a high tolerance for mutations.

The splice junction sequences of the 5′ end region of the two genes conform to the now well-defined consensus sequences (Mount, 1982) and therefore represent locations of high homology between the two genes. As expected from the studies of the *Xenopus* vitellogenin gene family (Wahli *et al.*, 1980; Germond *et al.*, 1983) or of other gene families (Konkel *et al.*, 1978; Lomedico *et al.*, 1979), homologous introns have less similarity than the coding sequences. Our sequence analysis of two *Xenopus* and chicken introns show that they compare like random sequences.

The alignment of the two homologous 5′ end flanking sequences reveals several blocks of significant homology of $>70\%$. The homologous stretches occur in the same order but the sequences separating them vary somewhat in length in the two genes. It appears likely that such conserved sequences are involved in the control of the transcription machinery. One can conceive different functions: multiple contacts made by RNA polymerase II, recognition specificity for auxiliary transcription factors or information capable of marking the chromatin structure around the 5′ end terminus of the gene (McKnight, 1982). In the vitellogenin system one also has to consider sequences that are involved in hormonal control of gene expression. It is now of interest to look for identical conserved regions in other vitellogenin genes. It will be important to test experimentally the functional significance of such potential regulator sequences, especially the conserved palindromic structure.

Materials and methods

In vitro transcription assay

The *in vitro* transcription reactions were carried out as described recently (Wahli *et al.*, 1982) in whole-cell extracts from HeLa cells prepared according to Manley *et al.*(1980).

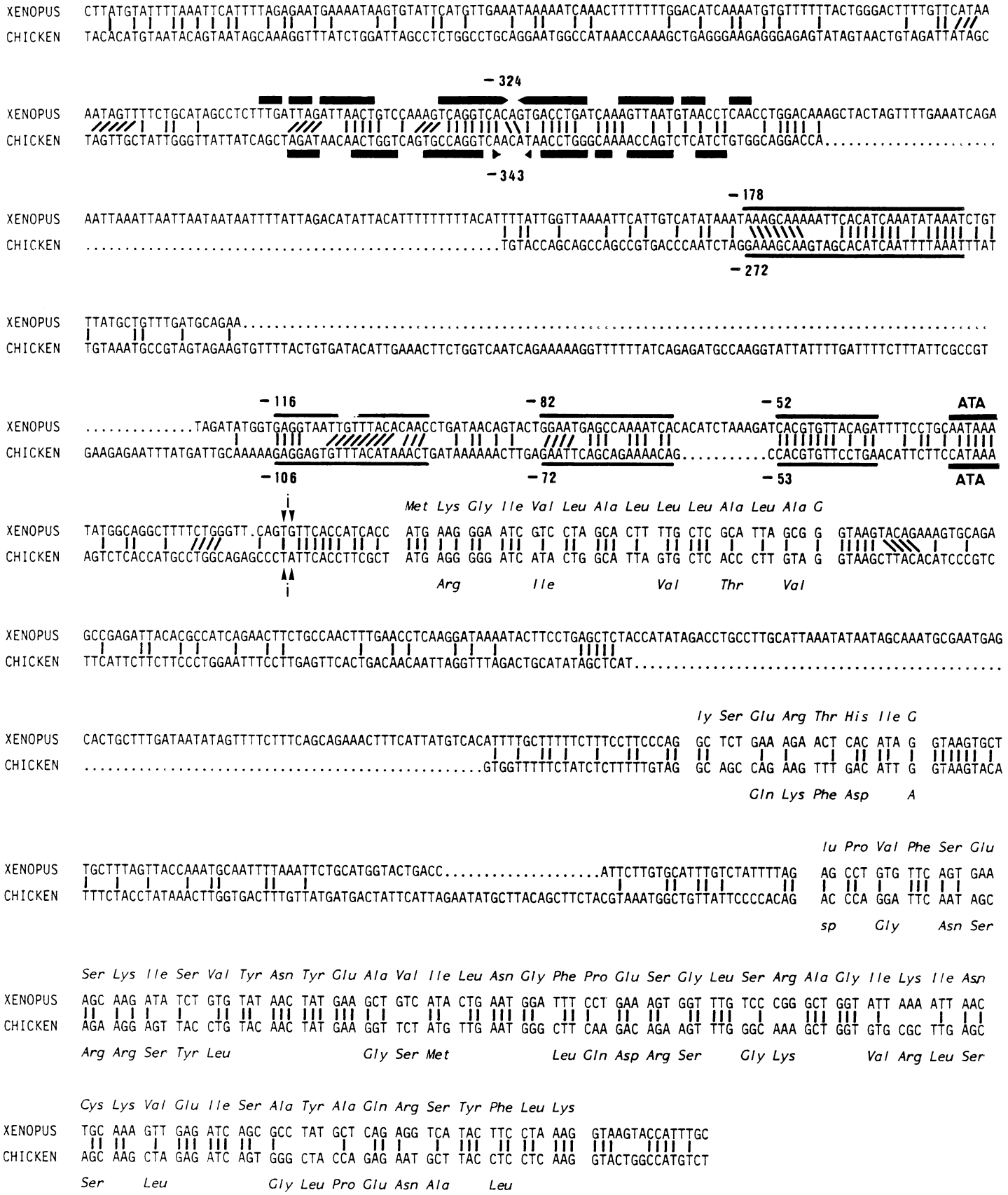

Fig. 7. Comparative display of the nucleotide sequence of the 5′ end region of the *X. laevis* A2 and of the chicken vitellogenin gene. The nucleotide sequence of the non-coding strand (mRNA strand) is given from the 5′ to the 3′ direction. The putative amino acid sequence of part of the *Xenopus* vitellogenin is given above the coding sequence. Amino acid substitutions in the chicken sequence are indicated under the chicken nucleotide sequence. Blocks of homology in the 5′ end flanking region are indicated by thin horizontal lines. The region of homology presenting palindromic structure is indicated with thick lines. (i) with arrows indicates the transcription initiation site. Dots were inserted to achieve optimal alignment of the sequences.

Table II. Side chain representation in the amino acids of the first three exons of the *Xenopus* and chicken vitellogenin genes

	No. of amino acids	Side chain				Homology	
		acidic	basic	uncharged polar	non-polar	identical amino acids (%)	similar amino acids (%)
		(% in parentheses)					
Exon 1							
Xenopus	13	–	1 (8)	1 (8)	11(84)	62	92
chicken	13	–	1 (8)	2(15)	10(77)		
Exon 2							
Xenopus	7	1(14)	2(29)	3(43)	1(14)	43	57
chicken	7	1(14)	1(14)	3(43)	2(29)		
Exon 3							
Xenopus	51	5(10)	6(12)	19(37)	21(41)	37	73
chicken	51	5(10)	7(14)	22(43)	17(33)		

Preparation of nuclear and cytoplasmic RNA

Poly(A)-containing RNA was purified from the liver of estrogen-treated *X. laevis* females as described earlier (Wahli *et al.*, 1976; Ryffel *et al.*, 1977). Nuclear RNA from livers of *X. laevis* males and estrogen treated females was prepared as described (Germond *et al.*, 1983). Total polysomal RNA from livers of estrogen-treated roosters was prepared as described by AB *et al.* (1976). Total cellular RNA from control rooster liver was isolated according to Auffray and Rougeon (1980).

S1 nuclease analysis

S1 mapping with *Xenopus* DNA fragments was carried out according to the procedure of Berk and Sharp (1977) as modified by Weaver and Weissmann (1979). 5′ end-labelled DNA probes were prepared from DNA fragments treated with calf intestinal phosphatase and then labelled with [γ-^{32}P]ATP by polynucleotide kinase (Maniatis *et al.*, 1975). Between 10 000 and 30 000 c.p.m. of the end-labelled DNA probe ($\sim 10^7$ c.p.m./pmol of ends) were co-precipitated with 0.1 – 5 μg of RNA, and tRNA to 10 μg. The precipitated nucleic acids were resuspended in 80% formamide, 400 mM NaCl, 1 mM EDTA, 40 mM Pipes (pH 6.4), heated at 70°C for 3 min and hybridized at 48°C for 12 – 15 h. Samples were then diluted 20 times with cold S1 nuclease buffer containing 300 mM NaCl, 4.5 mM $ZnSO_4$, 30 mM sodium acetate (pH 4.6), 20 μg/ml single-stranded salmon sperm DNA and S1 nuclease (see figure legends for the amount of Units used; 1 Unit digests 1 μg of single-stranded DNA in 30 min; Vogt, 1980). After digestion for 1 h at 30°C, the digestion products were directly precipitated with ethanol, resuspended and analysed on 6% or 8% acrylamide sequencing gels (Sanger and Coulson, 1978). For the chicken gene, S1 nuclease mapping was performed essentially as described for *Xenopus*.

DNA sequencing

The DNA sequencing strategy described by Messing *et al.* (1981) was applied to sequence the 5′ end of the *Xenopus* and chicken vitellogenin genes. The phage M13 mp7 was used as vector for *Xenopus* gene fragments and the phages M13 mp8 and 9 for chicken gene fragments. Sequencing was done by the chain termination method (Sanger *et al.*, 1977). Stretches of the *Xenopus* DNA sequence given in Figure 7 were also sequenced according to Maxam and Gilbert (1980). The sequences were analysed using a computer program prepared by Staden (1982a, 1982b).

mRNA sequencing procedure

The sequencing of *Xenopus* vitellogenin A2 mRNA was performed essentially as described by Hamlyn *et al.* (1978). Using a specific primer (see text), mRNA was transcribed with reverse transcriptase (Life Sciences, St. Petersbourg) in chain termination reactions. 0.14 pmol (4 ng) of primer were denatured at 100°C for 1 min in H_2O. 0.4 pmol (1 μg) of purified vitellogenin mRNA and buffer at 1.6 x final concentration (see below) were added. Hybridization was carried out at 65°C for 30 min. The 6.5 U of reverse transcriptase, cold deoxynucleotide triphosphates (except dATP) to a final concentration of 14 μM each and 10 μCi [α-^{32}P]dATP (400 Ci/mmol, final concentration of 0.8 μM) were added to the mixture. This mixture was distributed into five tubes. To four of them dideoxynucleotide triphosphates were added to final concentrations of 6.25 μM ddATP, 25 μM ddGTP, 25 μM ddCTP and 50 μM ddTTP, respectively. The dATP concentration in the A reaction was increased by adding another 5 μCi of [α-^{32}P]dATP. The fifth tube served as a control reaction for non-specific stops (results not shown). Reverse transcription was performed in 6 μl volumes at 42°C for 10 min [final concentrations of the buffer: 140 mM KCl, 10 mM $MgCl_2$, 20 mM β-mercaptoethanol and 100 mM Tris-HCl (pH 8.3)]. After raising all dNTP concentrations to 100 – 150 μM, synthesis was continued for another 10 min, in order to minimize non-specific stops due to low dNTP concentrations. The reactions were stopped by addition of 1.5 volumes of deionized formamide, containing 9 mM EDTA, 0.09% xylene cyanol and 0.06% bromphenol blue and analyzed on a 6% sequencing gel (Sanger and Coulson, 1978).

The 5′ end of the chicken vitellogenin mRNA was sequenced essentially as described for *Xenopus* with the following modifications: the buffer was 10 mM $MgCl_2$, 7.5 mM KCl, 8 mM dithiothreitol, 50 μg/ml actinomycin D and 50 mM Tris-HCl (pH 8.4). 13 μM ddGTP, 13 μM ddATP, 13 μM ddTTP, 2.5 μM ddCTP and 66 μM dGTP, 66 μM dATP, 66 μM dTTP, 22 μM dCTP and 50 μCi [α-^{32}P]dCTP were used per reaction (5 μl). 2 pmol of primer were denatured at 100°C for 3 min and hybridized with 7 μg polysomal RNA containing ~0.02 pmol of vitellogenin mRNA, by cooling the mixture down to room temperature within 30 min. 0.5 U of reverse transcriptase were used per reaction.

Acknowledgements

We wish to thank S.Gerber-Huber, O.Hagenbuechle and M.Katzenmeier for gifts of materials, S.Gerber-Huber, O.Hagenbuechle, G.Ryffel, U.Schibler and R.Wittek for helpful comments on the manuscript. We are also grateful to R.Staden for the computer program to analyse DNA sequences and to L.Vuillet for adapting it to our system, to B.ten Heggeler, N.Buechi and H.Pagel for help in the preparation of the manuscript. This research was supported by the Swiss National Science Foundation, the Etat de Vaud and the Netherlands Foundation for Chemical Research (S.O.N.) with financial aid from the Netherlands Organization for the Advancement of Pure Research.

References

AB,G., Roskam,W.G., Dijkstra,J., Mulder,J., Willems,M., van der Ende,A. and Gruber,M. (1976) *Biochim. Biophys. Acta,* **454**, 67-78.
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.
Auffray,C. and Rougeon,F. (1980) *Eur. J. Biochem.,* **107**, 303-314.
Benoist,C. and Chambon,P. (1981) *Nature,* **290**, 304-310.
Berk,A.J. and Sharp,P.A. (1977) *Cell,* **12**, 721-732.
Brock,M.L. and Shapiro,D.J. (1983) *J. Biol. Chem.,* **258**, 5449-5455.
Buetti,E. and Diggelmann,H. (1981) *Cell,* **23**, 335-345.
Dierks,P., van Ooyen,A., Mantei,N. and Weissmann,C. (1981a) *Proc. Natl. Acad. Sci. USA,* **78**, 1411-1415.
Dierks,P., Wieringa,B., Marti,D., Reiser,J., van Ooyen,A., Meyer,F., Weber,H. and Weissmann,C. (1981b) in Brown,D.D., Fox,C.F. (eds.), *Developmental Biology Using Purified Genes, ICN-UCLA Symposium on Molecular and Cellular Biology,* Academic Press, NY, pp. 347-366.
Dierks,P., van Ooyen,A., Cochran,M.D., Dobkin,C., Reiser,J. and Weissman,C. (1983) *Cell,* **32**, 695-706.
Doehmer,J., Baringa,M., Vale,W., Rosenfeld,M.G., Verma,I.M. and Evans, R.M. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 2268-2272.
Fromm,M. and Berg,P. (1982) *J. Mol. Appl. Genet.,* **1**, 457-481.
Gerber-Huber,S., Felber,B.K., Weber,R. and Ryffel,G.U. (1981) *Nucleic Acids Res.,* **9**, 2475-2494.
Germond,J.E., ten Heggeler,B., Schubiger,J.L., Walker,P., Westley,B. and Wahli,W. (1983) *Nucleic Acids Res.,* **11**, 2979-2997.
Grosschedl,R. and Birnstiel,M.L. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 7102-7106.
Grosveld,G.C., de Boer,E., Shewmaker,C.K. and Flavell,R.A. (1982a)

Nature, **295**, 120-126.
Grosveld,G.C., Rosenthal,A. and Flavell,R.A. (1982b) *Nucleic Acids Res.,* **10**, 4951-4971.
Hagenbuechle,O., Santer,M., Steitz,J.A. and Mans,R.S. (1978) *Cell,* **13**, 551-564.
Hamlyn,P.H., Brownlee,G.G., Cheng,C.-C., Gait,M.J. and Milstein,C. (1978) *Cell,* **15**, 1067-1075.
Hynes,N.E., Kennedy,N., Rahmsdorf,V. and Groner,B. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 2038-2042.
Konkel,D.A., Tilghman,S.M. and Leder,P. (1978) *Cell,* **15**, 1125-1132.
Kurtz,D.T. (1981) *Nature,* **291**, 629-631.
Lomedico,P., Rosenthal,N., Efstratradis,A., Gilbert,W., Kolodner,R. and Tizard,R. (1979) *Cell,* **18**, 545-558.
Maniatis,T., Jeffrey,A. and Kleid,D.G. (1975) *Proc. Natl. Acad. Sci. USA,* **72**, 1184-1188.
Manley,J.L., Fire,A., Cano,A., Sharp,A.P. and Gefter,M.L. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 3855-3859.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.,* **65**, 499-560.
McKnight,S.L. (1982) *Cell,* **31**, 355-365.
McKnight,S.L. and Kingsbury,R. (1982) *Science (Wash.),* **217**, 316-324.
McKnight,S.L., Gavis,E.R., Kingsbury,R. and Axel,R. (1981) *Cell,* **25**, 385-398.
Meijlink,F.C.P.W. (1983) Thesis, University of Groningen, pp. 50-51.
Mellon,P., Parker,V., Gluzman,Y. and Maniatis,T. (1981) *Cell,* **27**, 279-288.
Messing,J., Crea,R. and Seeburg,P.H. (1981) *Nucleic Acids Res.,* **9**, 309-321.
Moss,T. and Birnstiel,M.L. (1979) *Nucleic Acids Res.,* **6**, 3733-3743.
Mount,S.M. (1982) *Nucleic Acids Res.,* **10**, 459-472.
Page,M.J. and Parker,M.G. (1983) *Cell,* **32**, 495-502.
Pelham,H.R.B. (1982) *Cell,* **30**, 517-528.
Renkawitz,R., Beng,H., Graf,T., Matthias,P., Grez,M. and Schuetz,G. (1982) *Cell,* **31**, 167-176.
Robins,D.M., Paek,I., Seeburg,P.H. and Axel,R. (1982) *Cell,* **29**, 623-631.
Ryffel,G.U., Wahli,W. and Weber,R. (1977) *Cell,* **11**, 213-221.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA,* **74**, 5463-5467.
Sanger,F. and Coulson,A.R. (1978) *FEBS Lett.,* **87**, 107-110.
Sassone-Corsi,P., Corden,J., Kedinger,C. and Chambon,P. (1981) *Nucleic Acids Res.,* **9**, 3941-3958.
Sollner-Webb,B. and Reeder,R.H. (1979) *Cell,* **18**, 485-499.
Staden,R. (1982a) *Nucleic Acids Res.,* **10**, 2951-2961.
Staden,R. (1982b) *Nucleic Acids Res.,* **10**, 4731-4751.
Struhl,K. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 4461-4465.
Vogt,V.M. (1980) *Methods Enzymol.,* **65**, 248-255.
Wahli,W. and Dawid,I.B. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 1437-1441.
Wahli,W., Wyler,T., Weber,R. and Ryffel,G.U. (1976) *Eur. J. Biochem.,* **66**, 457-465.
Wahli,W., Dawid,I.B., Wyler,T., Jaggi,R.B., Weber,R. and Ryffel,G.U. (1979) *Cell,* **16**, 535-549.
Wahli,W., Dawid,I.B., Wyler,T., Weber,R. and Ryffel,G.U. (1980) *Cell,* **20**, 107-117.
Wahli,W., Dawid,I.B., Ryffel,G.U. and Weber,R. (1981) *Science (Wash.),* **212**, 298-304.
Wahli,W., Germond,J.E., ten Heggeler,B. and May,F.E.B. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 6832-6836.
Wallace,R.A. (1978) in Jones,R.E. (ed.), *Vertebrate Ovary,* Plenum Publishing, NY, pp. 469-502.
Wang,S.-Y. and Williams,D.L. (1980) *Biochemistry (Wash.),* **19**, 1557-1563.
Weaver,R.F. and Weissmann,C. (1979) *Nucleic Acids Res.,* **7**, 1175-1193.
Wiley,H.S. and Wallace,R.A. (1981) *J. Biol. Chem.,* **256**, 8626-8634.
Wilks,A., Cato,A.C.B., Cozens,P.J., Mattaj,I. and Jost,J.-P. (1981) *Gene,* **16**, 249-259.
Wiskocil,R., Bensky,P., Dower,W., Goldberger,R.F., Gordon,J.I. and Deeley,R.G. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 4474-4478.

Note added in proof

Recently, Geiser *et al.* (1983) *J. Biol. Chem.,* **258**, 9024-9030, have obtained the same organization of the 5′ end region of the chicken vitellogenin gene as in this article.