12. Medical Research Council, *Annual Report, 1978–1979* (London, England, 1979), p. 4.
13. M. Massenet, *Rapport sur L'Emploi Scientifique. Rapport au Premier Ministre* (La Documentation Française, Paris, 1978), p. 83.
14. Projet de Loi de Finances pour 1980, *La Recherche Scientifique et Technique en France en 1980* (Imprimerie Nationale, Paris, 1979), document annexe, p. 135.
15. *Le Monde*, 18 January 1980, p. 3.
16. Commissariat General du Plan, *Rapport de la Commission Recherche* (La Documentation Française, Paris, 1976), p. 40.
17. Natural Sciences and Engineering Research Council of Canada, *A Five Year Plan for the Programs of the Natural Sciences and Engineering Research Council* (Ottawa, November 1979), chapter 3, p. 3 and chapter 4, p. 2.
18. Science Council of Canada, *University Research in Jeopardy. The Threat of Declining Enrollment* (report 31, Ottawa, 1979).
19. National Research Council, *Nonfaculty Doctoral Research Staff in Science and Engineering* (Washington, D.C., 1978).
20. R. Radner and C. Kuh, *Preserving a Lost Generation: Policies to Assure a Steady Flow of Young Scholars Until the Year 2000. A Report and Recommendations* (Carnegie Council on Policy Studies in Higher Education, Berkeley, Calif., October 1978).
21. F. Press, "Universities in the national R & D efforts," remarks at Florida State University, Tallahassee, 7 April 1978, p. 3 (duplicated).
22. R. Atkinson, "University research and graduate education," remarks to the Annual Meeting of the Eastern Association of Graduate Schools, Albuquerque, N.M., 7 March 1977.
23. R. E. Klitgaard, *The Decline of the Best? An Analysis of the Relationships Between Declining Enrollments, Ph.D. Production and Research* (discussion paper 65D, Kennedy School of Government, Cambridge, Mass., 1979).
24. Although not precise, projections and forecasts have been used in Canada [J. Holmes, "The age structure and anticipated retirement and replacement demand for full time faculty by province and university" (Statistics Canada, Ottawa, 15 May 1978)], Germany (*6*), the United States (*5*, *20*), and the United Kingdom [Department of Education and Science, "Higher education in the 1990's" (London, February 1978); "Future trends in higher education" (London, March 1979)]. In France in 1977, the Delegation General a la Recherche Scientifique et Technique reported, "At present, personnel statistics are almost entirely lacking, and the few that are compiled do not serve the most useful purposes."
25. Committee of Vice Chancellors and Principals, "University development, selected statistical material relating to staff and students" (London, 14 September 1979), p. 2; Bundesministerium für Bildung und Wissenschaft, *Leitforstellung für die Künftige Förderung des Wissenschaftlichen Nachwuchses* (Bonn, 30 May 1979), p. C; M. Zur-Muelhlen, "The age structure of Canadian university teachers and its implications," notes prepared for the annual meeting of the Canadian Association of University Business Officers and the National Association of College and University Business Officers, Montreal, 14 to 21 July 1978.
26. European Research Foundation, *Opportunities for Scientific Talent and Mobility of Scientists in Europe* (Cologne, September 1979).
27. Supported by NSF contract PRM 7927259.

Vitellogenesis and the Vitellogenin Gene Family

Walter Wahli, Igor B. Dawid, Gerhart U. Ryffel, Rudolf Weber

One of the early steps in embryogenesis is the appearance and establishment of the primordial germ cell population. These cells then enter a complex course of successive events, eventually producing the completely differentiated gametes in the sexually mature organism (*1*). The differentiation of the oocyte itself can be divided into stages according to various morphological and biochemical characteristics. In the frog *Xenopus laevis*, the premeiotic oogonia measure 5 to 10 micrometers in diameter whereas the mature oocytes reach a size of 1.3 to 1.4 millimeters. This dramatic growth is to a large extent a result of massive deposition of yolk protein (*2*). Experimental work over the past 15 years has demonstrated that the yolk proteins that are deposited in the growing oocyte are derived from a common precursor named vitellogenin. In *Xenopus*, as in all oviparous vertebrates, vitellogenin is synthesized in the liver of the mature female under the control of estrogen, is secreted into the bloodstream, transported to the ovary, selectively taken up by the oocytes, and cleaved into the yolk proteins lipovitellin and phosvitin (*3*). In this article we review earlier findings and present some new data on yolk protein production and the genes encoding the major yolk proteins in *X. laevis*, an organism that has proved useful for studying various aspects of vitellogenesis.

Vitellogenin and Its Relation to the Yolk Proteins

Vitellogenin in *X. laevis* occurs as a dimer of two subunits of about 200,000 daltons each (*4–7*). These polypeptides represent a heterogeneous population of related molecules (see below), but it is not known whether the vitellogenin dimer comprises identical or different subunits. Vitellogenin in *X. laevis* blood consists of about 12 percent lipids, 1.5 percent phosphate, and 1 percent carbohydrate; in addition, calcium appears to be attached to the protein phosphate groups (*8*). Lipidation, phosphorylation, and glycosylation of vitellogenin occur prior to its secretion from the parenchymal liver cells, but very little is known about the mechanisms and the sites of these rapid and extensive posttranslational modifications (*9*).

Serum vitellogenin is a very stable protein with a half-life of about 40 days in males or in ovariectomized females. Its uptake by the oocytes is controlled by gonadotropins that also induce ovulation of the fully grown oocytes. Vitellogenin reaches the oocytes through the blood capillaries of the follicular theca and diffuses to the surface of the oocytes through channels between the surrounding follicle cells (*8*). The oocytes selectively take up vitellogenin by pinocytosis. This uptake is about 50 times faster as compared to that of other serum proteins (*8*, *10*). The molecular basis of this preferential uptake of a specific protein is not known but has been attributed to a vitellogenin receptor on the oocyte membrane. There is some evidence that the uptake is not strictly species-specific (*11*). Oocytes from which the follicular theca has been removed take up vitellogenin in vitro, grow at a similar rate as in vivo, reach normal size, acquire the typical pigment pattern, and initiate maturation after stimulation by progesterone (*12*). This culture method may allow studies on the regulation of oocyte growth and hormone-controlled maturation, and vitellogenin processing in the oocyte.

After its uptake by the oocyte, vitellogenin is converted into lipovitellin and phosvitin by specific proteolytic cleavages. Lipovitellin contains 22 percent lipid and is composed of two types of subunits with molecular sizes of 115,000 and 31,000 daltons. The large subunit contains almost no phosphate and the small subunit is significantly phosphory-

W. Wahli and I. B. Dawid are at the Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205. W. Wahli's present address is Institut de Biologie Animale, University de Lausanne, CH-1005 Lausanne, Switzerland. G. U. Ryffel and R. Weber are at the Department of Zoology, University of Bern, CH-3012 Bern, Switzerland.

0036-8075/81/0417-0298$01.50/0

lated. Phosvitin has a molecular size of 35,000 of which serine residues and bound phosphate represent 56 and 10 percent, respectively (*13*). In addition, small phosphoproteins of 14,000 and 19,000 daltons named phosvettes I and II have recently been identified (*14*). We discuss later the question of arrangement of the mature subunits within the vitellogenin molecule and the fact that vitellogenins of distinct structure have been identified in *X. laevis*. The cleavage products of vitellogenin are incorporated into the crystalline lattice of the yolk platelets, which account for about 90 percent of the protein content of mature oocytes (*8, 10*). The yolk proteins of the platelets do not turn over but are stored throughout oogenesis and provide a reservoir of amino acids, lipids, and inorganic phosphate for embryonic development.

Hormonal Induction of Vitellogenin Synthesis

Vitellogenin is found normally in egg-producing females but not in males. However, the liver of male *Xenopus* can be induced by estrogen to synthesize and excrete large amounts of this protein, which accumulates in the blood (Fig. 1) (*3*). This induction in male liver parenchymal cells is a useful feature for the study of the cell and molecular biology of differential gene expression as regulated by a hormone. Other steroids, such as testosterone, progesterone, cortisol, and dexamethasone do not induce the synthesis of vitellogenin (*8*). Ultrastructure analysis of hepatocytes during the induction process has revealed all the changes typical for the activation of a cell to produce and secrete large amounts of proteins (*15–17*). Vitellogenin is first detected within 12 hours after hormone administration; its synthesis increases over an 8- to 15-day period, and after 15 days it declines gradually to zero level

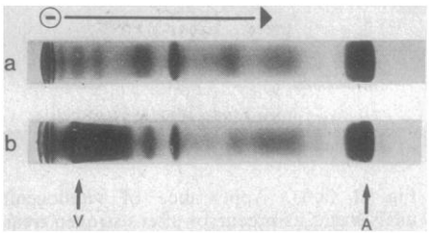


Fig. 1. Vitellogenin accumulation in the serum of an estrogen-treated male *X. laevis*. Serum samples from a control male (a) and from a male stimulated for 7 days with 1 mg of 17β-estradiol (b) were analyzed by electrophoresis in nondenaturing 5.5 percent polyacrylamide gels. The positions of vitellogenin (*V*) and albumin (*A*) are given. The arrow indicates the direction of electrophoresis.

(*5, 18, 19*). Nevertheless, the state of the liver with respect to estrogen induction remains altered. When the hormone is given again (secondary stimulation), the lag period is shorter and the response is more rapid (*5*). In both the primary and secondary stimulation, induction occurs in fully differentiated hepatocytes which do not proliferate in response to the stimulus (*20*). Moreover, the parenchymal liver cell population responds homogeneously by synthesizing vitellogenin (*21*). These latter characteristics are valuable in the study of the hormonal response and lend certain advantages to a study of this system as compared to other hormone-regulated systems.

Summary. Vitellogenin is synthesized under estrogen control in the liver, extensively modified, transported to the ovary, and there processed to the yolk proteins lipovitellin and phosvitin. In the frog *Xenopus laevis* there are at least four distinct but related vitellogenin genes. The two genes A1 and A2 have a 95 percent sequence homology in their messenger RNA coding regions, and contain 33 introns that interrupt the coding region (exons) at homologous positions. Sequences and lengths of analogous introns differ, and many introns contain repetitive DNA elements. The introns in these two genes that have apparently arisen by duplication have diverged extensively by events that include deletions, insertions, and probably duplications. Rapid evolutionary change involving rearrangements and the presence of repeated DNA suggests that the bulk of the sequences within introns may not have any specific function.

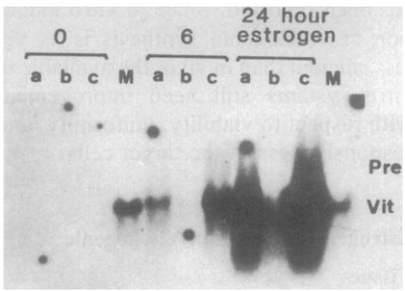
To study the kinetics of the induction process, methods for in vitro culture of liver tissue have been developed. Direct induction of vitellogenin synthesis has been obtained in organ culture (liver cubes) of male *Xenopus* liver, and successful induction has also been reached in primary monolayer cultures of liver cells (*20–23*). The important conclusions from these studies are that estrogen alone is sufficient to elicit vitellogenin production and that no DNA synthesis in the target cells is required for the induction process. No important difference between induction in vivo and in vitro has been reported. Since in vitro induction of vitellogenin synthesis is as yet less efficient than in vivo, the available in vitro systems still need improvement with respect to viability, uniformity, and responsiveness of the target cells.

Estrogen Receptors in Vitellogenic Tissues

It is generally thought that estrogens act on target cells by first binding to cytoplasmic receptors. The estrogen receptor is then translocated to the nucleus, binds to chromatin at specific acceptor sites, and consequently influences gene activity (*24*). The estrogen receptor system of *Xenopus* liver, however, appears to differ somewhat from this model. Only very low levels of receptor have been found in male *Xenopus* (200 sites per cell compared to some 10,000 sites per cell in the chick oviduct) and about half of this receptor is found in nuclei. After estrogen treatment of the male, there is a tenfold increase in the level of nuclear receptor. This increase is much greater than can be accounted for by translocation of the preexisting cytoplasmic receptor and probably represents new synthesis. The maximum level of 1000 sites per nucleus is reached within 12 hours, after 8 days the level starts to decline slowly, and after 30 days—when vitellogenin synthesis has stopped completely—the level of receptor is still about three times higher than in control (nontreated) males. It is possible that the increased levels of receptors are required for high rates of vitellogenin mesenger RNA (mRNA) production, and the increased levels at the end of primary stimulation may play a role in the more rapid increase of vitellogenin synthesis during secondary stimulation (*25*).

Development of Estrogen Dependence of Vitellogenin Synthesis

Vitellogenin is first inducible at the metamorphic stage 62, whether the tadpoles develop normally or at a rate increased by thyroxine. The transition to the inducible state does not occur in larvae developmentally arrested in prometamorphosis for up to 6 months (*26*). Estrogen treatment during metamorphosis can affect protein synthesis in the liver beginning at stage 53, before vitellogenin is inducible. Thus, different estrogen responses do not appear synchronously. Low levels of the nuclear estrogen receptor (100 sites per nucleus) are present as early as stage 53, but consist-



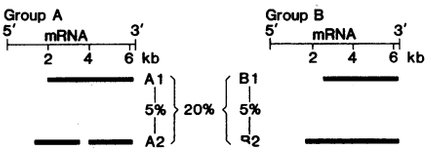


Fig. 2 (left). Appearance of vitellogenin mRNA and its precursor after estrogen treatment. Nuclear RNA from a control male and males treated for 6 and 24 hours with 1 mg of 17β-estradiol was extracted and fractionated on a poly(U) (polyuridylic acid) Sepharose column; 20 μg of unfractionated RNA (a), 20 μg of RNA lacking poly(A) (b), and 4 μg of poly(A)-containing RNA (c) were separated on a methylmercury agarose gel, transferred to diazobenzyloxymethyl paper, hybridized with a probe specific for A1 vitellogenin mRNA [the cloned cDNA pXlvc 23, nick-translated with ^{32}P-labeled deoxycytidine triphosphate (dCTP)], and exposed for autoradiography. The marker lanes (*M*) contain 1 or 0.5 ng of vitellogenin mRNA. For details see (*32*). The positions of the vitellogenin mRNA (*Vit*) and its precursor (*Pre*) are given. Fig. 3 (right). Relationship between the different cloned vitellogenin cDNA's. The cloned cDNA's can be organized into two main groups of sequences. A and B, which differ in about 20 percent of their nucleotides. Each main group contains two subgroups A1 and A2 or B1 and B2 which differ from each other in about 5 percent of their sequence. The mRNA maps give a scale, and the horizontal black bars below indicate the portions of the four mRNA's that have been cloned as cDNA's (*30, 34*).

ent with the hypothesis that increased levels of receptors are required for induction of vitellogenin synthesis, estrogen increases the level of nuclear receptors only from stage 62 onward (*27*).

Accumulation of Vitellogenin Messenger RNA During Hormone Induction

Vitellogenin mRNA was purified from the liver of estrogen-stimulated *Xenopus* by virtue of its high concentration, large size, and poly(A) (polyadenylic acid) content (*6, 28*). The mRNA has been identified by several criteria, most convincingly by in vitro translation to 200,000-dalton polypeptides that could be precipitated with vitellogenin antibodies (*5, 16, 29, 30*). The mRNA molecules comprise about 6300 nucleotides (*6, 28*).

Complementary DNA (cDNA) synthesized on purified vitellogenin mRNA has been used as a probe to titrate the accumulation of the mRNA during hormonal stimulation. Vitellogenin mRNA is not detected in unstimulated males, but 10 days after stimulation more than 30,000 vitellogenin mRNA molecules have accumulated per cell (*31*). There is a lag period of several hours (the minimum reported is 4.5 hours) before vitellogenin mRNA can be detected.

About 40 days after primary stimulation vitellogenin mRNA can no longer be detected in liver RNA. Secondary stimulation leads to an immediate and more efficient response; vitellogenin mRNA molecules have been detected after 1 hour, the rate of accumulation is much faster, and close to 100,000 mRNA's per liver cell have accumulated after about 1 week (*31*). These results show that in both primary and secondary stimulation, the rate of vitellogenin synthesis is dependent on the concentration of cytoplasmic mRNA.

The rapid accumulation of vitellogenin transcripts after estrogen induction is illustrated in Fig. 2, which also shows that molecules larger than the mRNA occur in the nucleus. These molecules are almost certainly precursors to mature mRNA, as discussed below (*32*). Together, the results obtained on vitellogenin induction along with other data support the current model for steroid hormone action; this model holds that these hormones regulate gene expression primarily at the transcriptional level. That does not exclude the possibility of additional effects such as differential mRNA stabilization for which there is some evidence for vitellogenin mRNA in *Xenopus* (*6, 23*).

The Vitellogenin Gene Family

Complementary DNA copies of vitellogenin mRNA have been cloned in *Escherichia coli* with the use of plasmid vectors (*33*). The analysis of a number of cDNA clones by different techniques has led us to the conclusion that vitellogenin is encoded in a family of at least four expressed genes (*34*). What had been assumed to be a homogeneous mRNA preparation in fact consists of four distinct although related molecules, all of which encode vitellogenin polypeptides (see below). Figure 3 shows the relations between the four vitellogenin mRNA molecules that have been identified. There are two main groups of sequences, named A and B, which differ in approximately 20 percent of their nucleotides. The A and B groups are each composed of two subgroups of sequences (A1 and A2; B1 and B2) which differ in about 5 percent of their nucleotides. These sequence comparisons were done by hybrid melting experiments. The four mRNA's have a similar length and accumulate in high amounts in individual, stimulated animals (*30, 34*).

To analyze the protein product of the four genes, the four mRNA's were isolated by hybridization to separate cDNA clones representing the four distinct mRNA's. The purified A1, A2, B1, and B2 mRNA's were then translated in vitro; all direct the synthesis of 200,000-dalton polypeptides that were immunoprecipitable with antibody to vitellogenin (*30*). Analysis of the cyanogen bromide cleavage products of the four vitellogenins synthesized in vitro has shown that A and B proteins differ substantially in their primary structure, whereas the A1 and A2 and the B1 and B2 polypeptides show only minor differences (*35*). These results correlate with the relations found between the different vitellogenin mRNA's. The translation experiments also show that each of the four proteins contains a serine-rich part corresponding to the phosphoprotein segment located in the carboxyl-proximal half of the vitellogenin polypeptide. Recent work on the structural relationships between vitellogenin and the yolk proteins also supports the conclusion that vitellogenin is composed of at least two very distinct polypeptides (*14*). Two main forms of *Xenopus* vitellogenin having the following organization are suggested by Wiley (*14*): NH_2-lipovitellin I-phosvitin-lipovitellin II-COOH, and NH_2-lipovitellin I-(phosvette I, phosvette II)-lipovitellin II-COOH. Phosvettes are thought to arise by an additional cleavage site in the serine-rich portion of one of the forms of vitellogenin beyond those cleavages that give rise to phosvitin in the other form. Variants of these main forms may result in the additional heterogeneity observed in vitellogenin and its cleavage products (*30, 36*).

The presence of four distinct mRNA's and thus four genes correlates with the observations of differences between related yolk proteins. We do not understand the physiological basis for the apparent need for multiple vitellogenins in *Xenopus*. However, there is reason to believe that variant vitellogenins and thus multiple genes for this protein class are required, because chickens also contain at least two vitellogenins that are precursors to distinct yolk proteins (*37*). Insect yolk proteins differ in overall structure from vertebrate vitellogenins, but it is notable that multiple genes en-

coding distinct proteins have been identified in these animals as well (*38*). Gene families encoding distinct but related proteins are a common feature of eukaryotic genomes (*39, 40*). The vitellogenin gene family in *Xenopus*, like other gene families, has apparently arisen by gene duplication as is discussed below.

Structural Organization of Vitellogenin Genes

Technical advances in DNA cloning technology (*41*) have made it possible to construct a *Xenopus* DNA library composed of 15 kilobase pairs (kb) DNA fragments cloned in the bacteriophage lambda vector Charon 4. Sets of overlapping DNA fragments have been isolated that contain the two A genes in their entirety together with long stretches of flanking sequences. So far, there is no evidence for linkage of genes A1 and A2 after analyzing regions of 11 to 31 kb between the two genes (*42*). Other groups of related genes are often linked, and it remains open whether the vitellogenin genes are also linked but separated by relatively long stretches of DNA. Figure 4 summarizes the structural organization of the coding regions in both genes as determined by R loop and RNA-DNA hybrid analysis in the electron microscope (*42*). In both genes, the 6-kb coding regions are interrupted 33 times by introns (intervening sequences in DNA) resulting in total gene lengths of 21 kb for gene A1 and 16 kb for gene A2. While both genes have the same mean exon length (portion of DNA that codes for the mRNA) (0.175 kb), they have different average intron lengths (0.45 kb in gene A1 and 0.31 kb in gene A2). Heteroduplex analysis of the two genes indicates that introns interrupt the structural sequences at homologous positions. Therefore, we conclude that genes A1 and A2 are probably the result of a duplication of an ancestral gene with a similar exon-intron organization as the extant genes.

Introns are transcribed along with exons into nuclear precursors of vitellogenin mRNA (*32*). Such large precursors have been detected in estrogen-induced liver but are absent from uninduced tissue (*32*) (Fig. 2). The analysis of nuclear RNA molecules revealed the presence of presumed large precursors that contain many and possibly all introns, and of apparent processing intermediates from which some introns had been spliced out. Comparison of these putative intermediates suggests that the splicing order of different introns does not follow a single pathway (*32*).

Repeated Sequences Around and Within the Vitellogenin Genes

In order to detect and localize sequences that are repeated many times in the *Xenopus* genome in the vicinity of the A1 and A2 vitellogenin genes, restriction endonuclease fragments of cloned DNA's were hybridized with in vitro labeled total *X. laevis* genomic DNA. Cloned fragments homologous to single or low copy genomic sequences fail to react under these conditions because the labeled DNA segments are too dilute in the probe; in contrast, fragments homologous to repeated DNA react and can be detected (*43*). Figure 5 illustrates that several of the cloned fragments give a positive signal and therefore contain one or more copies of a repeated sequence. Figure 6 summarizes the results of this analysis and gives the position of the fragments containing repeated DNA. Obviously, repeated sequences are not restricted to the flanking regions, but also occur at several positions within the two genes. The intragenic repeated sequences must be located within introns, since experiments with cloned cDNA's and mRNA's have shown that these sequences are present in very low copy number (*34, 44*) (cDNA lane in Fig. 5). We do not know how many different types of repeated sequences occur within and around the two genes, but identical repeats are not found at similar positions and in the same orientation in genes A1 and A2 since only homologous exons form stable heteroduplexes (*42*).

Gene A1 has a longer intron mean length and has more repeats in introns of its 5′ proximal half as compared to gene A2. Some of the length differences between analogous introns most likely reflect the acquisition or loss of repetitive sequences. This suggestion is supported by the observed variation in the sequence of introns due to apparent polymorphism. We have previously described length heterogeneity in intron 11 of the A1 gene as follows: "long" and "short" cloned DNA fragments differed by about 350 base pairs at this position (*42*). This length difference is not a cloning artifact since both long and short form sequences are present in uncloned DNA. In an Eco RI digest of the long clones an Eco RI fragment of 1.5 kb covers this region while in short clones it

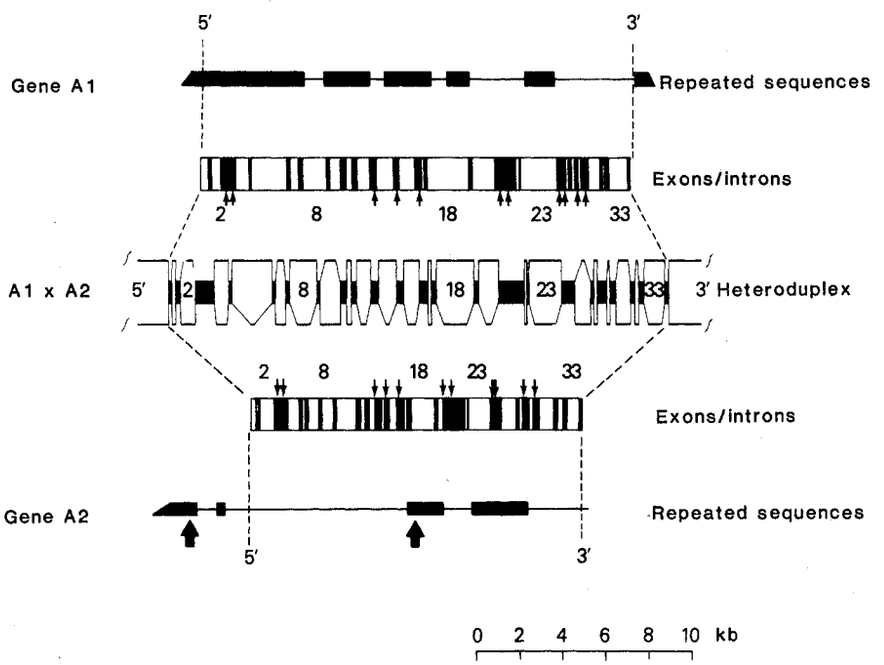
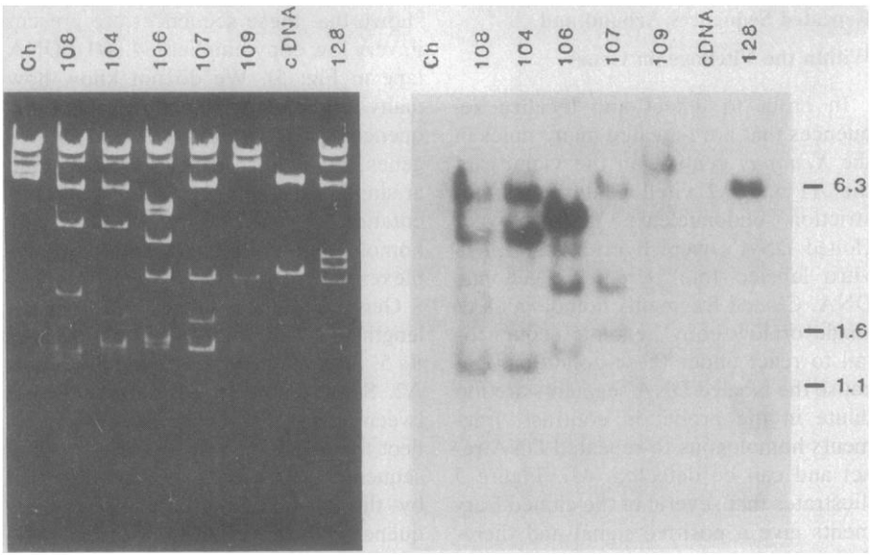
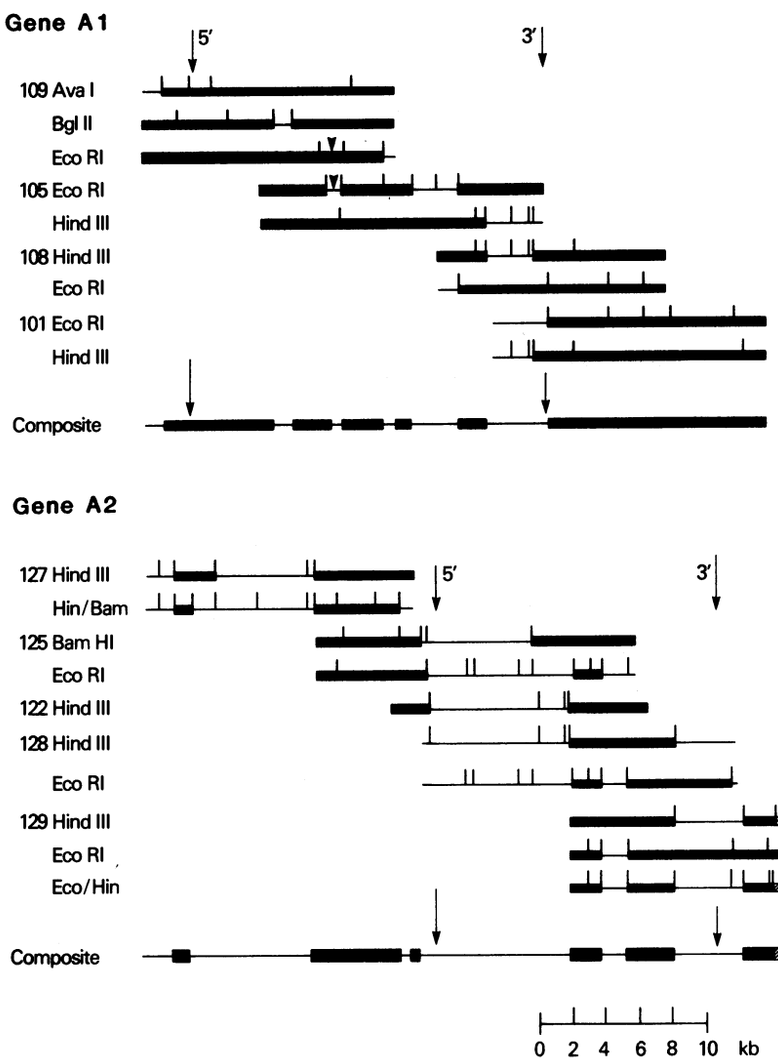


Fig. 4. Exon-intron maps of the two vitellogenin A genes and maps of regions containing repetitive DNA. Exons are defined as DNA regions that encode mature mRNA sequences; introns are DNA regions that interrupt the coding sequence and are not represented in mature mRNA (*46*). The R loop molecules formed between genomic vitellogenin DNA clones and vitellogenin mRNA's were analyzed to provide exon-intron maps. The small arrows in the R loop maps indicate very small introns of about 100 base pairs. In the heteroduplex molecules, homologous exons formed duplexes while analogous introns differing in sequence and length were seen as heterology loops (*42*). Regions where middle repetitive DNA was found are indicated in the top and bottom maps as heavy bars. These heavy bars give a maximal estimate of the size of the repetitive regions since the repeats are probably all shorter than the restriction fragments in which they were localized (Fig. 6). The large arrows in the repeated sequences map of gene A2 give the positions of a particular repeat of about 260 base pairs found in opposite orientation outside and inside the gene (Figs. 7 and 8). Since the exons represent low copy number DNA, the repeats are localized in introns.





is a 1.15-kb fragment. Figures 5 and 6 show that the long 1.5-kb fragment contains a sequence repeated elsewhere in the genome, whereas the 1.15-kb fragment does not. From these results two conclusions can be derived. First, introns allow for polymorphisms that are not expressed at the level of protein or mRNA (*45*). Second, variation in introns may involve insertions or deletions of many nucleotides and can lead to the insertion of a repetitive sequence into an intron.

Inverted repeats which are known to form "hairpin" or "stem-loop" structures have attracted attention as possible sites of signal functions. We have observed a number of small hairpin loops within several introns and in the flanking regions of the vitellogenin A1 and A2 genes (*42*). We also detected an inverted repeat element in the 5′ flanking region and at a position within the A2 gene. Figure 7 shows a heteroduplex molecule between two nonoverlapping fragments of the A2 gene that are inserted into the lambda vector in opposite orientations. The two fragments do not pair except in a small region which corresponds to an inverted repeat sequence. From micrographs like the one shown and from cross-hybridization experiments between restriction fragments we have located one member of the inverted repeat about 2.8 kb from the 5′ end of the gene and the other to a position within the gene most probably in intron 18 (Figs. 4 and 8). As seen in Fig. 6, the restriction

Fig. 5 (top). Detection of middle repetitive DNA in the vitellogenin genes and their flanking sequences. Eco RI endonuclease digests of cloned genomic DNA's (0.6 μg) were separated by electrophoresis in a 1.2 percent agarose gel. As controls, Eco RI digests of cloned cDNA (pXlvc23) and Charon 4 DNA (*Ch*) were included. The DNA fragments were transferred to a nitrocellulose membrane filter and hydridized with labeled *Xenopus* genomic DNA (10^5 cpm/cm^2, nick-translated with ^{32}P-dCTP to 3×10^8 cpm/μg) in 50 percent formamide, 4× SET, 5× Denhardt solution, 0.1 percent sodium pyrophosphate, 0.2 percent sodium dodecyl sulfate (SDS), yeast RNA at 250 μg/ml, and calf thymus DNA at 100 μg/ml for 72 hours at 37°C. The filter was washed in 50 percent formamide, 4× saline sodium citrate (SSC), and 0.2 percent SDS at 37°C and exposed for autoradiography. Only fragments containing repeated DNA can be visualized by this method (*43*). Molecular sizes of markers are given in kilobase pairs. The left panel shows the gel after staining with ethidium bromide, and the right panel shows the autoradiogram. Fig. 6 (bottom). Occurrence of middle repetitive DNA within and around the vitellogenin genes. Endonuclease digests of cloned genes A1 and A2 including their flanking regions were analyzed for repeated DNA as described in the legend to Fig. 5. Fragments containing repeated DNA are represented by thick lines and fragments present at low copy number by thin lines. The numbers at the left are the designations of the different cloned DNA's represented in each map. The small vertical bars indicate the sites of the different restriction endonucleases specified at the left of each line. Arrows indicate the tentative 5′ and 3′ ends of the genes (*42*). The arrowheads in gene A1 indicate the long and short forms of the Eco RI fragment that exhibits length heterogeneity in intron 11 (*42*). The hatched region in the 3′ flanking region of gene A2 indicates a fragment in which the presence or absence of repeated DNA could not be clearly determined. The composites give a summary of the results and are incorporated in part into Fig. 4.

fragments that contain this inverted element are repeated at a moderately high level in the *Xenopus* genome. We have no direct evidence that the inverted element is in fact the dispersed repeated element, but this appears the most likely interpretation. Therefore, it appears that repeated DNA elements may occur both in the vicinity of and within structural genes in opposite orientations.

Conclusion: Introns as Silent DNA Sequences

The two closely related vitellogenin A genes most likely arose by a gene duplication because the mRNA coding regions are highly homologous and are interrupted by introns at homologous positions. The presence of introns in the ancestral gene could be explained by Gilbert's model of DNA shuffling (*46*). After the duplication event the intron-exon arrangement and the primary sequence of the exons were largely conserved while intron sequence and size drifted rapidly. Similar observations have been made in other duplicated genes that have been studied (*47*). While at least one spliceable intron is required for the expression of certain genes (*48*), it appears that the majority of introns of a gene and the majority of the sequences within an intron may be phenotypically silent. Such an interpretation would explain their rapid drift in duplicated genes. As has been proposed, such drifting could have a beneficial effect by reducing homology between duplicated genes, thus diminishing the target size for recombination events that might result in the loss of a gene (*49*). Our data on the vitellogenin genes suggest that intron divergence involves not only point mutations but also large-scale deletions, insertions, and possibly duplications resulting in substantial length differences between analogous introns (Fig. 4). Although the first few genes that were analyzed contained only single-copy DNA within their introns, it is becoming clear that middle repetitive DNA may be present in these regions (Fig. 4) (*50*). In the vitellogenin genes, a repeat may be present in only one of two analogous introns and analogous introns may contain nonhomologous repeat elements. A change in intron length directly related to insertion or deletion of a repeated element is apparent in intron 11 of polymorphic forms of gene A1; thus it appears that repeats can enter or leave introns without influencing the gene product. Our results suggest that repeated DNA may be mobile in the genome of vertebrates as has been demonstrated in yeast and *Drosophila* (*51*). Rearrangements involving repeated sequences may be a significant factor in the divergence of introns in the vitellogenin genes and perhaps other genes as well.

Introns are transcribed as part of the primary gene products in all cases studied so far, including vitellogenin (*32, 52*). If repeats are widely distributed within introns we expect that nuclear RNA has a sequence organization just like genomic DNA, where single copy and repeated sequences are interspersed. Such organization has been reported for nuclear RNA of various organisms (*53*). Since different repeats may be present in the introns of different genes, the type and concentration of repeat transcripts in nuclear RNA would reflect differential gene activity. Such an interpretation may be sufficient to explain the different types and concentrations of repeat transcripts found in nuclear RNA at different stages of development or in different tissues (*54*). An additional feature of nuclear

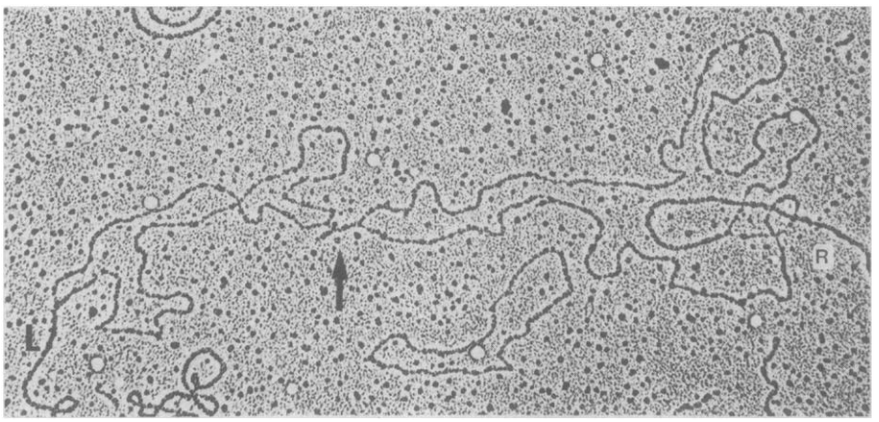


Fig. 7. Inverted repeat in two nonoverlapping clones of gene A2. A heteroduplex was formed between λXlv121, which contains an almost complete A2 gene, and λXlv126, which contains mostly flanking sequences inserted into the vector in opposite orientation (see Fig. 8). The double-stranded regions formed by the left (*L*) and right (*R*) arms of the phage vector which continue beyond the figure stabilize the heteroduplex. Only a small duplex of about 260 base pairs can be seen within the cloned *Xenopus* DNA (arrow) revealing a repeat present in opposite orientation in λXlv126 and λX1v121. The positions of this repeat are shown in Fig. 8, which gives further details.

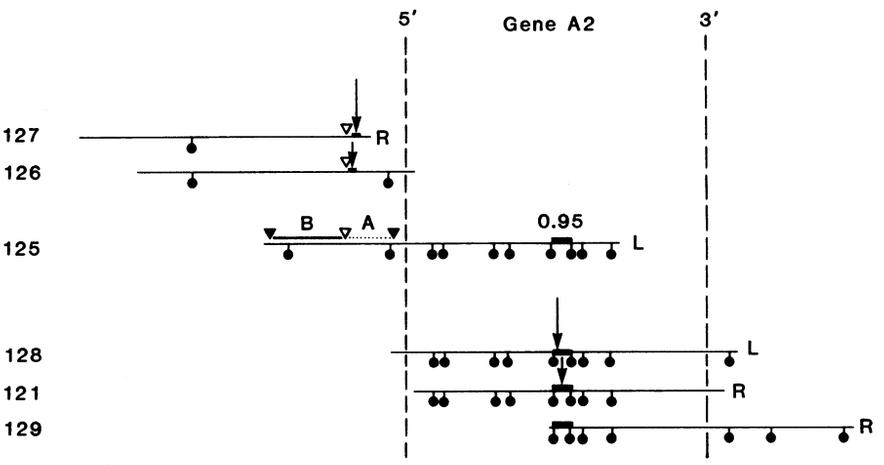


Fig. 8. Localization of a repeat element that occurs in opposite orientation within and outside the A2 gene. Heteroduplexes were formed between λXlv126 and 121 (Fig. 7) and between λXlv127 and 128. In each member of a pair, the *Xenopus* DNA is inserted in opposite orientation. A small duplex of about 260 base pairs was found, and the positions of the repeat are indicated by arrows. The repeat occurs about 2.8 kb upstream from the tentative 5′ end of the gene, just at the right of a Sal I site (▽). Its second position in the middle of the gene is probably within intron 18 (Fig. 4). In a complementary approach, two 5′ flanking DNA fragments, A and B, were purified from λXlv125 DNA. Fragment A is a Hind III (▼)/Sal I (▽) fragment containing the repeat as mapped by electron microscopic heteroduplex analysis. Fragment B is the Sal I (▽)/Hind III (▼) fragment to the left of the A fragment and was used as a control. The two fragments were labeled in vitro by nick translation and hybridized separately to Eco RI (●) digests of λXlv126, 125, 121, and 129. An Eco RI fragment (heavy bar) of 0.95 kb located within the gene hybridized to the A probe but not to the B probe, confirming the mapping of the inverted repeat by electron microscopy.

RNA may be explained by reasoning along similar lines. We have shown in the vitellogenin gene A2 that an element of the same repeat family is present in opposite orientation in the 5′ flanking region and within an intron (Figs. 4 and 8). If we hypothesize from this finding that repeats of the same family may occur in introns of different genes in opposite orientations it follows that the nuclear transcripts will have regions of intermolecular complementarity, as demonstrated in HeLa cell RNA (*55*), and that both strands of elements of one repeated family would be transcribed, as observed in sea urchins (*54*).

While it is clear that exon-intron boundaries are highly conserved, suggesting specific functional requirements for these sequences (*39, 56*), we believe that the bulk of intron sequences may be phenotypically silent. These DNA stretches might be considered "selfish DNA" in the meaning of a recent hypothesis (*57*), but it is debatable whether the presence of such DNA is the result of intragenomic selection or of random forces (*58*). The finding of repeated DNA sequences in introns and the suggestion that at least some of these repeats may be mobile elements is in concert with the properties of selfish DNA (*57*). We believe that the application of the concept of phenotypically silent DNA to much of the sequences in introns can rationalize some aspects of genome organization including the interspersion of repeats and single copy DNA in transcription units and the widespread and cell type-specific expression of both strands of many repeats in nuclear RNA.

References and Notes

1. E. Davidson, *Gene Activity in Early Development* (Academic Press, New York, 1976).
2. J. N. Dumont, *J. Morphol.* **136**, 153 (1972).
3. M. Gruber, E. S. Bos, G. Ab. *Mol. Cell. Endocrinol.* **5**, 41 (1976); E. W. Bergink, R. A. Wallace, J. A. Van De Berg, E. S. Bos, M. Gruber, G. Ab, *Am. Zool.* **14**, 1177 (1974); R. A. Wallace and E. W. Bergink, *ibid.*, p. 1159; J. R. Tata, *Cell* **9**, 1 (1976); G. U. Ryffel, *Mol. Cell. Endocrinol.* **12**, 237 (1978); J. R. Tata and D. F. Smith, *Recent Prog. Horm. Res.* **35** (1979).
4. E. W. Bergink and R. A. Wallace, *J. Biol. Chem.* **249**, 2897 (1974); D. F. Smith, T. M. Penning, A. Q. Ansari, K. A. Munday, M. Akhtar, *Biochem. J.* **174**, 353 (1978).
5. M. V. Berridge, S. R. Farmer, C. D. Green, E. C. Henshaw, J. R. Tata, *Eur. J. Biochem.* **62**, 161 (1976).
6. W. Wahli, T. Wyler, R. Weber, G. U. Ryffel, *ibid.* **66**, 457 (1976).
7. B. K. Follett and M. R. Redshaw, in *Physiology of Amphibia*, B. Lofts, Ed. (Academic Press, New York, 1974), vol. 2, pp. 219–308.
8. R. A. Wallace, in *The Vertebrate Ovary: Comparative Biology and Evolution*, R. E. Jones, Ed. (Plenum, New York, 1978), pp. 469–502.
9. R. A. Wallace, *Biochim. Biophys. Acta* **74**, 505 (1963); *ibid.* **215**, 176 (1970); M. J. Clemens, *Prog. Biophys. Mol. Biol.* **28**, 71 (1974).
10. R. A. Wallace and J. N. Dumont, *J. Cell. Physiol.* **72**, 73 (1968).
11. R. A. Wallace, R. A. Deufel, Z. Misulovin, *Comp. Biochem. Physiol. B* **65**, 151 (1980).
12. R. A. Wallace and Z. Misulovin, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5534 (1978).
13. E. W. Bergink and R. A. Wallace, *J. Biol. Chem.* **249**, 2897 (1974); D. H. Ohlendorf, G. R. Barbarash, A. Trout, C. Kent, L. J. Banaszak, *ibid.* **252**, 7992 (1977).
14. H. S. Wiley, thesis, University of Tennessee (1979).
15. E. W. Bergink, M. T. Tseng, J. L. Wittliff, *Eur. J. Cell Biol.* **14**, 362 (1977).
16. J. A. Lewis, M. J. Clemens, J. R. Tata, *Mol. Cell. Endocrinol.* **4**, 311 (1976).
17. J. K. Skipper and T. H. Hamilton, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2384 (1977).
18. R. A. Wallace and D. W. Jared, *Can. J. Biochem.* **46**, 953 (1968); J. L. Wittliff and F. T. Kenney, *Biochim. Biophys. Acta* **269**, 485 (1972).
19. D. J. Shapiro, H. J. Baker, D. T. Stitt, *J. Biol. Chem.* **251**, 3105 (1976).
20. C. D. Green and J. R. Tata, *Cell* **7**, 131 (1976).
21. L. J. Wangh, J. A. Osborne, C. C. Hentschel, R. Tilly, *Dev. Biol.* **70**, 479 (1979).
22. L. J. Wangh and J. Knowland, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3172 (1975); B. K. Felber, G. U. Ryffel, R. Weber, *Mol. Cell. Endocrinol.* **12**, 151 (1978).
23. D. J. Shapiro and H. J. Baker, in *Ontology of Receptors and Mode of Action of Reproductive Hormones*, T. H. Hamilton and J. L. Clark, Eds. (Raven, New York, 1978).
24. K. R. Yamamoto and B. M. Alberts, *Annu. Rev. Biochem.* **45**, 721 (1976).
25. B. Westley and J. Knowland, *Cell* **15**, 367 (1978); *Biochem. Biophys. Res. Commun.* **88**, 1167 (1979); B. Westley, *Differentiation* **15**, 67 (1979).
26. S. Huber, G. U. Ryffel, R. Weber, *Nature (London)* **278**, 65 (1979); F. E. B. May and J. Knowland, *Dev. Biol.* **77**, 419 (1980); two other papers investigating in vitro induction of vitellogenin synthesis reported a somewhat earlier appearance of competence for estrogen-dependent vitellogenin synthesis [J. Knowland, *Differentiation* **12**, 47 (1978); J. K. Skipper and T. H. Hamilton, *Science* **206**, 693 (1979)].
27. F. E. B. May and J. Knowland, *Dev. Biol.*, in press; personal communication.
28. D. J. Shapiro and H. J. Baker, *J. Biol. Chem.* **252**, 5244 (1977); W. Wahli, T. Wyler, R. Weber, G. U. Ryffel, *Eur. J. Biochem.* **86**, 225 (1978).
29. K. D. Lanclos and T. H. Hamilton, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3934 (1975); S. R. Farmer, E. C. Henshaw, M. V. Berridge, J. R. Tata, *Nature (London)* **273**, 401 (1978).
30. B. K. Felber *et al.*, *Eur. J. Biochem.* **105**, 17 (1980).
31. G. U. Ryffel, W. Wahli, R. Weber, *Cell* **11**, 213 (1977); H. J. Baker and D. J. Shapiro, *J. Biol. Chem.* **252**, 8428 (1977); *ibid.* **253**, 4521 (1978).
32. G. U. Ryffel, T. Wyler, D. B. Muellener, R. Weber, *Cell* **19**, 53 (1980).
33. W. Wahli, G. U. Ryffel, T. Wyler, R. B. Jaggi, R. Weber, I. B. Dawid, *Dev. Biol.* **67**, 371 (1978); D. F. Smith, P. F. Searle, J. G. Williams, *Nucleic Acids Res.* **6**, 487 (1979).
34. W. Wahli, I. B. Dawid, T. Wyler, R. B. Jaggi, R. Weber, G. U. Ryffel, *Cell* **16**, 535 (1979).
35. R. B. Jaggi, B. K. Felber, S. Maurhofer, R. Weber, G. U. Ryffel, *Eur. J Biochem.*, in press.
36. H. S. Wiley and R. A. Wallace, *Biochem. Biophys. Res. Commun.* **85**, 153 (1978); D. H. Ohlendorf, G. R. Barbarash, A. Trout, C. Kent, L. J. Banaszak, *J. Biol. Chem.* **252**, 7992 (1977).
37. S.-Y. Wang and D. L. Williams, *Biochemistry* **19**, 1557 (1980).
38. T. G. Warren and A. P. Mahowald, *Dev. Biol.* **68**, 130 (1979); T. Barnett, C. Pachl, J. P. Gergen, P. C. Wensink, *Cell* **21**, 729 (1980).
39. I. B. Dawid and W. Wahli, *Dev. Biol.* **69**, 305 (1979).
40. E. O. Long and I. B. Dawid, *Annu. Rev. Biochem.* **49**, 727 (1980).
41. F. R. Blattner *et al.*, *Science* **196**, 161 (1977); N. Sternberg, D. Tiemeier, L. Enquist, *Gene* **1**, 255 (1977); W. D. Benton and R. W. Davis, *Science* **196**, 180 (1977); T. Maniatis, R. C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, G. K. Sim, A. Efstratiadis, *Cell* **15**, 687 (1978).
42. W. Wahli and I. B. Dawid, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1437 (1980); ———, T. Wyler, R. Weber, G. U. Ryffel, *Cell* **20**, 107 (1980).
43. C.-K. J. Shen and T. Maniatis, *Cell* **19**, 379 (1980).
44. H. J. Widmer, R. B. Jaggi, R. Weber, G. U. Ryffel, *Eur. J. Biochem.* **99**, 23 (1979).
45. Similar observations have been reported for the ovalbumin gene [E. C. Lai, S. L. C. Woo, A. Dugaiczyk, B. W. O'Malley, *Cell* **16**, 201 (1979)].
46. W. Gilbert, in *Eucaryotic Gene Regulation ICN-UCLA Symposia on Molecular and Cellular Biology*, R. Axel, T. Maniatis, C. F. Fox, Eds. (Academic Press, New York, 1979), vol. 14, pp. 1–12.
47. D. A. Konkel, J. V. Maizel, Jr., P. Leder, *Cell* **18**, 865 (1979); A. van Ooyen, J. van den Berg, N. Mantel, C. Weissmann, *Science* **206**, 337 (1979); A. Efstratiadis *et al.*, *Cell* **21**, 653 (1980); P. Lomedico *et al.*, *ibid.* **18**, 545 (1979); F. Perler, A. Efstratiadis, P. Lomedico, W. Gilbert, R. Kolodner, J. Dodgson, *ibid.* **20**, 555 (1980).
48. D. H. Hamer and P. Leder, *Cell* **18**, 1299 (1979); P. Gruss, C. I. Lai, R. Dhar, G. Khoury, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4317 (1979).
49. D. C. Tiemeier *et al.*, *Cell* **14**, 237 (1978).
50. M. Cochet, F. Gannon, R. Hen, L. Maroteaux, F. Perrin, P. Chambon, *Nature (London)* **282**, 567 (1979).
51. S. S. Potter, W. J. Brorein, Jr., P. Dunsmuir, G. M. Rubin, *Cell* **17**, 415 (1979); E. Strobel, P. Dunsmuir, G. M. Rubin, *ibid.*, p. 429; R. Levis, P. Dunsmuir, G. M. Rubin, *ibid.* **21**, 581 (1980); S. Potter, M. Truett, M. Phillips, A. Maher, *ibid.* **20**, 639 (1980); J. R. Cameron, E. Y. Loh, R. W. Davis, *ibid.* **16**, 739 (1979).
52. S. M. Tilghman, P. J. Curtis, D. C. Tiemeier, P. Leder, C. Weissmann, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1309 (1978); A. J. Kinniburgh, J. E. Mertz, J. Ross, *Cell* **14**, 681 (1978); D. R. Roop, J. L. Nordstrom, S. Y. Tsai, M.-J. Tsai, B. W. O'Malley, *ibid.* **15**, 671 (1978); J. L. Nordstrom, D. R. Roop, M.-J. Tsai, B. W. O'Malley, *Nature (London)* **278**, 328 (1979); P. Chambon *et al.*, in *Eucaryotic Gene Regulation*, R. Axel, T. Maniatis, C. F. Fox, Eds. (Academic Press, New York, 1979), pp. 259–279; A. Halbreich, P. Pajot, M. Foucher, C. Grandchamp, P. Slonimski, *Cell* **19**, 321 (1980).
53. M. J. Smith, B. R. Hough, M.E. Chamberlin, E. H. Davidson, *J. Mol. Biol.* **85**, 103 (1974); J. E. Darnell and R. Balint, *J. Cell. Physiol.* **76**, 349 (1970); G. R. Molloy, W. Jelinek, M. Salditt, J. E. Darnell, *Cell* **1**, 43 (1974); D. S. Holmes and J. Bonner, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1108 (1974).
54. R. H. Scheller, F. D. Costantini, M. R. Kozlowski, R. J. Britten, E. H. Davidson, *Cell* **15**, 189 (1978).
55. N. Fedoroff, P. K. Wellauer, R. Wall, *ibid.* **10**, 597 (1977).
56. R. Breathnach, C. Benoist, K. O'Hare, F. Gannon, P. Chambon, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4853 (1978); F. Crick, *Science* **204**, 264 (1979); I. Seif, G. Khoury, R. Dhar, *Nucleic Acids Res.* **6**, 3387 (1979); M. R. Lerner, J. A. Boyle, S. M. Mount, S. L. Wolin, J. A. Steitz, *Nature (London)* **283**, 220 (1980).
57. W. F. Doolittle and C. Sapienza, *Nature (London)* **284**, 601 (1980); L. E. Orgel and F. H. C. Crick, *ibid.*, p. 604.
58. G. Dover, *ibid.* **285**, 618 (1980).

59. We thank B. Kay, R. Goldberger, M. Singer, R. Wallace, and B. Westley for comments on the manuscript. F. E. B. May and J. Knowland for communicating unpublished observations, and D. Muellener and M. Rebbert for excellent technical assistance. Supported in part by the Swiss National Science Foundation, the Swiss Academy of Sciences, and the "Stiftung zur Forderung der wissenschaftlichen Forschung an der Universitat Bern."