Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 6832–6836, November 1982
Biochemistry

Vitellogenin genes *A1* and *B1* are linked in the *Xenopus laevis* genome

(gene libraries/*in vitro* transcription/gene and genome duplication)

WALTER WAHLI*, JACQUES-EDOUARD GERMOND*, BEATRICE TEN HEGGELER* AND FELICITY E. B. MAY†‡

*Institut de Biologie animale, Université de Lausanne, Palais de Rumine, CH-1005 Lausanne, Switzerland; and †Abteilung fuer Zell- und Entwicklungsbiologie, Universitaet Bern, Baltzerstrasse 4, CH-3012 Bern, Switzerland

Communicated by Igor B. Dawid, August 8, 1982

ABSTRACT **Genomic clones containing the *Xenopus laevis* vitellogenin gene *B1* have been isolated from DNA libraries and characterized by heteroduplex mapping in the electron microscope, restriction endonuclease analysis, and *in vitro* transcription in a HeLa whole-cell extract. Sequences from the 3′-flanking region of the previously isolated *A1* vitellogenin gene were found in the 5′-flanking region of this *B1* gene. Thus, the two genes are linked, with 15.5 kilobase pairs of DNA between them. Their length is about 22 kilobase pairs (*A1* gene) and 16.5 kilobase pairs (*B1* gene) and they have the following arrangement: 5′–*A1* gene–spacer–*B1* gene–3′. The analysis of heteroduplexes formed between the two genes revealed several regions of homology. Both genes are in the same orientation and, therefore, are transcribed from the same DNA strand. The possible events by which the vitellogenin gene family arose in *Xenopus laevis* are discussed.**

The transcription of the vitellogenin genes, which code for the precursor of the main yolk proteins, is strongly controlled by estrogen in the liver of oviparous vertebrates (1). The understanding of the molecular basis of this hormone-regulated gene expression requires knowledge of the fine structure and arrangement of the vitellogenin genes. We have reported that in *Xenopus laevis*, vitellogenin is encoded by at least four genes, called *A1*, *A2*, *B1*, and *B2*, all of which are under estrogen control (2, 3). Although there is only about 80% sequence homology between the coding regions of the *A* and *B* genes, there is about 95% homology between A1 and A2 and between B1 and B2 mRNA coding sequences (2). The two *A* genes have been isolated (4) and analysis of their structural organization has suggested that they arose from a duplication event (5). Although flanking regions of these two *A* genes have been analyzed, we found no evidence for linkage between them. We now report on the isolation of the *B1* gene and its flanking regions, whose characterization reveals a close linkage between *A1* and the *B1* vitellogenin gene.

MATERIALS AND METHODS

Gene Libraries of *Xenopus laevis*. The first library was constructed as described by Maniatis *et al.* (6). DNA from early embryos was partially digested with the restriction endonucleases *Alu* I and *Hae* III and was ligated with *Eco*RI linkers to phage λ Charon 4 arms. The recombinant molecules were packaged *in vitro*, and the phages were amplified before storage (4). The second library was prepared by similar procedures from liver DNA of a single mature female, partially digested with the restriction endonuclease *Eco*RI (7). After *in vitro* packaging, the phages were not amplified but were directly plated out and then screened as described (4).

Restriction Endonuclease Analysis and Subcloning of DNA Fragments. Digestions with the different restriction endonucleases were performed as indicated by the suppliers (Bethesda Research Labs, Rockville, MD; New England BioLabs, Beverly, MA). Gel electrophoresis, Southern transfer, nick-translation, and hybridization were all performed by standard procedures. The *Eco*RI fragments of the genomic clone λXlv 201 (see Fig. 1) were subcloned in the *Eco*RI site of pBR322, which had been dephosphorylated (8).

***In Vitro* Transcription Assay.** Whole-cell extracts from HeLa cells were prepared as described by Manley *et al.* (9) and were a kind gift of O. Hagenbuechle (ISREC, Lausanne). *In vitro* transcription incubations were done in 8.3-μl reaction mixtures containing 500 μM each of ATP, GTP, and CTP; 40 μM UTP with 1–5 μCi (1 Ci = 3.7×10^{10} becquerels) of [α-^{32}P]UTP; 10 mM creatine phosphate; 0.3–0.4 μg of template DNA; and 5 μl of HeLa cell extract added as the last component of the mixture. Incubations were for 60 min at 30°C. RNA was purified by phenol/chloroform extraction, glyoxylated for 10 min (10), and analyzed by gel electrophoresis.

Heteroduplex Analysis in the Electron Microscope. Heteroduplexes between *B1* cloned DNAs were prepared for electron microscopic analysis as described (4). Heteroduplexes between *A1* and *B1* clones were formed and analyzed as follows: DNA (2.5 μg/ml each) of two clones were denatured for 5 min at 65°C in the presence of 70% deionized formamide/0.5 M NaCl/5 mM EDTA/10 mM Tris·HCl, pH 8.0, and then were annealed in the same buffer for 60 min at 35°C, followed by 30 min at 30°C. The sample was adjusted to contain (final concentrations) 0.5 μg of DNA per ml, 25% (vol/vol) formamide, 0.7 M urea, 0.08 M NaCl, 10 mM EDTA, 0.1 M Tris·HCl (pH 8.5), and 40 μg of cytochrome *c* per ml and was spread onto a hypophase of 5% recrystallized formamide. pBR322 and phage ØX174 DNAs were included as double- and single-strand length standards, respectively.

RESULTS

Isolation of the *B1* Vitellogenin Gene. The study of cDNA clones has shown that vitellogenin in *Xenopus laevis* is encoded by four related but distinct genes (2). The amplified and unamplified *Xenopus laevis* gene libraries were screened with cloned cDNA corresponding to B1 mRNA under conditions of stringency in which hybridization occurred to *B* gene sequences but not to *A* gene sequences. Five clones from the 3′ region of the *B1* gene were isolated from the first library (λXlv 205–209; Fig. 1) in which sequences from the 5′ region seemed to be underrepresented. Clones with the 5′ region of the gene (λXlv 201–204; Fig. 1) were isolated from the unamplified library.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviation: kb, kilobase(s).

‡ Present address: Institut National de la Santé et de la Recherche Médicale (U 148), 60, rue de Navacelles, F-34100 Montpellier, France.

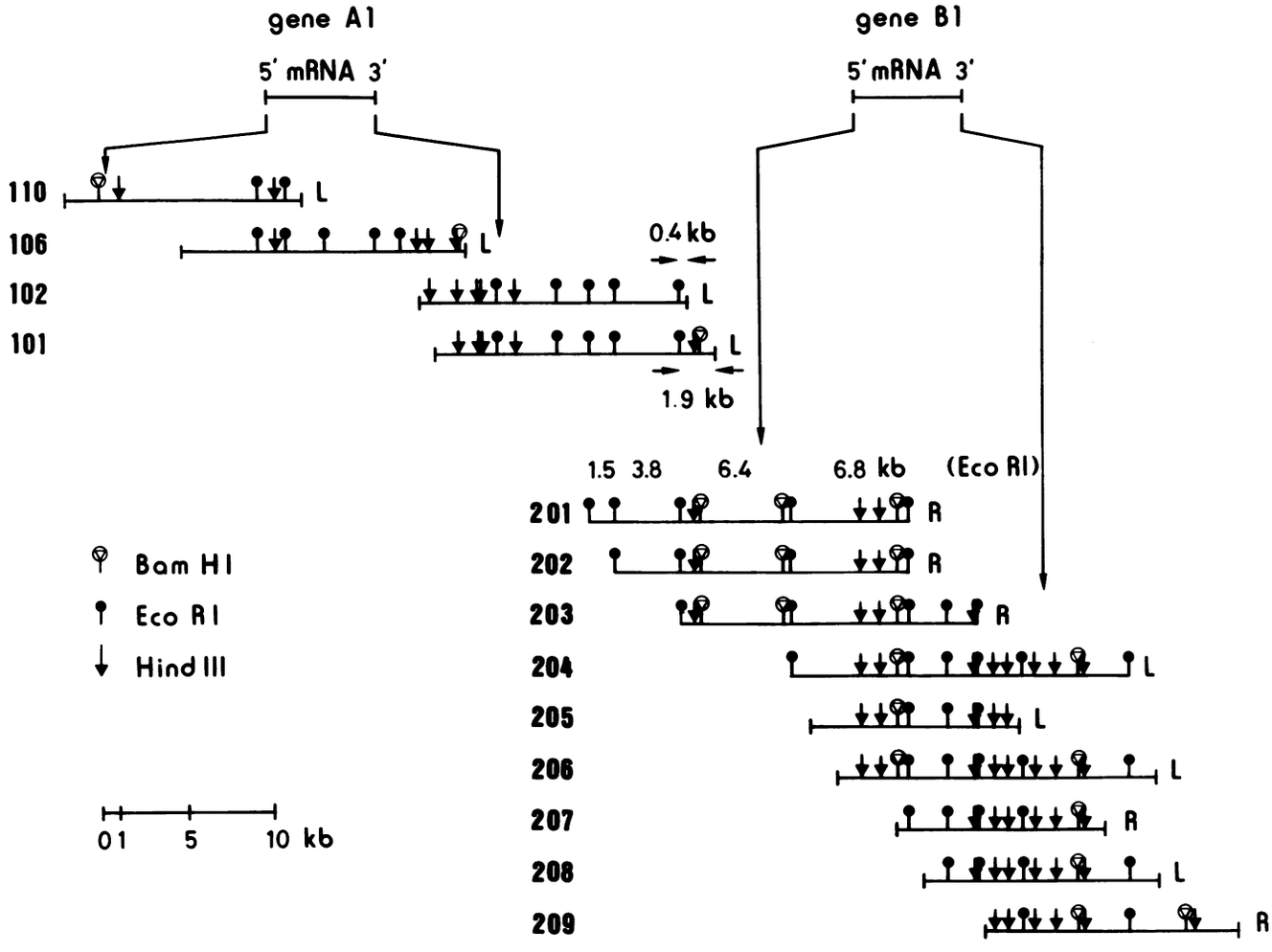

FIG. 1. Maps of overlapping cloned fragments of the *A1* and *B1* vitellogenin genes. Clones λXlv 101, 102, 106, and 110 were isolated and characterized as described (4, 5). Clones λXlv 201–209 were isolated from two different libraries (see text). Vitellogenin mRNA is shown at the top at the same scale as the genomic DNAs. The 5′ ends of the genes were localized by *in vitro* transcription and the 3′ ends, by hybridization with 3′ end proximal cDNA clones. L and R refer to the left and right arms of the λ Charon 4 vector and indicate the polarity of the inserted *Xenopus* DNA. The size of the *Eco*RI fragments help to understand the results presented in Fig. 2.

Those genomic clones were shown to contain *B1* sequences by dot hybridization/hybrid melting experiments as described (4). The overlapping *B1* clones (λXlv 201–209; Fig. 1) were aligned by a combination of experiments including hybridization with cloned cDNAs, heteroduplex mapping by electron microscopy, and localization of the cleavage sites for several restriction endonucleases (Fig. 1). The clones cover ≈38 kilobases (kb) of DNA. With no exception so far, the maps of the restriction endonuclease cleavage sites agree well in the overlapping regions of the different clones irrespective of the library from which they originate.

Linkage Between the *A1* and *B1* Vitellogenin Genes. Clones containing the *A1* gene have been isolated and characterized (4, 5). Comparative analysis of the restriction maps revealed an identity between the 3′-flanking region of the *A1* gene (λXlv 101 and 102) and the 5′-flanking region of the *B1* gene (λXlv 201–203) (Fig. 1). This result suggests linkage between the two genes. This possibility could not be tested by heteroduplex analysis between the *A1* and *B1* clones available because, as indicated in Fig. 1, the *Xenopus* genomic DNA fragments to be analyzed were inserted in opposite orientations in the vector λ Charon 4. Therefore, linkage between the two genes was demonstrated in the following hybridization experiment.

DNA from the clone λXlv 201 was digested with *Eco*RI endonuclease, and the four restriction fragments of 1.5 kb, 3.8 kb, 6.4 kb, and 6.8 kb (Fig. 1) were subcloned in pBR322. The subcloned fragments were radioactively labeled with ^{32}P by nick-translation and hybridized to a Southern blot transfer of *Eco*RI endonuclease digests of the *A1* clones λXlv 102 and 101 and of the *B1* clones λXlv 201 and 202 as a control (see Fig. 1 for orientation). The results are shown in Fig. 2. The 1.5-kb probe hybridized to itself (λXlv 201) and to the 1.5-kb fragment of the *A1* clones λXlv 102 and 101. The 3.8-kb probe hybridized to the 3.8-kb fragment of the two *B1* clones and of the two *A1* clones. The 6.4-kb probe hybridized to the 6.4-kb fragment in clones λXlv 201 and 202 and to the 0.4-kb and 1.9-kb fragments in the clones λXlv 102 and 101, respectively (see also Fig. 1). As expected, the 6.8-kb probe hybridized only to the homologous sequences in λXlv 201 and 202. This experiment gives strong evidence for linkage between the *A1* and *B1* genes.

Mapping of the 5′ End of the Linked Genes by *in Vitro* Transcription. The position of the transcription initiation sites of the linked *A1* and *B1* genes—that is, their 5′ ends—were mapped by *in vitro* transcription in a whole-cell extract of HeLa cells (9).

Fig. 3 shows the results obtained with the *A1* gene. An *A1* clone (λXlv 110) was digested with the endonuclease *Ava* I, and the mixture of fragments, λ Charon 4 sequences included, was

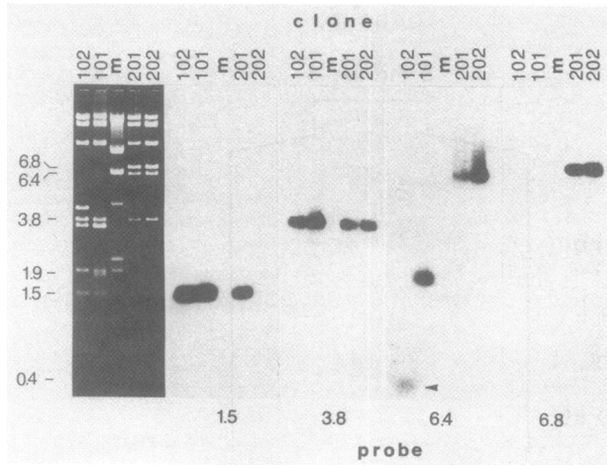

FIG. 2. Determination of the linkage between the *A1* and *B2* genes by hybridization between subcloned *Eco*RI fragments and *Eco*RI digests of genomic clones. The genomic clones λXlv 101, 102, 201, and 202 (Fig. 1) were cleaved with *Eco*RI; the fragments were separated on a 0.8% agarose gel and stained with ethidium bromide (*Left*). The digests from parallel gels were transferred to nitrocellulose filters and hybridized with the following labeled *Eco*RI fragments of λXlv 201 subcloned in pBR322: 1.5-kb, 3.8-kb, 6.4-kb, and 6.8-kb fragments (see Fig. 1). Hybridization and washing were as described (4). The arrowhead points to the 0.4-kb short band that gives a weak signal. m, Phage λ DNA digested with *Hin*dIII as size marker.

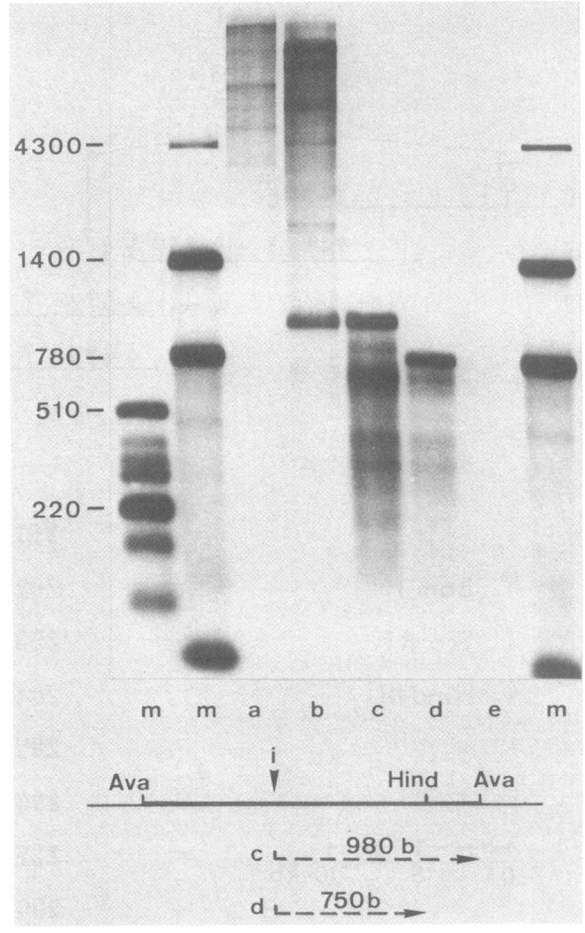

FIG. 3. Gel analysis of the *in vitro* transcription products of the vitellogenin *A1* gene. Lanes: a, transcription in Manley's extract in the presence of an *Ava* I digest of λ Charon 4 DNA; b, *Ava* I digest of λXlv 110 DNA; c, isolated 5'-end 1.6-kb *Ava* I fragment; d, isolated 1.6-kb *Ava* I fragment truncated by *Hin*dIII; e, isolated 1.3-kb *Ava* I fragment, which is adjacent to the 1.6-kb *Ava* I fragment in the 3' direction; m, size markers (base pairs). The scheme represents the run-off transcripts and the position of the initiation site (i) of *in vitro* transcription. b, Bases.

used for *in vitro* transcription. Run-off transcripts were detected by gel electrophoresis and autoradiography. Fig. 3, lane b, shows an *A1*-specific 980-base transcript that was not found when an *Ava* I digest of λ Charon 4 alone was transcribed (Fig. 3, lane a). To determine which *Ava* I restriction fragment bears the transcription initiation site, two contiguous *Ava* I fragments of 1.6 and 1.3 kb were isolated. These fragments mapped in the region where the putative 5' end of the gene had previously been localized by heteroduplex and R-loop analysis in the electron microscope (5). The 1.6-kb *Ava* I fragment generated the run-off product of 980 bases (Fig. 3, lane c and scheme) found after transcription of the whole λXlv 110 *Ava* I digest (Fig. 3, lane b). When shortened by *Hin*dIII endonuclease digestion, this fragment generated a discrete RNA transcript corresponding to the expected size of 750 bases (Fig. 3, lane d and scheme). In contrast, the 1.3-kb fragment, which is contiguous to the 1.6-kb fragment on the 3' side, was a poor template for *in vitro* transcription and gave no discrete products (Fig. 3, lane e).

Fig. 4 shows the results obtained with the *B1* gene. In a preliminary experiment, the putative 5' end of the *B1* gene was mapped within the 6.4-kb *Eco*RI endonuclease fragment (Fig. 1), with hybridization of a subcloned DNA fragment containing the 5' end of the related *A2* vitellogenin gene. The 6.4-kb *Eco*RI fragment of the *B1* gene was subcloned, isolated, and tested in the *in vitro* transcription system. A run-off transcript of 1,600 bases was synthesized (Fig. 4, lane a) and, based on this length, the position of the transcription initiation site was localized (Fig. 4, and scheme). Confirmation of this position was obtained from transcripts of the same *Eco*RI fragment truncated with *Bam*HI (Fig. 4, lane b and scheme) or *Ava* I (Fig. 4, lane c and scheme). The putative 3' end of the *A1* gene has been mapped earlier (ref. 5; Fig. 1), and the putative 3' end of the *B1* gene was localized by hybridization of ^{32}P-labeled 3'-end cDNA clones to different restriction endonuclease digests of *B1* clones.

Taken together, these results define the arrangement of the two linked genes on the continuous 67 kb of isolated DNA: 5'–2.5 kb of flanking sequences–gene *A1* of 22 kb–intergene distance of 15.5 kb–gene *B1* of 16.5 kb–11 kb of flanking sequences–3'. The two genes are in the same orientation and, therefore, are transcribed from the same DNA strand.

Electron Microscopic Analysis of the Relatedness Between the Two Linked Genes. The divergence between coding sequences of the *A* and *B* genes has been reported to be at least 20% based on hybrid melting experiments with cloned cDNAs (2), and the difference between corresponding introns in the *A1* and *B1* genes is expected to be much higher than 20%. Heteroduplexes were formed between the related *A1* and *B1* genes and analyzed in the electron microscope to gain information on the distribution of the most conserved regions within the genes and their flanking regions. Because of the extent of divergence between the two genes, stable heteroduplexes could be obtained only under much less stringent conditions than those usually used in such analyses. Heteroduplexes were formed between the following λXlv clones, vector DNA included: 109·201, 106·204, 104·207. These three pairs cover the whole *A1* and *B1* genes, including flanking sequences (Fig. 1; ref. 4). As an example, Fig. 5 shows a heteroduplex between λXlv 106 and 204. The *A1*·*B1* heteroduplexes contain regions of sufficient sequence homology to form stable associations interspersed with numerous often asymmetric heterology loops, where the divergence is such that pairing is no longer possible.

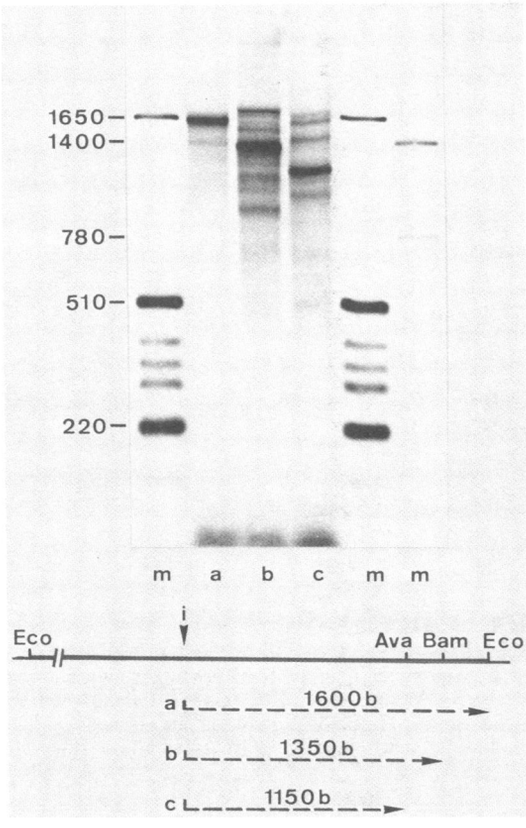

FIG. 4. Gel analysis of the *in vitro* transcription products of the vitellogenin *B1* gene. Transcription in Manley's extract in the presence of the isolated 5'-end 6.4-kb *Eco*RI fragment of λXlv 201 (lane a), the 6.4-kb *Eco*RI fragment truncated by *Bam*HI (lane b) and by *Ava* I (lane c), and size markers (lanes m). The scheme represents the run-off transcripts and the position of the initiation site (i) of *in vitro* transcription. b, Bases.

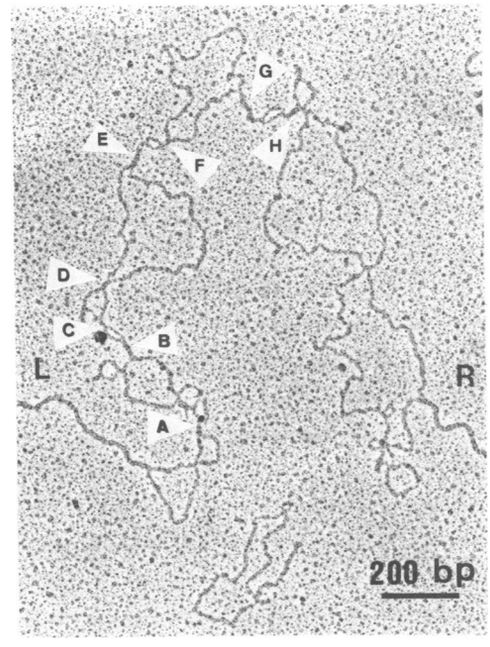

FIG. 5. Electron micrograph of an heteroduplex between the *A1* genomic clone λXlv 106 and the *B1* genomic clone λXlv 204 (Fig. 1). L and R refer to the left and right arms of the λ phage Charon 4 vector, which are only partially shown. The heterology loops proximal to the R and L arms are at the boundary between phage Charon 4 and *Xenopus* DNA. Arrowheads: Small regions of homology (A–H paired regions) between *A1* and *B1* sequences interspersed with symmetric or asymmetric heterology loops of various sizes representing regions of insufficient homology to form stable duplexes. bp, Base pairs.

From each of the three pairs of heteroduplexes, 15 molecules were measured, and the regions of homology were mapped and are shown schematically in Fig. 6. Several regions of homology, preferentially localized in the 5' side, can be seen between the two genes. Although there is at least one region of significant homology in the 5'-flanking region, no such region can be detected in the 3'-flanking sequences. Control experiments were performed with DNAs where we do not expect to find paired regions in the cloned genomic material (pairs λXlv 101·201 and 101·202 containing genomic DNAs with homologous regions inserted in opposite orientation in the vector). No paired regions were found between these DNAs, indicating that the hybrids we observe in Fig. 5 are true homologies and not collapsed regions due to the low stringency used.

DISCUSSION

The complete *B1* vitellogenin gene including its flanking regions was isolated from two different *Xenopus laevis* gene libraries. Sequences from the 3' region of the gene were obtained from an amplified library (4) in which sequences from the 5' region were underrepresented. Therefore, the latter were isolated from an unamplified library (7). The three clones containing the 5' end of the *B1* gene and flanking sequences (λXlv 201, 202, and 203) are recombinants which grow very poorly. This characteristic may well explain why these DNA fragments were not found in the amplified gene library. Presumably they were diluted by fast-growing phages in the amplification step. Unexpected frequencies of particular gene sequences in different libraries have been observed by us and by others (7, 11–14).

So far no differences have been observed in the maps of the restriction endonuclease cleavage sites in the overlapping sequences of clones isolated from the embryonic DNA library and the liver cell DNA library. This suggests, for the gene regions where direct comparison is possible, that no gross rearrangement takes place during metamorphosis when the genes become inducible (15, 16). *In vitro* transcription from DNA fragments bearing the 5' end of the genes has proved to be very useful to localize with precision the 5' end of the *A1* and *B1* genes. Because the promoters of the two genes are quite strong in whole-cell extracts and are faithfully used (unpublished data), they may be good candidates to study *in vitro* transcription in homologous systems derived from induced and uninduced *Xenopus laevis* hepatocytes.

Studies on DNA content, number of chromosomes, and immunological distances between albumins of different species of the genus *Xenopus* led to the suggestion that a whole genome duplication occurred in *Xenopus laevis* about 30 million years ago (17, 18). Such an event would have produced pairs of related genes, each one on a separate chromosome, and pairs of closely related genes have indeed been observed in *Xenopus laevis* (2, 4, 7, 19, 20). However, the extent of divergence between *A* and *B* vitellogenin coding sequences (20%) suggests that the *A* and *B* sequences first arose by the duplication of an ancestral vitellogenin gene before the genome duplication. If that is the case, the combination of the two duplications would have resulted in the following situation: one *A* gene and one *B* gene linked on one chromosome and the other *A* and *B* genes linked on a different chromosome. The divergence between the *A1* and *A2* genes and between the *B1* and *B2* genes of only about 5%, combined with the described linkage between the *A1* and *B1* genes, agrees well with this prediction.

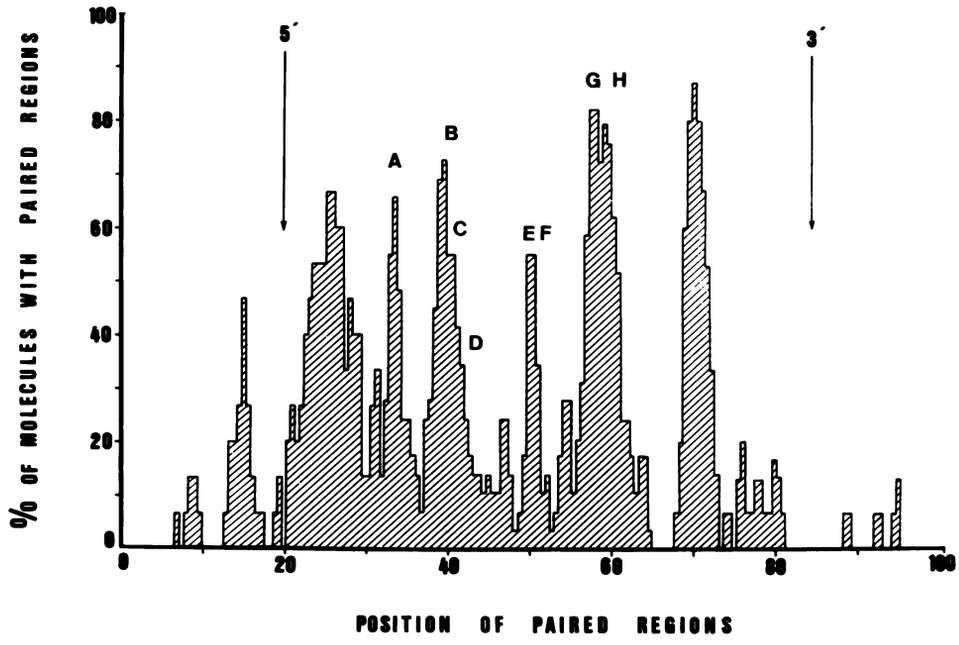

FIG. 6. Diagram showing the regions of homology between the *A1* and *B1* genes. The individual maps of 45 heteroduplex molecules between three pairs of genomic clones covering the whole genes with flanking regions were drawn and summarized (λXlv 109·201, 106·204, 104·207; see Fig. 1 and ref. 4). The total length of the genomic DNA of the two genes covered by the three overlapping pairs was normalized (100%), and the positions of the paired regions were plotted. The position of the 5′ and 3′ ends are indicated by arrows. The letters A–H indicate the position of the paired regions observed in Fig. 5.

The analysis in the electron microscope reveals that, even under conditions of low stringency, only short regions of the two genes were homologous enough to pair. These regions presumably represent exon sequences, which are known to diverge less rapidly than intron sequences as already observed for the *A1* and *A2* genes (5). Rapid divergence, especially in introns, would reduce the possibilities for unequal crossing over and might have played an important role in maintaining linkage between the *A1* and *B1* genes. Members of other multigene families coding for proteins have been found to form relatively stable clusters. The linked *A1* and *B1* genes, though producing mRNAs of similar length (6.3 kb) and sequence, are very different in size, *A1* being about 6 kb longer than *B1*. This observation again underlies the importance of deletions and insertions in the rapid divergence of the introns of related genes (5, 21).

The vitellogenin genes *A1* and *B1*, as well as *A2* and *B2*, are expressed coordinately after estrogen stimulation (2, 3). Nothing is known about the influence, if any, of the structural organization of the vitellogenin gene family on its regulation by estrogen. This coordinate expression might be a selective pressure that helped to maintain the two genes in tandem. It will be of interest in this context to study the arrangement and localization of the two other members (*A2*, *B2*) of the gene family.

We are grateful to Dr. Otto Hagenbuechle for the gift of HeLa cell extracts, Mr. Markus Haenggi for technical assistance, Dr. Guy Mayor for the computer program to measure DNA molecules, and Dr. John Knowland for critical reading of the manuscript. This work was supported by the Etat de Vaud and by the Swiss National Science Foundation.

1. Wahli, W., Dawid, I. B., Ryffel, G. U. & Weber, R. (1981) *Science* **212**, 298–304.
2. Wahli, W., Dawid, I. B., Wyler, T., Jaggi, R. B., Weber, R. & Ryffel, G. U. (1979) *Cell* **16**, 535–549.
3. Felber, B. K., Maurhofer, S., Jaggi, R. B., Wyler, T., Wahli, W., Ryffel, G. U. & Weber, R. (1980) *Eur. J. Biochem.* **105**, 17–24.
4. Wahli, W. & Dawid, I. B. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1437–1441.
5. Wahli, W., Dawid, I. B., Wyler, T., Weber, R. & Ryffel, G. U. (1980) *Cell* **20**, 107–117.
6. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* **15**, 687–701.
7. May, F. E. B., Weber, R. & Westley, B. R. (1982) *Nucleic Acids Res.* **10**, 2791–2806.
8. Ullrich, A., Shine, J., Chirgwim, J., Pictet, R., Tischer, E., Rutter, W. J. & Goodman, H. M. (1977) *Science* **196**, 1313–1319.
9. Manley, J. L., Fire, A., Cano, A., Sharp, P. A. & Gefter, M. L. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3855–3859.
10. McMaster, G. K. & Carmichael, G. C. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4835–4838.
11. Perrin, F., Cochet, M., Gerlinger, P., Cami, B., LePennec, J. P. & Chambon, P. (1979) *Nucleic Acids Res.* **6**, 2731–2748.
12. Fritsch, E. F., Lawn, R. M. & Maniatis, T. (1980) *Cell* **19**, 959–972.
13. Dolan, M., Sugarman, B. J., Dodgson, J. B. & Engel, J. D. (1981) *Cell* **24**, 669–677.
14. Herrlich, P., Hynes, N. E., Ponta, H., Rahmsdorf, U., Kennedy, N. & Groner, B. (1981) *Nucleic Acids Res.* **9**, 4981–4995.
15. Knowland, J. (1978) *Differentiation* **12**, 47–51.
16. Huber, S., Ryffel, G. U. & Weber, R. (1979) *Nature (London)* **278**, 65–67.
17. Bisbee, C. A., Baker, M. A., Wilson, A. C., Hadji-Azami, I. & Fischberg, M. (1977) *Science* **195**, 785–787.
18. Thiébaud, C. H. & Fischberg, M. (1977) *Chromosoma* **59**, 253–257.
19. Westley, B., Wyler, T., Ryffel, G. U. & Weber, R. (1981) *Nucleic Acids Res.* **9**, 3557–3574.
20. Widmer, H. J., Andres, A. C., Niessing, J., Hosbach, H. A. & Weber, R. (1981) *Dev. Biol.* **88**, 325–332.
21. Konkel, D. A., Maizel, J. V., Jr., & Leder, P. (1979) *Cell* **18**, 865–873.