Isolation of two closely related vitellogenesis genes, including their flanking regions, from a Xenopus laevis gene library

(recombinant DNA/restriction endonuclease mapping/gene family)

WALTER WAHLI AND IGOR B. DAWID

Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205

Communicated by Maxine F. Singer, December 27, 1979

ABSTRACT A gene library of Xenopus laevis was constructed from embryonic DNA partially digested with restriction endonucleases Hae III and Alu I and joined to the phage λ Charon 4 cloning vector with EcoRI linkers. Nucleotide sequences from three of the four related vitellogenesis genes have been isolated. Two of the genes (called A1 and A2) were isolated in their entirety together with long stretches of flanking sequences. These two closely related vitellogenesis genes have lengths of about 21 and 16 kilobases, but both produce a vitellogenin mRNA of 6.3 kilobases.

In amphibians and other oviparous vertebrates, vitellogenin, the precursor of the yolk proteins, is synthesized in the liver under the control of estrogen. We and others have shown earlier that, in Xenopus laevis, hormonal induction is followed by extensive accumulation of stable vitellogenin mRNA in the hepatocytes (1, 2); in fully induced liver, vitellogenin mRNA accounts for almost 50% of the poly(A)-containing RNA (3, 4). Consequently, 70–90% of the newly synthesized protein is vitellogenin (refs. 5 and 6; for reviews see refs. 7 and 8). Cloning experiments involving cDNA derived from purified vitellogenin mRNA have led us to the conclusion that the vitellogenin mRNA population is composed of four related RNAs of identical length (9, 10). We called the four mRNAs A1, A2, B1, and B2. The A and B groups exhibit a sequence difference of about 20%; the sequence difference between A1 and A2 or B1 and B2 mRNA sequences is only about 5%. All four mRNAs code for vitellogenins of similar molecular weights and are expressed simultaneously upon hormone treatment in individual animals (10, 11). Hybridization of cloned cDNA to Southern blots of uncloned genomic DNA digests showed that the different RNAs are transcribed from distinct genes (10). The biological implications of this “variant repetition” (12) are not known. We now report on the construction of a X. laevis gene library and on the isolation of the two A genes, including their flanking sequences.

MATERIALS AND METHODS

Preparation of a X. laevis Gene Library. The Xenopus library was constructed by using procedures described by Maniatis et al. (13). X. laevis embryonic DNA with a mean length of over 100 kilobases (kb) (gift from Steven McKnight) was prepared from embryos obtained from three independent matings. The DNA was fragmented by partial digestion with restriction endonucleases Alu I and Hae III. The DNA pooled from four independent Alu I and four independent Hae III digests was fractionated on 10–35% sucrose gradients (13), and DNA fragments of 15–24 kb were collected. This DNA was treated with EcoRI methylase (gift from Robert Rubin and Paul Modrich; see ref. 14) in 0.1 M Tris-HCl, pH 8.0/10 mM EDTA/6 μM S-adenosyl-L-methionine/5 mM dithiothreitol/200 μg of bovine serum albumin per ml/105 μg of DNA per ml/77 ng (2 nM) of EcoRI methylase per ml for 80 min at 37°C (14). After two phenol extractions and two ethanol precipitations, the methylated DNA was dissolved in 5 mM Tris-HCl, pH 7.5, and dialyzed for 7 hr against the same buffer. Before joining to the Xenopus DNA, octameric EcoRI linkers (Collaborative Research, Waltham, MA) were phosphorylated to convert the 5′-hydroxyl ends to 5′-phosphate ends (15). After the kinase reaction at 37°C, the reaction mixture was cooled to 16°C over 6 hr to allow renaturation of the linkers. The reaction mixture was then added to the Xenopus DNA and the ligation reaction was conducted at 8°C for 27 hr in 66 mM Tris-HCl, pH 7.6/10 mM MgCl2/1 mM spermidine/15 mM dithiothreitol/1 mM ATP/12 μg of EcoRI linkers per ml/150 μg of Xenopus DNA per ml/20 Weiss units of phage T4 ligase (New England Biolabs) per ml. The reaction mixture was extracted with phenol, and the DNA was precipitated with ethanol, dissolved in 100 mM Tris-HCl, pH 8.0/10 mM EDTA, heated 10 min at 60°C, loaded on 10–35% sucrose gradients, and centrifuged as described (13). DNA fragments bigger than 15 kb were collected and digested with a large excess of EcoRI. The left and right arms of the Charon 4 vector were prepared as described (13), and a reaction mixture containing 300 μg of arms and 125 μg of Xenopus DNA per ml in ligase buffer without ATP, dithiothreitol, and ligase was heated for 1 hr at 42°C to anneal the phage cohesive ends. After cooling down to 8°C, ATP, dithiothreitol, and ligase were added to 1 mM, 15 mM, and 100 Weiss unit/ml, respectively, and incubated for 24 hr. The in vitro packaging extracts were prepared and handled as described (15). The packaging reaction (3 μg DNA per tube) was performed exactly as described (15) and the mixture was plated on a fresh overnight culture of Escherichia coli KB02 at a density of 7–10 × 109 plaque-forming units (PFU) per 15-cm-diameter L broth plate. After an overnight incubation at 37°C the top agar containing the amplified phages was scraped and handled as described (13). The experiments were carried out under P2–E1K conditions in accordance with the NIH Guidelines for Recombinant DNA Research.

Screening of the Library. Screening was done on 150-mm L broth or NZCYM plates (13) containing 104 PFU by the in situ hybridization method (16). Plaques were lifted by two sequential applications of 132-mm BA85 Schleicher and Schuell nitrocellulose filters to plates precooled to 4°C. The first application was for 10–15 min and the second for 20–25 min, both at room temperature. The filters were dried for 1 hr at room temperature, treated with alkali, neutralized, and baked for 6 hr at 80°C. The filters were preincubated at room temperature

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.

Abbreviations: NaCl/Cit, 0.15 M NaCl/0.015 M sodium citrate; kb, kilobase(s) or kilobase pairs; PFU, plaque-forming unit.
for 10–15 hr in 50% formamide/0.6 M NaCl/200 mM Tris-HCl, pH 8.0/20 mM EDTA/0.1% sodium pyrophosphate/5 × Denhardt solution (17)/250 μg of E. coli tRNA per ml/100 μg of sheared denatured salmon sperm DNA per ml/0.2% sodium dodecyl sulfate (4 ml per filter). Vitellogenin cDNA clones (9, 10) were used as hybridization probes after nick translation to a specific activity of 0.5–1 × 10⁶ cpm/μg (18). Hybridization was performed with 1.5–2 × 10⁶ cpm of probe per 132-mm filter for 24 hr at 37°C, in the same solution as above (2.7 ml per filter) in a sealed plastic bag containing up to 25 filters. The filters were soaked sequentially in the hybridization solution in a petri dish before being introduced into the plastic bag. After hybridization, the filters were washed three times for 75 min each in 50% formamide/0.3 M NaCl/0.03 M sodium citrate/0.1% sodium dodecyl sulfate at room temperature, and one time for 45–50 min at 37°C in the same solution, followed by another wash of 75 min at room temperature. After being rinsed in 0.30 M NaCl/0.03 M sodium citrate (=2X standard saline/citrate, NaCl/Cit) the filters were dried and exposed with an intensifying screen without preflashing of the films. An exposure of 10–15 hr resulted in strong positive signals from the plaques containing vitellogenin sequences. Positive plaques were purified by two or three further screening cycles with platings at much lower density. Phages containing vitellogenin sequences were grown in liquid cultures and DNA was prepared essentially as described (13).

Complexity Analysis of the Library. The procedures described by Britten et al. (19) and Galau et al. (20) were used as follows. DNA from X. laevis blood cells or from total library was sheared to an average length of 300 nucleotides in a Virtis blender. Single-copy tracer was prepared from blood cell DNA by 68°C incubation of the denatured DNA to a Cot (concentration of DNA in mol of nucleotide per liter × incubation time in sec) of 1500 followed by isolation of the single-stranded material on hydroxyapatite. This DNA was then annealed to a Cot of 13,000, at 68°C, and the reassociated material was collected on hydroxyapatite, recovered by melting at 98°C, and again annealed to a Cot of 13,000. A sample of this DNA was labeled with [3H]ATP and [3H]CTP by gap translation. For measurement of reassociation rates, labeled DNA was mixed with at least a 1000-fold excess of sheared unlabeled blood cell DNA or library DNA and annealed in 0.4 M phosphate buffer at 68°C to the desired Cot. The samples were then fractionated on hydroxyapatite columns. Cot values were corrected to standard salt concentration (19). Cot values for reassociation of library DNA are expressed in terms of the concentration of Xenopus DNA inserts. The extent of reaction was normalized for reactivity of the tracer, which ranged from 57% to 78% in different experiments.

Restriction Endonuclease Analysis. Digestions with the different restriction endonucleases were done as recommended by the supplier (Bethesda Research Labs, Bethesda, MD).

Table 1. X. laevis DNA library

| DNA: embryonic, A1, A2, Hae III nonlimit digests, octamer | EcoRI linkers |
| Cloning vector: λ Charon 4 |
| Packaging efficiency of extracts for λ wild type: 4 × 10⁷ PFU/μg |
| λ Charon 4: 4 × 10⁸ PFU/μg |
| Packaging efficiency of Xenopus DNA: 1.5 × 10⁶ PFU/μg |
| Number of independent recombinant phage obtained: 0.9–1 × 10⁶ |
| Number of recombinant phage needed for a complete library: 8.85 × 10⁸ |
| Complexity of cloned DNA: similar to uncloned DNA (see Fig. 1) |
| Mean length of the cloned Xenopus DNA fragments: 16 kb |

![FIG. 1. Reassociation kinetics of X. laevis and library DNA. Labelled single-copy DNA was prepared from X. laevis blood cell DNA and annealed with an excess of unlabelled blood cell DNA (●, O) or library DNA (▲). O and ● represent two separate experiments.](image)

Heteroduplex Analysis. The DNAs (1.5 μg/ml of each clone) were denatured at 65°C for 5 min in 70% formamide/0.3 M NaCl/5 mM EDTA/10 mM Tris-HCl, pH 8.0 and subsequently reannealed in the same buffer for 90 min at 35°C. A sample was diluted 1:10 to a final concentration of 40% (vol/vol) formamide/1.8 M urea/0.1 M Tris-HCl, pH 8.5/10 mM EDTA/50 μg of cytochrome c per ml and spread on 5% formamide. Simian virus 40 and phage φX174 DNAs were included as double- and single-stranded length markers. Pictures were taken in a Philips EM 400 electron microscope.

RESULTS

Construction and Characterization of a Gene Library from X. laevis. The construction of this Xenopus library followed the strategy described by Maniatis et al. (13) and was made
possible by the development of convenient λ cloning vectors (21) and in situ packaging systems (22). In addition, the rapid in situ plaque hybridization procedure of Benton and Davis (16) allows the screening for single copy sequences in the complex Xenopus genome. The characteristics of our library are given in Table 1. The mean length of the cloned Xenopus DNA fragments was determined by electron microscopy from total library DNA and is 16 kb. This agrees well with the mean length of 15.9 kb obtained from 23 different inserts containing vitellogenin sequences (see below). Given the size of the X. laevis genome of 3.1 × 10⁹ base pairs per haploid set (23) and the mean length of the cloned fragments, a library of 8.85 × 10⁸ independent recombinant phages is needed to find any single copy sequence with a probability of 99% (24). Because we obtained 0.9–1 × 10⁶ independent recombinant phages, and because the single copy complexity of the library is similar to that of uncloned Xenopus DNA (Fig. 1), the library can be considered as complete. However, due to the assumptions made in this kind of analysis and the limited sensitivity of the measurements, we cannot exclude the possibility that certain regions of the genome are underrepresented in our library.

Isolation of Vitellogenin Gene Sequences. Our cloning experiments of vitellogenin cDNA (9) have led to the conclusion that vitellogenin in X. laevis is encoded in four related genes (10). In a first screening of the library a mixed probe of cDNA clones representing all four genes was used to isolate sequences from the different genes. Eighteen recombinant phages were obtained in the first screen, which tested 4 × 10⁸ phages. The clones were named λXIV (λ X. laevis vitellogenin) and numbered. To attribute each cloned genomic sequence to one of the four genes, we used the spotting technique applied earlier to demonstrate sequence divergence between different cDNA clones (10). The cloned genomic DNAs were spotted onto a nitrocellulose filter (16 spots per clone), and the filter was cut into four parts containing 4 spots per clone and hybridized separately to labeled cDNA probes representing the four genes. The filter sections were cut again into four parts and these strips were washed under conditions of increasing stringency. After washing the whole filter was assembled again and autoradiographed. Thus, each cloned DNA was tested with each of the four labeled probes at four levels of stringency. The results are shown in Fig. 2. Under the least stringent conditions only a few if any A/B crosshybrids are seen due to the presence of a 20% difference between the A and B structural sequences. Thus,
restriction endonuclease digests

**Fig. 4.** Length polymorphism in the A1 gene. The figure shows restriction endonuclease digests of several cloned DNAs separated on an agarose gel. Lane 1, XXlv106, HindIII; lane 2, 106, HindIII-Xho I double digest; lane 3, 106, Xho I; lane 4, XXlv105, Xho I; lane 5, 105, HindIII-Xho I double digest; lane 6, 105, HindIII; lane 7, ADNA digested with HindIII, marker; lane 8, 105, HindIII; lane 9, 106, HindIII (see also Fig. 3). Complete arrows point to homologous HindIII-Xho I double digest fragments showing length heterogeneity: 1.55 kb in XXlv106, lane 2, and 1.2 kb in XXlv105, lane 5. Arrowheads indicate homologous intragenic HindIII fragments showing length heterogeneity: 8.3 kb in XXlv106 and 7.95 kb in XXlv105.

**FIG. 5.** Heteroduplex between XXlv123 and XXlv127. L and R indicate the left and right arms of the phage vector, which continue beyond the figure. The two large deletion loops at the boundary between phage and Xenopus DNA are due to the fact that the two clones overlap in only part of their length. The arrow points to the small deletion loop caused by length heterogeneity within the cloned Xenopus sequences.

additional clones not shown in Fig. 2. XXlv108, 109, 110, 128, and 129 were found in later screenings with labeled cDNA clones as probe, and XXlv126 and 127 were found by screening the library with the 5ʹ proximal HindIII fragment of XXlv125; the latter screen is an example of "chromosomal walking" along the DNA strand in the 5ʹ flanking region. The ten A1 clones and the nine A2 clones cover 38 and 42.5 kb, respectively. The vitellogenin mRNA of 6.3 kb (9, 25) is drawn in the same scale at the top of the figure. The arrows originating at the 5ʹ and 3ʹ ends of the mRNA point to the positions on the genomic clones beyond which A1 and A2 gene sequences diverge sufficiently so that no A1-A2 heteroduplexes form. The results of this heteroduplex mapping as well as R-loop mapping with mRNA that confirmed these positions will be presented in detail elsewhere. Considering the sequence divergence points as tentative 5ʹ and 3ʹ ends of the structural genes, the lengths of the A1 and A2 genes are 21 and 16 kb, respectively. XXlv128 was found in a round of screening with a cDNA clone corresponding to the 3ʹ end of the A2 gene. This recombinant phage turned out to be a fortunate find because it contains a complete copy of the A2 gene, assuming that our tentative assignments of the ends of the gene are correct.

As shown in Fig. 3, the restriction maps derived from different clones agree well in their overlapping regions, with some exceptions. XXlv103 and 106 have an additional HindIII and BamHI site, respectively. Also in XXlv106, the large intragenic HindIII fragment (8.3 kb) is 350 base pairs longer than the corresponding fragment (7.95 kb) in XXlv105 (Fig. 3 and Fig. 4, arrows in lanes 8 and 9). Xho I/HindIII double digests of these two clones show that the resulting small Xho-Hind fragment measures 1.55 kb in XXlv106 and 1.2 kb in XXlv105 (Fig. 4, arrows in lanes 3 and 5). XXlv109 and XXlv110 are identical to XXlv106 in this region, and XXlv107 is identical to XXlv105. Electron microscopic analysis of heteroduplexes between these two types of clones shows a small deletion loop in the region corresponding to the small Xho I/HindIII fragment. This length difference appears to occur within an intron (unpublished data). A similar kind of observation was made in the 5ʹ flanking sequence of the A2 gene. XXlv126 and 127 have an additional BamHI site to the left of the Sal I site. Due to this additional site the 3.35-kb BamHI fragment from this region in XXlv124 and 125 is replaced by two BamHI fragments of 1.4 kb and 2.3 kb (total 3.7 kb) in XXlv126 and 127. Heteroduplex analysis between the two types of clones revealed, in analogy to the sit-
ution described for the A1 gene, a small deletion loop at the position of the additional BamHI site in λXMv126 and 127 (Fig. 5).

DISCUSSION
This study illustrates some of the advantages of the gene library approach in the isolation of large genes forming a small family. In a single screening experiment with a mixed cloned cDNA probe we isolated sequences of three of the four members of the family. We then focused our work on the complete isolation of the two A genes, including their flanking sequences. The isolation and characterization of two large sets of overlapping clones demonstrate that, in the region studied, our library behaves as expected and appears to be effectively complete. Also, the identity of the restriction maps in their overlapping parts suggests that the cloned sequences are indeed representative of uncloned genomic DNA. Because the A1 and A2 structural gene sequences are closely related (10), it would have been more difficult because of crosshybridization to reach such a strong conclusion from hybridization experiments using restriction digests of total uncloned genomic DNA. Taking into account that the cloned DNA was derived from embryos pooled from three matings, the level of heterogeneity in the restriction pattern that could be due to population polymorphism is relatively low. Such polymorphism may become more apparent as the sites of more enzymes are mapped. The length heterogeneity found at a specific position within the A1 gene and in the 5' flanking region of the A2 gene is not fully understood. At the present time we cannot decide with certainty whether we deal with population polymorphism or with a cloning artifact. If the latter were true, the results would strongly suggest localized deletion (or insertion) hot spots, because these length variations occur in several cloned DNAs at the same position. Because of the above requirement and because both length differences appear to occur in regions that do not code for protein, we believe that the effect is most likely due to polymorphism in the Xenopus population.

We had suggested earlier by blotting experiments in which uncloned DNA was hybridized with cDNA clones that the vitellogenin genes are split (10), as are numerous other eukaryotic genes (for a review, see ref. 12). The finding that the A1 and A2 genes are 3.3 and 2.5 times longer than the mRNA supports and extends these observations. Studies on nuclear RNA molecules that are interpreted as precursors to vitellogenin mRNA have shown the presence of 12 introns in a portion of the A1 gene (26). The arrangement and sizes of the many introns in the two A genes will be presented in a future publication. The vitellogenin mRNA population is homogenous in size (1, 25), and R-loop experiments in the electron microscope with A1 and A2 cDNA clones have more directly demonstrated that A1 and A2 mRNAs are of similar length (9, 10). The length difference of 5 kb found between the two genes must therefore be accounted for by differences in the introns. It is not yet known if the vitellogenin genes are linked. In analogy to some other multigene families (27, 28), it would not be surprising if they are linked. From the present data we can already conclude that, if the two genes are linked, the intergene distance must be larger than 31 kb if the order is (5')A1-A2(3') and larger than 11.5 kb if the order is (5')A2-A1(3').

We thank Drs. S. McKnight, R. Rubin, P. Modrich, F. Blattner, and L. Enquist for gifts of materials, Drs. L. Enquist and R. Kramer for advice, Dr. T. Maniatis for sending us his work before publication, and Dr. M. Singer for comments on the manuscript. This work was supported in part by a fellowship from the Swiss Academy of Sciences to W.W.