Volume 12 Number 20 1984 Nucleic Acids Research

The primary structure of the larval β_1-globin gene of *Xenopus laevis* and its flanking regions

Wolfgang Meyerhof[1], Stella Klinger-Mitropoulos[1], Jürg Stalder[2], Rudolf Weber[2] and Walter Knöchel[1]

[1]Institut für Molekularbiologie und Biochemie, Freie Universität Berlin, Arnimallee 22, D-1000 Berlin 33, FRG, and [2]Zoologisches Institut, Abteilung für Zell- und Entwicklungsbiologie, Universität Bern, Baltzerstrasse 4, CH-3012 Bern, Switzerland

Received 23 July 1984; Revised and Accepted 26 September 1984

ABSTRACT

We present the complete nucleotide sequence of the larval β_I-globin gene of Xenopus laevis including 240 nucleotides of the 5' flanking region and 594 nucleotides beyond the polyadenylation site. The site of transcription initiation was mapped by S_1 nuclease, and the site of polyadenylation was determined by comparison with corresponding cDNA clones.

The larval Xenopus β_I-gene shows the same internal structure as the β-globin genes of higher vertebrates, viz. 3 exons interrupted by 2 intervening sequences. The first intervening sequence, which is of exceptional length, spans over 564 nucleotides and interrupts the coding sequence at amino acid 30, whereas the second one comprises 968 nucleotides and is located between the amino acids 104 and 105. The second intervening sequence contains a long inverted repeat of almost perfect homology.

The 5' flanking region contains a TATA- and a CAAT-box at positions -33 and -58, respectively. An additional TATA-box is located at -197 and two more CAAT-boxes occur at positions -105 and -237.

INTRODUCTION

Sequential changes in hemoglobin synthesis are a common feature of vertebrate development and reflect differential expression of the globin genes which encode the polypeptide subunits of hemoglobin. In the South African Clawed Toad (Xenopus laevis) the switch from larval to adult hemoglobin occurs at metamorphosis (1) and is characterized by the complete replacement of the larval by adult globin chains (2, 3). Therefore the Xenopus globin system represents an attractive model for investigating expression of developmentally regulated genes.

Analysis of cDNA clones, derived from anemic animals, revealed four larval and four adult globin sequences, which in each stage comprise two closely related α- and two closely re-

© IRL Press Limited, Oxford, England.

lated ß-sequences (4). In contrast to higher vertebrates, whose α- and ß-globin genes are arranged in clusters on different chromosomes, the α- and ß-genes of Xenopus laevis are linked (5, 6). According to a recent study the Xenopus laevis globin gene family comprises at least 12 genes, which are symmetrically arranged in two clusters. Each cluster carries the linked adult α- and ß-genes which at the 5' end are flanked by two larval α-genes and at the 3' end by two larval ß-genes (7).

To date the nucleotide sequences are available of cDNA clones corresponding to the adult α- and ß-sequences as well as those of the larval α_I and $ß_I$-sequences (8-12). Moreover, the complete sequence of the adult $ß_I$-gene was reported by Patient et al. (13), and part of the adult α_I-gene was sequenced by Partington and Baralle (14).

Comparative studies on the fine structure of Xenopus laevis globin genes may provide information on the role of DNA sequences in the developmental control of gene expression. Since the Xenopus laevis globin genes are as yet the only globin genes available from lower vertebrates, they are also of interest in the study of gene evolution. To elucidate further the molecular basis of developmentally controlled expression of the Xenopus laevis globin genes, it was of interest also to analyze the structure of a larval gene with its flanking regions. In this paper we now present the complete nucleotide sequence of the larval $ß_I$-gene including 240 nucleotides of the 5' and 594 nucleotides of the 3' flanking regions together with the amino acid sequence of the corresponding polypeptide.

MATERIALS AND METHODS

Isolation of phage and plasmid DNAs

λ recombinant phage DNA was isolated by the glycerol step gradient procedure (15). Plasmid DNAs and phage M13 RF DNAs were prepared by the rapid alkaline extraction method (16) and further purified on CsCl gradients. Single stranded M13 phage DNA was prepared by polyethylene glycol precipitation and subsequent phenol extraction. Restriction enzyme digestions and ligation of DNAs were performed according to the recommenda-

tions of the suppliers (BRL, Boehringer).

Subcloning

λXG4 DNA was digested with Eco RI and ligated into the Eco RI site of plasmid pUC8. After propagation of bacteria the DNA of the clone containing the larval $ß_I$-globin gene was isolated and a restriction map was established. For subcloning in M13 phages appropriate restriction fragments were electrophoretically separated on low melting temperature agarose gels and ligated without removal of agarose (17) into the corresponding cleavage sites of M13 mp8 or M13 mp9 DNA (18).

Sequence analysis

Nucleotide sequences were determined by the dideoxy chain termination method (19) using sets of deletion mutants which had been isolated by nuclease Bal 31 digestion (20).

Mapping mRNA start point

Cytoplasmic RNA was extracted from red blood cells of anemic tadpoles (4) at stages 53 - 54 (21). Polyadenylated RNA was purified by oligo(dT)-cellulose chromatography (22). Radiolabelled probes for nuclease S_1 mapping (23) were prepared by the prime cut method according to the protocol given by Ackerman (24). We used a DNA fragment extending from the Pvu II site in exon I to the 5' Eco RI site (clone 841P1). DNA was digested with Hinf I, which cuts at position -47. The labelled DNA strand was isolated from a denaturing polyacrylamide gel and annealed (approximately 5×10^4 cpm) in 50 % formamide at 44 ^oC to RNA isolated from a total of 10 tadpoles. S_1-nuclease digestions (BRL) were performed at 20 ^oC or 37 ^oC, respectively. Protected fragments were run together with control reactions and with sequencing reactions as size markers on 0.4 mm polyacrylamide gels (25).

Preparation, identification and sequencing of cDNA clones

cDNA clones from mRNA of tadpole red blood cells were prepared exactly as described for adult animals (9). Clones pVD12 and pVC9 were identified to encode larval $ß_I$-globin structural gene sequences by hybridization with a labelled probe derived from plasmid p41 (see Figure 1). Restriction fragments to be sequenced were either 5' end labelled by polynucleotide kinase with ^{32}P-γ ATP or 3' end labelled by terminal transferase with

^{32}P-cordycepin and, after cleavage with a second restriction enzyme, submitted to the chemical sequencing reactions (26).

RESULTS AND DISCUSSION

Sequencing strategy

The larval β_I-globin gene of *Xenopus laevis* was isolated from a genomic DNA library (27). Figure 1 shows the restriction map of the recombinant phage λXG4 (7). The 5.1 KB Eco RI fragment containing the gene was cloned in plasmid p41. Four fragments have been subcloned in phage M13 mp8 to obtain initial data of nucleotide sequences. According to the restriction map clone 841P1 contains the 2.5 KB Pvu II/Eco RI fragment, clone 841P2 the 1.4 KB Pvu II fragment, clone 841B1 the 3.5 KB Bgl II/Eco RI fragment and clone 841B2 the 1.6 KB Bgl II/Eco RI fragment, respectively. The inserts of the clones 841B1 and 841B2 were trimmed with the exonuclease Bal 31 from their 3' and 5' ends and recloned into M13 mp8 or M13 mp9. Overlapping clones of both strands were sequenced using the dideoxy chain termination method (19).

841P1 was used to prepare the labelled probe for mapping the mRNA start point and 841P2 was isolated to determine the nucleotide sequence surrounding the Bgl II site.

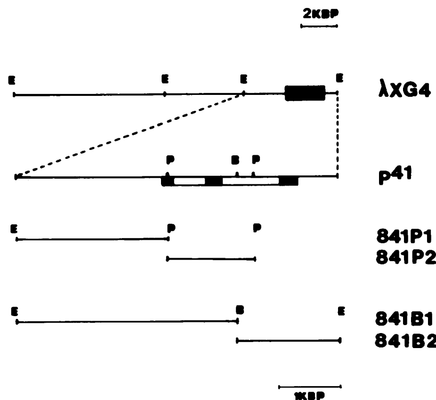

Figure 1: Genomic clone (λXG4) and derived subclones used for sequencing.
Restriction enzyme cleavage sites are (E) Eco RI, (B) Bgl II, (P) Pvu II. The solid boxes in p41 represent exons, and the open boxes indicate intervening sequences.

-240 -230 -220 -210 -200 -190 -180 -170

TCACAATTGCCCTCATTTGATGTGGTTTAAGTTTCATGTTGTTATAAAGAATCAACTTTACAATTTAAGAACTATATGG

-160 -150 -140 -130 -120 -110 -100 -90

CATTCCACATATACAAAAGATATATTAGCTTAAGGTTAAAAATTTATTTTGAAGGCAATAGGGTGGGGTGGAGGAAAAAA

-80 -70 -60 -50 -40 -30 -20 -10

AATATGACACAGCAGAAATGCACAATGGGTGTGACTCAGCATGGCCATATAAAGCAAGGCCAACAACTCAAAGGAACAGC

Hinf I

* 10 20 30 40 50 VAL HIS LEU SER ALA

AGCCTCTTACTTCTGCCATTTGAAGTCTTAGCAGCTACTCCCATCTACAGCCACC ATG GTG CAC CTG TCA GCT

Pvu II

ASP GLU LYS SER ALA ILE ASN ALA VAL TRP SER LYS VAL ASN ILE GLU ASN ASP GLY HIS

GAT GAG AAA TCC GCC ATT AAC GCC GTA TGG TCA AAG GTC AAC ATT GAA AAT GAT GGC CAT

ASP ALA LEU THR ARG 156 166 176 186 196 206

GAT GCC CTG ACC AG GTAACTTGAAGCACATTGCTAATTACTTACAAATAGTTTTATTTTAGGGGTTTTGAGAC

216 226 236 246 256 266 276 286

TGTTTGTATATGTGTTTGTTTGTATATGTGTTGATTACAAAAGCAGCAACATTAAACATAATCTCCTTATATGCTTGAAG

296 306 316 326 336 346 356 366

TTAATTTGGCCACATCCGTTGCGCTTCCTCTGTACCTGTAAATTTAACTTTTTCCATGTTGTGAATTTCAACATTAAATG

376 386 396 406 416 426 436 446

TAGCATGGTAATATAATGAAGGATAATTTAAAACAAAATATTTACATTCTGGGACAATGATGCCCAGCCAGAGGACCTTT

456 466 476 486 496 506 516 526

AGATGGTGCTGGACTATGAGAAGCACCATTTATATGTTGTAGAAACTAGTAGTTGCAGTCCAACAACACTCCTGTTGCCC

536 546 556 566 576 586 596 606

CATGAGTGATGAAATACATTAAAGCCTATTTTCTATACACTGATATAAAGGGTTTTGCATGATATAGTAAATATTGAGAA

616 626 636 646 656 666 676 686

AATATTAAGAATAAATACATTTATTTTTTGTTAGTGTAAATATTTCAGTACTCTTTAACATTGATAAGCTATTTATAACT

696 706 LEU LEU VAL VAL PHE PRO TRP THR GLN ARG TYR PHE SER

CATCAAATTTAAATTATTTCTACAG G CTG CTG GTT GTG TTT CCC TGG ACC CAG AGG TAT TTC AGC

SER PHE GLY ASN LEU SER ASN VAL ALA ALA ILE SER GLY ASN ALA LYS VAL ARG ALA HIS

AGC TTT GGA AAC CTA TCC AAT GTG GCT GCC ATC TCT GGA AAT GCC AAG GTT CGT GCC CAT

```
GLY LYS LYS VAL LEU SER ALA VAL ASP GLU SER ILE HIS HIS LEU ASP ASP ILE LYS ASN
GGC AAG AAG GTT TTG TCT GCC GTC GAT GAA TCC ATC CAC CAT TTA GAT GAT ATC AAG AAC

PHE LEU SER VAL LEU SER THR LYS HIS ALA GLU GLU LEU HIS VAL ASP PRO GLU ASN PHE
TTC CTC TCT GTA CTG AGC ACG AAG CAC GCT GAG GAA CTC CAC GTG GAC CCT GAA AAC TTC

LYS        944       954       964       974       984       994       1004
AAG GTACGTGTTATCACCTTATAATGTTCTTAGGCACACTTTTTCTGTTCAGGGTTTTTATATTATTTGAATTACAATA

     1020      1030      1040      1050      1060      1070      1080      1090
TATATAATTCAACACAATTCTGTATGAGAATAATAAAGAATCCCCCCAGTATTTATTTCATGATTACAGAATCTTTAATT

     1100      1110      1120      1130      1140      1150      1160      1170
AACAAAGTGCACCAGCAGAGCTAACCATATCAACCAATGAGCAATTTGTTTTGATCAGTGTACTTCAAATTAAAATATAA

     1180      1190      1200      1210      1220      1230      1240      1250
AGTCAAAGATCTGATTGGTTGCTATGGGTACCTTGTTGTGCTAGTAGATCAGTCTTTTAGTTTTTCATACTAATTTATTA
      Bgl II
     1260      1270      1280      1290      1300      1310      1320      1330
AAATGGAATCATTAATATCGACTAAAATGTAATGGAGGATAATTTGTTTTTGACAAATAGGTTTGTTATTATTATTTTTT

     1340      1350      1360      1370      1380      1390      1400      1410
CTTTTAATTGTTGACACATTATGGTTTATTTATAGGGTTTGTCAAACAGATTCAGTTTACTATAGTTTCGATTACGTTTT

     1420      1430      1440      1450      1460      1470      1480      1490
ATAATAACGGCCTCCCTTTAATGCACACATAGACACAGCTGTGCTTGAGAACTATAGTGAGATTTGGTTTGCTATTCGTG
                                    Pvu II
     1500      1510      1520      1530      1540      1550      1560      1570
TGTGCTTGCATCTAAAATGTACTCATGAGTTTTAAATTCATTTTAAGGTTTCGTTAATATAAATTTGTAATAAATATAAA

     1580      1590      1600      1610      1620      1630      1640      1650
TTACTTTCTGTTTGTATAATACACTTCCTAATTTGCATAAGTGAGTAGAGTGTCTCATTTATAAACAGTGGGCATATTTG

     1660      1670      1680      1690      1700      1710      1720      1730
CACCTGGGCAGTAACCCATAGAAACTATTCACTGGCTGGCTTTTTTTTTTTTGTCAGTTGCATTCAGTTGGTTTCATGGGT

     1740      1750      1760      1770      1780      1790      1800      1810
TACTGCCCAGGTGCACATTTGCCCAGTGTTTATAAATGAGCCCAATGCATTTTAAAATTAGCAGGCTAATAAGAAGAAAA

     1820      1830      1840      1850      1860      1870      1880      1890
GCGTTATTCTTTTATTATCAAAGGAAGGCTTTTCATATCTTAATGTTCTTTACCACTGCAATAAGTCTAACTCATGTTTG
```

```
     1900     ARG LEU ALA ASP VAL LEU VAL ILE VAL LEU ALA GLY LYS LEU GLY ALA
TCACTTTCACAG  CGT CTT GCT GAT GTG CTG GTG ATT GTC TTG GCT GGC AAA CTG GGA GCT

ALA PHE THR PRO GLN VAL GLN ALA ALA TRP GLU LYS PHE SER ALA GLY LEU VAL ALA ALA
GCC TTC ACT CCT CAA GTC CAG GCT GCC TGG GAG AAA TTC AGC GCT GGG CTG GTG GCT GCT

LEU SER HIS GLY TYR PHE          2040      2050      2060      2070      2080
CTT AGC CAT GGC TAC TTC TAA  AGAATTTCATTCCATGCAAAGCCCAATATCTGCTCGCCTGCAACAAAGA

       2090      2100      2110      2120      2130      2140      2150      2160
CAACTCTCACAATAAGTGCATTGCTTGTTGTCAATAAAGATTTCTTGTCAAACTGCTAAAAAAAATTGTTTTTCAATTAT

       2170      2180      2190      2200      2210      2220      2230      2240
TTCAAGAAGAATTAGAACATGTGTTTACCTAATACTGACAAGGTAAATTATTTCATACCAAAATACAATCTATTTTTATG

       2250      2260      2270      2280      2290      2300      2310      2320
ATATTGGTAATAAGATATTAAATTGTTATCAGATATTTAAACTTTAAGGAACCGTATATTAATGTGAATGTGAATGATAA

       2330      2340      2350      2360      2370      2380      2390      2400
AAAAAACATCCCATCCCTAATGTCATCATTACTGAAATATTATATAACACAGTGTAGATTTAATTACATTTTATTCAGTA

       2410      2420      2430      2440      2450      2460      2470      2480
TGTCACAAAATACCAACATTACTCTAACTTTAGCATTCCACAGCAATCTTGATATTATTGTTGCAATACCAGTTAAATAC

       2490      2500      2510      2520      2530      2540      2550      2560
ATACACAAATTATTTTTGCTATTCTAGTTAGATATATCTATGGTTTACGGCTAAATTCTATACTATATATTCCCTTAGCA

       2570      2580      2590      2600      2610      2620      2630      2640
CCTAACTATCTTTCTTTGGTATTTAGTTGCATTGTTTGCATCCCTTGTATCCCTTCAGAGTGAAACACAAGTAGCACCTT

       2650      2660      2670      2680      2690      2700      2710      2720
TTGGGGGACACGGATCAAACCTACCATGTGGTGCAGCATATAAATGTTTCTCCATTTACCAATTCTAACAAAAACTGATT

       2730
TGCATGGAATTC
      └────┘
       Eco RI
```

Figure 2: Nucleotide sequence of the Xenopus laevis larval β_I-globin gene.
The nucleotide sequence of the anticoding strand is shown from 5' to 3' including the flanking regions. The mRNA start point is indicated by an asterisk. The mRNA 3' terminus is indicated by a vertical arrow. Nucleotides which diverge from pXGßT1 (11) are underlinded, and putative signal sequences are boxed (for explanation see text). Within the coding region the amino acid sequence is given above the nucleotide sequence. Horizontal arrows indicate an inverted repeat within IVS II. Restriction sites, referred to in the text, are indicated.

Mapping of the larval $ß_I$-globin gene

The complete nucleotide sequence and the derived amino acid sequence of the larval $ß_I$-globin gene is shown in Figure 2. The start point of the mRNA was determined by S_1 nuclease mapping. From single-stranded DNA of the M13 clone 841P1 we synthesized a complementary DNA strand uniformly labelled with ^{32}P dCTP. This probe extends from the universal 17 nucleotide M13 primer to the Hinf I site located 14 nucleotides upstream of the TATA box. S_1 digestions of hybrids with larval globin mRNA were carried out at 20 °C and 37 °C using various amounts of enzyme (20, 200, 2000 and 10 000 U). In contrast to incubations at 37 °C more reproducible results were obtained at 20 °C with 2000 U of enzyme. As shown in Figure 3, a protected fragment of 71 nucleotides was found. This fragment extends from the Pvu II site to an A residue (designated by an asterisk in Figure 2). This corresponds to the rule that transcription is usually initiated at an adenine (28).

To determine the 3' end of exon III we compared the gene sequence with those from our cDNA clones pVD12 and pVC9 derived from larval $ß_I$-globin mRNA (unpublished) and with the published cDNA sequence pXGßT1 (11). However, the precise location of the 3' end of exon III was not possible, because 19 nucleotides 3' to the polyadenylation signal we find a T followed by a tract of eight A residues in the gene sequence; a T also precedes the poly(A) tail of the mRNA sequence. Therefore, each of the A residues or the preceding T could be used as the poly(A) addition site. However, since exon III of the adult $ß_I$-gene also ends with CTA and the A was suggested to be the polyadenylation site (13), the first A in the anticoding strand of the larval gene sequence may, indeed, have the same function.

The sequence of 2972 nucleotides shown in Figure 2 thus includes 240 nucleotides of the 5' flanking region, 2138 nucleotides of the gene and 594 nucleotides of the 3' flanking sequence.

The exons

The exons of the gene perfectly match to the sequence of our larval cDNA clones pVD12 and pVC9 (unpublished). In contrast, we note 3 base changes compared with the cDNA sequence

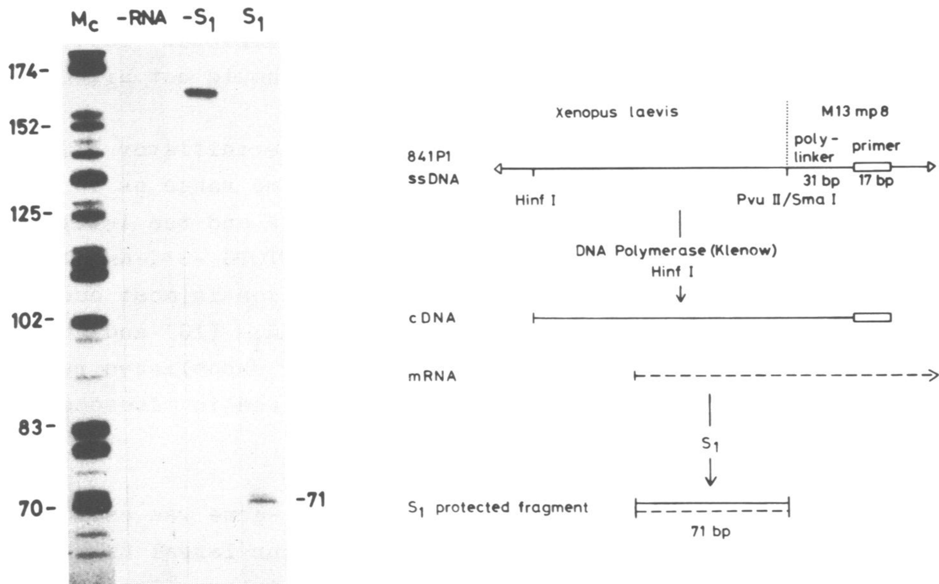

Figure 3: Determination of the mRNA start point.
The experimental strategy is schematically outlined. Polyadenylated RNA was isolated from red blood cells of anemic tadpoles. As labelled probe we used a single stranded transcript of template DNA from clone 841P1, which was isolated from a denaturing polyacrylamide gel after Hinf I cleavage. Including the 17 bp primer the M13 mp8 sequence to the Pvu II/Sma I cleavage site contains 48 nucleotides. Lane M_C shows the C-track of a sequencing reaction of a DNA fragment used as size marker (lengths in nucleotides). Lane S_1 shows the protected fragment of 71 nucleotides extending from the Pvu II site to the mRNA start point. Lane $-S_1$ shows the labelled probe without S_1 digestion. Lane -RNA shows a control without addition of RNA prior to incubation at annealing conditions and subsequent S_1 digestion.

pXGßT1 recently published by Banville et al. (11), which, based on the restriction map, was suggested to be identical to the larval $ß_2$-globin cDNA clones described by Widmer et al. (4) and therefore corresponds to the larval $ß_I$-gene (7): The first exchange in triplett 107 is a G⟶C transversion resulting in an amino acid exchange Gly⟶Ala, the second one is a silent mutation at the third position of codon 138 (C⟶G) and the third one is a T⟶C transition in the 3' untranslated region. These differences in sequence may reflect polymorphism in the <u>Xenopus laevis</u> population. The amino acid exchange from Gly_{107} to

Ala_{107} is probably not of physiological importance. Although it involves one of the $\alpha 1 \beta 1$ contact sites in hemoglobin (29), both Gly and Ala are neutral amino acids, which should not affect protein conformation.

The 5' untranslated region, exclusive the initiator ATG, comprises 55 nucleotides, which is in the same range as in other ß-globin genes (30, 31). Between the cap site and the initiator ATG we find at position +10 a sequence 5'-CTTCTG -3'. An almost identical sequence occurs at a similar position in most eucaryotic mRNAs (32, 31) including both the adult β_I- (13) and the adult α_I-globin genes of Xenopus laevis (our unpublished result). This sequence element is assumed to be involved in ribosome binding (33).

The intervening sequences

The internal structure of the larval β_I-gene was established by comparing its sequence with those of our larval β_I-globin cDNA clones; this allows positioning of the intervening sequences within the gene. Like all known ß-globin genes of vertebrates the Xenopus laevis β_I-globin gene contains two intervening sequences. IVS I interrupts the coding region at codon 30. This is also the case for the chicken embryonic ρ- and ε-genes (30). Although two different splicing frames are possible for generation of the mRNA sequence, we favour the frame shown in Figure 2, because it is consistent with the observation that intervening sequences as a rule start with GT and end with AG (34).

A striking feature of the larval β_I-globin gene is the length of IVS I which consists of 564 nucleotides. The unusual length of IVS I confirms previous measurements of R-loops (7) which showed that very large IVS I sequences are a common feature in all larval Xenopus laevis globin genes. Large intervening sequences of α-globin genes have already been reported for the embryonic human zeta-gene (35) and the chicken π'-gene (36), respectively. Whereas the IVS I of the human zeta-gene contains a several fold reiterated short DNA sequence, there is no evidence of repeated sequences within IVS I of the larval β_I-gene; this also holds true for the IVS I of the chicken π'-gene (36). Furthermore, there is no evidence of insertion ele-

ments within the IVS I of the larval Xenopus $ß_I$-gene, as it was described for the IVS II of the goat γ-globin gene (37).

The second intervening sequence (IVS II) interrupts the coding region between amino acid positions 104 and 105. Its length of 968 nucleotides is similar to the IVS II of several mammalian ß-like globin genes (31) and is only slightly longer than the IVS II of the adult Xenopus laevis $ß_I$-gene (13). The IVS II of the larval Xenopus laevis $ß_I$-globin gene contains an almost perfect inverted repeat of 50 nucleotides (indicated by arrows in Figure 2) which even extends for another 30 nucleotides although at a reduced homology.

The 5' flanking region

The DNA region preceding the site of transcription initiation of polymerase II genes contains highly conserved sequence motifs which are essential for gene expression (for review see: 38). A first motif is the TATA-box which is located in most eukaryotic polymerase II dependent genes about 30 nucleotides upstream the site of transcription initiation. In the larval Xenopus $ß_I$-globin gene this sequence is found at position -33, and an additional TATA-box occurs further upstream at position -197.

The CCAAT-box is another highly conserved sequence which in many eukaryotic genes occurs 70 - 90 nucleotides in front of the transcription start. The larval Xenopus $ß_I$-gene does not contain the complete CCAAT-box, but, as already known from other eukaryotic genes, a modified CAAT-box, of which three copies occur at positions -58, -105 and -237 respectively.

Interestingly, absence or modification of CCAAT homologies in ß-globin genes are known to occur in pseudogenes like the rabbit $ß_2$-gene (39) and a lemur ß-pseudogene (40). The CCAAT-box is also modified in the human δ-globin gene which is only expressed at a very low level (41).

On the other hand, recent studies of the Herpes simplex virus thymidine kinase gene promoter revealed that the adjacent sequences of the CCAAT-box may be even more important for transcriptional efficiency (42). In view of this, the homology of the sequence 5' GGGTGGGGTGGAG 3' immediately following the CAAT-signal at -105 with the sequence 5' GGGTGGGGAGGAG 3' at corre-

sponding positions in the embryonic chicken ρ- and ε-globin genes is striking. The conserved sequence of the embryonic ß-like chicken globin genes is part of an inverted repeat which may represent a promoter element (30). In the larval Xenopus $ß_I$-gene the -105 CAAT-motif and the adjacent conserved sequence are flanked by an inverted repeat of 6 nucleotides. The resulting sequence 5' TATTTTNNNNNCAATAGGGTGGGGTGGAGNNNNNAAAATA 3' shows an obvious homology with the putative promoter element of the embryonic chicken globin genes.

In some eukaryotic genes the -100 and further upstream regions contain sequences which modulate transcriptional efficiency (43). In the avian and mammalian ß-globin genes there are short homologous sequences within this region (44, 45), yet such motifs are lacking in both the larval and the adult ß-genes of Xenopus laevis.

As mentioned above the larval Xenopus $ß_I$-gene contains a second TATA-box (at -197) preceded by a further CAAT-box (at -237). Similar findings were reported for the human ε-, ß- and γ-globin genes (46-48). In these instances alternative sites of transcription initiation were found. On the other hand, no additional transcription start has as yet been detected in the larval Xenopus $ß_I$-gene. Even if the labelled probe only extends to the Hinf I cleavage site at -47 (-44 on the coding strand), additional further upstream initiation should result in another S_1 protected fragment of 115 (71+44) nucleotides. This was not observed. However, we would like to mention that the sensitivity of our S_1-mapping experiment should still be improved to exclude minor initiation sites.

The 3' flanking region

Although various conserved sequences, including inverted repeats, were found in the 3' flanking regions of many eukaryotic genes, the functional significance of these elements is as yet not well understood. Apart from the putative polyadenylation signal AATAAA the larval Xenopus $ß_I$-gene contains additional conserved sequences in the 3' flanking region. In fact, the sequence 5' TTTCAAGAA 3' (boxed in Fig. 2) which shows some homology to a sequence of the Xenopus laevis U2 snRNA genes is also present in the 3' region of many eukaryotic polymerase II genes

(49). Moreover, the partially overlapping sequences 5' TGTGTTTAC 3' and 5' TACCTAATA 3', located 63 nucleotides downstream of the polyadenylation signal, are closely related to sequence motifs discovered in the 3' flanking regions of human interferon genes (50-53), the rabbit ß-globin gene (54), chicken ovalbumin gene (55) and even the SV40 early functions (56). However, the functional significance of these sequences remains to be demonstrated.

REFERENCES

1. Just, J.J., Schwager, J. and Weber, R. (1977) Wilhelm Roux's Archives 183, 307-323.
2. Hentschel, C.C., Kay, R.M. and Williams, J.G. (1979) Dev. Biol. 72, 350-363.
3. Hosbach, H.A., Widmer, H.J., Andres, A.-C. and Weber, R. (1982) in Embryonic Development Part A: Genetic Aspects, Burger, M.M. and Weber, R. (eds.) Alan R. Liss, New York, 115-125.
4. Widmer, H.J., Andres, A.-C., Niessing, J., Hosbach, H.A. and Weber, R. (1981) Dev. Biol. 88, 325-332.
5. Jeffreys, A.J., Wilson, V., Wood, D., Simons, I.P., Kay, R.M. and Williams, J.G. (1980) Cell 21, 555-564.
6. Patient, R.K., Banville, D., Brewer, A.C., Elkington, J.A., Greaves, D.R., Lloyd, M.M. and Williams, J.G. (1982) Nucleic Acids Res. 10, 7935-7945.
7. Hosbach, H.A., Wyler, T. and Weber, R. (1983) Cell 32, 45-53.
8. Williams, J.G., Kay, R.M. and Patient, R.K. (1980) Nucleic Acids Res. 8, 4247-4258.
9. Knöchel, W., Meyerhof, W., Hummel, S. and Grundmann, U. (1983) Nucleic Acids Res. 11, 1543-1553.
10. Kay, R.M., Harris, R., Patient, R.K. and Williams J.G. (1983) Nucleic Acids Res. 11, 1537-1542.
11. Banville, D., Kay, R.M., Harris, R. and Williams, J.G. (1983) J. Biol. Chem. 258, 7924-7927.
12. Andres, A.-C., Hosbach, H.A. and Weber, R. (1984) Biochim. Biophys. Acta 781, 294-301.
13. Patient, R.K., Harris, R., Walmsley, M.E. and Williams, J.G. (1983) J. Biol. Chem. 258, 8521-8523.
14. Partington, G.A. and Baralle, F.E. (1981) J. Mol. Biol. 145, 463-470.
15. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning. Cold Spring Harbor Laboratory.
16. Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523.
17. Crouse, G.F., Frischauf, A. and Lehrach, H. (1980) Meth. Enzym. 101, 78-89.
18. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
19. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
20. Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E. and Surrey, S. (1982) Proc. Natl. Acad. Sci. USA 79, 4298-4302.

21. Nieuwkoop, P.D. and Faber, J. (1956) Normal table of Xenopus laevis (Daudin) North-Holland, Amsterdam.
22. Aviv, H. and Leder, P. (1973) Proc. Natl. Acad. Sci. USA 69, 1408-1412.
23. Weaver, R.F. and Weissmann, C. (1979) Nucleic Acids Res. 7, 1175-1193.
24. Ackerman, E. (1983) EMBO J. 2, 1417-1422.
25. Sanger, F. and Coulson, A.R. (1978) FEBS Lètters 87, 107-110.
26. Maxam, A.M. and Gilbert, W. (1980) Meth. Enzym. 65, 499-560.
27. May, F., Westley, B. and Weber, R. (1982) Nucleic Acids Res. 10, 2791-2807.
28. Lewin, B. (1980) Gene Expression, Eukaryotic Chromosomes, 2nd. edn. pp. 668-675, John Wiley and Sons, Inc.
29. Fermi, G. (1975) J. Mol. Biol. 97, 237-256.
30. Dodgson, J.B., Stadt, S.J., Choi, O.R., Dolan, M., Fischer, D.H. and Engel, J.D. (1983) J. Biol. Chem. 258, 12685-12692.
31. Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) Cell 21, 653-668.
32. Baralle, F.E. and Brownlee, G.G. (1978) Nature 274, 84-87.
33. Hagenbüchle, O., Sauter, M., Steitz, J.A. and Mans, R.J. (1978) Cell 13, 551-563.
34. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) Proc. Natl. Acad. Sci. USA 75, 4853-4857.
35. Proudfoot, N.J., Gil, A. and Maniatis, T. (1982) Cell 31, 553-563.
36. Engel, J.D., Rushing, D.J., McCeme, K.C. and Dodgson, J.B. (1983) Proc. Natl. Acad. Sci. USA 80, 1392-1396.
37. Schon, E.A., Cleary, M.L., Haynes, J.R. and Lindgrel, J.B. (1981) Cell 27, 359-369.
38. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
39. Lacy, E. and Maniatis, T. (1980) Cell 21, 545-553.
40. Jeffreys, A.J., Barrie, P.H., Harris, S., Fawcelt, D.H., Nugent, Z.J. and Boyd, A.C. (1982) J. Mol. Biol. 156, 487-503.
41. Spritz, R.A., DeRiel, J.K., Forget, B.G. and Weissman, S.M. (1980) Cell 21, 639-646.
42. McKnight, S. and Kingsbury, R. (1982) Science 217, 316-324.
43. Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. USA 77, 7102-7106.
44. Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. and Weissmann, C. (1983) Cell 32, 695-706.
45. Hardison, R.C. (1983) J. Biol. Chem. 258, 8739-8744.
46. Allan, M., Grindlay, G.J., Stefani, L. and Paul, J. (1982) Nucleic Acids Res. 10, 5133-5147.
47. Allan, M., Lanyon, W.C. and Paul, J. (1983) Cell 35, 187-197.
48. Grindlay, G.J., Lanyon, W.G., Allan, M. and Paul, J. (1984) Nucleic Acids Res. 12, 1811-1820.
49. Mattaj, I.W. and Zeller, R. (1983) EMBO J. 2, 1883-1891.

50. Taya, Y, Devos, R., Tavernier, J., Cheroutre, H., Engler, G. and Fiers, W. (1982) EMBO J. 1, 953-958.
51. Nagata, S., Mantei, N. and Weissmann, C. (1980) Nature 287, 401-408.
52. Lawn, R.M., Gross, M., Houck, C.M., Franke, A.E., Gray, P.V. and Goeddel, D.V. (1981) Proc. Natl. Acad. Sci. USA 78, 5435-5439.
53. Degrave, W., Derynck, R., Tavernier, J., Haegeman, G. and Fiers, W. (1981) Gene 14, 137-143.
54. Van Ooyen, A., Van den Berg, J., Mantei, N. and Weissmann, C. (1979) Science 206, 337-344.
55. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucleic Acids Res. 8, 127-142.
56. Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Henverswyn, H., Van Herreweghe, J., Volckaert, G. and Ysebaert, M. (1978) Nature 273, 113-120.