J Mol Evol (1988) 28:64–71

Journal of Molecular Evolution

Primary Structure and Evolutionary Relationship between the Adult α-Globin Genes and Their 5′-Flanking Regions of *Xenopus laevis* and *Xenopus tropicalis*

Jürg Stalder,[1] Urs Wirthmüller,[1] James Beck,[2] Andreas Gruber,[1] Wolfgang Meyerhof,[2]* Walter Knöchel,[2] and Rudolf Weber[1]

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

Summary. To investigate the evolution of globin genes in the genus *Xenopus*, we have determined the primary structure of the related adult α_I- and α_{II}-globin genes of *X. laevis* and of the adult α-globin gene of *X. tropicalis*, including their 5′-flanking regions. All three genes are comprised of three exons and two introns at homologous positions. The exons are highly conserved and code for 141 amino acids. By contrast, the corresponding introns vary in length and show considerable divergence. Comparison of 900 bp of the 5′-flanking region revealed that the *X. tropicalis* gene contains a conserved proximal 310-bp promoter sequence, comprised of the canonical TATA and CCAAT motifs at homologous positions, and five conserved elements in the same order and at similar positions as previously shown for the corresponding genes of *X. laevis*. We therefore conclude that these conserved upstream elements may represent regulatory sequences for cell-specific regulation of the adult *Xenopus* globin genes.

Key words: *Xenopus laevis* — *Xenopus tropicalis* — Adult α-globin genes — Gene organization — Conserved upstream sequences

Introduction

Phylogenetic analysis of albumins (Bisbee et al. 1977) together with studies on DNA content (Thiébaud and Fischberg 1977) and chromosome numbers (Tymowska and Fischberg 1982) have led to the hypothesis that *X. tropicalis* might be closely related to the common ancestor of the genus *Xenopus* from which *Xenopus laevis* has evolved by genome duplication some 30 million years (Myr) ago. This view is further supported by the recent demonstration that duplicated genes coding for α- and β-globin (Jeffreys et al. 1980) as well as actin (Stutz and Spohr 1986) are present in *X. laevis*, whereas only one copy of those genes is found in *X. tropicalis*. Further evidence for gene duplication in *X. laevis* has been presented for the genes coding for vitellogenin (Wahli and Dawid 1980), albumin (Westley et al. 1981), ribosomal proteins (Bozzoni et al. 1982), and calmodulin (Chien and Dawid 1984).

Comparative studies on the fine structure of similar genes provide a powerful tool for assessing the evolutionary relationships of descendent species (Wilson et al. 1987) and for identifing sequence elements of putative functional signifiance. The *Xenopus* globin gene family is most suitable for such studies, because it shows the same arrangement of the linked α- and β-genes in *X. laevis* and *X. tropicalis* (Jeffreys et al. 1980). Moreover, there is evidence that the duplicated α- and β-globin genes of *X. laevis* are expressed in a highly coordinated and stage-specific fashion (Widmer et al. 1981).

From a recent comparison of the amino acid sequences deduced from the cDNA sequences corresponding to the mRNAs of the adult α- and β-globin genes of *X. tropicalis* and various members of the *X. laevis* group, Knöchel et al. (1986) have proposed that divergence into *X. tropicalis* and *X. laevis* may have occurred already by 110–120 Myr ago and that

Offprint requests to: J. Stalder
* *Present address:* Institut für Zellbiochemie und klinische Neurobiologie, Universitätskrankenhaus Eppendorf, Martinistrasse 52, D-2020 Hamburg 20

genome duplication may have taken place in the *X. laevis* group some 40–60 Myr ago.

In continuation of earlier studies on the evolution of globin genes in the genus *Xenopus* and to localize putative regulatory elements we have sequenced the two related adult α-globin genes of *X. laevis* and the corresponding gene of *X. tropicalis* including their 5′-flanking regions. Sequence analysis reveals highly conserved exons and more diverged introns, and several conserved sequence elements within 900 bp of the 5′-flanking regions of these genes.

Materials and Methods

Nucleotide Sequence Determination. Restriction fragments of the adult α-globin genes were isolated from genomic DNA clones and subcloned into M13mp8 and M13mp9 RF DNA (Messing and Vieira 1982). Where indicated, inserts of the subclones were trimmed with the exonuclease Bal31 (BRL) from their 3′ and 5′ ends and recloned into M13 phages. DNA sequences were determined on both strands by the dideoxy chain termination method (Sanger et al. 1977).

Mapping of the Transcription Start Sites. The sites of transcription initiation were mapped with either S1 nuclease or primer extension. For S1 nuclease mapping, cytoplasmic 9S poly(A)$^+$ RNA was isolated from erythroblasts of anemic adult toads (Widmer et al. 1981) according to Ryffel and McCarthy (1975) and annealed to restriction fragments spanning the putative transcription initiation sites, labeled at their 5′ ends with [γ^{32}P]ATP and polynucleotide kinase (Maniatis et al. 1982) to a specific activity of 1×10^8 cpm/μg. Hybrids obtained after incubation of 40 ng of 9S poly(A)$^+$ RNA with 5×10^4 cpm of labeled DNA for 3 h at 45°C were digested with 500 or 1000 units of S 1 nuclease (Sigma) for 90 min at 20°C (Maniatis et al. 1982).

For primer extension, a 75-bp fragment extending from the PvuII to the HindIII site in exon 1 of the *X. tropicalis* gene was labeled at its 5′ end with [γ^{32}P]ATP and polynucleotide kinase. Hybridization to cytoplasmic RNA and elongation of the primer with reverse transcriptase was carried out according to Meyerhof et al. (1986).

DNA Sequence Comparisons. For nucleotide sequence comparisons the computer program described by Sege et al. (1981) was used. Two-dimensional dot matrix analysis was carried out with a computer program developed by one of us (A. Gruber, unpublished). The program is similar to others and compares two DNA sequences in a two-dimensional plot. It scores matches at a window setting of 6. Only perfect matches of six nucleotides between each sequence will score a match. If two sequences are identical, a diagonal line with a positive slope will be produced. Deletions and insertions as well as mismatches are detected by breaks in the diagonal. Matches at different positions in the two sequences create points on both sides of the diagonal.

Results and Discussion

Sequencing Strategy

The adult α-globin gene of *X. tropicalis* was isolated as recombinant phage λXtG 27 by screening a genomic library (Stutz and Spohr 1987) with the adult α-globin cDNA clone 1C 10 (Knöchel et al. 1986). The adult α_I- and α_{II}-globin genes of *X. laevis* were isolated previously as genomic clones λXG 127 and λXG 143, respectively (Hosbach et al. 1983). Restriction fragments of the adult α-globin genes of both species and the 5′-flanking region of the *X. tropicalis* gene were subcloned into M13 phages for sequencing. The BamH1 fragment on the 3′ side of the *X. laevis* adult α_{II} gene was trimmed with the exonuclease Bal31 from its 3′ and 5′ ends and recloned into M13 phages in both orientations. Clones of both strands were sequenced using the dideoxy chain termination method (Sanger et al. 1977). The nucleotide sequence of intron 2 of the *X. tropicalis* gene was determined from one strand of eight independent clones. All individual sequencing reactions agreed fully. The 5′ ends and the 5′ flanking regions of the adult α_I and α_{II} genes of *X. laevis* were sequenced previously (Stalder et al. 1986).

Gene Organization

The nucleotide sequences of the related adult α-globin genes of *X. laevis* and of the corresponding *X. tropicalis* gene are shown in Fig. 1. The sites of transcription initiation of the *X. laevis* genes, marked as position +1 in Fig. 1, were mapped by S1 nuclease digestion of hybrids between cytoplasmic poly(A)$^+$ 9S RNA from erythroblasts and endlabeled genomic DNA fragments. For the *X. tropicalis* gene the site of transcription initiation was determined by primer extension. Figure 2 shows that transcription initiation of the *X. laevis* α_I gene most likely occurs at an A residue 128 bp 5′ to the HindIII site of exon 1 (Bendig and Williams 1983), whereas in the α_{II} gene the localization of the 5′ end is somewhat less precise, probably due to the AT-rich sequence around the site of transcription initiation. Analogous to the α_I gene, we locate the start point of mRNA synthesis of the α_{II} gene at the A residue 129 bp 5′ to the HindIII site in exon 1. In the case of the *X. tropicalis* gene, the PvuII–HindIII fragment from exon 1 labeled at its 5′ end was used as primer and elongated with reverse transcriptase. From the resulting fragment the site of transcription initiation was localized at the A residue 128 bp 5′ to the HindIII restriction site in exon 1.

To determine the 3′ ends and to locate the intervening sequences, the three α gene sequences were compared to those of their corresponding complementary DNAs (cDNAs), using the clones pXG6C1 (Kay et al. 1983) and XGAD11 (Knöchel et al. 1983) for the *X. laevis* α_I and α_{II} genes, respectively, as well as the clones 1C10 and 1G6 (Knöchel et al. 1986) for the *X. tropicalis* α gene. Whereas the sequences of the *X. tropicalis* gene and the α_I gene of *X. laevis* are identical to the corresponding cDNAs,

X. laevis adult α_I-globin gene **A**

```
     -20              +1               20               40               60
tataaaagaggactgtctttgaagattttgcatctggtgtagactgcttcttgcacaacacaaacagcaaccATGCTTCTTTCAGCCGATGACAAGAAAC

     80              100              120              140              160
ACATCAAGGCAATTATGCCTGCTATCGCTGCCCATGGCGACAAATTTGGGGGAGAAGCTTTGTACAGgtaaattatactctgaatgaatgctgttaacca

     180              200              220              240              260
actgctaatattaactggtttattgtcagatggcccttgataacattgctgaacggattaaacaagccatttgcttaatagtaaaataactctaatgtat

     280              300              320              340              360
ctctttcatttcattgtttctctacagGATGTTCATAGTCAACCCCAAGACCAAAACTTACTTCCCTAGTTTTGACTTCCACCACAATTCAAAACAGATC

     380              400              420              440              460
AGTGCTCATGGCAAGAAAGTTGTGGATGCTCTGAATGAAGCTTCCAACCATTTGGATAACATCGCTGGAAGCATGAGCAAGCTGAGTGACCTCCATGCCT

     480              500              520              540              560
ATGACCTGAGAGTGGACCCTGGCAACTTCCCAgtaagtcctattcaacaatatatatataagtgtaaagggtaaaagtgatatatagcagtatgtaatgc

     580              600              620              640              660
tctttttaaacagatatgcagtatgtaggttaaaacagatcttaataataaatacaaaaaaatgtaaacagtggttttaggtaaaccaggtataagcctt

     680              700              720              740              760
taatcacactgtatgcacagcttatttaatacgatatcaagtctaacttatataaataccccaggaaggcactatgtcctgtttaatttgtaatcttata

     780              800              820              840              860
tactgctaattaaatggcctttatgtttgctcttatgtacccttactctggtttctttttttttcagTTGCTGGCCCATAATATATTGGTGGTTGTTG

     880              900              920              940              960
CTATGAACTTCCCTAAGCAGTTTGATCCTGCAACCCATAAGGCCCTGGATAAGTTCTTGGCTACCGTATCTACTGTTCTGACATCCAAATATCGTTAAgg

     980             1000             1020             1040             1060
ctcagcaacaacagcagcagaagtctcaacatcagacatcagttaattatatgcaatcaaactgacaaagcttgtgaaagaatgttctgaaataaacatt

    1080             1100             1120             1140             1160
ttaaccattatcctaaagttgcatttattactgtctgcacatggttaattaaaaaatatacatatagtgtttactcctaatatgacaaagttttcagaaa
        *
```

X. laevis adult α_{II}-globin gene **B**

```
     -20              +1               20               40               60
tatataagaggactgtcactttagcttttgcatttggtgtcgactgcttatcgcaacaacacaaacagcaacaATGCTTCTTTCAGCTGATGACAAGAAA

     80              100              120              140              160
CACATCAAGGCAATTATGCCTTCCATAGCCGCTCATGGCGACACATTTGGTGGAGAAGCTTTGTACAGgtaaattatgctctgaattaatgttgtcaacc

     180              200              220              240              260
aactgctaatattaataaagatcagtttagtgtttaataacatagctgaatgggttaaacaagccatttgctaaaaaggtgcctgatggctgatggatag

     280              300              320              340              360
taaaattactataatttatccctttattctcatttctccgcagGATGTTCTTGGTTAACCCTAAGACCAAAACCTACTTTCCTAGTTTTGACTTCCACC

     380              400              420              440              460
ACAATTCAAAACAGATCACTTCTCATGGCAAGAAAGTCGTCGATGCTCTGAATGAAGCTGCCAACCATTTGGATAACATTGCTGGAAGCATGAGCAAGCT

     480              500              520              540              560
GAGCGACCTCCATGCCTATGACCTGAGAGTGGATCCGGGCAACTTCCCAgtaagtcctttgcaatgcgtattcttctctatatcgctctgtcattggtat

     580              600              620              640              660
atgtatactttagaaggtaaacagatatgcagtatgtagggtcaaagagaactaaatcataaatactagcattaacctaatacatgtaaagcccttaatc

     680              700              720              740              760
acattgtgtacacagtttatctctatgtatttaattaaatattaaatagatcttattaagaacttgtataaataaagatggctctaagttctattcaagt

     780              800              820              840              860
tggtatcctatatattgccaactaaatggcctttatgcttgctgtcatatatccttactctgatttctttctttcctttgcagTTGCTGGCTCATAATT

     880              900              920              940              960
TGCTGGTGGTTGTTGCTATGCACTTCCCTAAGCAGTTTGATCCTGCAACCCATAAGGCCCTGGATAAGTTCCTGGCTACCGTATCTACTGTTCTGACTTC

     980             1000             1020             1040             1060
CAAATATCGTTAAggctcagcagcaactgcagcagcaatttcaacatcaaaatcagcaaatcacatgcaatccaactgacacaacttgtgaaagaatgtt

    1080             1100             1120             1140             1160
ctgaaataaacatttaacctttatcctaaagttgcatttatctttttcttttattactgaaagcaaatggttaattaaatatatatacaggtatggg
                      *
```

X. tropicalis adult α-globin gene **C**

```
     -20              +1               20               40               60
tataaagaggactgtcgcagtagctttagcatctggtgtggactgcttattgcacaacacaaacaacaATGCATCTTACAGCTGATGACAAGAAACACAT

     80              100              120              140              160
CAAGGCCATTTGGCCTTCTGTAGCTGCTCATGGTGACAAATATGGCGGAGAAGCTTTGCACAGgtaaattatgctctgaattaatgctgttaaccaattg

     180              200              220              240              260
ctaaaatgaattaggataaatttattgtcagatggtccttaataaagttgctgaatgggttaaacatgccatttccttaaaaggtgcctgatggatatta

     280              300              320              340              360
aaattatccctttactctcatttttttctacagGATGTTCATGTGTGCTCCCAAGACCAAAACCTACTTTCCTGATTTTGACTTCAGCGAACATTCAAAA

     380              400              420              440              460
CACATCTTGGCTCATGGCAAGAAAGTTTCGGATGCTCTGAATGAGGCTTGCAACCATCTGGACAACATTGCCGGATGCCTGTCCAAGCTGAGTGACCTCC

     480              500              520              540              560
ATGCCTATGACCTGAGAGTGGATCCAGGCAACTTCCCAgtaagtgctttgcaatgtgtattataaattttctatatttctaattttatatatatttgga

     580              600              620              640              660
gtgtaaataatatataaaaggtatatatactgcgtagactgttttgttttatctttatgtatttaataagatatcaagtccagctaattaagagctta

     680              700              720              740              760
tgtaaataccccccaaaatgactctatgtcctgtttaatttgaaacactctatactacttaacttaacttaatgacctttatgcttgctaccatgcacacc

     780              800              820              840              860
tattctgattgtctttctttctttcttttgcagTTGCTGGCCCATCAAATTCTGGTGGTTGTTGCTATCCATTTCCCTAAGCAGTTTGACCCTGCAACCC

     880              900              920              940              960
ATAAGGCCCTGGACAAGTTCCTGGTTTCCGTATCTAATGTTCTGACATCCAAATATCGTTAAggctcagcagtaacagtagcagaagtttggacatcaga

     980             1000             1020             1040             1060
catcagttaatgacaaacaatcaaactgacacagcttgtgaaagaatgttctgaaataaacattttaaattattctaaagttgccttctttcattatt
                                                                  *
```

Fig. 1. Nucleotide sequences of the adult α_I- (**A**) and α_{II}- (**B**) globin genes of *X. laevis* as well as the adult α-globin gene of *X. tropicalis* (**C**). The sequences are shown from the TATA box (underlined) 5′ to the genes and are extended on the 3′ side beyond the polyadenylation site (marked with an asterisk) as determined by comparison with the corresponding cDNAs. The +1 position indicates the start site of mRNA synthesis. Capital letters are used for exon sequences and small letters for introns and nontranslated regions. The polyadenylation signal in the 3′ nontranslated region is underlined.

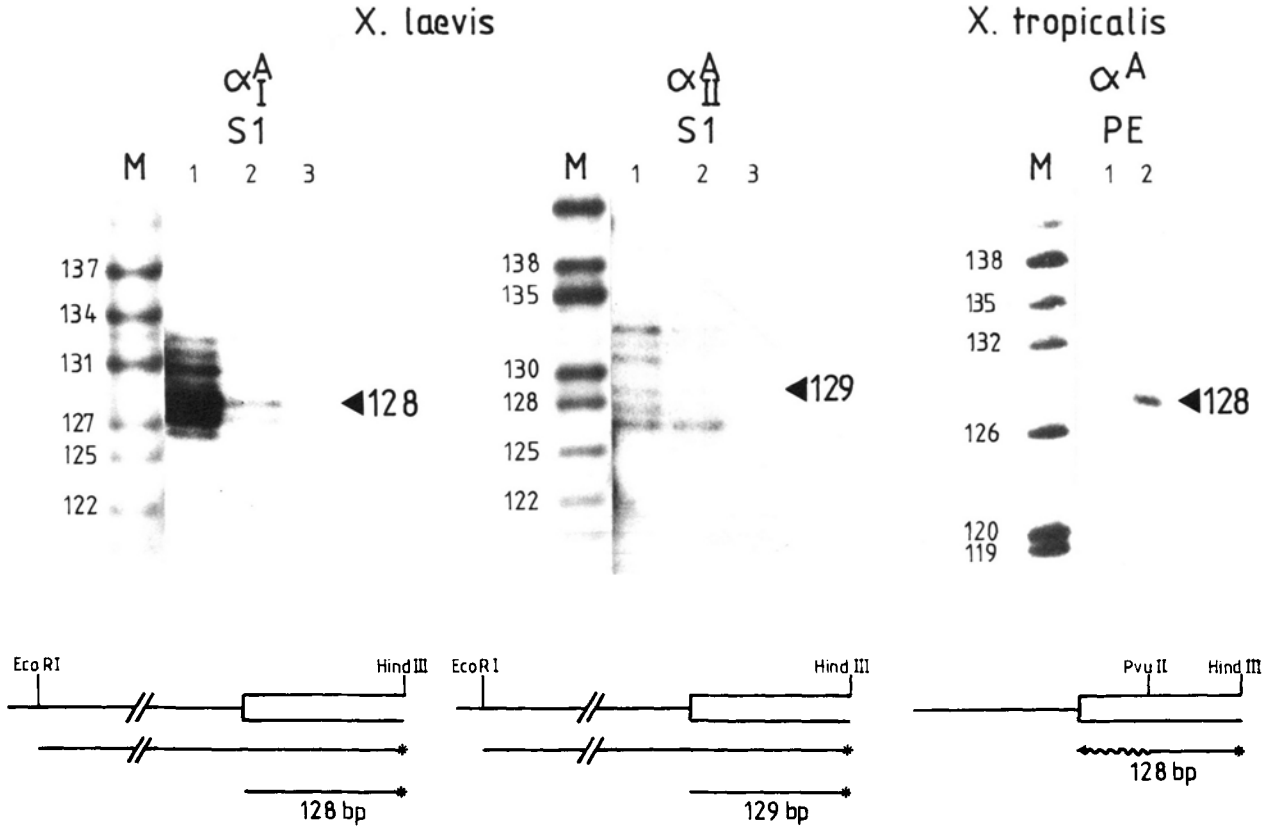

Fig. 2. Determination of the mRNA transcription start site. The transcription initiation sites of the adult α_I (α_I^A)- and α_{II}(α_{II}^A)-globin genes of *X. laevis* were mapped with S1 nuclease. In each case an EcoRI–HindIII fragment spanning the 5′ end of the corresponding gene was labeled at its 5′ end with ^{32}P (marked with an asterisk). Lanes 1 and 2 show the fragments protected from S1 nuclease digestion using 500 U (lane 1) or 1000 U (lane 2) of S1 nuclease after annealing the probes to red blood cell RNA from anemic toads. Lane 3 shows the digestion of the probe without added RNA but with 1000 U of S1 nuclease. The transcription initiation site of the adult α-globin gene of *X. tropicalis* was mapped with primer extension (PE). Lane 2 shows the extension product obtained with reverse transcriptase of the PvuII–HindIII primer labeled at the 5′ end with ^{32}P (marked with an asterisk) and annealed to red blood cell RNA of anemic toads. Lane 1 shows the result of primer extension without added RNA. In lane M, sequencing reactions were used as size markers. The arrowheads indicate the site of transcription initiation in base pairs 5′ to the HindIII sites in exon 1 (see text).

Table 1. Organization of the adult α-globin genes of *X. laevis* ($Xl\alpha_I^A$, $Xl\alpha_{II}^A$) and *X. tropicalis* ($Xt\alpha^A$) in base pairs

Region	$Xl\alpha_I^A$	$Xl\alpha_{II}^A$	$Xt\alpha^A$
5′ Noncoding	41	42	41
Exon 1	95	95	95
Intron 1	160	176	170
Exon 2	205	205	205
Intron 2	337	335	295
Exon 3	129	129	129
3′ Noncoding	111	110	111
Total length	1078	1092	1046

we noticed seven base changes between the *X. laevis* α_{II} gene and its cDNA. However, only two of the base changes lead to amino acid substitutions.

The exon–intron boundaries were determined according to the consensus sequence 5′ exon/GT intron AG/exon 3′ (Breathnach et al. 1978).

The organization of the related adult α-globin genes of *X. laevis* and of the corresponding *X. tropicalis* gene is shown in Table 1. All three genes contain three exons of 95, 205, and 129 bp, respectively, which together code for 141 amino acids. The coding sequences are interrupted by two introns at homologous positions. The smaller first intron interrupts the coding sequence within codon 31 and the larger second intron between codons 99 and 100. In contrast to the constant size of the corresponding exons, the corresponding introns of the three genes show some variation in length. The length of the 5′ leader as well as the 3′ trailer sequences is almost identical in the three *Xenopus* genes.

Except for the greater intron length in the *Xenopus* genes, the overall organization of these genes is very similar to that of the human α_I and α_{II} genes (Liebhaber et al. 1980; Michelson and Orkin 1983) as well as to the chicken α^A and α^D genes (Dodgson and Engel 1983). This comparison shows that the overall organization of the adult α-globin genes is not only conserved within the genus *Xenopus* but also between lower and higher vertebrates.

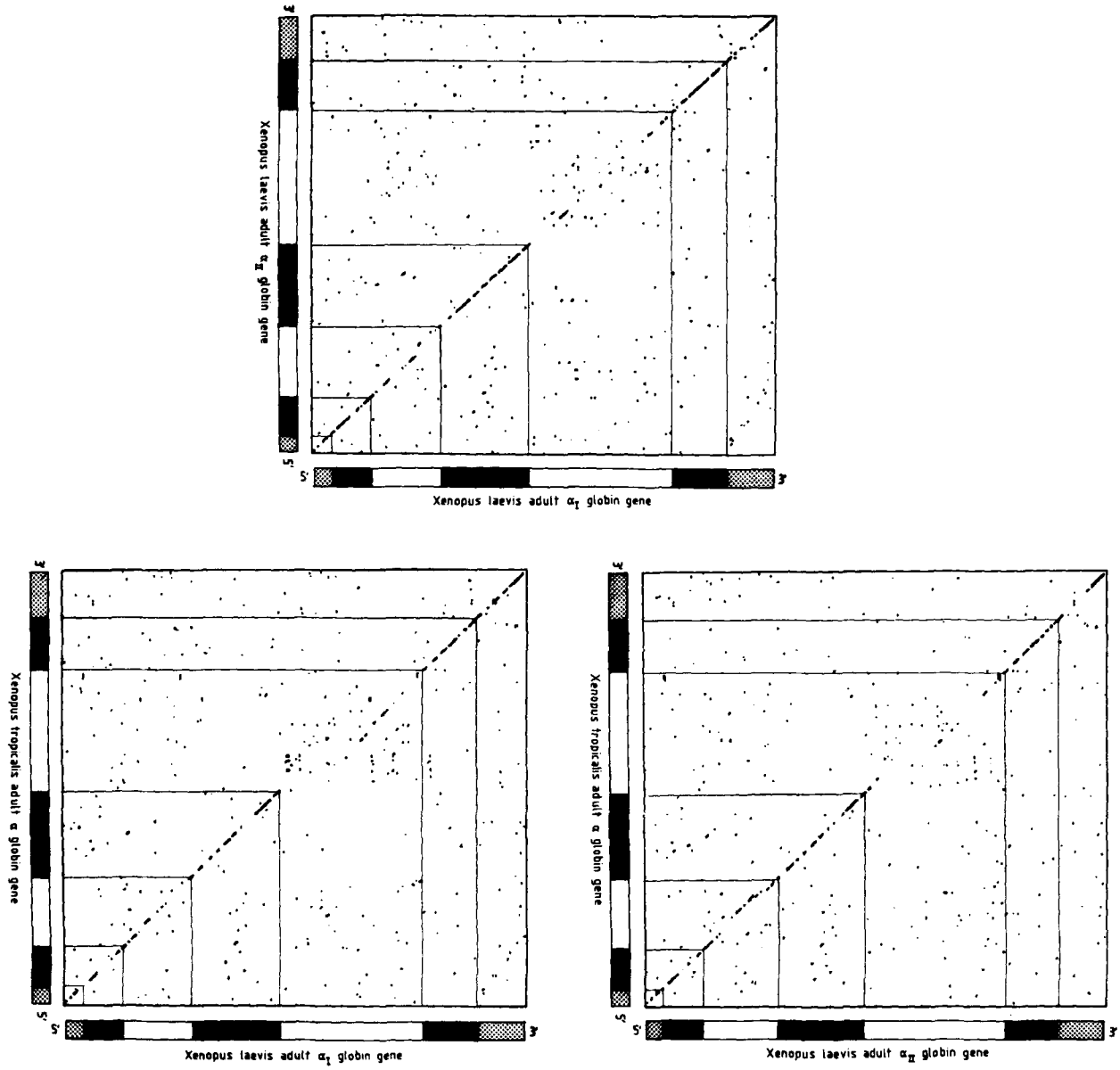

Fig. 3. Similarity between the *Xenopus* adult α-globin genes by two-dimensional dot matrix analysis. Each dot represents a 6-bp perfect match between a pair of globin gene sequences. Similarities at corresponding positions result in a series of dots in the diagonal.

Table 2. Nucleotide sequence similarity in percentages between the adult α-globin genes of *X. laevis* ($Xl\alpha_I^A$, $Xl\alpha_{II}^A$) and *X. tropicalis* ($Xt\alpha^A$)

Region	$Xl\alpha_I^A/Xl\alpha_{II}^A$	$Xl\alpha_I^A/Xt\alpha^A$	$Xl\alpha_{II}^A/Xt\alpha^A$
5′ Noncoding	88.6	92.1	89.5
Exon 1	90.2	84.8	85.9
Intron 1	84.0	78.2	82.6
Exon 2	93.2	83.4	82.9
Intron 2	60.4	64.2	72.0
Exon 3	94.4	89.7	89.7
3′ Noncoding	87.4	87.6	82.9

Table 3. Similarity boxes within the 5′-flanking regions of the adult α-globin genes of *X. laevis* ($Xl\alpha_I^A$, $Xl\alpha_{II}^A$) and *X. tropicalis* ($Xt\alpha^A$)

5′-Flanking region	$Xl\alpha_I^A$ length	Identical nucleotides	$Xt\alpha^A$ length	Identical nucleotides	$Xl\alpha_{II}^A$ length
5′ Proximal	289 bp	235	310 bp	225	287 bp
Box I	22 bp	18	20 bp	19	22 bp
Box II	23 bp	20	38 bp	18	23 bp
Box III	61 bp	48	64 bp	44	61 bp
Box IV	53 bp	44	53 bp	42	56 bp
Box V	109 bp	67	86 bp	64	111 bp

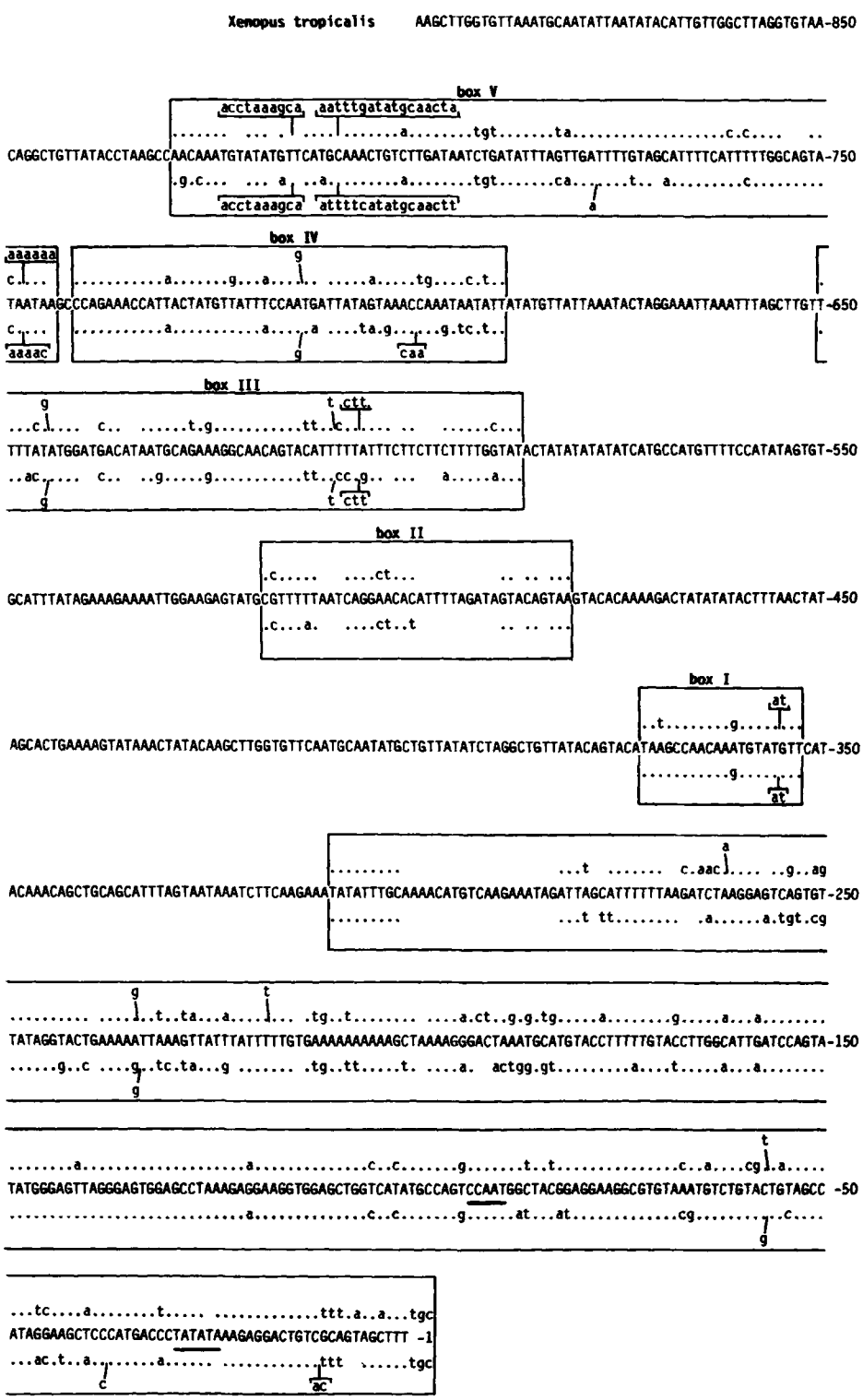

Fig. 4. Similarity boxes within the 5′-flanking regions of the adult α_I- and α_{II}-globin genes of *X. laevis* and the corresponding gene of *X. tropicalis*. From position −900 to −1, the *X. tropicalis* upstream sequence is aligned to obtain maximal matching with the similarity boxes 5′ to the *X. laevis* genes. For comparison the sequence 5′ to the *X. tropicalis* gene is taken as reference. Above and below the sequences of the similarity boxes 5′ to the adult α_I- and α_{II}-globin genes of *X. laevis*, respectively, are shown. Identical bases are marked by dots. The TATA and the CCAAT motifs are underlined. The boxed areas represent conserved sequences 5′ to the three genes.

Sequence Similarities

In a previous comparison of the nucleotide sequences of adult α-globin cDNA clones, Knöchel et al. (1986) reported a divergence of 7.6% between the related α sequences of *X. laevis*, and of 14.7% and 15.1% between the α sequence of *X. tropicalis* and the *X. laevis* α_I and α_{II} sequences, respectively. To assess similarities among the adult *X. laevis* α-globin genes and between these and the corresponding *X. tropicalis* gene, we have compared the sequences pairwise by two-dimensional dot matrix analysis. In such plots similar sequences are detected as stretches of dots on the diagonal. From Fig. 3 it is evident that in all cases the sequence similarity is greater between exons than between introns. However, the sequence conservation between the

three genes is not evenly distributed over the entire coding region: Table 2 shows that exon 3 is more conserved than exons 1 and 2. The higher conservation within exon 3 is also evident in the deduced amino acid sequences, suggesting that subdomains of the adult α-globin polypeptides may be under different functional constraints. The 5′ and the 3′ noncoding regions show a similar degree of similarity as do the exons. All three genes contain the putative polyadenylation signal AATAAA at homologous positions in the 3′ noncoding region (cf. Fig. 2).

In contrast to the exons, the introns display less sequence similarity, which in all three genes is lowest in intron 2. However, short sequence elements ranging between 10 and 30 bp are highly conserved in both introns (Fig. 3). Particularly striking is a sequence of 41 nucleotides at the very 5′ end of intron 1, of which 36 nucleotides are identical in the three *Xenopus* genes. Because this sequence is absent in the adult human or adult chicken α-globin genes it may not be of general functional significance. On the other hand, that shorter sequences of 5–8 bp may be conserved in introns between *Xenopus* species and higher vertebrates as are splice junctions is not excluded.

Conserved Sequence Elements in the 5′-Flanking Region

We have previously shown that the 5′-flanking regions of the coordinately expressed adult α-globin genes of *X. laevis* contain several conserved sequences of different length (Stalder et al. 1986). Moreover, the 5′-flanking regions include a functional promoter and at least some regulatory elements for cell-specific gene expression (J. Stalder, manuscript submitted). It was of interest, therefore, to extend a sequence comparison to the upstream region of the corresponding gene of *X. tropicalis*.

To this end the available 900 bp of the *X. tropicalis* upstream region were compared to the similarity boxes in the upstream regions of the *X. laevis* α genes. Figure 4 shows that by alignment of the *X. tropicalis* upstream sequence with the similarity boxes of *X. laevis* for maximal matching, several regions of similarity can be detected. The proximal 310 bp of the 5′-flanking region of the *X. tropicalis* gene, which include a TATA box 25–30 bp upstream of the transcription initiation site and a CCAAT motif around position −90, show a remarkable sequence similarity to the corresponding regions of the *X. laevis* α_I and α_{II} genes (Table 3). In addition, further upstream five stretches of DNA comprising between 20 and 86 bp are very similar to five previously identified regions of similarity within the 5′-flanking regions of the *X. laevis* genes (boxes I–V). The similar regions are separated from each other by sequences showing less than 50% similarity. Although the extent of similarity between corresponding conserved regions 5′ to the *X. laevis* and *X. tropicalis* genes may vary, it is noteworthy that they occur in the same order and are located at similar positions. The fact that these regions are conserved over some 120 Myr suggests that they may indeed represent essential elements for promoter activity and cell-specific regulation of the adult *Xenopus* α-globin genes.

Acknowledgments. We thank Mrs. M. Kummer for typing the manuscript. This investigation was supported by the Swiss National Science Foundation (Grant 3.409-0.86), the Deutsche Forschungsgemeinschaft (Kn 200/2-3), and Fonds der Chemischen Industrie.

References

Bendig MM, Williams JG (1983) Replication and expression of *Xenopus laevis* globin genes injected into fertilized *Xenopus* eggs. Proc Natl Acad Sci USA 80:6197–6201

Bisbee CA, Baker MA, Wilson AC, Hadji-Azimi I, Fischberg M (1977) Albumin phylogeny for clawed frogs (*Xenopus*). Science 195:785–787

Bozzoni I, Tagnoni A, Perandrei-Amaldi P, Beccari E, Buongiorno-Nardelli M, Amaldi F (1982) Isolation and structural analysis of ribosomal protein genes in *Xenopus laevis*. J Mol Biol 161:353–371

Breathnach R, Benoist C, O'Hare K, Gannon F, Chambon P (1978) Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon–intron boundaries. Proc Natl Acad Sci USA 75:4853–4857

Chien Y-H, Dawid IB (1984) Isolation and characterization of calmodulin genes from *Xenopus laevis*. Mol Cell Biol 4:507–513

Dodgson JB, Engel JD (1983) The nucleotide sequence of the adult chicken α-globin genes. J Biol Chem 258:4623–4629

Hosbach HA, Wyler T, Weber R (1983) The *Xenopus laevis* globin gene family: chromosomal arrangement and gene structure. Cell 32:45–53

Jeffreys AJ, Wilson V, Wood D, Simons JP, Kay RM, Williams JG (1980) Linkage of adult α- and β-globin genes in *X. laevis* and gene duplication by tetraploidization. Cell 21:555–564

Kay RM, Harris R, Patient RK, Williams JG (1983) Complete nucleotide sequence of a cloned cDNA derived from the major adult α-globin mRNA of *X. laevis*. Nucleic Acids Res 11: 1537–1542

Knöchel W, Meyerhof W, Hummel S, Grundmann U (1983) Molecular cloning and sequencing of mRNA coding for minor adult globin polypeptides of *Xenopus laevis*. Nucleic Acids Res 11:1543–1553

Knöchel W, Korge E, Basner A, Meyerhof W (1986) Globin evolution in the genus *Xenopus*: comparative analysis of cDNAs coding for adult globin polypeptides of *Xenopus borealis* and *Xenopus tropicalis*. J Mol Evol 23:211–223

Liebhaber SA, Goossens MJ, Wai Kan Y (1980) Cloning and complete nucleotide sequence of human 5′-α-globin gene. Proc Natl Acad Sci USA 77:7054–7058

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning. Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY

Messing J, Vieira J (1982) A new pair of M13 vectors for

selecting either DNA strand of double-digit restriction fragments. Gene 19:269–276

Meyerhof W, Köster M, Stalder J, Weber R, Knöchel W (1986) Sequence analysis of the larval β_{II}-globin gene of *Xenopus laevis*. Mol Biol Rep 11:155–161

Michelson AM, Orkin SH (1983) Boundaries of gene conversion within the duplicated human α-globin genes. J Biol Chem 258:15245–15254

Ryffel GU, McCarthy BJ (1975) Complexity of cytoplasmic RNA in different mouse tissues measured by hybridization of polyadenylated RNA to complementary DNA. Biochemistry 14:1379–1385

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467

Sege RD, Söll D, Ruddle FH, Queen C (1981) A conversational system for the computer analysis of nucleic acid sequences. Nucleic Acids Res 9:437–444

Stalder J, Meyerhof W, Wirthmüller U, Gruber A, Wyler T, Knöchel W, Weber R (1986) Conserved sequences and cell specific DNase I hypersensitive sites upstream of the coordinately expressed α_I- and α_{II}-globin genes of *Xenopus laevis*. J Mol Biol 188:119–128

Stutz F, Spohr G (1986) Isolation and characterization of sarcomeric actin genes expressed in *Xenopus laevis* embryos. J Mol Biol 187:349–361

Stutz F, Spohr G (1987) A processed gene coding for a sarcomeric actin in *Xenopus laevis* and *Xenopus tropicalis*. EMBO J 6:1989–1995

Thiébaud C-H, Fischberg M (1977) DNA content in the genus *Xenopus*. Chromosoma 59:253–257

Tymowska J, Fischberg M (1982) A comparison of the karyotype, constitutive heterochromatin, and nucleolar organizer regions of the new tetraploid species *Xenopus epitropicalis* Fischberg and Picard with those of *Xenopus tropicalis* Gray (Anura, Pipidae). Cytogenet Cell Genet 34:149–157

Wahli W, Dawid IB (1980) Isolation of two closely related vitellogenin genes, including their flanking regions from a *Xenopus laevis* gene library. Proc Natl Acad Sci USA 77: 1437–1441

Westley B, Wyler T, Ryffel G, Weber R (1981) *Xenopus laevis* serum albumins are encoded in two closely related genes. Nucleic Acids Res 9:3557–3574

Widmer HJ, Andres A-C, Niessing J, Hosbach HA, Weber R (1981) Comparative analysis of cloned larval and adult globin cDNA sequences of *Xenopus laevis*. Dev Biol 88:325–332

Wilson AC, Ochman H, Prager EM (1987) Molecular time scale for evolution. Trends Genetics 3:241–247

Received March 1, 1988/Revised April 18, 1988