Molecular cloning and sequencing of mRNAs coding for minor adult globin polypeptides of *Xenopus laevis**

Walter Knöchel, Wolfgang Meyerhof, Sabine Hummel and Ulrich Grundmann

Institut für Molekularbiologie und Biochemie, Freie Universität Berlin, Arnimallee 22, D-1000 Berlin 33, FRG

Received 11 January 1983; Revised 10 February 1983; Accepted 14 February 1983

ABSTRACT

Globin mRNA was isolated from immature red blood cells of an adult *Xenopus laevis* female, mRNA/cDNA hybrids were integrated in the Pst I cleavage site of pBR 322 by G/C tailing, and cloned in *Escherichia coli* strain HB 101. By restriction site analysis as well as hybridization behaviour we identified two clones coding for minor adult α and β globin chains. Nucleotide sequence analysis and derived amino acid sequences are presented.

INTRODUCTION

Like it is known from birds and mammals, the South African clawed toad, *Xenopus laevis*, produces different globin polypeptides during his life cycle. The pattern of chains observed at tadpole stages is replaced at metamorphosis by the pattern of adult chains (1, 2), which has been found to consist at least of four components, that is two major chains (designated in this work as α^1 and β^1) and two minor chains (α^2 and β^2) (3,4). During the last years a number of investigators have described cDNA cloning of globin mRNA from *Xenopus* immature red blood cells in bacterial plasmids (2, 4, 5, 6). Nucleotide sequence analysis of the α^1 and β^1 (7, 8, 9) sequence have been reported, but neither nucleotide nor amino acid sequence data are known for the minor chains. The investigation of these sequences should be interesting for the following reason. In case of *Xenopus laevis* it is assumed, that a genome duplication occurred about 30 million years ago (10, 11). Moreover it has been found that α^1 and β^1 as well as α^2 and β^2 globin genes are linked and that the gene linkage group coding for major and that coding for minor chains are located on different genomic fragments (12, 13). This arrangement is very different from that in

higher vertebrates, where the gene clusters coding for α and β chains are separated and located on different chromosomes. Since in *Xenopus* there is a close homology between both types of α as well as between both types of β genes, an ancestral gene should first have duplicated to α and β sequences followed much later by a genome duplication, giving rise to genes coding for major and minor sequences. A comparison of the exon nucleotide sequences and resulting amino acid sequences of the four genes allows to analyse the divergence of the mRNAs and polypeptides encoded by them. The data contribute to the evaluation of the *Xenopus* globins and they may help to understand evolutionary development of the globin genes in *Xenopus* by such fundamental processes like gene or genome duplication.

MATERIALS AND METHODS

Isolation of globin mRNA

An adult female of *Xenopus laevis* was made anemic by two subcutaneous injections of 0.5 ml phenylhydrazine (5 mg/ml) at a 48 h interval (14). 15 days later the blood was collected by cardiac puncture. Blood cells were washed and lysed as reported by Battaglia and Melli (3). Polysomes were isolated by centrifugation at 100,000 g. Polysomal RNA was prepared by phenol/chloroform/isoamyl alcohol extraction and mRNA was isolated by three successive runs on oligo(dT)-cellulose (15).

Synthesis of cDNA/mRNA hybrids and cloning in pBR 322

Globin cDNA/mRNA hybrids were prepared according to Zain et al. (16) with the following modifications. RNA concentration was raised to 1 μ g/100 μ l and incubation was done in the presence of 30 units of human placental ribonuclease inhibitor (17) supplied by Bolton Biologicals, St. Paul, Minnesota. dC elongation of hybrids and dG elongation of Pst I cleaved plasmid pBR 322 were performed with terminal transferase as originally reported by Roychoudhury et al. (18) and the modifications described previously (19). Hybrid plasmids were prepared according to Wensink et al. (20). *E. coli* strain HB 101 was transformed using the Ca shock method (21).

Colony screening and identification of different globin sequences

600 bacterial colonies grown on tetracycline containing medium were transferred to nitrocellulose paper (22). Filters were hybridized with ^{32}P -labelled globin cDNA (23), subsequently washed and submitted to autoradiography. Recombinant plasmids of 50 clones showing autoradiographic signals of varying intensities were isolated by the rapid alkaline extraction procedure (24) and digested with various restriction enzymes using buffers and conditions as recommended by commercial suppliers.

The restriction patterns were compared with those reported by Widmer et al. (4). For further analysis of the clone bank the plasmid inserts of individual clones which had been identified to carry an α or β globin sequence were released with Pst I and ^{32}P -labelled by nick translation using a commercial kit (Amersham Buchler) and the protocol enclosed. The total set of 600 clones was hybridized with these probes and subsequently washed at different salt concentrations with increasing stringencies (4).

Determination of nucleotide sequences

Restriction fragments of the inserted plasmid DNAs to be sequenced were selected and isolated as shown in Fig. 1 and 2. They were either 5'-end labelled by polynucleotide kinase with (γ - ^{32}P)-ATP or 3'-end labelled by terminal transferase with 3'-(^{32}P)-cordycepin 5'-triphosphate using commercial kits (NEN) and the protocols enclosed. After cleavage with a second restriction enzyme the fragments were separated by gel electrophoresis and submitted to the chemical sequencing reactions of Maxam and Gilbert (25). Two electrophoresis runs of each reacted fragment allowed sequence reading for up to 250 nucleotides.

RESULTS

To obtain cloning of complete or nearly complete globin mRNA sequences we applied the mRNA/cDNA cloning procedure originally reported by Zain et al. (16). This method confines manipulations on cDNA to a minimum and therefore an artificial breakdown of sequences by nuclease attack may be limited. Transformation of *E. coli* strain HB 101 with globin mRNA/cDNA inte-

grated by G/C joining in the Pst I cleavage site of plasmid pBR 322 yielded 600 colonies, more than 60% hybridized with ^{32}P -labelled globin cDNA as judged by visible autoradiographic signals. Plasmid DNAs of 50 randomly selected clones were isolated according to the rapid alkaline extraction procedure described by Birnboim and Doly (24). Lengths of the integrated sequences were analysed by Pst I cleavage and subsequent gel electrophoresis. 30 plasmids which carried inserts of more than 550 nucleotides were further investigated for restriction sites with the following enzymes: Eco R I, Hind III, Bam H I, Hae III, Pvu II, Pst I, Hinf I, Hpa II and Hha I. According to cleavage sites the recombinant plasmids could be classified into three groups. By comparison of the restriction patterns with those reported by Widmer et al. (4) nine clones were identified to carry an α^1 and five to carry an α^2 globin nucleotide sequence. We failed however, to distinguish among the remaining 16 clones between β^1 and β^2 sequences on the basis of a Hinf I site, which was found by above authors to exist in β^1 but not in β^2 . All recombinant plasmids of the β group investigated by us exhibited this restriction site. Since genetic polymorphism may well account for such a difference, we characterized the cloned sequences by their hybridization behaviour. For this purpose a sequence, identified as α^2 , as well as a β sequence were released by Pst I cleavage and ^{32}P -labelled by nick-translation. Hybridization of all recombinants with these probes and subsequent washing at increasing stringencies clearly demonstrated that both the α and the β group contained at least two types of nucleotide sequences, that is, those which retain most of radioactivity and those which severely loose radioactivity from nitrocellulose filters during the washing procedure. With respect to the 30 clones from which the plasmids had been isolated, the results of the hybridization experiment fully substantiated our findings for the 14 α clones gained by restriction sites analysis and revealed two types of β sequences, because four out of the 16 β clones did nearly not hybridize with the nick-translated probe under most stringent conditions.

For nucleotide sequence analysis we selected one α^2 sequence (clone XG AD 11) and three β sequences (clones XG AE 10,

Figure 1: Restriction map and sequenced restriction fragments of α^2 globin sequence cloned in XG AD 11. The arrows show directions and terminal points of nucleotide reading. One fragment was isolated after Hpa II cleavage of the plasmid (cleavage site in pBR 322 at nucleotide 3549 (26)), 5'-end labelling and subsequent cleavage with Hae III.

XG AB 9 and XG AF 1). XG AE 10 was out of the group of 12 clones which strongly hybridized with the labelled β probe under stringent conditions and by sequence analysis it was proven to be completely identical with the β^1 sequence published by Williams et al. (8) (data not shown). XG AB 9 and XG AF 1 were out of the four clones which hybridized to the β^1 probe under non-stringent but failed to hybridize under stringent conditions. As nucleotide sequences of these two clones were completely identical, we conclude that both recombinant plasmids carry the sequence coding for the adult minor β globin (data for XG AB 9 are not shown). Figures 1 and 2 show the restric-

Figure 2: Restriction map and sequenced restriction fragments of β^2 globin sequence cloned in XG AF 1. The arrows show directions and terminal points of nucleotide reading.

A) NUCL. SEQ. XEN. L. α^2 GLOBIN TGTGACTGCTTATGCAACAACAACAACAGCACA ATG
 B) NUCL. SEQ. XEN. L. α^1 GLOBIN* TGCACAACAACAACAACAGCAACC ATG
 C) AM. ACID SEQ. XEN. L. α^2 GLOBIN
 D) AM. ACID SEQ. XEN. L. α^1 GLOBIN*
 E) AM. ACID SEQ. CHICKEN α^A GLOBIN**

A) CTT CTT TCA GCT GAT GAC AAG AAA CAC ATC AAG GCA ATT ATG CCT CTG ATA GCG GCT CAT²⁰
 B) CTT CTT TCA GCC GAT GAC AAG AAA CAC ATC AAG GCA ATT ATG CCT GCT ATC GCT GCC CAT
 C) LEU LEU SER ALA ASP ASP LYS LYS HIS ILE LYS ALA ILE MET PRO SER ILE ALA ALA HIS
 D) LEU LEU SER ALA ASP ASP LYS LYS HIS ILE LYS ALA ILE MET PRO ALA ILE ALA ALA HIS
 E) VAL LEU SER ALA ALA ASP LYS ASN ASN VAL LYS GLY ILE PHE THR LYS ILE ALA GLY HIS

A) GGC GAC AAA TTT GGT GGA GAA GCT TCG TAC AGG ATG TTC TTG GTT AAC CCI AAG ACC AAA³⁰
 B) GGC GAC AAA TTT GGT GGA GAA GCT TTG TAC AGG ATG TTC ATA GTC AAC CCC AAG ACC AAA
 C) GLY ASP LYS PHE GLY GLY GLU ALA SER TYR ARG MET PHE LEU VAL ASN PRO LYS THR LYS
 D) GLY ASP LYS PHE GLY GLY GLU ALA LEU TYR ARG MET PHE ILE VAL ASN PRO LYS THR LYS
 E) ALA GLU GLU TYR GLY ALA GLU THR LEU GLU ARG MET PHE THR THR TYR PRO PRO THR LYS

A) ACC TAC TTT CCT AGT TTT GAC TTC CAC CAC AAT TCA AAA CAG ATC ACT TCT CAT GGC AAG⁴⁰
 B) ACT TAC TTC CCT AGT TTT GAC TTC CAC CAC AAT TCA AAA CAG ATC AGT GCT CAT GGC AAG
 C) THR TYR PHE PRO SER PHE ASP PHE HIS HIS ASN SER LYS GLN ILE THR SER HIS GLY LYS
 D) THR TYR PHE PRO SER PHE ASP PHE HIS HIS ASN SER LYS GLN ILE SER ALA HIS GLY LYS
 E) THR TYR PHE PRO HIS PHE ASP LEU SER HIS GLY SER ALA GLN ILE LYS GLY HIS GLY LYS

A) AAA GTG GTG GAT GCT CTG AAT GAA GCT GGC AAC CAT TTG GAT AAC ATT GCT GGA AGC ATG⁵⁰
 B) AAA GTT GTG GAT GCT CTG AAT GAA GCT TCC AAC CAT TTG GAT AAC ATC GCT GGA AGC ATG
 C) LYS VAL VAL ASP ALA LEU ASN GLU ALA ALA ASN HIS LEU ASP ASN ILE ALA GLY SER MET
 D) LYS VAL VAL ASP ALA LEU ASN GLU ALA SER ASN HIS LEU ASP ASN ILE ALA GLY SER MET
 E) LYS VAL VAL ALA ALA LEU ILE GLU ALA ALA ASN HIS ILE ASP ASP ILE ALA GLY THR LEU

A) AGC AAG CTG AGC GAC CTC CAT GCC TAT GAC CTG AGA GTG GAT CCG GGC AAC TTC CCA TTG⁶⁰
 B) AGC AAG CTG AGT GAC CTC CAT GCC TAT GAC CTG AGA GTG GAC CCT GGC AAC TTC CCA TTG
 C) SER LYS LEU SER ASP LEU HIS ALA TYR ASP LEU ARG VAL ASP PRO GLY ASN PHE PRO LEU
 D) SER LYS LEU SER ASP LEU HIS ALA TYR ASP LEU ARG VAL ASP PRO GLY ASN PHE PRO LEU
 E) SER LYS LEU SER ASP LEU HIS ALA HIS LYS LEU ARG VAL ASP PRO VAL ASN PHE LYS LEU

A) CTG GCT CAT AAT TTG CTG GTG GTT GTT GCT ATG CAC TTG CCT AAG CAG TTT GAT CCT GCA⁷⁰
 B) CTG GCC CAT AAT ATA TTG GTG GTT GTT GCT ATG AAC TTT CCT AAG CAG TTT GAT CCT GCA
 C) LEU ALA HIS ASN LEU LEU VAL VAL VAL ALA MET HIS PHE PRO LYS GLN PHE ASP PRO ALA
 D) LEU ALA HIS ASN ILE LEU VAL VAL VAL ALA MET ASN PHE PRO LYS GLN PHE ASP PRO ALA
 E) LEU GLY GLU CYS PHE LEU VAL VAL VAL ALA ILE HIS HIS PRO ALA ALA LEU THR PRO GLU

A) ACC CAT AAG GCC TTG GAT AAG TTC TTG GCT ACC GTA TCT ACT GTT CTG ACI TCC AAA TAT⁸⁰
 B) ACC CAT AAG GCC CTG GAT AAG TTC TTG GCT ACC GTA TCT ACT GTT CTG ACA TCC AAA TAT
 C) THR HIS LYS ALA LEU ASP LYS PHE LEU ALA THR VAL SER THR VAL LEU THR SER LYS TYR
 D) THR HIS LYS ALA LEU ASP LYS PHE LEU ALA THR VAL SER THR VAL LEU THR SER LYS TYR
 E) VAL HIS ALA SER LEU ASP LYS PHE LEU CYS ALA VAL GLY THR VAL LEU THR ALA LYS TYR

A) CGT TAA GGCTCAGCAGCAACTGCAGCAGCAATTTCACATCA_AATCAGCAATCAGATCAATCCAACTGACACA
 B) CGT TAA GGCTCAGCAACAACAGCAGCAGCAAGTCTCAACATCAGACATCAGTTAATTATGCAATCAAACTGACAAA
 C) ARG
 D) ARG
 E) ARG

A) ACTGTGAAAGATTTGTTCTGAAATAACATTTTAACCTTT POLY(A)
 B) GCTGTGAAAGATGTTCTGAAATAACATTTTAACCTTT POLY(A)

Figure 3: Nucleotide sequence of XG AD 11 and derived α^2 globin amino acid sequence of *Xenopus laevis*. For comparison the mRNA nucleotide and amino acid sequence of α^1 globin reported by Kay et al. (9)*, as well as chicken α^A globin amino acid sequence (27)** are included. Divergences between nucleotides or amino acids are indicated by underlining.

A) NUCL. SEQ. XEN. L. β^2 GLOBIN
 B) NUCL. SEQ. XEN. L. β^1 GLOBIN*
 C) AM. ACID SEQ. XEN. L. β^2 GLOBIN
 D) AM. ACID SEQ. XEN. L. β^1 GLOBIN*
 E) AM. ACID SEQ. CHICKEN β GLOBIN**

AAGCTCAGAATAAACGCTCAACTTTGGCC ATG

10
 A) 66T TTT ACA GCA CAT GAT CGT CAG CTG ATC AAC AGC ACC TGG GGC AAA CTA TGT GCC
 B) GLY LEU THR ALA HIS ASP ARG GLN LEU ILE ASN SER THR TRP GLY LYS LEU CYS ALA
 C) VAL HIS TRP THR ALA GLU GLU LYS GLN LEU ILE THR GLY LEU TRP GLY LYS VAL ASN VAL
 20
 A) ATA GGA AAA GAG GCC CTT GGC AGG CTG CTG AAI ACT TTT CCC TGG ACC CAA AGA
 B) AAG ACT ATT GGA CAA GAG GCC CTT GGA CGT CTG CTG TGG ACT TAT CCC TGG ACC CAA AGA
 C) ILE GLY LYS GLU ALA LEU GLY ARG LEU LEU ASN THR PHE PRO TRP THR GLN ARG
 D) LYS THR ILE GLY GLN GLU ALA LEU GLY ARG LEU LEU TRP THR TYR PRO TRP THR GLN ARG
 E) VAL GLU CYS GLY ALA GLU ALA LEU ALA ARG LEU LEU ILE VAL TYR PRO TRP THR GLN ARG
 30
 A) TAC TTT AGT TCT TTT GGG AAC CTC AGC AGT GCT GAA GCC ATC TTC CAC AAT GAA GCT GTG
 B) TAC TTT AGT TCT TTT GGG AAC CTC AAC AGT GCT GAT GCC GTC TTC CAC AAT GAG GCT GTG
 C) TYR PHE SER SER PHE GLY ASN LEU SER SER ALA GLU ALA ILE PHE HIS ASN GLU ALA VAL
 D) TYR PHE SER SER PHE GLY ASN LEU ASN SER ALA ASP ALA VAL PHE HIS ASN GLU ALA VAL
 E) PHE PHE ALA SER PHE GLY ASN LEU SER SER PRO THR ALA ILE LEU GLY ASN PRO MET VAL
 40
 A) GCT GCT CAT GGT GAA AAG GTG GTG ACA TCT GTT GGA GAG GCC ATC AAA CAC ATG GAC GAC
 B) GCT GCT CAT GGT GAA AAG GTG GTG ACA TCT ATT GGA GAG GCC ATC AAG CAC ATG GAT GAC
 C) ALA ALA HIS GLY GLU LYS VAL VAL THR SER VAL GLY GLU ALA ILE LYS HIS MET ASP ASP
 D) ALA ALA HIS GLY GLU LYS VAL VAL THR SER ILE GLY GLU ALA ILE LYS HIS MET ASP ASP
 E) ARG ALA HIS GLY LYS LYS VAL LEU THR SER PHE GLY ASP ALA VAL LYS ASN LEU ASP ASN
 50
 A) ATA AAG GGC TAT TAT GCT CAG CTG AGC AAG CAC CAC TCA GAA ACC CTA CAT GTG GAT CCA
 B) ATA AAG GGA TAT TAT GCT CAG CTG AGC AAA TAC CAC TCA GAG ACC CTA CAT GTG GAT CCA
 C) ILE LYS GLY TYR TYR ALA GLN LEU SER LYS HIS HIS SER GLU THR LEU HIS VAL ASP PRO
 D) ILE LYS GLY TYR TYR ALA GLN LEU SER LYS TYR HIS SER GLU THR LEU HIS VAL ASP PRO
 E) ILE LYS ASN THR PHE SER GLN LEU SER GLU LEU HIS CYS ASP LYS LEU HIS VAL ASP PRO
 60
 A) GCC AAC TTC AAG CGC TTC GGT GGC TGC TTG TTC ATT ACC CTG GCT CAC CAC TTT GAT GAA
 B) TTG AAC TTC AAG CGC TTC GGT GGC TGC TTG TCT ATT GCC TTG GCT CGC CAC TTC CAT GAA
 C) ALA ASN PHE LYS ARG PHE GLY CYS LEU PHE ILE THR LEU ALA HIS HIS PHE GLY GLU
 D) LEU ASN PHE LYS ARG PHE GLY CYS LEU SER ILE ALA LEU ALA ARG HIS PHE HIS GLU
 E) GLU ASN PHE ARG LEU LEU GLY ASP ILE LEU ILE VAL LEU ALA ALA HIS PHE SER LYS
 70
 A) GAA TAT ACA CCA GAA ITG CAT GCT GCA TAT GAA CAT CTG TTT GAT GTA ATT GCC GAC GCC
 B) GAA TAT ACA CCT GAG CTA CAT GCT GCC TAT GAA CAT CTT TTT GAT GCC ATT GCC GAC GCC
 C) GLU TYR THR PRO GLU LEU HIS ALA ALA TYR GLU HIS LEU PHE ASP VAL ILE ALA ASP ALA
 D) GLU TYR THR PRO GLU LEU HIS ALA ALA TYR GLU HIS LEU PHE ASP ALA ILE ALA ASP ALA
 E) ASP PHE THR PRO GLU CYS GLN ALA ALA TRP GLN LYS LEU VAL ARG VAL VAL ALA HIS ALA
 80
 A) CTT GGC AGG GGT TAC CAC TGA ACCAGCCTCAAGAACACCTGAATGGAGTCGCTAAGCTACATAGAAG
 B) CTT GGC AAG GGT TAC CAC TAA ACCAGCCTCAAGAACACCCGAATGGAGTCGCTAAGCTACATAATACC
 C) LEU GLY ARG GLY TYR HIS
 D) LEU GLY LYS GLY TYR HIS
 E) LEU ALA ARG LYS TYR HIS
 90
 A) AACCTACAGTACACTTTTACATAATGTTTTCCTCCCAAAATGTAGCCATTCTATATAGCTCTCTAATAAAGAAG
 B) AACCTACA CTTTACAAAATGTTGTGCC CCAAAATGTAGCCATTCTGATCTGCTCTAATAAAAAAGAAAG
 100
 A) TTTCTTCACATTCT POLY(A)
 B) TTTCTTCACATTCT POLY(A)

Figure 4: Nucleotide sequence of XG AF 1 and derived partial β^2 globin amino acid sequence of *Xenopus laevis*. For comparison the nucleotide sequence and complete amino acid sequence of β^1 globin, reported by Williams et al. (8)*, as well as the chicken β globin amino acid sequence (28)** are included. Divergences between nucleotides or amino acids are indicated by underlining. The 145 amino acids of the *Xenopus* β^1 sequence have been aligned to the 146 amino acids of the chicken β sequence beginning at the C-terminus.

tion maps of the α^2 sequence from XG AD 11 and the β^2 sequence from XG AF 1 as well as sequencing strategies. Most stretches of the two sequences were analysed in both directions. After subtracting the G/C tails (on an average 20-30 nucleotides each) and the poly(A) tail (16 and 18 nucleotides, respectively) the α^2 sequence had a length of 575 and the β^2 sequence a length of 510 nucleotides. Sequencing data and the amino acid sequences derived by them are given in figures 3 and 4. The α^2 sequence comprises 39 nucleotides (including the start codon ATG) at the 5'-end, the complete sequence coding for 141 amino acids, and 113 nucleotides (including the stop codon TAA) at the 3'-end. The β^2 sequence contains the information for 124 amino acids beginning with amino acid 22 or 23, respectively (see legend to figure 4), and has 138 nucleotides (including the stop codon TGA) at the 3'-end. For comparison the nucleotide and amino acid sequences of Xenopus α^1 (9) and of β^1 (8) as well as the nucleotide derived amino acid sequences of chicken α^A (27) and β globins (28) are included in figures 3 and 4.

DISCUSSION

The comparison between nucleotides of the α^1 sequence and those of the α^2 sequence reveals a divergence of 7.6% (32/423) for the coding and 14.9% (17/114) for the 3' untranslated region the deletion of one nucleotide in the 3' untranslated region of the α^2 sequence has been taken into account. The comparison of the 5' untranslated region reveals a divergence of 16.7% (4/24), but as this value is mainly due to three nucleotides at the 5'-end of the α^1 sequence (where cloning artifacts most likely arise) it should be taken as preliminary unless it is confirmed by sequencing data from longer cDNA clones or genomic DNA fragments. The corresponding values for the β sequences are 10.8% (40/372) for the coding and 13.7% (19/139) for the 3' untranslated part, including the deletion and insertion of one nucleotide, each, and the insertion of GTACA in the β^2 sequence. The overall divergence for the coding and 3' untranslated sequence is 9.1% (49/537) for the α chains and 11.5% (59/511) for the β chains. These numbers do roughly correspond

to divergences of 6% and 8%, which had been determined from melting curves of cloned α - and β sequences by Widmer et al. (4). With regard to these experiments it has to be taken into consideration that G/C tails, which are introduced by the cloning procedure, take part in the hybridization reaction. This could explain that the actual values being calculated from the divergence of nucleotide sequences are somewhat higher. On the other hand, as complete data for the 5'-end of α^1 and β^2 sequences are not yet known, the divergences of the complete sequences may also slightly differ from our values.

The mutations in the coding regions of the α sequences as well of the β sequences mostly occur at third positions of base triplets. However, since in some instances also the first and second positions are concerned, there are resulting amino acid exchanges. We observed eight exchanges within the 141 amino acids of the α chains (5.7%), and 15 exchanges within the 124 amino acids of the β chains being compared (12.1%). It is interesting to note, that divergence of β sequences is much higher than that of α sequences, and that this difference, which was already observed for the overall nucleotide sequence divergences (see above), is even more evident at the level of polypeptides. A comparison of the *Xenopus* globin amino acid sequences with those of α^A and β globin chains of chicken revealed a 60.3% (85/141) homology for the α chains and a 53.1% (77/145) homology for the β chains (under the precondition, that the 145 amino acids of the *Xenopus* β^1 chain are aligned to the 146 amino acids of chicken β chain beginning from the C-terminus and that homologies with major as well with minor components are considered). Most interestingly, six out of eight amino acid differences between *Xenopus* major and minor α (70, 112) as well as β chains (49, 54, 136, 143) are such that the minor chains are identical to the chicken sequences (two are vice versa (α 29, β 35)). From this observation two alternative models of gene conservation might be deduced. First, the minor components after having been arisen by genome duplication were subjected to the same evolutionary changes as chicken globin chains. Then the major chains could resemble the ancestral genes before genome duplication or could have undergone further

evolutionary exchanges at these amino acid positions. The second possibility is that the minor components resemble the ancestral genes and that the globin sequence at these amino acid positions has been conserved during evolution to birds. We believe this alternative to be more likely, since amphibia had diverged from birds much earlier than genome duplication in *Xenopus laevis* could occur. Except for position α 70, identical amino acids are found at the other five positions in α and β chains of goose (29) and α chain of pheasant (27). Moreover, four of these six amino acids (α 70, α 112, β 49 and β 54) are also found in mouse globins (for nucleotide and derived amino acid sequences (30, 31)). The major genes in *Xenopus* should instead have undergone evolutionary exchanges to adapt to the special physiological requirements in this species. This theory is supported by the different cDNA clone numbers, indicating that the major gene transcripts are actually present at higher amounts than those of the minor genes. However, the question remains, why the minor genes are nevertheless transcribed. Investigations on the physical and functional properties of *Xenopus* globins are necessary to solve this problem. Furthermore, sequencing of the α and β globin genes of *Xenopus tropicalis*, a rather primitive species, which did not undergo genome duplication, would be of valuable help to elucidate the evolutionary development in the *Xenopus* species.

ACKNOWLEDGEMENTS

This investigation was supported by grants from the Deutsche Forschungsgemeinschaft (Sonderforschungsbereich Embryonale Entwicklung und Differenzierung). We are indebted to the authors of ref (9) for sending a preprint of their article before publication.

*This article is dedicated to Prof. Dr. Heinz Tiedemann on the occasion of his 60th birthday

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. Battaglia, P. and Melli, M. (1977) *Dev. Biol.* 60, 337-350.
4. Widmer, H.J., Andres, A.C., Niessing, J., Hosbach, H.A. and Weber, R. (1981) *Dev. Biol.* 88, 325-332.
5. Humphries, P., Old, R., Coggins, L.W., McShane, T., Watson, C. and Paul, J. (1978) *Nucl. Acids Res.* 5, 905-924.
6. Kay, R.M., Harris, R., Patient, R.K. and Williams, J.G. (1980) *Nucl. Acids Res.* 8, 2691-2707.
7. Richardson, C., Capello, J., Cochran, M.D., Armentrout, R.W. and Brown, R.D. (1980) *Dev. Biol.* 78, 161-172.
8. Williams, J.G., Kay, R.M. and Patient, R.K. (1980) *Nucl. Acids Res.* 8, 4247-4258.
9. Kay, R.M., Harris, R., Patient, R.K. and Williams, J.G. (1983) *Nucl. Acids Res.*, 11, 1537-1542
10. Bisbee, C.A., Baker, M.A., Wilson, A.C., Hadji-Azami, I. and Fischberg, M. (1977) *Science* 195, 785-787.
11. Thiebaud, C.H. and Fischberg, M. (1977) *Chromosoma* 59, 253-257.
12. Jeffreys, A.J., Wilson, V., Wood, D., Simons, J.P., Kay, R.M. and Williams, J.G. (1980) *Cell* 21, 555-564.
13. Hosbach, H.A., Wyler, T. and Weber, R. (1983) *Cell* 32, 45-53.
14. Thomas, N. and Maclean, N. (1975) *J. Cell Sci.* 19, 509-520.
15. Aviv, H. and Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* 69, 1408-1412.
16. Zain, S., Sambrook, J., Richard, J.R., Keller, W., Fried, M. and Dunn, R. (1979) *Cell* 16, 851-861.
17. Martynoff, G., Pays, E. and Vassart, G. (1980) *Biochem. Biophys. Res. Commun.* 93, 645-653.
18. Roychoudhury, R., Jay, E. and Wu, R. (1976) *Nucl. Acids Res.* 3, 101-116.
19. Knöchel, W. and Bladauski, D. (1981) *Wilhelm Roux's Archives* 190, 97-102.
20. Wensink, P.C., Finnegan, D.J., Donelson, J.E. and Hogness, D.S. (1974) *Cell* 3, 315-325.
21. Cohen, S.N., Chang, A.C.Y. and Hsu, L. (1972) *Proc. Natl. Acad. Sci. USA* 69, 1408-1412.
22. Grunstein, M. and Hogness, D.S. (1975) *Proc. Natl. Acad. Sci. USA* 72, 3961-3965.
23. Knöchel, W. and Bladauski, D. (1980) *Wilhelm Roux's Archives* 188, 187-193.
24. Birnboim, H.C. and Doly, J. (1979) *Nucl. Acids Res.* 7, 1513-1523.
25. Maxam, A.M. and Gilbert, W. (1980) *Meth. Enzym.* 65, 499-560.
26. Sutcliffe, J.G. (1979) *Cold Spring Harbor Symp. Quant. Biol.* 43, 77-90.
27. Knöchel, W., Wittig, B., Wittig, S., John, M.E., Grundmann, U., Oberthür, W., Godovac, J. and Braunitzer, G. (1982) *Nature* 295, 710-712.
28. Richards, R.J., Shine, J., Ullrich, A., Wells, J.R.E. and Goodman, H.M. (1979) *Nucl. Acids Res.* 7, 1137-1146.
29. Oberthür, W., Braunitzer, G. and Kalas, S. (1981) *Hoppe-Seyler's Z. Physiol. Chem.* 362, 1101-1112.
30. Konkelt, D.A., Tilghman, S.M. and Leder, P. (1978) *Cell* 15, 1125-1132.
31. Nishioka, Y. and Leder, P. (1979) *Cell* 18, 875-882.