

Characterization of Adult α - and β -Globin Genes in the Zebrafish

By Fung-Yee Chan, Judith Robinson, Alison Brownlie, Ramesh A. Shivdasani, Adriana Donovan, Carlos Brugnara, John Kim, Bing-Chuen Lau, H. Ewa Witkowska, and Leonard I. Zon

Developmental switching of hemoglobins (Hbs) occurs in most vertebrates, yet the cellular and molecular basis for this process remains elusive. The zebrafish is a new genetic and developmental system that can be used to study embryogenesis, and mutants with a variety of defects in hematopoiesis have recently been derived. To initiate our studies on Hb switching in this organism, we have characterized the globins expressed in the adult. Reversed-phase high performance liquid chromatography and mass spectrometric analyses of adult peripheral blood hemolysates showed that there are three major α globins and two β globins in circulating erythroid cells. In addition, we have isolated and characterized zebrafish adult α - and β -globin cDNA clones that

encode some of these globins. High levels of α - and β -globin gene expression were detected in adult erythroid cells, whereas embryonic erythroid cells expressed little, if any, of these RNAs. We have also shown that the α - and β -globin genes are tightly linked on the same chromosome and are arrayed in a 3'-5' to 5'-3' configuration, respectively. The characterization of these genes and regulatory elements in this globin locus will provide insight into the process of globin gene transcription. With these reagents, future studies of Hb switching in zebrafish mutants with defective hematopoiesis will be possible.

© 1997 by The American Society of Hematology.

VERTEBRATE HEMATOPOIESIS is characterized by developmentally regulated alterations in globin gene expression as the site of blood formation changes.¹ In humans, primitive blood islands of the yolk sac express embryonic-specific globins (such as ζ and ϵ), whereas definitive erythroid cells in the fetal liver and bone marrow predominantly express α globin with γ and β chains, respectively. Developmental switching is regulated at two levels. First, gene transcription is activated by cell-specific and stage-specific transcription factors acting through proximal cis-regulatory sites and the locus control regions (LCRs). Second, the embryo forms primitive and definitive hematopoietic cell populations at different stages of development.^{2,3} These regulatory mechanisms appear to be conserved throughout vertebrate evolution, although, in teleosts, control of hemoglobin (Hb) switching at the transcriptional level has not yet been shown.

Lampreys express only one globin chain and assemble Hb tetramers.⁴ Subsequently, during the evolution of fish, globin duplication is believed to have given rise to the α - and β -globin complex. In *Xenopus*, the α and β globins are linked on the same chromosome, whereas, in higher vertebrates,

the α - and β -globin loci are found on distinct chromosomes.¹ In most species studied to date, transcription proceeds in the same direction (5' to 3' orientation) for the globin genes in a given locus. The globin genes typically consist of three exons and two introns¹ and contain a classical promoter with TATA and CAAT boxes, as well as CACCC and GATA sequence motifs. Globin gene expression is regulated by LCRs,⁵⁻⁹ which are DNase-I-hypersensitive and confer position-independent, copy-number-dependent, high-level expression of reporter genes in either erythroid cell lines or in erythroid cells of transgenic mice. Recent observations have suggested that each subregion of the LCR has unique characteristics for developmental expression.¹⁰ An LCR has been shown in chickens, mice, sheep, and humans.⁵⁻⁹

Although adult globin proteins have been characterized in some teleosts, globin gene expression and switching have not been studied extensively. Nevertheless, globin switching in fish has been predicted to occur based on differences in the characteristics of primitive and definitive erythroid cells^{11,12} and on differences between larval and adult globin genes in lampreys.⁴ Finally, the results of studies using polyclonal antisera generated to adult and larval teleost blood show a switching event that correlates with the onset of erythropoiesis in the pronephros.¹¹⁻¹⁴

The zebrafish is a powerful new developmental genetic system for analysis of early events in organogenesis. The initial site of hematopoiesis in the zebrafish (and most teleosts) is the intermediate cell mass of Oellacher (ICM), which is established by 23 hours after fertilization and expresses globin,¹⁵ the ICM is the functional equivalent of the yolk sac in most other vertebrates¹⁶⁻¹⁸ and expresses the transcription factors GATA-1 and GATA-2.¹⁵ Embryonic hematopoiesis proceeds through day 10 of development, after which definitive hematopoiesis is initiated in the pronephros; maturation of pronephric progenitors to erythroblasts occurs by day 12. In the adult, the mesonephric kidney is the major site of hematopoiesis.

Here, we have characterized the globins expressed in the adult zebrafish. The isolation of these globin genes will allow future study of Hb switching in the zebrafish, particularly in mutants that affect hematopoiesis. Analysis of genomic clones for these globins shows that the α and β globins are linked on the same chromosome, as are the *Xenopus* globin loci, but in a 3'-5' to 5'-3' configuration, respectively. Studies of transcriptional regulation of the globin genes present

From the Departments of Biochemistry and Medicine, Division of Hematology/Oncology, Children's Hospital and Dana-Farber Cancer Institute, Department of Pediatrics, Harvard Medical School, Children's Hospital, Boston, MA; the Howard Hughes Medical Institute; and the Children's Hospital Oakland Research Institute, Oakland, CA.

Submitted October 11, 1995; accepted September 4, 1996.

Supported by National Institutes of Health (NIH) Grants No. 1P50 DK 49216 and 5RO1 HL48801. L.I.Z. is an assistant investigator of the Howard Hughes Medical Institute. H.E.W. and B.-C.L. are supported through the NIH Northern California Comprehensive Sickle Cell Center Grant No. HL20985. The VG BioQ mass spectrometer was purchased from a grant of the NIH Shared Instrumentation Program (RR06505; C.H.L. Shackleton, PI).

Address reprint requests to Leonard I. Zon, MD, Division of Hematology/Oncology, Children's Hospital, 300 Longwood Ave, Enders 650, Boston, MA 02115.

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. section 1734 solely to indicate this fact.

© 1997 by The American Society of Hematology.

0006-4971/97/8902-0017\$3.00/0

in this unusual locus may provide important clues to the early evolutionary programs that lead to globin switching.

MATERIALS AND METHODS

Animals. Adult zebrafish (an inbred AB strain obtained from E. Sullivan [Harvard University, Cambridge, MA] and wild-type from Ekkwill Waterlife Resources [Gibson, FL]) were maintained under standard conditions at 28°C.¹⁹ Embryos were derived from matings of adult fish and staged according to Westerfield.¹⁹

Cells. Blood cells were obtained from adult zebrafish by intracardiac puncture with a 10- μ L glass microhematocrit tube (Curtin Matheson Scientific, Houston, TX). The blood cells were washed once in 0.9 \times phosphate-buffered saline (PBS; Dulbecco's 10 \times ; Life Technologies, Inc, Gaithersburg, MD) and subjected to RNA preparation. Embryonic blood cells were collected by immersing the embryo in 0.9 \times PBS and cutting across the tail with a sharp scalpel (no. 15, Bard-Parker; Becton Dickinson AcuteCare, Franklin Lakes, NJ), allowing blood cells to pool in the PBS and collecting them into a drawn glass pipette by capillary action. About 3,000 embryonic red blood cells (RBCs) could be obtained per embryo in this manner.

RNA preparation. Total cellular RNA was prepared by guanidinium isothiocyanate extraction.²⁰ Adult blood (10 μ L) yielded about 2 μ g and blood from 100 embryos yielded about 2 μ g of total cellular RNA.

Cloning of a zebrafish adult α -globin cDNA. Plaques (500,000) from a zebrafish retina cDNA library (prepared from AB and Ekkwill zebrafish by J.R. in the Uni-ZAP cloning vector [Stratagene, La Jolla, CA]) were screened with a visual pigment gene (red opsin) under high stringency conditions at 6 \times SSC (1 \times SSC = 0.15 mol/L sodium chloride, 0.015 mol/L sodium citrate; pH, 7.0). One cDNA was noted to be a fusion clone that included the visual pigment homologue and another cDNA. This additional cDNA was subcloned into the *Eco*RI and *Xho*I site of the Bluescript vector (Stratagene) and subjected to dideoxynucleotide sequence analysis using the Sequenase Version 2.0 kit (US Biochemicals, Cleveland, OH) following standard protocols.²¹ A search of the basic local alignment search tool (BLAST) database²² showed that the cDNA encoded a full-length α globin. We refer to this globin as α_{A1} (α globin, Adult, gene 1).

Cloning of a genomic DNA clone for the adult α globin. Filters bearing 600,000 plaques from an amplified zebrafish genomic DNA library (kindly provided by Dr M. Kieran, Children's Hospital, Boston, MA) were screened for hybridization with the full-length α -globin cDNA. Filters were hybridized following standard protocols²¹ under high stringency at 6 \times SSC and were washed to 0.25 \times SSC. Four hybridizing plaques were isolated, and DNA was prepared using standard techniques.²¹ Digestion of each independent phage DNA with several restriction enzymes show similar patterns, and one phage (designated 2-1) was used for further analysis and subcloning. Southern blot analysis,²¹ using the full-length α -globin cDNA as probe, determined which restriction fragments contained the α -globin gene. A 2.1-kb *Pst*I fragment from one phage (designated 2-1) hybridized strongly and was subcloned into Bluescript (Stratagene). The fragment was sequenced, as described above, using synthetic oligonucleotide primers. The sequence was analyzed with the BLAST program²² for similarities to other genes.

Isolation of β -globin cDNAs. We isolated an adult β -globin cDNA as part of a random sequencing project of cDNAs from an adult zebrafish kidney cDNA γ ZAP II (Stratagene) library (kindly provided by Drs J. Rast and G. Litman, University of South Florida, St Petersburg, FL). 100 random cDNAs were sequenced, and one (clone 27) was determined to encode β -globin based on BLAST analysis.²² This β -globin cDNA is referred to as β_{A1} (β globin, Adult, gene 1).

— The sequence of one end of the 2.1-kb *Pst*I fragment from phage

2-1 showed homology to other fish β globins (based on BLAST analysis). Filters bearing 150 plaques from a zebrafish spleen cDNA library (kindly provided by Drs J. Rast and G. Littman) were hybridized with the *Pst*I fragment as probe under high stringency conditions (as described above). Numerous plaques hybridized to the probe, and four were isolated for subcloning. Each phage was subjected to polymerase chain reaction (PCR) amplification (94°C for 1 minute, 55°C for 1 minute, and 72°C for 1 minute; 25 cycles) with the primers directed to phage arms (5'-GGTGGCGACGACTCC-TGGAGCCCG-3' and 5'-TTGACACCAGACCACTGGTAATG-3'). The products were subcloned into the pCR II vector following the protocol provided with the TA Cloning kit (Invitrogen Corp, San Diego, CA). Sequence of three cDNA clones show identity to the α_{A1} -globin sequence, and one phage was very similar to the sequence of the β -globin gene on the *Pst*I fragment; this β globin (called β_{A2}) appears to represent a transcript from another highly homologous β -globin gene.

Reverse-transcription PCR (RT-PCR) analysis. RNA samples were analyzed by RT-PCR.²³ First-strand cDNA was prepared from 1.5 μ g of total RNA, using oligo (dT)₁₂₋₁₈ (Collaborative Research, Lexington, MA) as primer, and Moloney murine leukemia virus (M-MLV reverse transcriptase (Life Technologies, Inc). The reaction conditions (for a 40- μ L reaction volume) are as follows: 50 mmol/L Tris-HCL (pH, 8.3); 75 mmol/L KCl; 3 mmol/L MgCl₂; 1 mmol/L each deoxyguanosine triphosphate (dGTP), deoxycytidine triphosphate (dCTP), deoxyadenosine triphosphate (dATP), and deoxythymidine triphosphate (dTTP); GeneAmp deoxynucleotide triphosphates (Perkin Elmer-Cetus, Norwalk, CT); 10 mmol/L dithiothreitol; 80 U rRNasin Ribonuclease Inhibitor (Promega, Madison, WI); and 400 U M-MLV reverse transcriptase for 60 minutes at 50°C. A reaction with water replacing the RNA was included as the negative control.

One tenth of the RT product was used in PCR analysis. Amplifications were performed using primers designed from the sequences obtained (see below). The reaction conditions (for a 50- μ L reaction volume) are as follows: 10 mmol/L Tris-HCl (pH, 8.3); 50 mmol/L KCl; 1.5 mmol/L MgCl₂; 0.2 mmol/L each dGTP, dCTP, dATP, and dTTP; 200 ng of each primer; and 2.5 U Taq DNA polymerase (Boehringer Mannheim, Mannheim, Germany). A small amount of [α -³²P]dCTP was included as a tracer. Reactions were performed in a DNA Thermal Cycler (Perkin Elmer-Cetus). Thirty PCR cycles (denaturation at 94°C for 1 minute, primer annealing at 55°C for 1 minute, and primer extension at 72°C for 1 minute) were performed to detect expression of the specific genes. Expression of the constitutively expressed zebrafish EF-1 α was also examined to compare quality and amount of RNA in the reactions.

One fourth of the PCR sample was electrophoresed on a 5% polyacrylamide gel, and products were detected using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA).

The following primers were used in the PCR analysis (F, forward; R, reverse): for α -globin, 5'-CAAGGCTGTTGTTAGGC-3' (F) and 5'-GCACAGTGTGTTGTCAG-3' (R) (the primer pair amplifies cDNA fragment of length 447 bp, corresponding to nucleotides 60 to 507 of the cDNA sequence); for β -globin, 5'-GGCCTGTGG-GGAAAGCTC-3' (F) and 5'-GTTGTCGGGATCCACATGCAG-3' (R) (the primer pair amplifies cDNA fragment of length 269 bp, corresponding to nucleotides 62 to 331 of the cDNA sequence; these primers amplify both the zebrafish β globins, β_{A1} and β_{A2} [see Results]); and for EF-1 α , 5'-CGGTGACAACATGCTGGAGG-3' (F) and 5'-ACCAGTCTCCACACGACCCA-3' (R) (the primer pair amplifies cDNA fragment of length 220 bp, corresponding to nucleotides 651 to 870 of the cDNA sequence [GenBank L23807]).

Hb electrophoresis. Hb electrophoresis was performed as described previously for the mouse.²⁴ Whole blood from four adult zebrafish was incubated for 15 minutes at room temperature in an equal volume of 0.067 mol/L cystamine/0.33% ammonium hydrox-

ide/0.5 mmol/L dithiothreitol. The sample was loaded onto a Titan III-H cellulose acetate plate (Helena Laboratories, Beaumont, TX), and individual Hb species were resolved by electrophoresis in 0.018 mol/L Tris/0.1 mol/L boric acid/2 mmol/L EDTA (free acid; pH, 8.4) over 1 hour (300 V). The plate was stained with 1% Ponceau S in 5% trichloroacetic acid, washed repeatedly in 5% acetic acid, and photographed under indirect light.

High performance liquid chromatography (HPLC) and electrospray mass spectrometry of globin chains and proteolytic peptides. Hemolysates were prepared by washing RBCs with saline (3×) and freezing the RBCs. On thawing, 2 vol of water was added; the sample was centrifuged in an Eppendorf centrifuge at 16,000g for 15 minutes in the cold cabinet (temperature, 5°C), and supernatant was collected. Automated analytical reversed-phase HPLC of hemolysates was performed as described previously²⁵ using a Spectra Physics HPLC system (Thermo Separation Products, Fremont, CA). The elution gradient was based on a method of Shelton et al²⁶ but modified to accommodate higher hydrophobicity of the zebrafish versus human globins; it consisted of two linear steps: from 42% A to 36% A in 45 minutes and then to 25.6% A in 65 minutes, where A was 20% acetonitrile/0.1% trifluoroacetic acid (TFA) and B was 60% acetonitrile/0.1% TFA. The mass spectrometric techniques used for intact globin and tryptic peptides analyses were described in detail in a review article.²⁷ A triple quadrupole VG BioQ electrospray mass spectrometer (Micromass, Altrincham, UK) interfaced with a Michrom narrow-bore HPLC system (Michrom BioResources, Auburn, CA) was used. Mass spectrometer was controlled, and the data were analyzed using a MassLynx 2.1 software; intact protein spectra were deconvoluted to a real molecular mass using a MaxEnt software²⁸ (all software from Micromass). Cone voltage ramp (25 V to 75 V) was used for acquisition of electrospray mass spectra of intact globins within a 500 to 1500 mass-to-charge (m/z) range; the collision-induced dissociation (CID) experiments were performed at a constant cone voltage of 60 V. The conditions of proteolysis of globins and mass spectrometric analysis of tryptic peptides were as described before. Before HPLC/electrospray ionization mass spectrometry (ESMS) analysis, proteolytic peptide mixtures were on-line desalted on a peptide cartridge (Michrom BioResources).

RESULTS

Isolation of a zebrafish α -globin cDNA. Screening of a zebrafish retina cDNA library for visual pigment cDNAs identified a single phage clone that contained two fused cDNAs encoding a visual pigment protein and an unrelated cDNA. The unrelated cDNA was subcloned between the *EcoRI* and *Xho I* sites of Bluescript KSII[−] (Stratagene, CA) and was subjected to dideoxynucleotide sequencing.²¹ Searching of GenBank databases (BLAST)²² showed that the cDNA encoded an α -globin (Fig 1A). We refer to this globin as α_{A1} . Close homology to globins in other fish species was observed (Fig 1B).

Isolation of a zebrafish α -globin genomic DNA phage clone. To characterize the α -globin gene, the α_{A1} cDNA was used as a probe to screen a zebrafish genomic DNA library. A total of 500,000 independent phage were screened, and four independent phage clones were isolated. Restriction enzyme digestion of DNA prepared from the phage showed that three of these clones were identical. A fourth clone showed a distinct pattern of fragments from the restriction enzyme digestion. Southern blot analysis was performed on DNA prepared from one of the three identical clones (designated 2-1) using the α_{A1} cDNA probe. A hybridizing 2.1-kb *Pst I* fragment was subcloned and contained the α_{A1} pro-

moter, as well as exons 1 and 2 and part of exon 3 (Fig 2A). Sequence analysis of the entire fragment showed that it also contained sequences encoding a β -globin gene. The direction of transcription of the α - and β -globin genes, predicted by intron-exon structure, was surprisingly in a 3'-5' to 5'-3' configuration (Fig 2B). Each promoter contained cis elements including TATA and CAAT boxes (underlined in the figure). GATA and CACCC sites in the α -globin and β -globin promoters (marked with double lines; exons are boxed) are also present. The first and second introns for the α -globin are 150 and 95 nucleotides in length, respectively. The first intron for the β -globin gene, as determined by comparison with the cDNA sequence (see below), is 100 nucleotides in length. An additional β_{A1} -globin gene fragment, including intron 2 (which is 98 nucleotides in length), was isolated by PCR with primers derived from exon 2 and 3. The intron-exon boundaries are presented in Fig 2C.

Isolation of a β -globin cDNA. A random sequencing project of 100 cDNA clones from an adult kidney cDNA library yielded one cDNA (designated 27) that encodes a β -globin (called β_{A1}) which is identical in sequence to the exons of the β -globin gene in the 2.1-kb *Pst I* fragment of phage 2-1 (Fig 3A). These cDNA sequences were used to determine the intron-exon boundaries of the β_{A1} gene (Fig 2). Sequence comparisons (using the BLAST program)²² with other fish globins showed close homology (Fig 3B).

Other globins. The globins expressed in adult zebrafish have previously been studied by electrophoresis²⁹ (although these results are published only in abstract form without figures). To confirm these results, adult zebrafish peripheral blood was subjected to Hb electrophoresis. Two strong bands and one weaker band were evident (Fig 4), which suggested that other adult globin genes are expressed in adult zebrafish in addition to the α - and β -globin genes that we have isolated. To obtain other globin cDNAs, 150 plaques from a zebrafish spleen cDNA library were screened with the *Pst I* fragment of phage 2-1, and four clones were isolated and sequenced. The sequence of three cDNA clones matched the previously isolated adult α -globin cDNA. The fourth clone was shown to encode a β globin (Fig 5). This β -globin (called β_{A2}) was 96% identical to β -globin A, and the predicted amino acid sequences indicated 3 differences between the two globins (Fig 3B). The amino acid at position 105 is arginine in β_{A1} globin and lysine in β_{A2} -globin B. It is possible that this difference could account for distinct migration patterns on Hb electrophoresis, but, based on the similarity of charge, it is unlikely. Thus, the other bands may represent tetramers between other α -globin and β -globin chains. Southern analysis of phage 2-1 DNA showed hybridization of the α - and β -globin cDNAs to bands distinct from the α - and β -globin genes studied here (data not shown). In addition, we have isolated a weak hybridizing globin cDNA that is expressed at a low level in the adult and in the embryo (data not shown). Thus, it appears that multiple α and β -globins are expressed in the adult zebrafish.

HPLC and mass spectrophotometric analysis of adult Hbs in the zebrafish. A combination of chromatographic and mass spectrometric techniques was further used to characterize the adult Hbs of the zebrafish. Reversed-phase HPLC of a mixture of globins derived from a number of wild-type

A

```

1 GAATTGGCAGCAGCACTTTGAGAAAAAGAAAACGCAGCGATGAGTCTCTCTGATACGGAC 60
      M S L S D T D
61 AAGGCTGTGTGTTAAGGCCATCTGGGCTAAGATCAGCCCCAAGGCCGATGAAATTTGGTGCT 120
      K A V V K A I W A K I S P K A D E I G A
121 GAAGCCCTCGCCAGAATGCTGACCGTCTACCCCTCAAACCAAGACCTACTTTTCTCACTGG 180
      E A L A R M L T V Y P Q T K T Y F S H W
181 GCTGACTTGAGCCCTGGGTCTGGTCCCGTGAAGAAGCACGGAAAGACTATCATGGGTGCC 240
      A D L S P G S G P V K K H G K T I M G A
241 GTCCGCGAAGCTATTTTCAAAAATAGACGACCTTGTGGGAGGACTGGCCGCCCTGAGCGAA 300
      V G E A I S K I D D L V G G L A A L S E
301 CTCCATGCCCTTCAAGCTGCGTGTGACCCGCGCAACTTCAAGATCCTGTACACAAATGTC 360
      L H A F K L R V D P A N F K I L S H N V
361 ATTGTGGTCATCGCCATGCTCTTCCCTGCAGACTTCACCCCTGAGGTTCACGTGTCAGTC 420
      I V V I A M L F P A D F T P E V H V S V
421 GACAAGTTCTTTAATAACTTGGCCCTGGCTCTCTCTGAGAAGTACCGCTAAACCTCAAGT 480
      D K F F N N L A L A L S E K Y R
481 GGATATTCACTGACAACAACACTGTGCGGGACTCTTAACAACATCATGCATGATGTTGGAA 540
541 TGAAATTTCTCAATAAAAGGCACAGAACTAAAAAAAAAAAAAAAAAAAACTCGAG 598

```

B

	1		50
Zebrafish	M.SLSDTDKA	VVKAIWAKIS	PKADEIGAEA LAR..MLTVY PQTkTYFSHW
Goldfish	M.SLSDKDKA	VVKALWAKIG	SRADEIGAEA LGR..MLTVY PQTkTYFSHW
Common Carp	M.SLSDKDKA	AVKGLWAKIS	PKADDIGAEA LGR..MLTVY PQTkTYFAHW
Salmon	M.SLTARDKS	VVNAFWGKIK	GKADVVGAEA LGR..MLTAY PQTkTYFSHW
Trout I	M.SLTAKDKS	VVKAFWGIS	GKADVVGAEA LGRDMLTAY PQTkTYFSHW
Desert Sucker	M.SLSDKDKA	DVKIAWAKIS	PRADEIGAEA LGR..MLTVY PQTkTYFAHW
Electric Eel	M.SLTAKSKS	TVKAFWGIS	SRADDIGAEA FGR..MLTVY PETkTYFASW
Trout IV	M.SLSAKDKA	NVKAIWGKIL	PKSDEIGEQA LSR..MLVVY PQTkAYFSHW
Bluefin Tuna	MTTSLDKDKS	TVKALWGIS	KSADAIGADA LGR..MLAVY PQTkTYFSHW
	51		100
Zebrafish	ADLSPGSGPV	KKHGKTIMG	VGEAISKIDD LVGGLAALSE LHAFKLRVDP
Goldfish	SDLSPGSGPV	KKHGKTIMG	VGDAVSKIDD LVGALSALSE LHAFKLRIDP
Common Carp	ADLSPGSGPV	KKHGKVMIG	VGDAVSKIDD LVGGLAALSE LHAFKLRVDP
Salmon	ADLSPGSAPV	KKHGGVIMIG	IGNAVGLMDD LVGGMGLSD LHAFKLRVDP
Trout I	ADLSPGSGPV	KKHGGIIMIG	IGKAVGLMDD LVGGMGLSD LHAFKLRVDP
Desert Sucker	ADLSPGSGPV	KHGKVMIG	IGDAVTKFDD LLGGLASLSE LHASKLRVDP
Electric Eel	SDLSPGSAV	KKHGKTIMG	IAEAVGHIDD LTGGLASLSE LHAFKLRVDP
Trout IV	ASVAPGSAPV	KKHGITIMNQ	IDDCVGHMDD LFGFLTKLSE LHATKLRVDP
Bluefin Tuna	PDMSPGSGPV	KAHGKVMIG	VALAVTKIDD LITGLGDLSE LHAFKMRVDP
	101		146
Zebrafish	ANFKILSHNV	IVVIAMLFPA	DFTPEVHVSV DKFFNNLALA LSEKYR
Goldfish	ANFKILAHNV	IVVIGMLFPG	DFTPEVHMSV DKFFQNLALA LSEKYR
Common Carp	ANFKILAHNV	IVVIGMLYPG	DFTPEVHMSV DKFFQNLALA LSEKYR
Salmon	GNFKILSHNI	LVTLAHFPA	DFTPEVHIIV DKFLAALSAA LADKYR
Trout I	GNFKILSHNI	LVTLAHFPS	DFTPEVHIIV DKFLAASVAA LADKYR
Desert Sucker	SNFKILANCI	TVVIMFYLPG	DFTPEVHASV DKFFQNLALA LGQKYR
Electric Eel	ANFKILAHNL	IVVLALFFPA	DFTPEVHMAV DKFFQNLASA LSEKYR
Trout IV	TNFKILAHNL	IVVIAAYFPA	EFTPEIHLVS DKFLQQLALA LAEKYR
Bluefin Tuna	SNFKILSHCI	LVVVAKMFPA	EFTPDHVSLS DKFLASVALA LAERYR

Fig 1. (A) Nucleotide sequence of the zebrafish adult α_{A1} -globin cDNA. The amino acid sequence is listed underneath. **(B)** Comparison of the predicted amino acid sequence of adult α globins of zebrafish and other fish species (obtained from Genbank). The initiator methionine is placed in italics, because the sequenced polypeptides do not include this methionine. Our studies show that the zebrafish α globins do contain the N-terminal methionine, but the β globins do not.

A

1 CTGCAGAACA AGAGATCAGT TAAGCTGTGG GTTCCCCATA TACTAATTTA AGAGCACGGC TGTTTAAGTA TTAACATCAT TCAGTTAATG CAGACGTAC

101 CTGAAGTTGT CGGGATCCAC ATGCAGTTTC TCAGAGTGCA TCACACTGAG GGCGGCATAG GTGTTCTTGA CGTTGTCCAT GTTCTTGATG GCTCTCTCAA
--> Exon 2

201 GACCTCCCAT CACAGTCCTC CCATGAGCTG CCACCTTGGG GTTACCCATG ATCGCAGCGG GGCTTGACAG GTTGCCGAAT GTGGCGAAAT ATCTCTGAGT

301 CCAGGGATAC ACGATCAGAC ATCTTCAAAA ACAGAGGCAT TAAGAGCGCC ATTAAGTAAA CTGCATATAG TTGGCATCAT TTAACACAAA GACAACTTT

401 CTACAAAGTA AAGAGTGAT A^MCTGGATAG GGCCTGAGGT CCGATTTTAT CGATATTGAG CTTTCCCCAC AGGCCAAGGA TGGCTGTGCG CTCGGCATCT
--> Exon 1

501 GTCCACTCAA CCA^MTGTTAC GCTTTGGAT GTCTTAGGT GCTTGCTTAA AGTCCAGTAG GAATCACTAG CAGAAATTCA GAAACCGGGG TGCATTTTAT
--> Exon 1

601 TGATGTGTGA TGTCTCCGCC CAATAAAGA TGCAAAATCT TATTGGATAA TCTTGCTGAG ATGACGCATG CTTTGGACAT CACTGATCGG ACCCATCCCA
GATA motif CACCC motif TATA box

701 AAAAAGCTGT TTTGAAAAAG AAAAAAAGT CTAAACATG AACCAGCATT TTTTTTTTTC ATTTTGGTCG TGCCCATATT AATCATGCAA TTTCTATGC

801 CTAAATATGT TCTGTTAAGG TCTTAAACA AATATAAATT ATGCTTTCCA ATAATAATA TAATAATAAT GATAATAATA ATAATAACAA TAATAATAAT

901 AATAATAATA ATAATAATA TAATAATAAT AATAATAATA ATAATAATAA TGATAAGTTG GTGGGAAATT ATTATTGAA AACATTAAAC ACGAAATTC

1001 TGTTTTCAAGT AATTCTGAG CCAACTTTA CTACACTTAT TTGTTTGTTC ATAATTACAA TATCAAGGCA TTGACATTAA AATGTGACAA AACACCTTTT
GATA motif CACCC motif

1101 CAAATATTGC AATTGAAAA AAAAAAGATT ATTTACAAAT CTATAGTGAT TTAATAATCA TTTTCGAAAT ATAATTTTGT TCATGATTCT TAATGGCTTT

1201 TAATTTTTAC TTGAGTGTG CATGTGTAAT TGTGTACAC ATTAATGCA GCTAAATGTG CAAAGCTTGA TTAAAGCATA CGTGATGTAT GTGTAATAAA

1301 ACTGTTATCT CCAAT box CAATTAAACA AGATGCAGAA AATTGGCAGC TTTTGTGATTG GACACCAGAA GCTGGAGATC AGCAGAAATG CGCCGAAGAG GTGTGGTTTC

1401 GACTGATTTT CACCC motif ATGGGTTGCT CCTCGTTTGG TTTTATATAA AAAGTAGGAC ACATGTCTGTG TTCCCAACA GCTTTTTCTG AAGAAACCTC ACTTGAGAAA
alpha-globin M TATA box

1501 AAGAAGCGC AGC^MATGAGT CTCTCTGATA CGGACAAGGC TGTGTTTAAAG GCCATCTGGG CTAAGATCAG CCCCAGGCC GATGAAATTG GTGCTGAAGC

1601 CCTCGCCAGG TAAAGCTCAC CTATGTTTTT TTTTCTATT TGAATTAGGA TTAATGATTG ACGGAACCTT ATAGAAGTTT ATTATTATTA TTAAGAATAA

1701 TCTATGATAA CTTATAATGT GTAAAAACAT AAATAACCTA TGTATTTTTT CCTCCAGAA TGCTGACCGT CTACCTCAA ACCAAGACCT ACTTTTCTCA
--> Exon 2

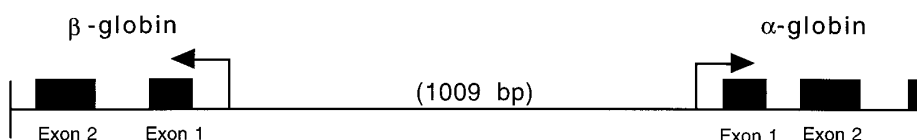
1801 CTGGGCTGAC TTGAGCCCTG GGTCTGGTCC CGTGAAGAAG CACGGAAAGA CTATCATGGG TGCCGTCGGC GAAGCTATTT CAAAAATAGA CGACCTTGTT

1901 GGAGGACTGG CCGCCCTGAG CGAACTCCAT GCCTTCAAGC TGCGTGTGTA CCCGGCCAAC TTCAAGGTAG GACGGGTCCA CAGCACGCA TAAAAATAAA

2001 TAAACGCTAG GTATTGCAGC ATTCATGTA TGGTAACGTT TTTCTGTTA TCTCTTTTCT A^MATCCTGTC ACACAATGTC ATTGTGGTCA TCGCCATGCT
--> Exon 3

2101 CTTCCTGCA G

B



C

α-globin

Exon 1	GCCTCGCCAG/GTAAAGCTCA.....TTCCCTCCAG/AATGCTGACC	Exon 2
	A L A R	M L T
Exon 2	CCAACTTCAAG/GTAGGACGGG.....TCTTTTCTAG/ATCCTGTCAC	Exon 3
	N F K	I L S

β-globin

Exon 1	GCCTATCCAG/GTATCACCTC.....GTTTTTGAAG/ATGCTGATC	Exon 2
	A L S R	C L I
Exon 2	ACAACTTCAAG/GTACGCTGTC.....TGTTCTGAG/CCTCTGCTG	Exon 3
	N F R	L L A

Fig 2. (A) Genomic sequence of the 2.1-kb *Pst* I fragment. The initiation codon for each globin is marked with an "M" and an arrow pointing towards the direction of transcription. The exons are boxed and labeled. CCAAT and TATA boxes are underlined (CCAAT with thin and TATA with thick lines). GATA and CACCC motifs are marked with double lines (GATA with thin and CACCC with thick lines). (B) 3'-5' to 5'-3' configuration of the globin genes. Arrows indicate the initiation site and direction of transcription for each globin. Black boxes represent the exons (*Pst* I fragment does not contain the 3' end of exon 3 of the α globin and exon 3 of the β globin). The 1009 bp of intervening sequences separate the two genes. (C) Intron-exon boundaries in the adult α_{A1}- and β_{A1}-globin genes.

A

```

1  GAATTCGGCACGAGCTGGACTTTTAAGCAAGCACCTAAGGACATCCAAAAGCGTAAACATG  60
    M
61  GTTGAGTGGACAGATGCCGAGCGCACAGCCATCCTTGGCCTGTGGGGAAAGCTCAATATC  120
    V E W T D A E R T A I L G L W G K L N I
121  GATGAAATCGGACCTCAGGCCCTATCCAGATGTCTGATCGTGTATCCCTGGACTCAGAGA  180
    D E I G P Q A L S R C L I V Y P W T Q R
181  TATTTGCCCACATTCGGCAACCTGTCAAGCCCCGCTGCGATCATGGGTAACCCCAAGTIG  240
    Y F A T F G N L S S P A A I M G N P K V
241  GCAGCTCATGGGAGGACTGTGATGGGAGGTCTTGAGAGAGCCATCAAGAACATGGACAAC  300
    A A H G R T V M G G L E R A I K N M D N
301  GTCAAGAACACCTATGCCGCCCTCAGTGTGATGCACCTCTGAGAACTGCATGTGGATCCC  360
    V K N T Y A A L S V M H S E K L H V D P
361  GACAACCTTCAGGCTTCTCGCTGATTGCATCACCGTTTGGCTGCCATGAAGTTTCGGCCAA  420
    D N F R L L A D C I T V C A A M K F G Q
421  GCTGGTTTCAATGCTGATGTCCAGGAGGCTGGCAGAAGTTTCTGGCTGTGGTCTGTTTCT  480
    A G F N A D V Q E A W Q K F L A V V V S
481  GCTCTGTGCAGACAGTACCCTAGAGTCTCATCGCCAATGAACGTCAGGTGTTTTCGAGA  540
    A L C R Q Y H
541  AGGTCACACCGCGCGGAGATGTTGAAACAAAGATCTAACTCTTTTGCAACAAGATCAATA  600
601  AAATAATGTTTAAAGCAAAAAAAAAAAAAAAAAAAAAAACTCGAG  645

```

B

	1				50
Zebrafish A1	M/EWIDAERT	AILGLWGKLN	IDEIGPQALS	RCLIVYPWIQ	RYFATFGNLS
Zebrafish A2	M/EWIDAERT	AILGLWGKLN	IDEIGPQALS	RCLIVYPWIQ	RYFATFGNLS
Goldfish	M/EWIDAERS	AIIGLWGKLN	PDELGPQALA	RCLIVYPWIQ	RYFATFGNLS
Common Carp	M/EWIDAERS	AIIALWGKLN	PDELGPEALA	RCLIVYPWIQ	RFFASYGNLS
Salmon	M/EWIDAERT	AILGLWGKLN	IDEIDPQALS	RCLIVYPWIQ	RYFATFGNLS
Trout IV	M/DWIDAERS	AIVGLWGKIS	VDEIGPQALA	RLIVSPWIQ	RHFSTFGNLS
Electric Eel	M/ELTEAQRG	ATVNLWGHLS	PDEIGPQALA	RLIVYPWIQ	RYFASFGNIS
Trout I	M/EWIDAESK	TISAVWGKVN	IDEIGPLALA	RVLIVYPWIQ	RYFGSFGNVS
Bluefin Tuna	M/EWTQQERS	IIAGFIANLN	YEDIGPKALA	RCLIVYPWIQ	RYFGAYGDLS

	51		*		100
Zebrafish A1	SPAAIMGNPK	VAAHGRTVMG	GLERAIKNMD	NVKNITYAALS	VMHSEKLHVD
Zebrafish A2	SPAAIMGNPK	VAAHGRTVMG	GLERAIKNMD	NIKNTYAALS	VMHSEKLHVD
Goldfish	SPAAIMGNPK	VAAHGRTVMG	GLERAIKNMD	NIKATYAPLS	VMHSEKLHVD
Common Carp	SPAAIMGNPK	VAAHGRTVEG	GLMRAIKMD	NIKATYAPLS	VMHSEKLHVD
Salmon	TPAAIMGNPA	VAKHGKTVMH	GLDRAVQNLD	DIKNAYTALS	VMHSEKLHVD
Trout IV	TPAAIMGNPA	VAKHGKTVMH	GLDRAVQNLD	DIKNYATLS	VMHSEKLHVD
Electric Eel	SAAAIMGNPK	VAAHGKVVVG	ALDKAVKNLN	NIKGTYAALS	TIHSEKLHVD
Trout I	TPAAIMGNPK	VAAHGKVVCG	ALDKAVKNMG	NILATYKSLS	ETHANKLFVD
Bluefin Tuna	TPDAIKGNAK	IAAHGVKVLH	GLDRAVKNMD	NINEAYSELS	VLHSDKLHVD

	101	*		*		148
Zebrafish A1	PDNFRLLADC	ITVCAAMKFG	QAGFNADVQE	AWQKFLAVVV	SALCRQYH	
Zebrafish A2	PDNFKLLADC	ITVCAAMKFG	QAGFNADIQE	AWQKFLAVVV	SALCRQYH	
Goldfish	PDNFRLLADC	ITVCAAMKFG	PSGFNADVQE	AWQKFLSVVV	SALCRQYH	
Common Carp	PDNFRLLADC	ITVCAAMKFG	PSGFSENVQE	AWQKFLSVVV	NALKRQYH	
Salmon	PDNFRLLADC	ITVCVAAKLG	PTVFSADIQE	AFQKFLAVVV	SALGRQYH	
Trout IV	PDNFRLLADC	ITVCVAAKLG	PAVFSADTQE	AFQKFLAVVV	SALGRQYH	
Electric Eel	PDNFRLLAES	FTVSVAMKLG	PSGFNAETQH	ALAKFLAEVV	SALGKQYH	
Trout I	PDNFRVLADV	LITVIAAKFG	.ASFTEPEIQ	TWQKFMKVVV	AAMGSRFY	
Bluefin Tuna	PDNFRILGDC	LITVIAANLG	DA.FIVETQC	AFQKFLAVVV	FALGRKYH	

Fig 3. (A) Nucleotide sequence of the zebrafish adult β_{A1} -globin cDNA. The amino acid sequence is listed underneath. **(B)** Comparison of the predicted amino acid sequence of the zebrafish adult β globins with β -globin sequences of other fish species. An asterisk (*) is used to designate the positions at which the amino acid sequences of zebrafish β_{A1} and β_{A2} globins differ (see Fig 5).



Fig 4. Hb electrophoresis of adult zebrafish blood. Two major and one minor band are detected.

adult zebrafish showed the presence of five HPLC peaks grouped in two distinct sets of low and high hydrophobicity (two and three peaks, respectively; see Fig 6). The same HPLC pattern was further observed in Hb samples obtained from individual families (data not shown). The relative ratio between amount of protein within each group was close to 1:1, suggesting that each set of HPLC peaks contained exclusively one type of globin (either α or β). The HPLC fractions were examined by ESMS, a technique that allows for measurement of molecular masses (M_r) of biomolecules with high accuracy (up to 0.01%; Fig 7).³⁰ Each of the first two HPLC peaks contained a single protein species with an

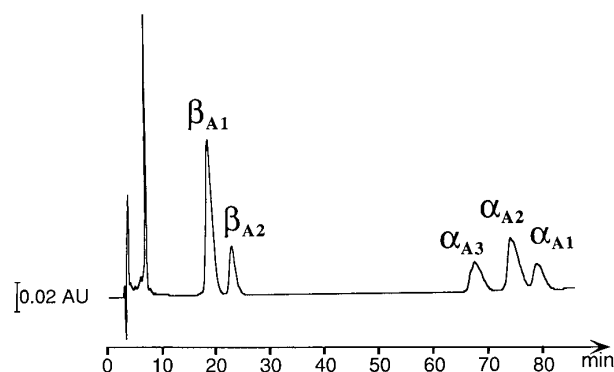


Fig 6. Reversed-phase HPLC of the hemolysate derived from a number of wild-type adult zebrafish (detection at 220 nm). The peak eluting at ~8 minutes contains heme molecules.

average molecular mass of 16257.4 Daltons and 16258.2 Daltons (in first and second peak, respectively). These molecular masses vary from those expected for known zebrafish β globins β_{A1} and β_{B2} ($M_r = 16389.1$) by -131 Daltons. The mass decrement of 131 Daltons can be rationalized by an elimination of one methionine from the protein sequence. We hypothesize that zebrafish β globins undergo a removal of the N-terminal methionine, a common processing of a nascent protein chain in eukaryotes,³¹⁻³³ and, thus, the desmethionyl β_{A1} and β_{B2} constitute the fast-eluting HPLC peaks. The three remaining HPLC peaks contained species whose molecular masses were close to the mass expected for the known α globin ($M_r = 15437.0$). Therefore, we putatively categorized the late-eluting species as adult zebrafish α globins. Their respective average molecular masses measured 15435.4, 15477.0, and 15434.0 Daltons and were

1	GGAATTTCGCGGGGCTGTAAACATGGTGTGAGTGGACAGATGCCGAGCGCACAGCCATCCT	60
	M V E W T D A E R T A I L	
61	TGGCCITGTGGGAAAGCTCAATATCGATGAAATCGGACCTCAGGCCCTATCCAGATGTCT	120
	G L W G K L N I D E I G P Q A L S R C L	
121	GATCGTGTATCCCTGGACTCAGAGATATTTCCGCCACATTTGGCAACCTGTCAAGCCCCGC	180
	I V Y P W T Q R Y F A T F G N L S S P A	
181	TGCGATCATGGGTAAACCCAAAGTGGCAGCTCACGGGAGGACTGTGATGGGAGGTCTCGA	240
	A I M G N P K V A A H G R T V M G G L E	
241	AAGAGCCATTAAGAACATGGACAACATCAAGAACACCTATGCCGCCCTCAGTGTGATGCA	300
	R A I K N M D N I K N T Y A A L S V M H	
301	CTCTGAGAACTGCATGTGGATCCCGACAACCTTCAAGCTTCTCGCTGACTGCATCACCGT	360
	S E K L H V D P D N F K L L A D C I T V	
361	TTGCGCTGCAATGAAGTTCCGCCAGGCTGGTTTCAATGCTGATATCCAGGAGGCTGGCA	420
	C A A M K F G Q A G F N A D I Q E A W Q	
421	GAAGTTTCTGGCTGTGGTCGTTTCCGCTCTGTGCAGACAGTACCACCTAGAGTCTCATCGC	480
	K F L A V V V S A L C R Q Y H	
481	CAATGAACGCCAGGTGTGTGTGGAGAAGGTGACACCGCGGAACGTGTTCAAACAAAGATC	540
541	TAACTCTTTTACAACACGGCCCGAATTC	568

Fig 5. Nucleotide and amino acid sequence of the β_{A2} -globin cDNA.

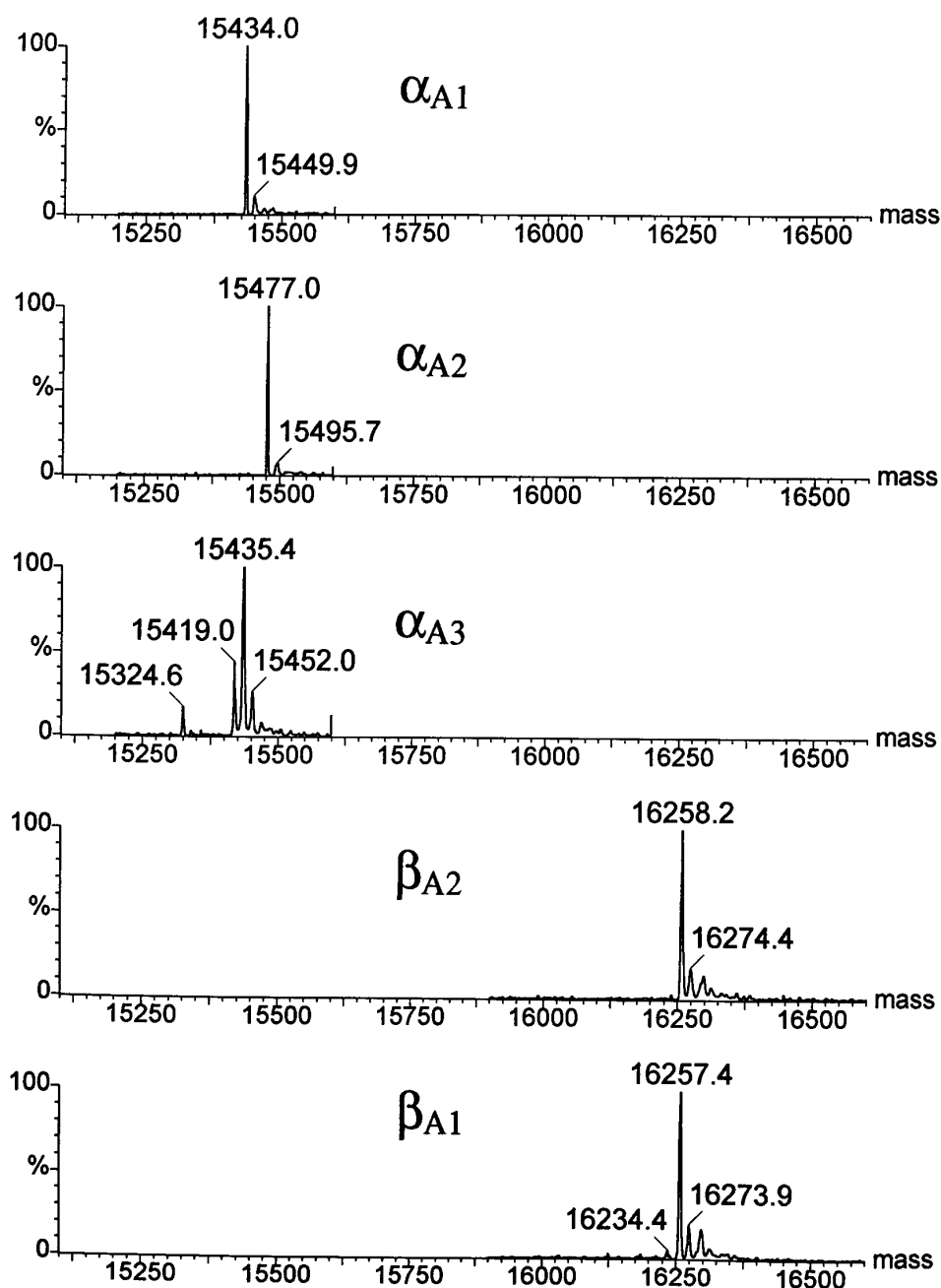


Fig 7. Electrospray ionization mass spectra of the zebrafish globins deconvoluted to real mass scale by MaxEnt software. Before ESMS analysis, globins were isolated by reversed-phase HPLC (Fig 6). Mass spectra of the HPLC fractions are shown in order of their elution; the bottom panel shows the mass spectrum of the least hydrophobic globin β_{A1} . Average molecular masses calculated for β_{A1} , β_{A2} , and α_{A1} globins on the basis of their known sequences are 16257.8, 16257.8, and 15437.0 Daltons, respectively.

subsequently assigned as α_{A3} , α_{A2} , and α_{A1} , respectively. It was noted that, whereas the α_{A2} and α_{A1} fractions looked pure and homogeneous in mass spectrometric analysis, the α_{A3} peak showed heterogeneity and contained at least two minor components (15324.6 and 15419.0 Daltons) in addition to the major species of 15435.4 Daltons. The molecular mass of the α_{A1} fraction showed a 3-Dalton decrement when compared with the value predicted for the known α_{A1} sequence. This mass difference slightly exceeds the method error of ± 1.5 Daltons typically observed for a 15-kD protein. At this point of analysis, we have not been able to establish whether a discrepancy in molecular mass is caused by an error in mass measurement or whether it reflects a real and

yet noncharacterized alteration in protein structure. In conclusion, reversed-phase HPLC and ESMS of adult zebrafish globins showed the presence of two β and at least three α globins.

Final confirmation of putative identities of zebrafish Hb components required further analysis, and it was performed on proteins isolated by reversed-phase HPLC. We attempted to verify that proteins eluting in the fast- and slow-HPLC peaks were indeed zebrafish α - and β -globin-related, respectively. In addition, we wished to establish which of the HPLC peaks contained the species consistent with the known sequences of the β_{A1} , β_{A2} , and α_{A1} globins. Two analytical approaches were used, mass spectrometric sequencing of

intact globins by means of CID in the gas phase³⁴ and peptide mapping. Although mass spectrometric sequencing of intact proteins is still in its early developmental stage, it already offers a great potential in evaluating differences between the very similar and generally known sequences. When intact protein (in a form of the multiply charged precursor ions) is subjected to collisions in the gas phase, some of the peptide bonds undergo cleavages, and protein fragments (product ions) encompassing various parts of the protein sequence are generated. Agreement between the measured and calculated molecular masses of the product ions points to the correspondence between their amino acid constitution and protein sequence. Unequivocal identification of a product ion requires that its charge state is independently established. However, even when a direct assessment of the ion charge states is not possible, a reliable assignment of product ions is still feasible provided that the following conditions are met: (1) more than one charge state is observed for a given fragment and/or (2) a series of fragments that differ by a single amino acid residue is observed. Interpretation of the protein CID data described below is based exclusively on analysis of the product ions that fulfilled at least one of the two criteria defined above.

Fragmentation patterns of both putative β globins were identical (spectrum of β_{A1} is shown in Fig 8). The major fragmentation pathway involved cleavages at the N-terminal side of the ⁵¹Pro; it is of interest that the same site in human β globin was previously found to be the most susceptible to CID.³⁵ A series of intense multiply charged y ions (a nomenclature of multiply charged product ions follows the conventions proposed by Roepstroff and Fohlman³⁶ and Biemann et al³⁷ for the singly charged product) corresponding to the zebrafish β -globin amino acid residues ⁴⁹Ser-¹⁴⁷His, ⁵⁰Ser-¹⁴⁷His, and ⁵¹Pro-¹⁴⁷His documented an agreement between the observed molecular masses and a constitution of the β_{A1} and β_{A2} globins within the sequence ⁴⁹Ser-C-terminus. Less intense cleavages were generated at several other sites, including the N- and C-terminal parts of the molecule, and proved to be of major diagnostic importance. A series of doubly charged b ions encompassing amino acid sequences ¹Val-¹⁰Ala through ¹Val-¹⁵Trp proved that there was no initiator methionine at the N-termini of zebrafish β globins. A series of triply charged y ions encompassing C-terminal sequences originating at ¹²⁶Asp and ¹²⁸Gln through ¹³⁰Ala (y22³ and y20³-y18³, respectively) indicated that β globin eluting in the first HPLC peak carried ¹²⁷Val (within y22³) characteristic for the β_{A1} globin. The y22³ ion diagnostic for the β_{A1} globin was not observed in the spectrum derived from the putative counterpart; however, the low intensity of this ion series in the β_{A2} spectrum failed to allow positive assignment of the more hydrophobic β globin as carrying ¹²⁷Ile. Thus, on the basis of protein CID data, two fast-eluting HPLC peaks were shown to carry zebrafish β globins, with the β_{A1} species being the less hydrophobic of the two.

All three α globins seemed to have followed the same fragmentation pathway in the gas phase (Fig 9). All the intense product ions were derived from the cleavages at the N-terminal site of the ¹¹⁵Pro; both the b-type ions covering the N-terminal sequence ¹Met-¹⁰⁸Val through ¹Met-¹¹⁴Phe

and the y-type ions encompassing the C-terminal part of the globin sequence ¹¹⁵Pro-¹⁴²Arg were clearly seen. In human α globin, the product ions derived from the fragmentations at the same proline residue are also very abundant,^{38,39} although they are accompanied by the ions generated at other cleavage sites whose counterparts we did not observe in the zebrafish α -globin spectra. It was the α_{A1} globin (eluting in the slowest HPLC peak) whose multiply charged product ions b108-b114 and y28 were consistent with the known sequence of zebrafish α globin (Fig 9, top panel). The b-type ions derived from the α_{A1} fraction showed a decrement in mass of 3.0 ± 1.6 Daltons ($n = 23$), whereas the y-type ions were measured accurately (mass difference between the observed and predicted values for y ions $+0.5 \pm 0.2$ Daltons; $n = 5$). This result is consistent with the discrepancy in molecular mass of the intact globin noticed previously (Fig 7). Full characterization of the α_{A1} protein is being currently performed to verify a possible variation in protein sequence. On the basis of its fragmentation spectrum, the α_{A2} globin whose intact molecular mass was 43 Daltons higher than that of α_{A1} globin was inferred to carry the mutation(s) within its first 108 amino acid residues because all the observed b-type ions had molecular masses higher than those of their α_{A1} counterparts by 43 ± 1.6 Daltons, whereas the y-type ions derived from the C-terminal part of the molecule measured exactly as did their α_{A1} counterparts (Fig 9, middle panel). The product ions derived from the α_{A3} globin could be explained as belonging to the b108-b114 and y28 categories provided that differences between the sequences of α_{A1} and α_{A3} chains were localized within the first 108 and the last 27 amino acid residues and resulted in a mass increment of 29.7 ± 2.3 Daltons and a mass decrement of 26.9 ± 0.4 Daltons, respectively. In the α_{A3} spectrum, in addition to the major fragment ions described above, we have noted the presence of the relatively low intensity ions (marked with dots in the bottom panel of Fig 9) whose molecular masses corresponded to the y28⁴ and y28³ species observed in the spectrum derived from the α_{A1} globin. Their appearance could be explained by assuming that a heterogeneity observed in the intact α_{A3} -globin fraction (Fig 2) was caused by the presence of the minor species related to, or derived from, the zebrafish α chain whose last 28 amino acid residues shared a sequence with the α_{A1} globin. Thus in conclusion, partial sequencing of the intact zebrafish α globins by CID in the source of electrospray mass spectrometer showed that the most hydrophobic of the three α globins had fragmentation characteristics consistent with the known α -globin sequence shown in Fig 1A. In addition, the differences between the α_{A1} , α_{A2} , and α_{A3} chains were mapped to specific regions of the α -globin molecule.

Peptide mapping of the β and α zebrafish proteins was further performed. HPLC-isolated globins were alkylated with iodoacetamide and proteolyzed with TPCK-trypsin, and the resulting mixtures of proteolytic peptides were analyzed by LC/ESMS. The peptides were designated as corresponding to the known α - and β -globin sequences when their measured average molecular masses matched the ones expected on the basis of the amino acid constitution. An identity of the β_{A1} globin was confirmed by finding the tryptic peptides encompassing the whole expected sequence (data

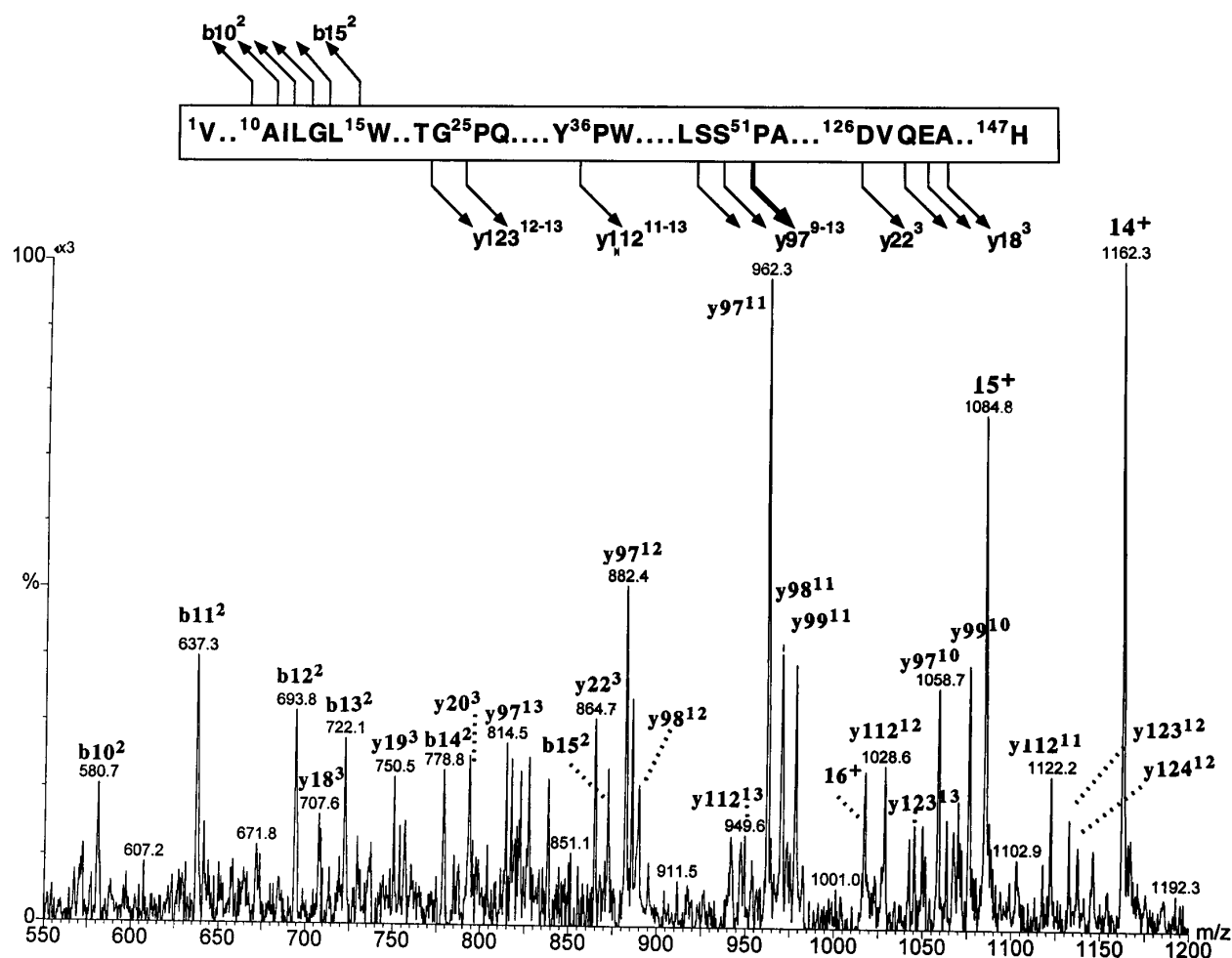


Fig 8. The figure shows the partial sequence of the β_{A1} globin and the postulated sites and products of cleavages generated by CID in the gas phase. Below, a portion of the electrospray ionization mass spectrum of the β_{A1} globin acquired under the conditions that induce partial protein fragmentation in the gas phase is shown. The spectrum shows multiply charged precursor molecular ions and product ions. The precursor ions represent intact protein and are marked with a number of positive charges that they carry (ie, 16^+ , 15^+ , 14^+). Product ions are derived from protein fragments produced in the course of CID. They are annotated with letters (b and y for ions encompassing the N- and C-termini of the molecule, respectively), followed by numbers representing positions within the sequence where the cleavages occur (numbering starts from the N- and C-terminus for the b and y ions, respectively). The number of positive charges that each product ion carries is given in superscript. Note that intensities of the ions within the m/z range of 550 to 875 are magnified 3 times.

not shown). Unequivocal identification of the protein eluting in the second HPLC peak as the β_{A2} globin was made on detection of the tryptic peptides T11 and T13 whose masses were consistent with the presence of ^{104}Lys and ^{127}Ile , respectively (data not shown). The preliminary analysis of tryptic peptides derived from three HPLC peaks carrying the α globins allowed us to confirm more than 60% of the sequence of the α_{A1} globin. It was also shown that the α_{A2} globin carried a difference within the tryptic peptide T7 (^{42}Thr — ^{58}Lys) resulting in a 14-Dalton increment in molecular mass of this fragment when compared with that of the corresponding α_{A1} peptide. Given that intact α_{A2} globin carries 43-Dalton mass increment over the α_{A1} , their respective sequences must vary in at least two different positions of the protein chain. In conclusion, the identities of both zebrafish β globins were confirmed, and their mobilities in reversed-phase HPLC were unequivocally established. It was also

shown that the most hydrophobic of all detected α globins corresponded to the already cloned zebrafish α globin. Two other α globins α_{A2} and α_{A3} were shown to differ from the α_{A1} by at least two amino acid substitutions.

Developmental expression of adult α and β globins. PCR primers were designed for zebrafish α -globin, β -globin (both β_{A1} and β_{A2}), and EF-1 α . RT-PCR analysis was performed on RNA samples derived from embryonic RBCs (24 hours, 2 days, 4 days, and 10 days), whole adult, adult spleen, and adult peripheral blood cells. As shown in Fig 10, very low levels of α - or β -globin RNA are expressed in embryonic erythrocytes. The level of these globins is often variable; occasionally, no signal is detected (as seen in the day-10 lane). The expression of the adult α or β globins is not detected in day-1 or day-2 embryos by in situ hybridization (data not shown). High levels of expression of each of these globins was observed in adult erythrocytes and the adult spleen.

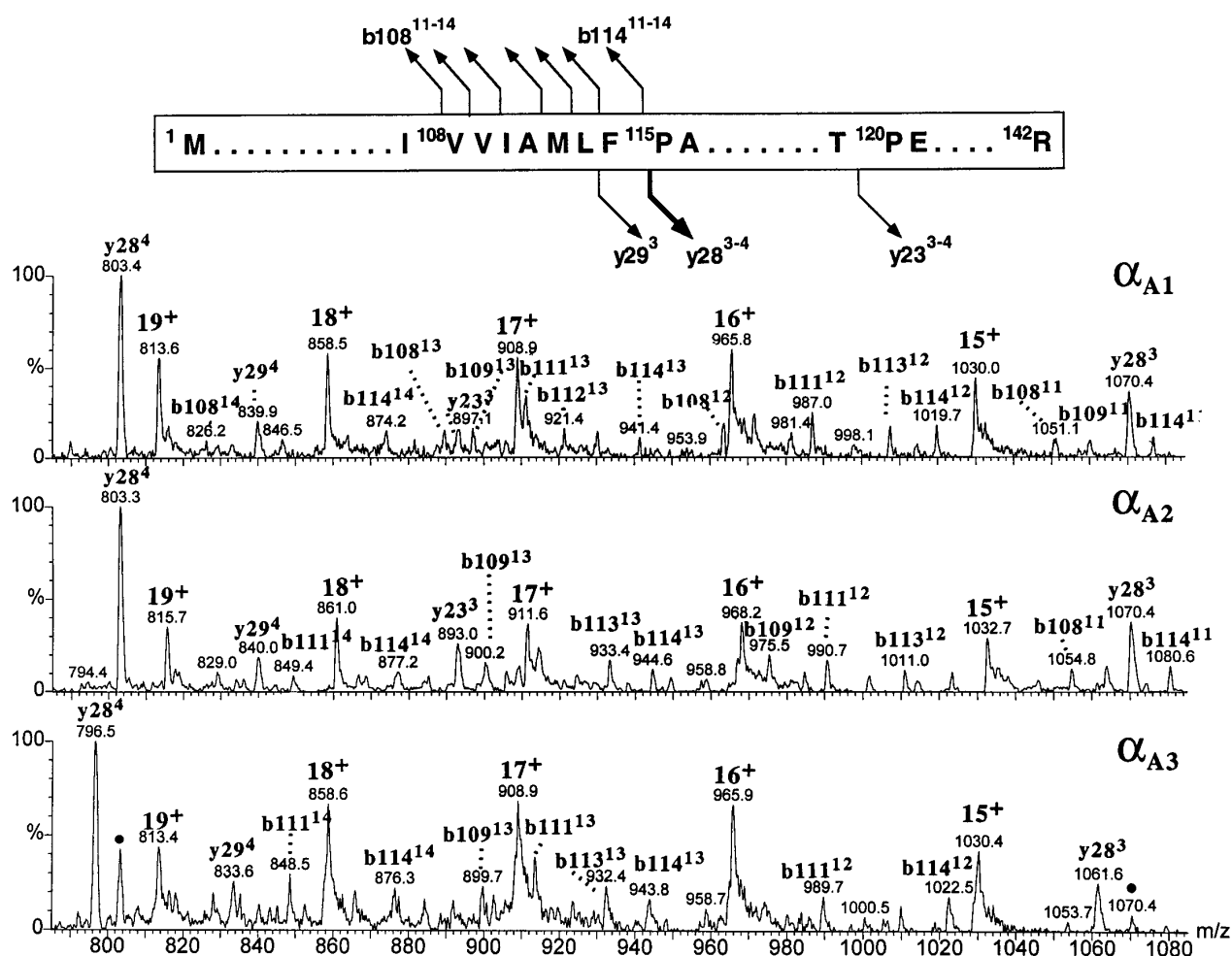


Fig 9. The figure shows the partial sequence of the α_1 globin and the postulated sites and products of cleavages generated by CID in the gas phase. Below, portions of the CID electrospray ionization mass spectra of the zebrafish adult α globins are shown. Identities of the product ions generated from the α_1 globin (upper panel) are based on correspondence between the measured and predicted m/z values for the series of y- and b-type ions observed in the spectrum. Tentative assignment of product ions observed in the spectra of α_{A2} and α_{A3} globins (middle and bottom panels, respectively) is based on the premise that all zebrafish α chains fragment in the gas phase as shown at the top of the figure. This assumption is reinforced by agreement between the molecular masses of the α_{A2} and α_{A3} globins measured for intact molecules (Fig 7) and calculated from the m/z values of the putative b- and y-fragment ions observed in the CID spectra. Ions marked with dots in the bottom panel are believed to be derived from the minor component of the α_{A3} -globin fraction and might indicate the presence of an additional α -type globin in adult zebrafish blood. Consult legend to Fig 8 for nomenclature used for peak annotation in the CID spectra.

DISCUSSION

Using a variety of molecular biological, chromatographic, and mass spectrometric techniques, we have shown the expression of three major α globins and two major β globins in the adult zebrafish. We have isolated cDNAs that encodes one of these α globins and both of the β globins. Studies are in progress to isolate cDNAs encoding the remaining globins. The sensitivity of the chromatographic and mass spectrometric techniques^{25,39} will be ideal for determining alterations in globin expression during development and in zebrafish hematopoietic mutants.

Together with those of prior studies, our results show that teleosts have distinct primitive and definitive erythrocyte lineages that express different globins.¹¹⁻¹⁴ Hb switching has been shown for the globins of the lampreys,⁴ suggesting that the switch mechanism existed in the primordial globin gene

locus before duplication that generated α - and β -globin loci. In zebrafish, primitive globins are initially expressed in the intermediate cell mass of Oellacher (ICM; A. Brownlie and L.I. Zon, unpublished data). Although the precise timing of Hb switching remains to be determined, definitive globins are thought to initiate transcription after pronephric hematopoiesis initiates. In adulthood, these globins are expressed in the circulating erythrocytes, mesonephric kidney, and spleen. More careful analysis of gene expression will determine the precise timing of the larval-adult switch in the zebrafish. Thus, in all vertebrate species studied, as the site of hematopoiesis changes during development, globin expression also switches.

The organization of α - and β -globin genes in the zebrafish is unusual compared with the loci of other vertebrates. A similar globin locus structure has recently also been described in the salmon (with a 3'-5' to 5'-3' configuration).⁴⁰

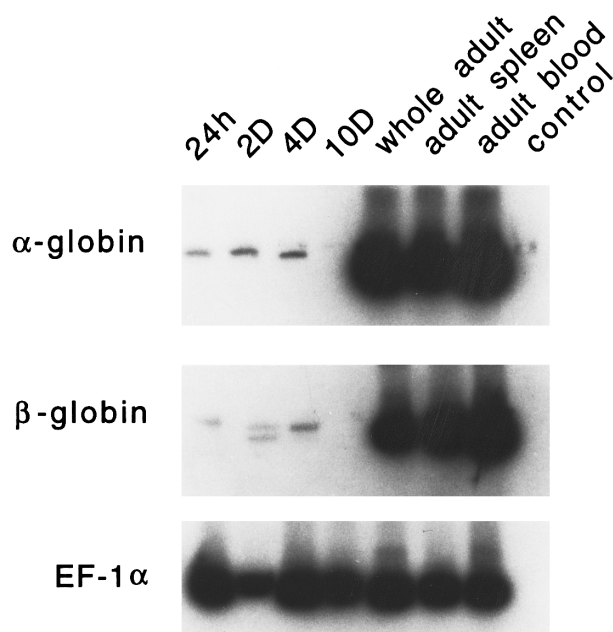


Fig 10. RT-PCR analysis of expression of the adult α - and β -globin mRNA in zebrafish embryonic blood at 24 hours, 2 days, 4 days, and 10 days; in a whole adult zebrafish; in a zebrafish adult spleen; and in zebrafish adult blood. EF-1 α is used as a control for RNA levels. A reaction with water (lane 4) is used as a negative control.

In salmon, the Bohr and non-Bohr globins are linearly arrayed in the same locus (T. McMorrow and F. Gannon, personal communication, June 1994). Because there are three major α globins and two major β globins expressed in the adult, and we have only described the α_{A1} - and the β_{A1} -globin gene here, other globin genes must exist. Southern blot analysis of the 2-1 phage clone suggests that the other two α -globin genes and one β -globin gene are present in this clone. We are in the process of isolating and characterizing these genes. The position of the embryonic globins in the locus remains to be determined. In the *Xenopus*, the embryonic globins flank the adult globins.⁴¹ Because the transcriptional units in zebrafish are in the opposite orientation, an inversion event during the evolution of teleosts may have positioned the embryonic globins together or intermixed them with adult globins such as that which occurs in chickens.⁴² Based on restriction-enzyme digestion patterns, one of the isolated zebrafish genomic DNA clones is distinct and includes another globin that is expressed at a low level in the embryo and the adult (F.Y. Chan and L.I. Zon, unpublished observations). More recently, we have isolated the major embryonic α - and β -globin cDNAs (A. Brownlie and L. Zon, unpublished data). The major embryonic and adult zebrafish globin genes map to the same zebrafish linkage group and are detected on a 550-kb *Mlu* II fragment (B. Paw, A. Brownlie, L.I. Zon, data not shown). A more detailed analysis of the structure of the globin locus can now be performed in the zebrafish.

The unusual structure of the teleost globin locus raises several points regarding gene transcription. The cyprinids (zebrafish and carp) have the smallest globin genes of any vertebrate isolated to date.⁴³ The introns are small in size,

and there is only kb of intervening sequences between the adult α_{A1} and β_{A1} globins. It is possible that other globins, embryonic and/or adult, are arrayed as tandem repeats. There is 9 kb of intervening sequences between the *Xenopus* α^A and β^A genes, which contains DNase-I-hypersensitive sites⁴⁴ that are likely to contain regulatory elements required for appropriate globin expression. We have not yet tested for DNase I hypersensitivity (or LCR elements) in the zebrafish locus. Our RT-PCR analysis shows that the α - and β -globin genes are coordinately expressed. It is possible that the two globin promoters in the 3'-5' to 5'-3' configuration would share common cis-elements or enhancer elements in the intervening sequences. Definition of the LCR elements of the zebrafish globin loci and the sequences that cis-regulate in the intervening sequence will lead to intriguing models of the genesis of Hb switching during vertebrate evolution.

ACKNOWLEDGMENT

We thank M. Kieran for providing the genomic DNA library, J. Rast and G. Littman for the spleen cDNA library, T. Thompson for the sample of adult blood RNA, and B. Paw for the adult spleen RNA. We thank B. Detrich for reading this manuscript and T. Gorr and A. Riggs (The University of Texas at Austin) for analyzing our sequences and family lineups. We thank C.H.L. Shackleton for his support in mass spectrometric studies. The cDNAs have been deposited in Genbank under accession no. U50379, U50380, and U50381. The *Pst*-I fragment of genomic DNA is accession no. U50382.

REFERENCES

1. Neinhuis AW, Maniatis T: Structure and expression of globin genes in erythroid cells, in Stamatoyannopoulos G, Nienhuis AW, Leder P, Majerus PW (eds): *The Molecular Basis of Blood Diseases*. Philadelphia, PA, Saunders, 1987, p 28
2. Dieterlen LF, Martin C: Diffuse intraembryonic hemopoiesis in normal and chimeric avian development. *Dev Biol* 88:180, 1981
3. Turpen JB, Knudson CM, Hoefen PS: The early ontogeny of hematopoietic cells studied by grafting cytogenetically labeled tissue anlagen: Localization of a prospective stem cell compartment. *Dev Biol* 85:99, 1981
4. Lanfranchi G, Pallavicini A, Laveder P, Valle G: Ancestral hemoglobin switching in lampreys. *Dev Biol* 164:402, 1994
5. Tuan D, Solomon W, Li Q, London IM: The "beta-like-globin" gene domain in human erythroid cells. *Proc Natl Acad Sci USA* 82:6384, 1985
6. Reitman M, Lee E, Westphal H, Felsenfeld G: An enhancer/locus control region is not sufficient to open chromatin. *Mol Cell Biol* 13:3990, 1993
7. Townes TM, Behringer RR: Human globin locus activation region (LAR): Role in temporal control. *Trends Genet* 6:219, 1990
8. Hug BA, Moon AM, Ley TJ: Structure and function of the murine beta-globin locus control region 5' HS-3. *Nucleic Acids Res* 20:5771, 1992
9. Li QL, Zhou B, Powers P, Enver T, Stamatoyannopoulos G: Beta-globin locus activation regions: Conservation of organization, structure, and function. *Proc Natl Acad Sci USA* 87:8207, 1990
10. Fraser P, Pruzina S, Antoniou M, Grosveld F: Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes Dev* 7:106, 1993
11. Iuchi I: The post-hatching transition of erythrocytes from larval to adult type in the rainbow trout, *Salmo gairdnerii* irideus. *J Exp Zool* 184:383, 1973
12. Iuchi I, Suzuki R, Yamagami K: Ontogenetic expression of

larval and adult hemoglobin phenotypes in the intergeneric salmonid hybrids. *J Exp Zool* 192:57, 1975

13. Iuchi I: Cellular and molecular bases of the larval-adult shift of hemoglobins in fish. *Zool Sci* 2:11, 1985

14. Iuchi I: Chemical and physiological properties of the larval and the adult hemoglobins in the rainbow trout, *Salmo Gairdnerii* Irdeus. *Comp Biochem Physiol* 44B:1087, 1973

15. Detrich HW, Kieran MW, Chan FY, Barone LM, Yee K, Rundstadler JA, Pratt S, Ransom D, Zon LI: Intra-embryonic hematopoietic cell migration during vertebrate development. *Proc Nat'l Acad Sci USA* 92:10713, 1995

16. Al-Adhami MA, Kunz YW: Ontogenesis of haematopoietic sites in brachydanio rerio. *Dev Growth Differ* 19:171, 1977

17. Colle-Vandeveld A: Sur l'origine du sang et des vaisseaux chez *Lebistes reticulatus* (Teleosteen). *J Embryol Exp Morph* 9:68, 1961

18. Oellacher J: Beitrage zur entwicklungsgeschichte der knochenfische nach beobachtungen am bachforelleneie. *Zool* 23:373, 1872

19. Westerfield M (ed): *The Zebrafish Book; A Guide for the Laboratory Use of Zebrafish (Brachydanio rerio)*. Eugene, OR, University of Oregon, 1989, p 1

20. Chomczynski P, Sacchi N: Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162:156, 1987

21. Sambrook J, Fritsch EF, Maniatis T (eds): *Molecular Cloning* (ed 2). Cold Spring Harbor, NY, Cold Spring Harbor Laboratory, 1989

22. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *J Mol Biol* 215:403, 1990

23. Kelley C, Yee K, Harland R, Zon L: Ventral expression of GATA-1 and GATA-2 in the *Xenopus* embryo defines induction of hematopoietic mesoderm. *Dev Biol* 165:193, 1994

24. Whitney JBI: Simplified typing of mouse hemoglobin (Hbb) phenotypes using cystamine. *Biochem Genet* 16:667, 1978

25. Witkowska HE, Bitsch F, Shackleton CHL: Expediting rare variant hemoglobin characterization by combined HPLC/electrospray mass spectrometry. *Hemoglobin* 17:227, 1993

26. Shelton JB, Shelton JR, Schroeder WA: High performance liquid chromatographic separation of globin chains on a large-pore C₄ column. *J Liquid Chromatogr* 7:1969, 1984

27. Shackleton CHL, Witkowska HE: Mass spectrometry in the characterization of variant hemoglobins, in Desiderio DM (ed): *Mass Spectrometry: Clinical and Biomedical Applications*, vol 2. New York, NY, Plenum, 1994, p 135

28. Ferrige AG, Seddon MJ, Green BN, Jarvis SA, Skilling J: Disentangling electrospray spectra with maximum entropy. *Rapid Commun Mass Spectrom* 6:707, 1992

29. Johnson RA, Hart NH: The ontogeny of hemoglobins in the teleost *Brachdanio rerio*. *Am Zool* 13:1348, 1973 (abstr)

30. Smith RD, Loo JA, Edmonds CG, Barinaga CJ, Udseth HR: New developments in biochemical mass spectrometry: Electrospray ionization. *Anal Chem* 62:882, 1990

31. Yoshida A, Lin M: NH₂-terminal formylmethionine- and NH₂-terminal methionine cleaving enzymes in rabbits. *J Biol Chem* 247:952, 1972

32. Wilson DB, Dintzis HM: Protein chain initiation in rabbit reticulocytes. *Proc Natl Acad Sci USA* 66:1282, 1970

33. Jackson R, Hunter T: Role of methionine in the initiation of haemoglobin synthesis. *Nature* 227:672, 1970

34. Loo JA, Edmonds CG, Smith RD: Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry. *Science* 248:201, 1990

35. Light-Wahl KJ, Loo JA, Edmonds CG, Smith RD, Witkowska HE, Shackleton CHL, Wu C-SC: Collisionally activated dissociation and tandem mass spectrometry of intact hemoglobin β -chain variant proteins with electrospray ionization. *Biol Mass Spectrom* 22:112, 1993

36. Roepstorff P, Fohlman J: Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *J Biomed Mass Spectrom* 11:601, 1984

37. Johnson RS, Martin SA, Biemann K: Collision-induced fragmentation of (M + H)⁺ ions of peptides—Side chain specific sequence ions. *Int J Mass Spectrom Ion Proc* 86:137, 1988

38. Bakhtiar R, Wu Q, Hofstadler SA, Smith RD: Charge state specific facile gas-phase cleavage of Asp 75-Met 76 peptide bond in the α -chain of human apohemoglobin probed by electrospray ionization mass spectrometry. *Biol Mass Spectrom* 23:707, 1994

39. Witkowska HE, Green BN, Morris M, Shackleton CHL: Intact protein electrospray ionization mass spectrometry can be the sole technique used for confirming the structure of a variant hemoglobin. *J Mass Spectrom Rapid Commun Mass Spectrom* S111, 1995

40. Wagner A, Deryckere F, McMorro T, Gannon F: Tail-to-tail orientation of the Atlantic salmon alpha- and beta-globin genes. *J Mol Evol* 38:28, 1994

41. Hosbach HA, Wyler T, Weber R: The *Xenopus laevis* globin gene family: Chromosomal arrangement and gene structure. *Cell* 32:45, 1983

42. Evans T, Felsenfeld G, Reitman M: Control of globin gene transcription. *Annu Rev Cell Biol* 6:95, 1990

43. Miyata M, Hirono I, Hoki T: Analysis of α -globin gene family of carp. I. α -Globin gene structure of no. 3, no. 6 and no. 7. *Nippon Suisau Gakkaishi* 59:1077, 1993

44. Stalder J, Meyerhof W, Wirthmuller U, Gruber A, Wyler T, Knochel W, Weber R: Conserved sequences and cell-specific DNase I hypersensitive sites upstream from the co-ordinately expressed alpha I- and alpha II-globin genes of *Xenopus laevis*. *J Mol Biol* 188:119, 1986