Molecular cloning of cDNA sequences coding for the major α - and β -globin polypeptides of adult *Xenopus laevis*

R.M.Kay, R.Harris, R.K.Patient and J.G.Williams

Imperial Cancer Research Fund, Mill Hill Laboratories, Burtonhole Lane, London, NW7 1AD, UK

Received 14 April 1980

ABSTRACT

This report describes the synthesis and cloning of almost complete DNA copies of the mRNAs encoding the major α -globin and major β -globin of *X.laevis*. Double-stranded globin cDNA was inserted into the PstI site of the plasmid pBR322 and two cloned recombinants (designated pXG6C1 and pXG8D2) were selected. These were shown to contain almost complete copies of *X.laevis* globin mRNA. Restriction enzyme maps were determined for each cDNA sequence using the established method of partial digestion of end labelled DNA. However, this procedure was modified such that isolation of individual DNA fragments was no longer required. Each plasmid was shown, by both hybrid arrested translation and filter selection of complementary RNA, to contain a sequence coding for one or other of the two major globin polypeptides. Sufficient DNA sequence information has been determined from each cDNA clone to demonstrate that pXG8D2 contains a β -globin sequence and pXG6C1 contains an α -globin sequence.

INTRODUCTION

Vertebrate development is accompanied by sequential changes in the haemoglobin phenotype which can be directly correlated with the regulated expression of a family of globin genes. Amphibia provide a convenient system in which to study the mechanisms involved in these changes since they occur during the metamorphic transition from larval to adult organism (1,2,3). Furthermore, globin gene structure has been analysed so far only in mammals and the chicken and in order to understand more fully the evolutionary development of this gene family, it is necessary to study more distantly related vertebrates. For these reasons we have chosen to investigate the molecular structure and expression of the globin genes of the amphibian, *Xenopus laevis*, the South African Clawed Toad.

Analysis of the composition of *X.laevis* haemoglobins has shown that the tadpole haemoglobin species are electrophoretically distinct from the adult forms (3) and that the transition between these states occurs during and after metamorphosis. Subsequent analysis of the globin polypeptide

composition of X.laevis blood in acid urea polyacrylamide gels revealed at least 5 electrophoretically distinct tadpole globins and at least 6 adult globins with two predominant species at each stage (4). Induction of anaemia in adult X.laevis was found to result in a wave of erythropoiesis (5) and mRNA isolated from the blood of anaemic animals was found to direct the synthesis of all the adult globin polypeptides (4,6). In a previous report (4) we described the cloning of cDNA sequences derived from adult X.laevis globin mRNA which allowed us to identify clones containing mRNA sequences encoding one major and one minor globin polypeptide. These recombinant plasmids contained only partial copies of the globin mRNA and we were not able to detect sequences coding for one of the two major adult globins. Similarly, Humphries et al. (7) have also reported the cloning of partial cDNA sequences derived from adult X.laevis globin mRNA but did not identify the polypeptides encoded in these sequences. Using more recently developed methods, we describe in this report the synthesis of almost complete double stranded cDNA copies of X.laevis adult globin mRNA sequences and their amplification in a plasmid vector. A detailed analysis of two cloned transformants shows that these contain almost full-length copies of the major adult α - and β -globin mRNA sequences.

MATERIALS AND METHODS

Animals

Adult female Xenopus laevis were obtained from Xenopus Ltd., Redhill, Surrey, England and were laboratory conditioned for at least 1 month before use.

Preparation of RNA

X.laevis erythroblast RNA was prepared from anaemic animals as previously described (4). PolyA+RNA was prepared by oligo dT cellulose chromatography and a 10S globin mRNA fraction was prepared from polyA+RNA by sucrose density gradient fractionation.

Synthesis of double stranded globin cDNA

A cDNA copy of X.laevis globin mRNA (10S fraction) was synthesized with AMV reverse transcriptase and a (dT)₁₄ primer using conditions previously described (8). The major cDNA product was estimated to have a size of 650±50 nts by alkaline agarose gel electrophoresis (9). We concluded that this cDNA was a complete copy of the globin mRNA since we find that X.laevis globin mRNA migrates in denaturing agarose gels with a size of 670±50 nts

(unpublished observation).

Single stranded cDNA was converted to a double stranded form with AMV reverse transcriptase using conditions similar to those described by Monahan *et al.* (10). Thus cDNA (1.5 μ g) was incubated in a total reaction volume of 800 μ l containing 8mM MgCl₂, 50mM Tris/HCl (pH 8.3), 20mM dithiothreitol and 400 μ M each of dATP, dCTP, dTTP and dGTP, and 200 units of AMV reverse transcriptase. The reaction was incubated at 46 $^{\circ}$ for 4 hr under autoclaved paraffin and then extracted first with phenol and then with chloroform:isoamyl alcohol (50:1) to remove tightly bound reverse transcriptase which protects the cDNA from S1 nuclease digestion (10). The reaction reproducibly converted 50-70% of the cDNA to double stranded (i.e., S1 resistant) DNA. The cDNA was passed over hydroxyapatite in 0.12M phosphate buffer (pH 6.8) at 69 $^{\circ}$ to remove unreacted single stranded cDNA and desalted through Sephadex G50. Analysis of the double stranded cDNA on alkaline agarose gels (9) showed that the cDNA was converted to a molecular size exactly twice that of the single stranded cDNA as would be expected for synthesis of a full-length double stranded molecule covalently linked at the 3' end of the cDNA. The double stranded cDNA (approx. 1 μ g of DNA) was stored frozen while small samples were assayed for optimum S1 nuclease digestion conditions for cleavage of the terminal hairpin.

Cleavage of the hairpin loop in the double stranded cDNA with S1 nuclease without degradation of the remainder of the molecule can be assayed by observing the conversion of double length cDNA to single length in alkaline agarose gels (11). Optimum conditions were found to be the following: double stranded cDNA (0.35 μ g) was incubated at 21 $^{\circ}$ for 30 min in a total volume of 1.4ml containing λ DNA (17 μ g), 30mM Na acetate (pH 4.5), 0.28M NaCl, 4.5mM Zn SO₄ and 500 units of S1 nuclease (Boehringer-Mannheim). After digestion the reaction was extracted first with phenol and then with phenol:chloroform:isoamyl alcohol (50:50:1). It was then passed over Sephadex G50 in 10mM Tris/HCl (pH 8.0), 1mM EDTA. The double stranded cDNA was tailed with poly dC using calf thymus terminal deoxynucleotide transferase (12). It was calculated that an average of 70 dC residues had been added to each 3' OH terminus.

Preparation of recombinant plasmids and transformation of E.coli Hb101

Plasmid pBR322 was digested with the restriction enzyme *Pst*I to yield a linear molecule, digestion being limited to the minimum level required to cause a 10²-fold reduction in transformation efficiency relative to super-

coiled pBR322 (reduction from 10^6 transformants to 10^4 transformants/ μ g). The linear plasmid was tailed with poly dG using terminal deoxynucleotide transferase. Approximately 30 dG residues were added to each 3' OH terminus. This procedure regenerates a single stranded PstI recognition sequence in the recombinant plasmids.

Plasmid DNA, tailed with poly dG was annealed with poly dC tailed double stranded cDNA at a molar ratio of 2:1 (assuming the cDNA was 600 nucleotides in length) and used in the transformation of E.coli strain Hbl01 to tetracycline resistance. In subsequent testing for ampicillin sensitivity, a 2 to 3 fold higher proportion of ampicillin sensitive transformants were found with the smallest colonies from the transformation suggesting some impairment of growth in these colonies. Overall, approximately 50% of the transformed clones were ampicillin sensitive, indicating an insertion in the PstI site and 450 of these were screened for the presence of X.laevis globin sequences.

In situ hybridisation of transformants

Tetracycline resistant, ampicillin sensitive colonies were grown overnight on nitrocellulose filters and prepared for hybridisation as described by Grunstein and Hogness (1975) (13). X.laevis erythrocyte polyA+RNA was labelled with (γ^{32} P)ATP and T_4 polynucleotide kinase to a specific activity of 10^7 cpm/ μ g and 10^6 cpm were hybridised to each filter. Colonies showing hybridisation to the RNA were detected by autoradiography. A total of 48 of the most strongly hybridising colonies were selected.

Purification of plasmids

Cleared lysates of chloramphenicol amplified bacterial cultures were prepared (14). Following phenol extraction, plasmid was isolated by CsCl/ethidium bromide centrifugation. For rapid purification of large numbers of plasmids, phenolised cleared lysates were subjected to phase separation in a dextran T500:polyethylene glycol (PEG) two phase system (15). Plasmid was precipitated from the PEG phase with CTAB (cetyl trimethyl ammonium bromide) and washed with 70% ethanol before dissolving in 10mM Tris/HCl (pH 8.0), 1mM EDTA. Rapid screening of plasmids for restriction enzyme digestion patterns was performed by concurrently digesting phenolised cleared lysates with RNase and the restriction enzyme. Products were resolved by agarose gel electrophoresis.

S1 nuclease mapping of recombinant plasmids

Recombinant plasmids were labelled with (32 P) by nick translation (16) in the absence of added DNase to a specific activity of 10^7 cpm/ μ g. A

sample of 4×10^5 cpm of plasmid was hybridised with an excess of X.laevis erythroblast RNA using conditions described by Berk and Sharp (17). Hybridisation was continued for 2 hr after which the solution was separated into two aliquots. One aliquot was heated to 100° for 2 min to denature hybrids and then both aliquots were diluted into 250 μ l each of S1 nuclease buffer (17) and incubated at 42° for 30 mins with 200 units of S1 nuclease. The digestion products were collected by ethanol precipitation and analysed by alkaline agarose gel electrophoresis.

Hybrid arrested translation (18)

Supercoiled plasmid DNA was linearized for hybridisation in one of three ways: (a) sonication, (b) restriction enzyme digestion with EcoRI or (c) alkali hydrolysis in 0.3M NaOH for 20 min at 98° . Alkali hydrolysis proved the most effective preparative method for plasmids purified by two phase separation because contaminating tRNA inhibited subsequent translation. Hybridisation and translation in a wheat germ S30 extract with (3 H) histidine were as previously described (4).

Filter hybridisation of mRNA

Plasmid DNA (10 μ g) was sonicated, bound to small diazobenzylloxymethyl paper discs (0.5cm) and X.laevis erythrocyte total cytoplasmic RNA (100 μ g) was hybridised to the filters as previously described (19). Hybridisations were performed at 50° for 4 hrs. Hybridised RNA was eluted from the filters, precipitated from ethanol twice and translated in a wheat germ S30 extract.

DNA sequencing

Restriction fragments to be sequenced were end-labelled with (32 P) by one of the following methods. 3' end-labelling of 5' overlapping ends was accomplished with AMV reverse transcriptase in the presence of one (α - 32 P)-dNTP and three unlabelled dNTPs. 5' end-labelling was performed on overlapping 5' ends by the exchange reaction of T4 polynucleotide kinase (20) in the presence of (γ - 32 P)ATP. Fragments were further digested with a second restriction enzyme and the products resolved in urea polyacrylamide preparative gels. Labelled fragments were electroeluted from gel slices and subjected to the chemical degradation method described by Maxam and Gilbert (21). Sequences were determined in 10% and 20% polyacrylamide urea gels, 0.4mm thick.

Biological Safety

All experiments involving genetic manipulation were carried out in accordance with the guidelines laid down by the British Genetic

Manipulation Advisory Committee.

RESULTS

Preparation and selection of cloned recombinants containing X.laevis globin cDNA sequences

Transformed E.coli colonies containing globin cDNA sequences inserted in the PstI site of pBR322 were prepared by standard procedures, details of which are given in Materials and Methods. A total of 450 colonies were screened by in situ filter hybridisation to (³²P) polyA+globin mRNA and 48 of the most strongly hybridising colonies were selected. Cleared lysates of each of the 48 clones were analysed on 0.8% agarose gels and a total of 19 containing the largest plasmids were chosen.

These plasmids were further purified and digested with HaeIII restriction endonuclease. All the plasmids showed digestion patterns which fell into one of two classes, examples of which are shown in Fig. 1, lanes

Figure 1. Restriction enzyme digests of cloned recombinant plasmids pXG6C1 and pXG8D2 resolved in 2% agarose. Lanes 1 and 8, marker pBR322 digested with AluI. Lanes 2,3 and 4 HaeIII digests of pBR322, pXG6C1 and pXG8D2 respectively. Lanes 5,6 and 7, PstI digests of pBR322, pXG6C1, pXG8D2 respectively.

3 and 4. There were 8 of the type shown in lane 3 and 11 of the type shown in lane 4. This observation is consistent with the cDNA sequences being derived from two globin sequences present in roughly similar amounts in the X.laevis globin mRNA.

Digestion of the plasmids with PstI restriction endonuclease cleaves out the inserted DNA sequence thus allowing an accurate estimate of the size of the insertion. Fig. 1, lanes 6 and 7, shows the digestion patterns of two plasmids pXG6C1 and pXG8D2, which contained inserts of approximately 640 nts and 670 nts respectively. Assuming that the combined length of the poly dG:dC tracts in the insert was about 50 nts, the lengths of the inserted segments in these recombinants were consistent with the presence of almost full length copies of X.laevis globin mRNA which we have observed to have a length of 670 ± 50 nts (unpublished observation).

Characterisation of selected recombinant plasmids

(a) Determination of the length of the globin cDNA sequence in the recombinants

The length of the globin cDNA sequences inserted in plasmids pXG6C1 and pXG8D2 cannot be determined from the size of the PstI digestion fragments because of an uncertainty in the length of the poly dG:dC tracts. We have previously reported (4) a procedure for avoiding this problem based on the S1 nuclease mapping method of Berk and Sharp (17). Briefly, recombinant plasmid, labelled with (32 P) by nick translation was hybridised to an excess of X.laevis globin mRNA under conditions which prevent reannealing of DNA. Hybrids were treated with S1 nuclease and the products analysed on alkaline agarose gels. Only sequences complementary to globin mRNA remain intact as a distinct band on an autoradiogram. Fig. 2 shows the results of an analysis of pXG6C1 and pXG8D2. The size of the protected fragments was compared with that of full length globin cDNA. The protected fragment in pXG6C1 was 595 nucleotides long and in pXG8D2 was 605 nucleotides long. Thus both fragments are slightly shorter than the cDNA and the globin mRNA.

(b) Restriction mapping and sequence orientation

A protocol has been developed to map restriction enzyme cleavage sites in the cDNA inserts which should have general application to mapping recombinant plasmids derived from pBR322. This is based on an established method of partial digestion of end-labelled DNA (22) but the new protocol eliminates the need to isolate individual restriction fragments.

Recombinant plasmids were cleaved with EcoRI restriction endonuclease, an enzyme which was known not to cleave in the globin cDNA. The linear

Figure 2. The length of *X.laevis* globin mRNA sequence in two cloned recombinant plasmids. Plasmid, labelled with (^{32}P) by nick translation and hybridised to *X.laevis* globin mRNA, was digested with S1 nuclease. Resistant hybrids were resolved in alkaline 2% agarose and detected by autoradiography.

Lane 1 *X.laevis* globin cDNA marker
 Lane 2 Plasmid pXG8D2
 Lane 3 Plasmid pXG8D2 denatured before S1 nuclease
 Lane 4 Plasmid pXG6C1
 Lane 5 Plasmid pXG6C1 denatured before S1 treatment

Positions for SV40 HinfI fragment markers are given with lengths in base pairs.

plasmid was end labelled with (^{32}P) using T4 polymerase or T4 polynucleotide kinase and subsequently digested with either HindIII or BamHI restriction endonucleases. The plasmid pBR322 is cleaved by HindIII 29 base pairs to the right of the EcoRI site (see Fig. 4) and BamHI cleaves 340 base pairs, also to the right of the EcoRI site. The insert in pXG6C1 is not cleaved by BamHI and thus secondary digestion with this enzyme generates two fragments, one 340 base pairs the other 4700 base pairs, both labelled at only one end. Partial digestion of this mixture of fragments with any restriction endonuclease and analysis in agarose gels

yields a pattern of labelled fragments, of which those lying between 648 and 1288 base pairs are due to cleavage within the cDNA insert. This is because the distance between the EcoRI and PstI sites in pBR322 is 648 base pairs and the length of the cDNA insert is a further 640 base pairs. The EcoRI-BamHI fragment of 340 base pairs does not interfere with the analysis and thus there is no requirement to isolate individual fragments before partial digestion mapping. Similarly, the insert in pXG8D2 is not cleaved by HindIII and the same procedure can be applied. In order to delineate the boundaries of the cDNA insert accurately, a partial PstI digest can be included and run in parallel. Although the absolute distance between cleavage sites is not particularly accurate with fragments of this length (± 30 base pairs), the relative positions of sites can be determined with great certainty. An example of the results obtained is shown in Fig. 3, and a restriction map of pXG6C1 and pXG8D2 prepared in this way is shown in Fig. 4.

The orientation of the cDNA sequences with respect to mRNA sequence was determined by hybridising (^{32}P) cDNA derived from the 3' end of X.laevis globin mRNA to restriction endonuclease digests of the plasmids resolved in agarose gels and blotted onto nitrocellulose filters (23). 3' specific cDNA was synthesized from X.laevis globin mRNA by AMV reverse transcriptase in the presence of limiting levels of one nucleotide triphosphate (24) and cDNA shorter than 300 nts was selected in sucrose gradients.

Fig. 5 shows the results of probing a HindIII digest of pXG6C1 and a BamHI digest of pXG8D2 with 3' specific cDNA and full length cDNA, the latter acting as a 3'+5' probe for comparison. BamHI restriction endonuclease cleaves once in pBR322 and once in the cDNA sequence of pXG8D2 producing two fragments (Fig. 5a, lane 2). Full length (^{32}P) cDNA hybridised equally to both fragments (Fig. 5c, lane 2). However, the 3' (^{32}P) labelled cDNA preferentially hybridised to the smaller fragment (Fig. 5b, lane 2) which is derived from the right half of the cDNA insert as drawn in Fig. 4b. HindIII restriction endonuclease cleaves once in pBR322 and 3 times in the cDNA sequence of pXG6C1 (Fig. 5a, lane 1). Hybridisation of full length cDNA to these fragments was weak but two bands could be detected (Fig. 5c, lane 1), the longer fragment (arrowed) deriving from the right hand side of the cDNA insert as drawn in Fig. 4a. The 3' cDNA hybridised preferentially to the smaller of these fragments (Fig. 5b, lane 1), the longer fragment being very weakly hybridised. Using the restriction map, the orientation

Figure 3. Restriction mapping of pXG6C1 and pXG8D2. Partial digestion products of end labelled DNA were resolved in 2% agarose and auto-radiographed. pXG6C1: lane 1 PstI, lane 2 HindIII, lane 3 HaeIII, lane 4 HinfI, lane 5 PvuII. pXG8D2: lane 1 HaeIII, lane 2 HinfI, lane 3 BamHI, lane 4 PstI.

of the cDNA sequences was deduced and is shown in Fig. 4.

(c) Identification of the globin polypeptide encoded by the cloned cDNAs

We have previously shown that X.laevis adult globin polypeptides are comprised of 6 electrophoretically resolved species, at least 3 of which have been identified as products of unique mRNA sequences (4). Two methods were used here to correlate the cloned cDNA sequences with individual globin polypeptides.

Hybrid-arrested translation as described by Patterson et al. (18), is based on the observation that when an mRNA is hybridised to a

An alternative method for identification of globin coding sequences in the recombinants is the filter hybridisation method described by Smith.

Figure 5. Orientation of globin sequences in pXG6C1 and pXG8D2. (a) Restriction endonuclease digestion fragments resolved in 2% agarose and stained with ethidium bromide lane 1, pXG6C1 cleaved with HindIII, lane 2, pXG8D2 cleaved with BamHI. (b) Autoradiogram of fragments in (a) blotted onto nitrocellulose and hybridised with short (3') *X.laevis* (³²P) globin cDNA. (c) As for (b) but hybridised with full length *X.laevis* (³²P) globin cDNA.

Searle and Williams (19). Fig. 6b shows the translation products of RNA selected by hybridisation to filter bound pXG6C1 (lane 1) and pXG8D2 (lane 2). We conclude that pXG6C1 contains sequences coding for the more rapidly migrating major globin polypeptide, and pXG8D2 contains sequences coding for the more slowly migrating major globin polypeptide.

Identification of α - and β -globin cDNA sequences in the cloned plasmids

In order to distinguish between α - and β -globins it was necessary to sequence either the globin proteins or the cloned DNA sequences coding for them. The latter approach is reported below. A number of regions in the globin molecule show a high degree of evolutionary conservation which can be used to identify α - and β -globins (25). One of the most conserved regions is the last few amino acids at the C-terminus. All α -globins studied so far terminate with the sequence Leu X Ser Lys Tyr Arg. This is

Figure 6. (a) Hybrid arrested translation of the two major adult globin polypeptides of *X.laevis* using cloned recombinant plasmids. Lanes 1 and 2 pBR322; lanes 3 and 4 pXG8D2; lanes 5 and 6 pXG1G4; lanes 7 and 8 pXG3B3; lanes 9 and 10 pXG6C1; lanes 11 and 12 pXG3H5. (b) Filter selection and translation of *X.laevis* globin mRNA using pXG6C1 (lane 1) and pXG8D2 (lane 2). Lane 3, total globin mRNA translation products. Translation products (labelled with (^3H) histidine) were resolved in acid-urea polyacrylamide and fluorographed.

distinct from the conserved β -globin terminal sequence,
Ala Leu Ala XX Tyr His.

Examination of the restriction maps of pXG6C1 (Fig. 4a) suggested that determination of the DNA sequence from the 3' HindIII site towards the 5' end of the cDNA would yield the position of the termination codon and C-terminal coding sequence (see arrow in Fig. 4a). Similarly for pXG8D2, the sequence 5' from the HinfI site (arrow Fig. 4b) should also yield the C-terminal coding sequence.

The DNA sequence obtained is shown in Fig. 7a. The positions of the termination codons was unambiguously obtained by reading the amino acid sequence in phase upstream from each possible termination site and comparing with the known globin protein sequences. Fig. 7b shows the homology of the sequence in pXG6C1 with other related and distant α -globins and pXG8D2 with other β -globins.

Comparison of the amino acid sequence from pXG6C1 with newt and human

(a)

pXG6C1

PHE LEU ALA THR VAL SER THR VAL LEU THR SER LYS TYR ARG
TTC TTG GCT ACC GTA TCT ACT GTT CTG ACA TCC AAA TAT CGT TAA GGCTCAG
AAG AAC CGA TGC CAT AGA TGA CAA GAC TGT AGG TTT ATA GCA ATT CCGAGTC

Hind III

CAACAACAGCAGCAGAAGTCTCAACATCAGACATCAGTTAATTATATGCAATCAAACGTACAAGCTT
GTTGTTGTCGTCGCTTCAGAGTTGTAGTCTGTAGTCAATTAATATACGTTAGTTTGACTGTTTCGAA

pXG8D2

LEU PHE ASP ALA ILE ALA ASP ALA LEU GLY LYS GLY TYR HIS
CTT TTT GAT GCC ATT GCC GAC GCC CTT GGC AAG GGT TAC CAC TAA ACCAGCC
GAA AAA CTA CGG TAA CGG CTG CGG GAA CCG TTC CCA ATG GTG ATT TGGTCGG

Hinf I

TCAAGAACACCCGAATGGAGTC
AGTTCTTGTGGGCTTACCTCAG

(b)

α globin amino acid sequences

<u>X.laevis</u> C terminus	PHE LEU ALA THR VAL SER THR VAL LEU THR SER LYS TYR ARG
<u>Newt</u>	" PHE LEU ASP VAL VAL GLY HIS VAL LEU THR SER LYS TYR ARG
<u>Human</u>	" PHE LEU ALA SER VAL SER THR VAL LEU THR SER LYS TYR ARG
<u>All other α</u>	" PHE LEU SER LYS TYR ARG

β globin amino acid sequences

<u>X.laevis</u> C terminus	LEU PHE ASP ALA ILE ALA ASP ALA LEU GLY LYS GLY TYR HIS
<u>Rana esculenta</u>	" HIS PHE CYS ALA VAL GLY ASP ALA LEU ALA LYS ALA TYR HIS
<u>Human</u>	" VAL VAL ALA GLY VAL ALA ASN ALA LEU ALA HIS LYS TYR HIS
<u>All other β</u>	" VAL ALA LEU ALA TYR HIS

Figure 7. (a) DNA sequence determined from the 3' regions of the cloned globin cDNA. DNA sequence was determined from the 3' HindIII restriction site in pXG6C1 and from the single HinfI site in the pXG8D2 insert in the direction shown by the arrows in Fig. 4. The derived C-terminal amino acid sequence is given in italics above the DNA sequence. (b) Comparison of C-terminal amino acid sequences of other α- and β-globins with the X.laevis sequences obtained in (a).

α-globins (26) showed considerable conservation. Of the 14 amino acids determined 11 were the same as newt and 13 the same as human. When comparing the conserved amino acids in all α-globins examined so far

(which are predominantly mammalian) we found that the six invariant amino acids are all present in the X.laevis sequence. Known β -globin sequences are much less conserved in the C-terminal region but the two invariant amino acids at the C terminus are present in the sequence derived from pXG8D2. As might be expected, the closest β -globin homology was observed with another anuran, the frog Rana esculenta (26) which has 8 out of 14 amino acids in common with X.laevis. Of the 6 amino acids found to be invariant in all other β -globins in this region, two were not conserved in the X.laevis sequence but, since the other β -globin sequences are almost entirely derived from mammals, this degree of divergence in X.laevis is not surprising.

DISCUSSION

Our approach to obtaining a more detailed knowledge of X.laevis globins has been to clone cDNA sequences derived from adult globin mRNA. Using a variety of techniques this report presents a detailed characterization of two recombinant plasmids, one shown to contain a copy of the most abundant α -globin mRNA sequence, the other a copy of the most abundant β -globin mRNA sequence. Based on the size of the S1 resistant globin DNA sequence in the two selected plasmids, pXG6C1 and pXG8D2, we estimate that no more than 30 nucleotides of cDNA sequence is missing. The synthesis and cloning of almost complete globin mRNA sequences from an amphibian source has not previously been reported.

Restriction mapping of the cDNA sequences was accomplished by a modification of the end-labelling, partial restriction enzyme cleavage approach which should be widely applicable to the rapid mapping of DNA sequences inserted in pBR322 and related plasmid vectors. The mapping showed that the globin sequences in pXG6C1 and pXG8D2 possessed no common restriction enzyme recognition sites. Furthermore, two independent methods were used to show that the globin sequence in pXG6C1 coded for the major globin polypeptide which migrated more rapidly in acid-urea polyacrylamide gels while pXG8D2 coded for the more slowly migrating major globin polypeptide. Each plasmid specifically inhibited the synthesis of only one polypeptide in a hybrid-arrested translation assay and selectively hybridised to only one mRNA in a filter binding assay. Our previous report describing cloned partial cDNA copies of X.laevis globin mRNA (4) showed that the more slowly migrating major species and one minor globin species were unique gene products because their synthesis could be

specifically inhibited by different cloned cDNA sequences. The isolation of pXG6C1 confirms that the polypeptide coded by the sequence in this plasmid is also unique and thus the remaining minor globin polypeptides coded for by X.laevis adult globin mRNA must arise from at least one other unique gene. We conclude that at least 4 different genes code for the globins of adult X.laevis blood. We did not identify any globin sequences in the clones selected in these experiments coding for minor globin polypeptides. Humphries et al. (7) have previously reported the cloning of partial cDNA sequences from adult X.laevis globin mRNA. However, insufficient data on the polypeptides encoded by these cDNA sequences or on the positions of restriction enzyme cleavage sites was presented to allow a correlation with the cloned sequences described here.

The preliminary DNA sequence data from each globin cDNA covering the termination codons has allowed an unambiguous designation of α - and β -globin sequence to pXG6C1 and pXG8D2 respectively. In the case of pXG6C1, the C-terminal amino acid sequence bears close homology with other α -globin sequences. For pXG8D2, the homology with other β -globins is also good, bearing 8 out of 14 amino acids in common with the frog Rana esculentā, including the functionally important amino acids Tyr His at the terminus of all β -globins (27).

Further DNA sequencing is in progress which will provide the primary nucleotide sequence of both α - and β -globin mRNA and also the α - and β -globin amino acid sequences, which are unknown. These full length globin cDNA sequences will also allow an investigation of the genomic organisation of the globin genes of X.laevis and such a study is in progress.

ACKNOWLEDGEMENTS

The authors wish to thank Dr. T. Beard for providing AMV reverse transcriptase, J. Jenkins for providing PstI and Dr. B. Griffin for providing other restriction enzymes.

REFERENCES

1. Maclean, N. and Jurd, R.D. (1972) *Biol. Rev.* 47, 393-437
2. Sullivan, B. (1974) *Chemical Zoology* (M. Florkin and B.T. Sheer, eds) 9, 77-122 Academic Press, N.Y.
3. Just, J.J., Schwager, J. and Weber, R. (1977) *Roux Arch. Develop. Biol.* 183, 307-323.
4. Hentschel, C.C., Kay, R.M. and Williams, J.G. (1979) *Dev. Biol.* 72, 350-363.
5. Thomas, N. and Maclean, N. (1975) *J. Cell Sci.* 19, 509-520.

6. Battaglia, P. and Melli, M. (1977) *Develop. Biol.* 60, 337-350.
7. Humphries, P., Old, R., Coggins, L.W., MacShane, T., Watson, C. and Paul, J. (1978) *Nuc. Acids Res.* 5, 905-924.
8. Williams, J.G. and Lloyd, M. (1979) *J. Mol. Biol.* 129, 19-35.
9. McDonnell, M., Simon, M. and Studier, F. (1977) *J. Mol. Biol.* 110, 119-146.
10. Monahan, J.J., McReynolds, L.A. and O'Malley, B.W. (1976) *J. Biol. Chem.* 251, 7355-7362.
11. McReynolds, L.A., Catterall, J.F. and O'Malley, B.W. (1977) *Gene* 2, 217-231.
12. Roychoudhury, R., Jay, E. and Wu, R. (1976) *Nuc. Acids Res.* 3, 101-116.
13. Grunstein, M. and Hogness, D.S. (1975) *Proc. Natl. Acad. Sci. USA* 72, 3961-3965.
14. Tewell, D.B. and Helinski, D.R. (1970) *Biochemistry* 9, 4428-4440.
15. Ohlsson, R., Hentschel, C.C. and Williams, J.G. (1978) *Nuc. Acids Res.* 5, 583-590.
16. Rigby, P.W.S., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* 113, 230-251.
17. Berk, A. and Sharp, P. (1977) *Cell* 12, 721-732.
18. Paterson, B.M., Roberts, B.E. and Knuff, C. (1977) *Proc. Natl. Acad. Sci. USA* 73, 3414-3422.
19. Smith, D.F., Searle, P.F. and Williams, J.G. (1979) *Nuc. Acids Res.* 6, 487-506.
20. Berkner, K.L. and Folk, W.R. (1977) *J. Biol. Chem.* 252, 3176-3184.
21. Maxam, A.M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* 74, 560-564.
22. Smith, H.O. and Birnstein, M.L. (1976) *Nuc. Acids Res.* 3, 2387-2398.
23. Southern, E.M. (1975) *J. Mol. Biol.* 98, 503-517.
24. Efstratiadis, A., Maniatis, T., Kafatos, F.C., Jeffrey, A. and Vournakis, J.N. (1975) *Cell* 4, 367-378.
25. Dayhoff, M.O. (1972) *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, Washington, D.C.
26. Fasman, G.D. (ed) *Handbook of Biochemistry and Molecular Biology* (1976) 3rd Edition. Proteins Vol. III 441-455.
27. Perutz, M.F. (1970) *Nature* 228, 726-739.