Volume 17 Number 20 1989 Nucleic Acids Research

Transcriptional and translational analysis of the human θ globin gene

Shui-on Leung, Emma Whitelaw and Nick J.Proudfoot

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

Received July 17, 1989; Accepted September 21, 1989

ABSTRACT

The human θ-globin gene in man appears to be functional, based on its sequence and evolutionary conservation. However its physiological role is unknown and furthermore its deletion in some individuals appears to have no effect on erythroid development. We have therefore analysed the transcriptional and translational competence of the θ globin gene to assess whether or not it is a silent or active globin gene. First, we demonstrate that θ globin mRNA is correctly spliced, by sequencing its cDNA. Second, using this θ cDNA, we generated synthetic θ globin mRNA and were able to demonstrate that this mRNA is translated into θ globin protein in wheat germ *in vitro* translation extracts. Similarly, the θ globin gene transfected into an erythroid cell line produces a protein product that comigrates with θ globin. Finally, we analysed the unusual promoter of the θ globin gene. The GC rich sequence directly adjacent to the multiple cap sites of θ globin mRNA functions as a promoter element in both erythroid and non-erythroid cell lines, while the more usual CCAAT and ATA box regions (found in all other globin genes) which are displaced by the GC rich promoter sequence, do not possess detectible promoter activity. Taken together, these results suggest that θ globin may have some as yet undetermined role in human erythropoiesis.

INTRODUCTION

A novel α-like globin gene, called θ, has recently been discovered at the 3' end of the human α globin gene cluster [1,2]. As shown in figure 1, this cluster contains a 5' terminal embryonic ζ globin gene and two adult α globin genes with three pseudogenes, ψζ, ψα1 and ψα2, in between. θ globin is then positioned at the extreme 3' end of this gene cluster, about 4 kb beyond the α1 globin gene. Ever since θ globin was discovered, first in orang-utan [1] and then in human [2], there has been uncertainty as to whether this gene plays a functional role in erythropoiesis or whether it is yet another α-like pseudogene.

Strong arguments have been put forward in favour of a functional role for θ globin. First the high degree of sequence conservation between θ globin gene

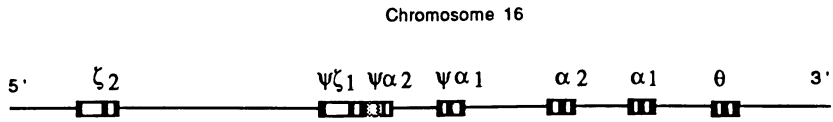


Figure 1
The diagram depicts the arrangement of different globin genes in the human α-globin gene cluster. The human θ-globin gene is located at the 3' end of the gene cluster.

sequences in primates argues that this gene has been positively selected during evolution [2,3]. Second, the θ gene contains three exons which, as with other α globin genes, encode a globin like polypeptide chain of 141 amino acids. No frame shifts or termination codons are present in these exons. However other aspects of the human θ globin gene argue against a functional role for this gene in erythropoiesis. Firstly, homozygous gene deletions of both the α and θ globin genes have been described in Man [4]. Surprisingly human foetuses with such globin gene deletions are viable until at least week 36 of gestation. This argues that the θ globin gene does not play a key role during development *in utero*. Furthermore it is unlikely that θ globin could form a normal haemoglobin protein structure since key amino acid residues required for haem binding and tetramer formation are absent in the θ globin polypeptide sequence [5]. Of course θ globin could play a quite different role in erythropoiesis that does not involve the normal process of oxygen transport.

We have previously demonstrated that the human θ globin gene is transcriptionally active, producing small amounts of mRNA in embryonic and foetal erythroid tissue [6]. Furthermore the human cell line K562, which has early erythroid characteristics [7], expresses significant levels of θ globin mRNA. Using K562 as a θ mRNA source, we were able to show that the 5' end of this mRNA is heterogeneous, initiating at multiple start sites over a 50 nucleotide region in a highly GC rich sequence. Such a promoter is very reminiscent of so-called housekeeping genes which, in contrast to θ globin, are generally expressed in all tissues. A further unusual feature of this gene's promoter is that upstream of the GC rich promoter region are standard globin gene promoter sequence elements, a CCAAT and TATA box. Indeed Hsu et al [8] have suggested that this region of the θ globin gene may act as a promoter for the θ globin gene, in contrast to our own data, which maps θ mRNA start sites to the GC region.

The experiments described in this paper further characterise the transcriptional

and translational competence of the θ globin gene. We first describe the cloning and sequence analysis of a θ globin cDNA. This reveals that the 3 exons of θ are spliced together as predicted to yield an uninterrupted reading frame. We then show that synthetic θ globin mRNA made by *in vitro* transcription of this θ cDNA is translated into a θ globin polypeptide. However the GC rich sequence at the 5' end of the mRNA markedly inhibits initiation of translation. We go on to show that a hybrid α/θ globin gene (with the α globin promoter attached to the θ globin coding sequence), when transfected into K562 cells, produces significant amounts of a protein product that comigrates with θ globin synthesised *in vitro*. The θ globin gene is therefore capable of producing a stable protein product in erythroid cells. Finally we have characterised the promoter of the θ globin gene by fusing it to the reporter gene Cat. We find that the GC rich promoter region is functional in both erythroid and non erythroid cell lines while the CCAAT/TATA promoter region is non functional. Taken together, these experiments demonstrate that the θ globin gene is both transcriptionally and translationally competent although at greatly reduced levels compared to the α globin gene.

MATERIALS AND METHODS

Construction of plasmids

α/θpSVed:

A hybrid gene containing the promoter of human α1-globin gene linked to the ATG initiation codon of the θ-globin gene so that it is actively expressed in HeLa cells was constructed as follows:

The two G residues 5' of the ATG codon of θ-globin gene in the sequence 5'-GGGATGGGC-3' were mutated using synthetic oligonucleotides into C residues to create an NcoI recognition site 5'-GCCATGGGC-3' which is found at an identical location in the human α1-globin gene. The 1.7 kb fragment containing the whole 1.2 kb 5' flanking sequence of θ-globin gene was removed by cutting the plasmid θpSVed [6] with EcoR1 and NcoI, and was replaced by the corresponding 1.2 kb EcoR1/NcoI fragment of α1pSVed [9] containing 500 bp of the α1-globin gene 5' flanking sequence including the α promoter. The resulting plasmid, α/θpSVed, contains the θ-globin genomic sequence linked at the ATG codon to 500 bp of α1-globin gene 5' flanking region containing the α1 promoter. A neomycin resistant gene was

included in the plasmid α/θpSVed [10]. This was used to generate stable cell lines expressing θ-globin.

SP6 constructs

The cDNA corresponding to the mRNA produced from the plasmid α/θpSVed contains the coding sequences of the θ-globin gene linked to 23 bp of α1 5' non coding sequence. θ cDNA containing various lengths of θ 5' flanking region were reconstituted as follows:-

A 57 bp HindIII/BssHII fragment spanning 23 bp and 34 bp of α1 5' flanking and θ coding sequence respectively was replaced with a 138 bp SmaI/BssHII fragment of the θ-globin gene containing 104 and 34 bp of θ 5' flanking and coding sequences, resulting in a θ cDNA containing 104 bp of θ 5' flanking sequence.

Bal31 digestion from the 5' SmaI site of the above cDNA created θ cDNA containing 84, 40 and 24 bp of θ 5' flanking sequences. The extent of deletion from the SmaI site was confirmed by DNA sequencing [11]. The respective cDNA thus produced was cloned between the BamHI and EcoR1 restriction sites of the vector pSP64 [12], resulting in the creation of α/θpSP64, θ_{104}pSP64, θ_{84}pSP64, θ_{40}pSP64 and θ_{24}pSP64 containing θ coding DNA linked to 23 bp of α1 5' sequence, as well as 104, 84, 40 and 24 bp of θ 5' flanking sequence respectively.

CAT Constructs

The ATG translational initiation codon of the θ-globin gene was removed together with 9 bases upstream by Bal31 digestion from the AccI site located 72 bp downstream of the ATG codon. The exact location of deletion was confirmed by DNA sequencing. After cutting with the restriction enzyme, XmnI, an 800 bp fragment spanning the sequence from the XmnI site to the Bal31 deletion site (9 bases upstream of the θ AUG) and containing the CCAAT/ATA sequence as well as the GC rich sequence of the θ-globin gene was cloned in the SmaI site 5' of the CAT gene in the plasmid pCAT0, forming the plasmid θ_{CAT0}. pCAT0 is a plasmid containing the chloramphenicol acetyltransferase (CAT) gene devoided of a promoter cloned into the BamHI site of pUC119 such that the 5' end of the CAT gene was adjacent to the SmaI site of the pUC119 polylinker [13,14].

A 600 bp fragment was discarded by cutting θ_{CAT0} with BamHI and KpnI, leaving the 200 bp GC rich sequence linked to the linearized plasmid, which upon recircularization, reconstituted the plasmid GC_{CAT0}.

To remove the GC rich sequence but retaining the upstream region including the CCAAT and ATA boxes of the θ-globin 5' flanking sequence, the plasmid θ_{CAT0} was cut with BamHI and HindIII such that a 2 kb fragment containing the CAT gene together with the 200 bp GC rich sequence of the θ-globin gene was replaced by a 1.8 kb SmaI/HindIII fragment of the plasmid pCAT0 containing the whole of CAT gene, thus resulting in the formation of plasmid ATA_{CAT0} with an upstream 800 bp θ 5' flanking sequence containing the CAAT/ATA box.

pIRV

The cotransfection control plasmid pIRV which expresses β-galactosidase contains the lacZ gene under the control of a mouse β-actin promoter.

All plasmids used for transfection were purified on CsCl gradients. Their quality as well as concentration were determined by agarose gel electrophoresis and spectrophotometry.

Cell Culture and Transfection

HeLa cells and K562 cells [15] were grown in DMEM medium supplemented with 10% foetal calf serum, 2 mM glutamine, 100 μg/ml penicillin and 100 U/ml streptomycin. Putko cells [16] were grown in RPMI medium containing the same supplements. K562 and Putko cells were split 1:5 every 2 days while HeLa cells were split 1:3 every 2 days.

HeLa cells were transfected with the standard calcium phosphate coprecipitation method.

K562 and Putko cells were transfected by electroporation method. 24 hours after cells were split 1:5, cells ($5x10^6$-10^7) were pelleted and resuspended in 1 ml of electroporation buffer (pH 7.1) containing 140 mM NaCl, 25 mM Hepes, 0.75 mM Na_2HPO_4 supplemented with 0.5% (w/v) PEG 6000. 50 μg of test plasmid, 5 μg of control plasmid and the resuspended cells were given a single shock of 1750 V at 25 μF using a BioRad Gene Pulser. The cells were either harvested after they were grown in half of their original volume for a further 40 hours or maintained in medium supplemented with 0.8 mg/ml of G418 in the establishment of stable transfected cell lines.

Enzyme Assay

Cellular lysates were prepared by sonication. The β-galactosidase activity of lysates was determined as described by Herbomel et al [17]. Cellular lysates corresponding

to equal amounts of β-galactosidase activity were used for the comparison of chloramphenicol acetyltransferase (CAT) activity. Lysates were heated to 65°C for 5 minutes before standard CAT assay was carried out as described by Gorman et al [13].

In vitro RNA Synthesis

SP6 constructs were linearized by digestion with EcoR1 which cut at the immediate downstream position of the inserted DNA. Linearized DNA templates (100 μg/ml) were transcribed in 40 mM Tris (pH 7.5), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM dithiothreitol, 100 μg/ml BSA and 500 μM each of rATP, rUTP and rCTP. Capping was achieved by the addition of dinucleotide cap analogue $m^7G(5')ppp(5')G$ in excess (500 μM) over rGTP (50 μM) in the *in vitro* transcription reaction. 50 U of HPRI (human placental RNase inhibitor) was present to inhibit the action of possible traces of RNase. Reactions were carried out in a final volume of 50 μl after the addition of 10 units of SP6 RNA polymerase (Amersham) at 37°C for 3-5 hours.

RNA recovered by ethanol precipitation was resuspended in 10 μl of TE. The amount and the integrity of RNA synthesised was assessed by analysing 1 μl of each RNA preparation in a 1% agarose gel.

In vitro Translation

In vitro translations with wheat germ extracts were performed according to the manufacturer's specifications (Amersham). Briefly, for a final reaction volume of 30 μl, 2 μl of the RNA preparations were mixed with a cocktail containing 1 μl of 1 M potassium acetate solution, 2 μl of 1 M amino acid mixture minus L-cysteine, 2.5 μl of 8.75 mCi/ml (^{35}S)-cysteine (ICN) or in some cases ^{14}C labelled amino acid mix (Amersham) and 15 μl of wheat germ extract (Amersham). The reaction mixture was incubated at 25°C for 1 hour and 5 μl of which was analysed either with a 15% SDS-PAGE or Triton-X-100 urea gel [18].

RESULTS

Isolation and characterisation of a θ globin cDNA clone

In order to study the expression of the θ globin gene, we attempted to isolate θ cDNA plasmids from erythroid cells (K562) which we knew expressed low levels of θ globin mRNA. However due to the close homology of the predominant α and ζ globin mRNAs we found it difficult to identify θ positive cDNA plasmids and

invariably isolated α or ζ globin cDNA plasmids. To circumvent this problem, we constructed a cDNA library from HeLa cells, which had been transfected with a hybrid α/θ globin gene. The α globin gene's promoter was used in place of θ since the α but not θ globin gene promoter is highly active in HeLa cells when replicated within an episomal plasmid [6].

The strategy employed to isolate the θ cDNA is outlined in Figure 2A. First the 5' end of the θ gene up to the initiation codon was replaced with an equivalent fragment from the α globin gene. This α/θ gene construct (in the transient expression vector pSVed [19]), was then transfected into HeLa cells and cytoplasmic RNA from these transfected cells was shown to contain significant levels of α/θ mRNA by S1 nuclease analysis (data not shown). A cDNA library was constructed from this RNA source and several θ gene positive colonies were identified. One of these was shown to be a full length α/θ cDNA plasmid containing 23 bp of α 5' non coding region, the whole θ globin coding sequence as well as a 3' terminal stretch of poly(A).

Although the θ globin gene appears to possess normal RNA splicing signals [8], it remained a possibility that its primary transcript may be spliced aberrantly and fail to form a functional, in-frame mRNA. We therefore sequenced the exon boundary regions of our α/θ cDNA plasmid to assess the specificity of the θ globin gene's splicing signals. As shown in figure 2B, the sequence of the α/θ cDNA reveals that the θ globin gene does indeed possess functional splice signals that result in the synthesis of an mRNA which encodes the θ globin polypeptide of 141 amino acids.

Translational analysis of θ globin mRNA

The α/θ cDNA plasmid described in the previous section was used to create synthetic θ globin mRNA. First the α/θ cDNA sequence was recloned into the SP6 promoter containing plasmid pSP64 [12]. The α 5' non coding region was then replaced with four different θ 5' non coding region fragments varying in size from 104 to 24 nucleotides (see Figure 3D). Since the 5' end of θ globin mRNA is heterogeneous, with Cap sites positioned from 50 to 5 nucleotides away from the initiation codon [6], we wanted to compare the relative translational efficiency of these different θ globin mRNAs. An α cDNA was also inserted into the same position in the pSP64 plasmid to provide an α globin mRNA control. Each pSP64 plasmid construct was truncated just beyond the poly(A) sequence in the cDNA insert and capped synthetic mRNA was synthesised *in vitro*. These were then added to a wheat germ *in vitro*

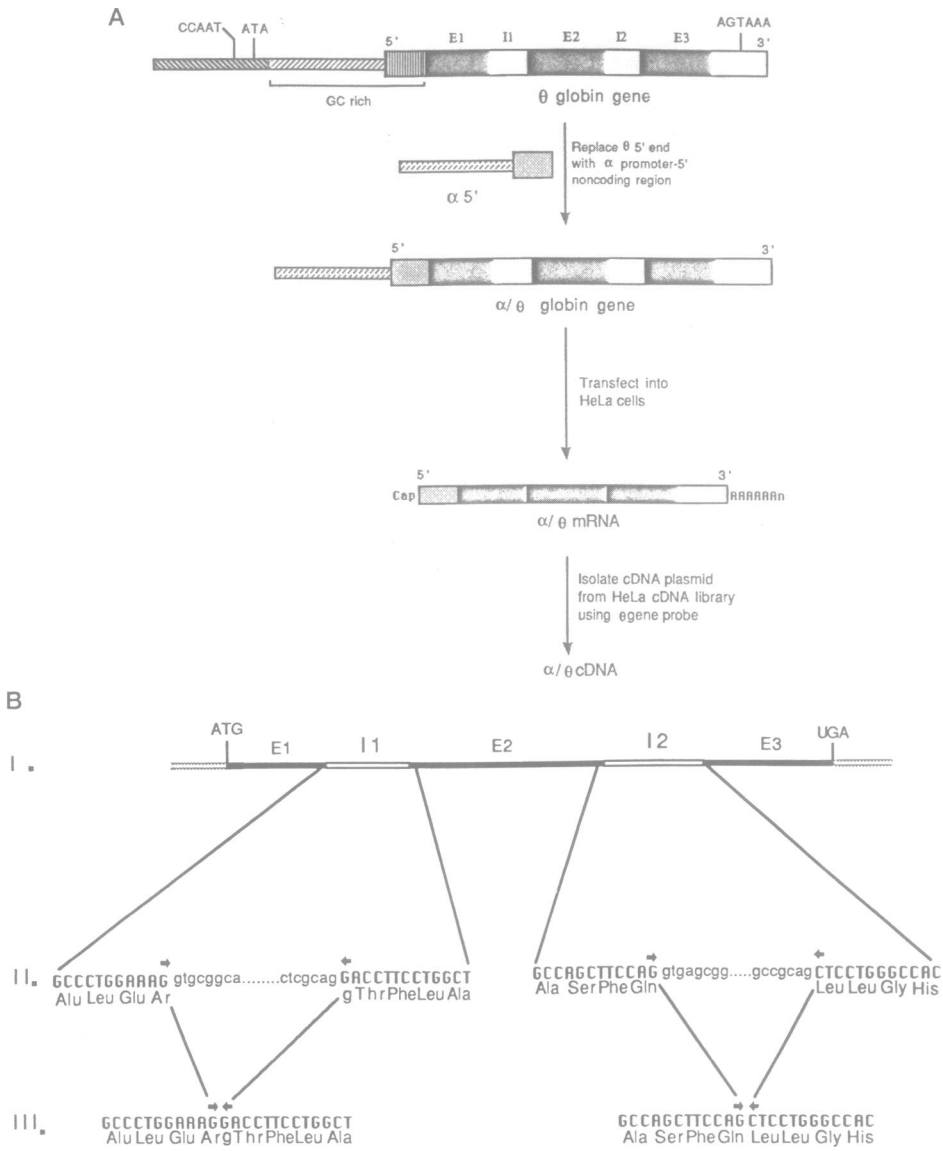


Figure 2

A. Schematic representation of the strategy to make a cDNA containing θ-globin coding and α-globin 5' non-coding sequences. Solid boxes represent the three exons of θ-globin gene, noted as E1, E2 and E3 respectively. The introns, noted as I1 and I2, as well as the 3' flanking sequences of the θ-globin gene, are shown as open boxes. The θ-globin 5' flanking sequences containing the non coding region ▥, the GC rich region ▨, and the upstream sequence containing the CCAAT/ATA consensus

translation extract and the translation products were labelled either with ^{14}C amino acids or ^{35}S cysteine and fractionated on polyacrylamide gels.

As predicted, the synthetic θ and α globin mRNAs do indeed produce identical sized polypeptides of 14 kd (Figure 3A). We then carried out a quantitative comparison of the amount of protein produced from the different θ cDNA constructs. Equal amounts of α/θ or θ synthetic mRNAs were synthesised *in vitro* as shown in figure 3B and then translated in wheat germ extracts as shown in figure 3C. As indicated the α/θ mRNA had the highest translational efficiency but this was nearly matched by the shortest θ mRNA. Clearly the longer θ mRNAs are less efficient, presumably due to the inhibitory nature of more extensive GC rich 5' non coding sequences. These experiments clearly demonstrate that θ mRNA can be translated and that the GC rich sequence present in the 5' non coding region of the longer θ mRNAs inhibits translation.

It is possible that although the θ globin gene can produce both mRNA and protein, the protein is highly unstable in erythroid cells and consequently is non functional. To test this possibility, we transfected the hybrid α/θ globin gene described above (which has an intact θ globin coding sequence) into K562 cells. This human erythroid cell line expresses all the functional globin genes except β and, as we have previously demonstrated, it also expresses low levels of θ globin mRNA [6]. The α/θ globin gene was placed in the plasmid vector pSVedNeo [10]. This vector expresses the neomycin resistance gene using the SV40 early promoter and polyadenylation

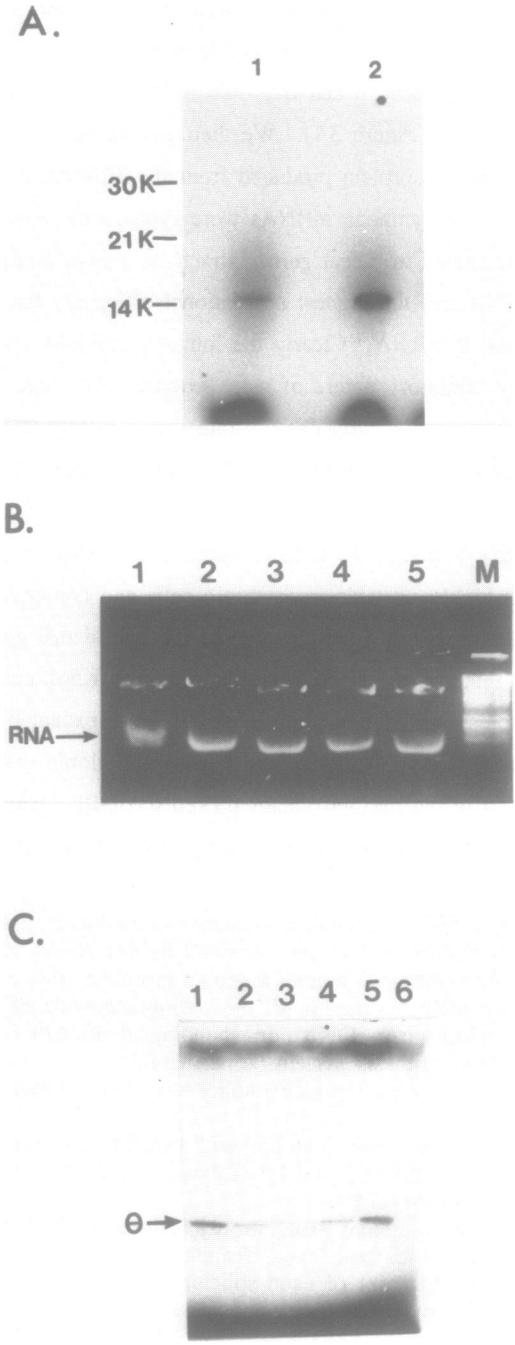
▨ , was replaced by a 500 bp promoter sequence immediately upstream of the ATG codon of the human α-globin gene to form the α/θ hybrid globin gene. The hybrid gene which is under the control of human α-globin promoter was expressed in HeLa cells, giving rise to α/θ mRNA carrying all the coding sequence of θ-globin with an α-globin 5' noncoding region. α/θ cDNA corresponding to the α/θ mRNA was then synthesised with a cDNA synthesis system (Amersham).

B. Sequence of θ cDNA across putative splice sites demonstrating that transcripts are correctly spliced.

I. Line diagram of the genomic θ-globin gene showing the translational initiation codon ATG, the three exons E1, E2 and E3 (solid lines), the two introns I1 and I2 as well as the termination codon TGA.

II. Sequences across the putative exon/intron junctions of the θ-globin gene. Only limited sequences are shown for both exons (capital letters) and introns (small letters). Arrows indicate the expected point of exon splicing. The corresponding amino acids coded for are indicated below the coding sequences.

III. cDNA sequence across the junctions indicates that the mRNA is correctly spliced. The point where arrows meet represents the junction of two exons.


A.
1
2
30K
21K
14K
B.
1
2
3
4
5
M
RNA
C.
1
2
3
4
5
6
θ

D

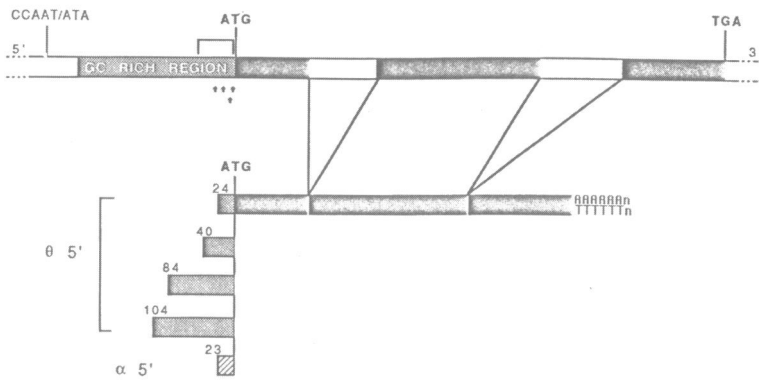


Figure 3

In vitro translation using RNA template from θ cDNA containing various lengths of 5' flanking sequence in wheat germ extract.

A. The *in vitro* translation product of α- (lane 1) and θ-globins (lane 2) from corresponding *in vitro* synthesized RNA was analyzed in a 15% polyacrylamide gel. The sizes of the two proteins are similar and they run at identical position, indicating that the θ-globin is of the same size as α-globin, i.e. 141 amino acids.

B. RNA synthesized *in vitro* by SP6 polymerase was analysed in a 1% agarose gel. Arrow indicates the bands representing different RNA resulting from the *in vitro* transcription of linearized α/θ (lane 1), θ_{104} (lane 2), θ_{84} (lane 3), θ_{40} (lane 4) and θ_{24} (lane 5) pSP64 plasmid. M is the size marker. It shows that the RNA is of similar amount and is not degraded.

C. *In vitro* translation product resulting from similar amount of *in vitro* transcribed RNA corresponding to α/θ (lane 1). θ_{104} (lane 2), θ_{84} (lane 3), θ_{40} (lane 4), θ_{24} (lane 5). Lane 6 represents the negative control containing no exogenous RNA in the *in vitro* translation reaction mixture. All products, indicated by an arrow, run at 14 kd.

D. The structure of cDNA used as templates for the *in vitro* synthesis of SP6 polymerase transcribed RNA.

The genomic sequence of θ-globin gene at the top indicates the CCAAT/ATA region, the GC rich region (hatched box), the exons (solid box), and the introns (open box). Arrows indicate the major transcriptional start sites as detected by 5' end mapping [6]. The bracketed region represents the region within which other minor transcriptional start sites are located. The different templates of θ cDNA used for the *in vitro* synthesis of RNA are illustrated in the lower part of the diagram. By Bal31 deletions (see Materials and Methods), 5' flanking sequences of θ-globin gene with various lengths of 24, 40, 84 and 104 bp were linked to the ATG codon of the θ cDNA, replacing the 23 bp of α-globin 5' non coding region in the original α/θ hybrid cDNA.

signals. Stably transfected K562 cells were selected using G418, which kills tissue culture cells unless they express the neomycin gene product [20]. Of 20 such neomycin resistant K562 clones, all showed enhanced levels of θ mRNA 3' ends using S1 nuclease analysis (data not shown) comparable to that of the endogenous α globin mRNA. Since the θ gene is driven off the more efficient α promoter these clones all produce substantially more θ mRNA sequence than the very low levels of endogenous θ globin mRNA.

Although the different globins synthesised in K562 cells are very similar in size, the use of Triton X100 urea polyacrylamide gel electrophoresis can be used to separate most of them [18]. We therefore wished to establish the position of the θ globin polypeptide with respect to the other globins using such a fractionation procedure. Blood lysates from adults (contains α and β globins), cord blood (contains α and γ globins) and from hydrop foetalis, a genetic disease in which all the α globin genes are deleted (contains ζ, ε and γ globins) were fractionated on a Triton urea gel and run with *in vitro* synthesised radioactive θ globin [see 21]. As shown in figure 4B, the θ globin comigrates with ζ globin in this fractionation system. Figure 4A then shows Triton X100 urea gel fractionation of K562 cell lysates, both normal K562 and stably transfected α/θ globin gene clones of K562 (several separate clones pooled together). It is evident that the ζ globin band just visible in normal K562 lysate becomes substantially more intense in the transfected K562 lysate, demonstrating that the transfected α/θ globin gene is expressing significant amounts of θ globin polypeptide in K562 cells. Evidently the θ globin polypeptide is perfectly stable in human erythroid cells. Unfortunately we were unable to obtain conditions in which θ and ζ globins separated on electrophoresis.

Transcriptional analysis of the θ globin gene's promoter

We wished to investigate the promoter sequences responsible for θ globin gene transcription. This question was of particular interest due to the unusual structure of the θ globin promoter region. As shown in Figure 2A, the θ globin gene possesses a CAAT and ATA box promoter-like sequence which is then displaced from the beginning of the gene by 200 bp of highly GC rich sequence. Our previous results [6] suggested that this GC rich region was the true promoter since heterogeneous θ mRNA Cap sites were found within and directly adjacent to this GC rich sequence. However such a GC rich promoter is normally associated with ubiquitously expressed, so-called housekeeping genes, while θ globin is apparently an early erythroid specific gene [6].

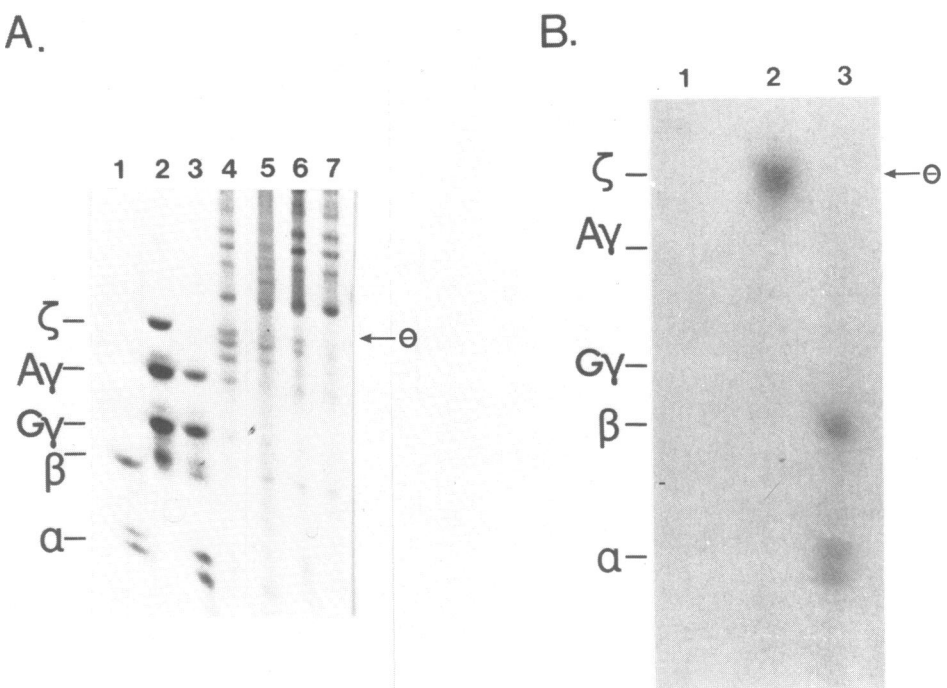
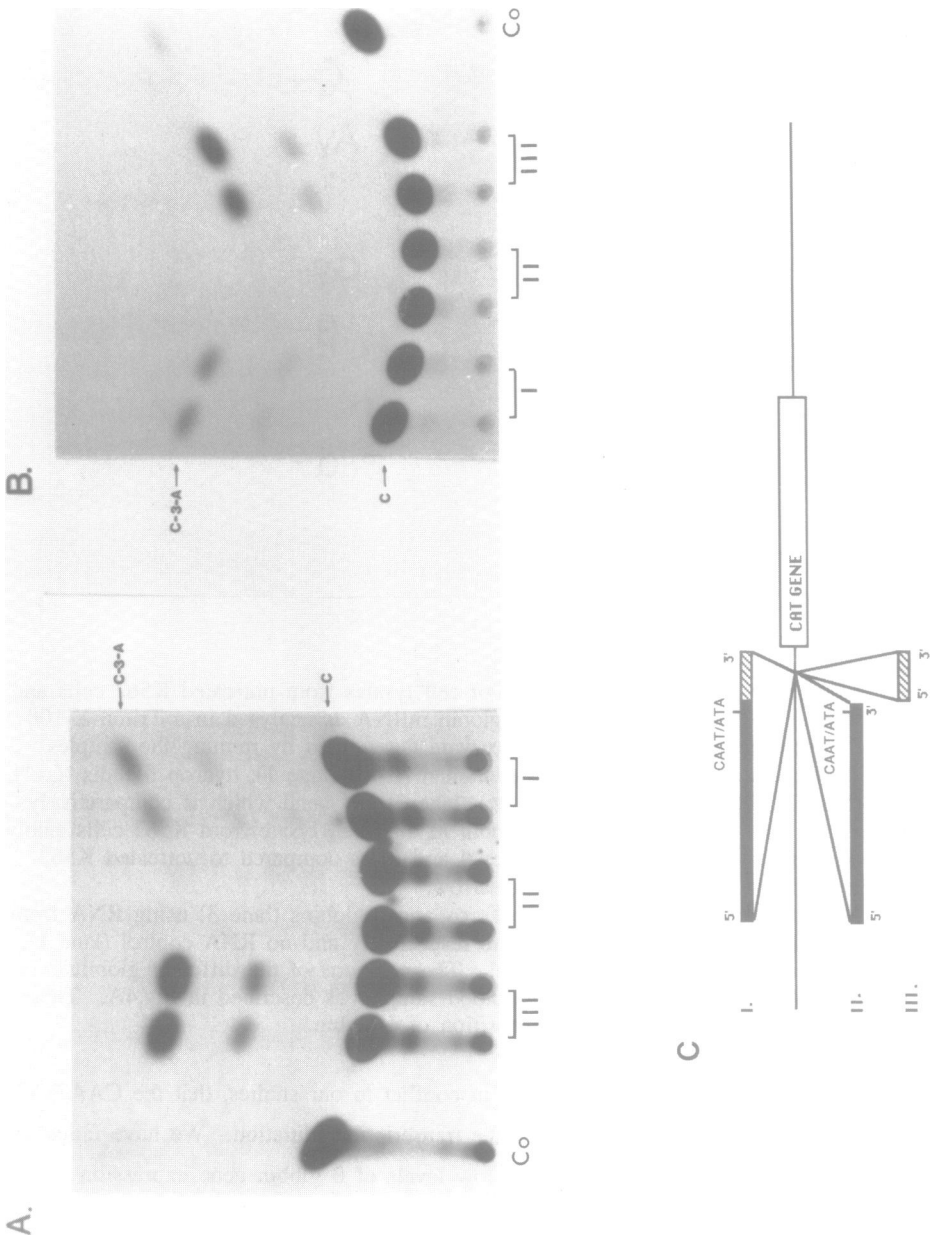


Figure 4
A. Comparison of the protein pattern of cell lysates from untreated K562 cells and K562 cells expressing "high" level of θ-globin mRNA as analysed in a Triton-X-100 urea gel. The locations of ζ, $^{A}\gamma$, $^{G}\gamma$, β, α were determined by running the samples alongside with blood proteins from adult human (α, β) (lane 1), hydrop foetalis (ζ, $^{A}\gamma$, $^{G}\gamma$, β) (lane 2), fetal cord blood ($^{A}\gamma$, $^{G}\gamma$, α) (lane 3). A band which is comparatively stronger and runs at the position of ζ-globin in total cell lysate from K562 cells stably expressing "high" level of θ-mRNA (lanes 4 and 6) as compared to untreated K562 cells (lanes 5 and 7) is marked with an arrow.
B. *In vitro* translated θ-globin (lane 2), α- and β-globins (lane 3) using RNA from *in vitro* transcription and adult reticulocyte respectively and no RNA control (lane 1), were analysed with Triton-X-100 urea gel. The positions of the different globins were determined by running the gel with the blood samples as described in Fig.4A. These data indicate that θ-globin runs at the position of ζ-globin.

Furthermore Hsu et al [8] have reported, in conflict to our studies, that the CAAT/ATA promoter region is the site of at least some transcription initiation. We have failed to detect such θ mRNAs. Due to the very low levels of θ globin gene expression observed in erythroid cells we elected to test the activity of the θ promoter linked to the reporter gene CAT, since the chloramphenicol acetyl transferase activity can be detected when very low levels of mRNA are present [13]. Figure 5C shows a diagram


A.
C-3-A
C
Co
I
II
III
B.
C-3-A
C
Co
I
II
III
C
CAT GENE
CAAT/ATA
CAAT/ATA
5'
3'
I.
II.
III.

of the three constructs we have investigated. As indicated three different θ promoter sequences were attached to the CAT gene, the whole promoter region (I) and either the CAAT/ATA region (II) or the GC rich region (III) separately. These three θ CAT constructs were transiently transfected into Putko or HeLa tissue culture cells. Putko cells are derived from K562 cells and still maintain erythroid characteristics (i.e. they express ζ, α, ε, γ and θ globin mRNAs). They are however more readily transfected than K562 cells. Each θ CAT plasmid was cotransfected with a β galactosidase expressing plasmid. Lysates from transfected Putko or HeLa cells were first assayed for β gal activity. Volumes of lysate yielding equal amounts of β gal activity were then assayed for CAT activity using thin layer chromatography as shown in Figure 5A (Putko) and 5B (HeLa). Each experiment is shown in duplicate. For Putko cells, very little background CAT activity was detectable without a promoter (Co). The whole θ promoter gave low level signals above this background, while the GC rich promoter region gave substantially higher levels of CAT activity. The CAAT/ATA promoter region gave just background CAT activity, indicating that this region of the θ 5' flanking region has no promoter activity. Very similar results were obtained with HeLa cells. A higher background (minus promoter) CAT activity was observed (Co). The whole θ globin promoter was somewhat higher than this background level while

Figure 5
A,B. Analysis of Putko cells (A) and HeLa cells (B) transfected with the three θ CAT plasmids shown in C. The position of the chloramphenicol and its acetylated products on the thin layer chromatogram are indicated. Co denotes transfection with the promoterless plasmid pCAT0 while I, II and III denote the three different θ CAT constructs depicted below.
C. Structure of the different CAT constructs.
Line diagram showing the regions of θ 5' flanking sequence analysed for promoter activity in the CAT assay. Solid rectangle represents the 800 bp upstream θ 5' flanking sequence containing the CCAAT/ATA sequence. The hatched rectangle represents the 200 bp of GC rich sequence of θ-globin gene.
I. Plasmid containing the 1 kb 5' flanking sequence of the θ-globin gene, including the GC rich sequence (hatched rectangle) and the CCAAT/ATA box (solid rectangle).
II. Plasmid containing an 800 bp fragment of upstream 5' flanking sequence of the θ-globin gene containing the CCAAT/ATA sequence but not the GC rich region.
III. Plasmid containing a 200 bp fragment of GC rich 5' flanking sequence of the human θ-globin gene. The respective fragments were cloned in the SmaI site of the plasmid pCAT0 [14]. The chloramphenicol acetyl transferase (CAT) gene is indicated as an open rectangle. The remaining sequence of the plasmid is represented by a straight line.

the GC rich promoter construct (III) gave a much higher level of CAT activity. Again the CAAT/ATA promoter construct (II) gave only background levels of CAT activity. We have also established that for the GC CAT construct (III) the CAT mRNA produced in Putko cells has heterogeneous 5' ends corresponding to the multiple Cap site of θ globin mRNA (data not shown).

The data described here allows us to draw several conclusions about the θ globin promoter. Firstly this promoter works when transfected into both erythroid and non erythroid cell lines in contrast to its early erythroid specificity *in vivo*. Second, the GC rich region is indeed the θ gene promoter, while the CAAT/ATA region appears to be devoid of promoter activity. Finally the lower level of transcription observed with the whole θ gene's promoter may suggest that the gene is subject to negative regulation by distal promoter sequence elements. However further characterization of the θ gene's 5' flanking region will be required to define this putative negative regulation.

DISCUSSION

The experiments described in this paper demonstrate that the θ globin gene is both transcriptionally and translationally functional. We show that the θ globin gene produces properly spliced mRNA and that this mRNA translates into a polypeptide of 141 aa. Furthermore the θ globin polypeptide is translated both *in vitro* and *in vivo* and appears to be a stable protein product when synthesised in erythroid cells.

The θ globin gene's promoter is different from all other globin genes studied to date. The promoter is a GC rich structure of some 200 bp that initiates the synthesis of θ globin mRNAs with heterogeneous 5' ends ranging from 5 to 50 nucleotides 5' to the initiation codon. Although in isolation this GC rich promoter region is relatively efficient, when it is part of the whole 5' flanking region, at least 10-fold lower transcription is observed. This suggests that the θ globin gene is subjected to negative regulation. Similarly negative regulation of the human ζ and ε globin genes has recently been described [14]. It is also surprising that θ globin although erythroid specific *in vivo* appears to function when transfected into either erythroid or non erythroid tissue culture cells. Since the θ globin gene's promoter closely resembles that of a ubiquitously expressed housekeeping gene, this may account for its lack of cell specificity in transfection experiments. Possibly *in vivo* when the θ globin gene is

part of the whole α globin gene cluster it is only expressed when the α like globins are in an active chromatin configuration. Such an active chromatin structure would only occur in erythroid tissue and thereby restrict the expression of θ globin to that tissue.

The ThyI gene of mouse may represent a similar case to θ globin in that the promoter of this gene is GC rich and initiates transcripts from multiple start sites [22,23]. However, like θ, this gene is tissue specific since ThyI surface antigen has only been detected in neurones, fibroblasts and lymphocytes [24].

In vitro translation of θ globin mRNA reveals that the θ mRNA with only 24 nucleotides of GC rich 5' non-coding region is translated relatively efficiently while θ mRNA with longer GC rich 5' non-coding regions are inefficiently translated. Since most θ globin mRNA has very short 5' non coding regions, less than 30 nucleotides [6], the inhibitory effect of the GC rich sequence on translation will only affect the rarer θ mRNAs with unusually long 5' non coding regions. Possibly the potential secondary structure of these GC rich sequences may be the cause of this inefficient translation. It should also be noted that the consensus sequence around the initiation codon of an mRNA proposed by Kozak [25] has a rather poor fit with θ globin mRNA which might further account for its translational inefficiency. We are also able to discount the possibility that the θ globin polypeptide is intrinsically unstable in erythroid cells since we show that K562 cells transfected with the θ globin gene produce high levels of exogenous θ globin. Clearly the key question yet to be addressed is can θ globin be detected in human erythroid tissues? Because of its low levels of expression and also because it comigrates with other globins in Triton X100 urea gels, we have so far failed to detect θ globin during normal human development. The availability of θ specific antibodies may ultimately provide sufficient sensitivity and specificity to prove the existence of θ globin protein in erythroid cells.

Finally the fact that the θ globin protein has amino acid differences to α globin that would preclude its association with haem or in tetramer formation suggest that this protein may have a very different physiology from other haemoglobins. A recent paper by Ley et al [26] has demonstrated that in some haemoglobinopathies such as sickle cell anaemia and β-thalassaemia, θ globin mRNA is detectable at elevated levels. Possibly these observations may throw some light onto the role of θ globin in erythroid cells.

ACKNOWLEDGEMENTS

We gratefully acknowledge Dr Bill Wood for provision of human blood samples and for advice on running Triton X100 urea gels. We are also grateful to Drs Doug Higgs, Bill Wood and John Clegg for advice during these experiments.

S.L. is an ICI Scholar. This work was supported by a Medical Research Council project grant (G8708393CB) to E.W. and N.J.P.

REFERENCES

1. Marks J., Shaw J-P. and Shen C-K.K., Nature 321, 785-788 (1986)
2. Nicholls R.D., Fischel-Ghodsian N. and Higgs D.R., Cell 49, 369-378 (1987)
3. Shaw J-P., Marks J. and Shen C-K.J., Nature 326, 717-720 (1987)
4. Fischel-Ghodsian N., Higgs R.D. and Beyev E.C., Nature 329, 397 (1987)
5. Clegg J., Nature 329, 465 (1987)
6. Leung S., Proudfoot N.J. and Whitelaw E., Nature 329, 551-554 (1987)
7. Rutherford T.R., Clegg J.B. and Weatherall D.J., Nature 280, 164-165 (1979)
8. Hsu S.L., Marks J., Shaw J.P., Tam M., Higgs D.R., Shen C.C. and Shen C-K.J., Nature 331, 94-96 (1988)
9. Higgs (et al) Nature 306, 398-400 (1983)
10. Proudfoot N., Nature 322, 562-565 (1986)
11. Maxam A.M. and Gilbert W., Meth.Enzymol. 65, 499-559 (1980)
12. Melton D.A., Krieg P.A., Rebagliati, Maniatis T., Zinn K. and Green M.R., Nucl.Acids Res. 12, 7035-7055 (1984)
13. Gorman C.M., Moffat L.F. and Howard B.H., Mol.Cell.Biol. 2, 1044-1051 (1982)
14. Lamb P., Watt P., Squire L. and Proudfoot N.J., manuscript in preparation (1989)
15. Lozzio C.B. and Lozzio B.B., Blood 45, 321-334 (1975)
16. Klein et al, J.Natl.Cancer Inst. 64, 725-738 (1980)
17. Herbomel P., Bourachot B. and Yanic M., Cell 39, 653-662 (1984)
18. Alter B. and Goff, Biochem.Biophys.Res.Comm. 94, 843-848 (1981)
19. Proudfoot N.J., Rutherford T.R. and Partington G.A., EMBO J. 3, 1533-1540 (1984)
20. Southern P.J. and Berg P., J.Molec.appl.Genet. 1, 327-341 (1982)
21. Weatherall & Clegg, The Thalassaemia Syndrome, 3rd ed., Blackwell Sci.Publs. (1981)
22. Giguere U., Isobe K.I. and Grosveld F., EMBO J. 4, 2017-2024 (1985)
23. Ingraham I.A. and Evans G.A., Mol.Cell.Biol. 6, 2923-2931 (1986)
24. Williams A.F. and Gagnon J., Science 216, 696-703 (1982)
25. Kozak M., Nucl.Acids Res. 15, 125-128 (1987)
26. Ley T.J., Maloney K.A., Gordon J.I. and Schwartz A.L., J.Clin.Invest. (in press) (1989)