The primary transcription unit of the human $\alpha 2$ globin gene defined by quantitative RT/PCR

Catherine M.Owczarek, Pita Enriquez-Harris and Nick J.Proudfoot*

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

Received October 9, 1991; Revised and Accepted January 20, 1992

ABSTRACT

We have set up an experimental system to map the primary transcription unit of the human $\alpha 2$ globin gene. The duplicated human α globin genes ($\alpha 2$ - $\alpha 1$) were linked to the α globin locus Positive Regulatory Element (PRE) and stably transfected into murine erythroleukaemia cells. We then developed a quantitative reverse transcriptase, polymerase chain reaction assay to map $\alpha 2$ primary transcripts using primer pairs derived from different parts of the $\alpha 2$ globin gene and its 3' flanking region. This approach has revealed the presence of steady state nuclear RNA past the poly(A) site of the $\alpha 2$ globin gene at approximately 40% of the level of unspliced intron transcript. Furthermore, these 3' flanking transcripts diminish 500 bp into the 3' flanking region, identifying this part of the $\alpha 2$ globin gene as the principal region of termination of transcription.

INTRODUCTION

Prokaryotic mRNAs form their 3' ends by transcriptional termination. In contrast, the 3' ends of eukaryotic mRNA are formed by endonucleolytic cleavage at a specific site in the transcript and subsequent polyadenylation. RNA polymerase II (Pol II) in higher eukaryotes appears to continue transcribing a gene past the poly(A) site until release of the polymerase from its DNA template is thought to eventually occur. The process of eukaryotic mRNA 3' end processing is now well characterised in terms of both *cis*- and *trans*-acting factors (1,2,3). However, the sequences and putative factors involved in the process of transcriptional termination remain somewhat less well defined. It has been established that in order for efficient termination to occur a functional poly(A) signal is necessary, for example in the human $\alpha 2$ globin gene (4), the SV40 early poly(A) region (5) and the mouse β major globin gene (6). Recently, a model of transcriptional termination in Pol II transcription units has been proposed (5,7). Once elongation of the transcription complex past the gene's poly(A) signal and 3' cleavage has occurred, the nascent RNA chain may be recognised by an RNA:DNA helicase in a similar fashion to bacterial factor rho (8), together with a 5' to 3' exonuclease. Removal of the nascent RNA by these enzymes may then catalyze termination at a downstream site.

A slowing down or pausing of the elongating polymerase complex would further enhance its eventual disengagement from the DNA template. Such 'pause signals' have been identified at the 3' end of several different genes: the human $\alpha 2$ globin gene (9), rabbit α globin gene (10) and mouse adenosine deaminase gene (11).

The principle difficulty in verifying models for transcriptional termination by polII is that primary transcripts which extend into the genes 3' flanking region up to the site of termination, are highly unstable and therefore hard to detect. Indeed, the reliance on nuclear run-off analysis (12) to plot polII density in the genes 3' flanking region may be questioned due both to the difficulty in obtaining signals above background level and because of the unphysiological nature of the technique. It may well be argued that in isolated nuclei, polII elongation complexes may fail to terminate properly and therefore continue transcription past the gene's poly(A) signals. We have sought to circumvent these difficulties by developing an alternative assay for the detection of primary transcripts beyond a genes poly(A) signals. To do this we have investigated the use of the highly sensitive polymerase chain reaction (PCR) (13) coupled with reverse transcription (14–17) to detect and quantitate steady state nuclear transcripts that read into the 3' flanking region of the human $\alpha 2$ globin gene.

Previous studies on transcription of the human $\alpha 2$ globin gene in erythroid cells showed that termination of transcription occurs soon after the poly(A) signal. The results were obtained using nuclear run-on analysis with large DNA probes so that the precise site of termination could not be established (4). More recently we have shown by transient transfection of human $\alpha 2$ globin constructs into HeLa cells that a transcripts extend over 300nt into the 3' flanking region up to a transcriptional pause site (9). Unfortunately it proved impossible to establish exactly where termination occurs beyond the pause site as repetitive sequences at this position obscured the nuclear run-off analysis.

In these studies we have employed gene constructs containing the duplicated human $\alpha 2$ and $\alpha 1$ globin genes with the full 3kb intergenic sequence between them. We have linked these a gene to the potent enhancer region normally positioned 50kb upstream of the human α globin gene cluster (18, 19) and then stably transfected this construct into murine erythroleukemia cells (MELC). Nuclear RNA purified from these transfected cells was analyzed by RT/PCR and has revealed that significant levels of

* To whom correspondence should be addressed

the α globin transcription read part the poly(A) signals up to the pause site (350nt downstream). Beyond this point transcript signals diminish to low levels indicating that this position is the physiological site of termination.

MATERIALS AND METHODS

DNA Constructs

pGW

A 3.8 kb HindIII-HindIII fragment, comprising the 3' half of the $\alpha 2$ globin gene, the intergenic region and the 5' half of the $\alpha 1$ globin gene (20) was cloned into the $\alpha 1$ globin gene in the construct Tag $\alpha 1$ pSVod (21) at the HindIII site in exon 2. The $\alpha 1$ globin gene in the construct Tag $\alpha 1$ pSVod contains a 13 bp tag in exon 1. Sequences upstream of the exon 2 HindIII site are essentially identical in $\alpha 2$ and $\alpha 1$ globin genes (22) and hence can be considered interchangeable. The resulting construct, pGW, therefore contains a tagged upstream α globin gene (henceforth known as $\alpha 2$ globin), the $\alpha 2$ - $\alpha 1$ intergenic region and a downstream α globin gene (henceforth known as $\alpha 1$ globin).

pGM and p Δ G

The $\alpha 2$ gene in pGW was replaced by either an $\alpha 2$ gene possessing an inactive poly(A) site to make pGM or a deleted promoter to make p Δ G. In detail the plasmids p $\alpha 2$ M3'PS (mutated poly(A) site) (4) and Δ PapSVod (deleted promoter region) (21) were digested with SfiI and SacI. Fragments extending from the SV40 origin to 3' of the α gene were isolated and inserted in place of the $\alpha 2$ gene in pGW.

pGW/GM/ Δ G-nPRE

In the second stage of construction, plasmids pGW, pGM and p Δ G were modified to contain the neomycin resistance gene and the human α globin Positive Regulatory Element (α -PRE, see refs. 18,19). The plasmid T ψ ζ -NeoPRE, a gift from Marc Pondel, contains the Tn5 neomycin phosphotransferase gene derived from pSV2Neo (23), as well as a 4.0 kb BamHI-linked fragment containing the α -PRE cloned into the BamHI site of T ψ ζ pSVodNeo (24). A SfiI-SalI fragment, comprising the α -PRE and the neomycin gene, was purified from this construct and then ligated into the pG-derivatives in place of the pBR322-derived tetracycline resistance gene. Plasmids were selected using kanamycin. A schematic diagram of the construct pGW-nPRE is illustrated in Figure 1A.

The plasmid pGEM α was subcloned by inserting a 3.8kb HindIII-HindIII fragment comprising the 3' half of the $\alpha 2$ globin gene, the intergenic region and the 5' half of the $\alpha 1$ globin gene, (as described above) into the HindIII site of the vector pGEM 7zf. A schematic diagram of the relevant regions of this construct is shown in Figure 1B.

Tissue Culture

Murine erythroleukaemia cells (MELCs) were maintained in RPMI 1640 medium supplemented with 15% foetal calf serum. Transfection into MELCs was carried out by electroporation. The cells were harvested in the exponential growth phase, washed with phosphate buffered saline (PBS) and suspended in 0.8 ml of electroporation buffer (25mM HEPES [N-2- hydroxyethyl-piperazine-N'-2-ethanesulphonic acid, pH 7.1], 140 mM NaCl, 0.75 mM Na₂HPO₄, 0.5% PEG 6000) and 50 μ g of test DNA was added. The mixture was electroporated with a Biorad Gene Pulser at 250V and 960 μ F, added to 25ml of medium and

incubated for 48 hours at 37°C in 5% CO₂. G418 (0.8 mg/ml) was then added to the medium and after about two weeks of G418 selection, resistant colonies were obtained. Induction of MELCs was performed by treatment with 5mM hexamethylene-bis-acetamide (HMBA) for three days.

RNA Purification

After induction, the MELCs were harvested, lysed in Nonidet P-40 detergent buffer and the cytoplasmic and nuclear fractions separated by centrifugation through a sucrose cushion. Cytoplasmic RNA was purified by phenol-chloroform extraction. Nuclear RNA was prepared by proteinase K incubation and phenol-chloroform extraction. After ethanol precipitation the total recovered nucleic acid was resuspended in 10mM Tris pH 7.5, 10mM MgCl₂ and treated with RNase free DNase I (Boehringer Mannheim) for one hour at 37°C to remove chromosomal DNA. The nuclear RNA was further purified by phenol-chloroform extraction, ethanol precipitation and then rinsed with 70% ethanol.

Synthesis of *in vitro* transcribed α globin RNA

The plasmid pGEM α was linearized with EcoRI and transcribed with SP6 RNA polymerase as described by Melton *et al.* (25). Template DNA was removed by digestion with RNase free DNase I (Boehringer Mannheim) for one hour at 37°C. The products were purified by extraction with phenol-chloroform, ethanol precipitation and then rinsed with 70% ethanol.

Oligonucleotides

The primers used for reverse transcription and subsequent amplification by the polymerase chain reaction were synthesized on an Applied Biosystems 380 B DNA synthesizer. The locations of the primer pairs in the α globin genes and the 3' flanking region of $\alpha 2$ globin are illustrated in Figure 3. The sequences of the primers and their positions on the α globin genes relative to the $\alpha 2$ cap site are shown in Table I.

S1 Nuclease Analysis of RNA

The method used for S1 analysis has been described previously (26). The S1 nuclease products were electrophoresed through a 4% polyacrylamide-8M Urea gel. Autoradiography was performed at -70°C with an intensifying screen.

Reverse Transcription and PCR

Nuclear RNA (50–200 ng) was incubated with Avian Myeloblastosis Virus Reverse Transcriptase (8U Life Sciences) for 45 minutes at 42°C in a total reaction volume of 10 μ l containing 1 $\times$ PCR buffer (20mM Tris-HCl pH 8.4, 5mM MgCl₂, 50mM KCl, 0.001% gelatin) 0.25mM of each dNTP, 10 U RNasin (Amersham) and 5 pmol of the appropriate anti-sense primer. The reaction was heated at 95°C for 10 minutes and then quick chilled on ice. 40 μ l of 1 $\times$ PCR buffer (with the appropriate pre-determined MgCl₂ concentration for the primer pair) containing one unit of Taq polymerase (Cetus) and 25 pmol of the upstream primer and a further 25 pmol of the downstream primer. The primers for amplification with PCR were end-labelled using [γ -³²P]ATP and T4 DNA polynucleotide kinase (New England Biolabs). Negative control reactions, which were incubated in the absence of AMV Reverse Transcriptase, were included for every primer pair. PCR was performed in a thermal cycler (Cetus) according to the following protocol: an initial denaturation at 95°C for 2 minutes was followed by 18–20 cycles

Table 1. Sequences of primers used in RT/PCR assay

Primer number (5' Primer)	Primer number (3' Primer)	Length of PCR product (bp)	Nucleotide Position of PCR product relative to $\alpha 2$ cap site
Upstream initiation			
5' 671 CCGCGCCCCAAGCATAAACC 3'	5' 672 CTCGCCAGCGTGCGGCCGA 3'	163($\alpha 2$) 150 ($\alpha 1$)	-40 - 110 3962 - 3911
Exon 1 ($\alpha 1$)			
670 GTCCCCACAGACTCAGAGAG	672 CTCGCCAGCGTGCGGCCGA	100	3812 - 3911
Intron 2			
733 GAGCGGCGGGCCGGGAGCGA	734 GTGCAGAGAAGAGGGTCAGTGC	138($\alpha 2$) 145 ($\alpha 1$)	458 - 595 4259 - 4403
Poly(A)			
432 CTGGACAAGTTCCTGGCTTCTG	433 CTTGGTCTGAGACAGGTAAAC	237	673 - 908
Post Poly(A)			
483 CTCTGTCCCGGAATGTGCCA	484 AGGACAGGGGATGGTTCAGCAC	160	856 - 1017
Pre-pause			
434 CTGCAGGCAGTGGCTCAGAAG	244 GTTTTGTTTTGTTTTGTTTTGTTTT	149	1032 - 1180
Pause site			
434 CTGCAGGCAGTGGCTCAGAAG	435 GGGGTGGGAATGAGAGAAATGT	212	1032 - 1243
Post-Pause			
1674 ACATTTCTCTCATTCCACCC	436 CTGCAGGAGACGTGCAGGAG	217	1222 - 1436
Pre-Alu			
436 GCAACTTCTGGAACGTCCTG	437 AAGCGGTGACAGGGTCTC	204	1547 - 1750
Post-Alu			
438 CTTCTCTGCCCAAGGCAGCT	439 AGGAGTGCATCAGGTCAGCA	120	2301 - 2420

of 1 minute denaturation, annealing the primers at 55°C for 1 minute and extending the primers for 3 minutes at 72°C. A further 5 minute extension at 72°C was performed to extend any remaining single stranded products. The amount of input RNA and the number of amplification cycles were held constant for a given set of experiments.

The PCR amplified products were separated on a 5% native polyacrylamide gel, stained with ethidium bromide, and the appropriate bands excised after visualization under UV light. Cold PCR product was sometimes added to the test PCR products to aid visualization. The radioactivity in each band was determined by Cerenkov counting.

Quantification of the RT/PCR Assay

At very high cycles of PCR it is possible to detect PCR products which derive from transcripts originating from a considerable distance into the 3' flanking region of the $\alpha 2$ globin gene. It was therefore desirable to ascertain whether the levels of these transcripts were significant. Furthermore, at saturating cycles, transcripts which are present in low abundance are over-represented, whilst transcripts which are present in high levels will reach a plateau and therefore be relatively under-represented. In order to quantitate the transcripts that we detected with saturating cycles of PCR we have used as a standard an *in vitro* transcribed SP6 RNA (pGEM α) (Figure 1). The SP6 α globin RNA contains the same sequences as the test RNA but the target sequence for each primer pair is present on the same RNA molecule in a single copy. This is reverse transcribed and amplified in parallel with the test RNAs. The values obtained after reverse transcription and amplification of the SP6 RNA were in a similar range to the values obtained from the test RNA. As the different primer pairs that we tested amplified with different efficiencies (data not shown) we could use this control RNA template to establish a value for the relative efficiency of

amplification for each primer pair. We have not attempted to determine precisely the number of molecules of RNA present at a given point but rather ascertain the relative levels of the different PCR products.

In the first step of the assay, cDNA is made from nuclear RNA using the specific 3' primers listed in Table 1, thus ensuring that only sense transcripts are being detected by the assay. These reactions are then used as templates for the polymerase chain reaction to amplify short regions from the respective cDNAs. The annealing temperature is set at 55°C which is below the T_m [as calculated by $T_m = 2(A+T) + 4(G+C)$] of all the primer pairs used in this analysis so that they could be directly compared. The PCR primers are labelled with [γ - 32 P]ATP to allow quantification of the PCR products. In performing this assay it proved important to minimize the differential loss of product during the steps, so the cDNA reactions and subsequent amplifications were carried out in the same reaction tubes. To improve reproducibility and minimize the level of background contamination from exogenously introduced plasmid sequences, master mixes were prepared for all steps of the reaction. Stringent procedures, as described by Kwok (27), to minimize carry-over and plasmid contamination were observed. The ratios of the different PCR products relative to each other remained constant from experiment to experiment (data not shown) when normalized to the SP6 control RNA but the absolute values of incorporated radioactivity varied which indicated that the amplification efficiencies of the primer pairs were varying slightly. It was therefore necessary to perform the SP6 control in parallel each time a quantitative RT/PCR assay was carried out. In order to determine the maximal RNA input to ensure that we stayed in the linear range of the amplification curve we added serial dilutions of RNA to the reverse transcriptase reactions and subjected them to subsaturating cycles of PCR. To determine the optimum number of cycles of PCR for quantitation of a given

Figure 1. A. Schematic representation of the plasmid pGW-nPRE, the construction of which is described in Materials and Methods. The human α globin genes are depicted as solid boxes and the direction of transcription is shown with an arrow. The SV40 origin sequence (solid box), Neomycin resistance (Neo) gene (hatched box) and the α globin Positive Regulatory Element or PRE (open box) are indicated. B. Diagram of the template (pGEM α) used for *in vitro* transcribed α globin RNA which was used as a control for quantification of RT/PCR reactions. The sequences transcribed are shown flanked by HindIII restriction sites which were used to clone the fragment into the vector pGEM7zf.

RNA sample, parallel samples of the test RNA were subjected to increasing amplification so that we could determine the cycle number at which amplification started to reach saturation. This was determined for each RNA sample tested using the primer pair that detects transcription from exon I of the $\alpha 1$ globin gene.

RESULTS

Analysis of human α globin gene expression in stably transfected MELCs by S1 nuclease digestion

The plasmid pGW-nPRE contains the duplicated human α -globin gene with the whole 3.0kb intergenic sequences between them, linked to the strong enhancer like sequence α -nPRE (18, 19) (normally 50kb upstream of the a gene locus). Mutations in the upstream a gene were also made that have either a mutated poly(A) signal (pGM) or deleted promoter (pΔG). These three a gene constructs were stably transfected into MEL cells and complex pools of clones were obtained (see Materials and Methods).

To demonstrate that human α globin transcripts were present in these MEL cell pools and that correct α globin mRNA 3' ends were being formed, S1 nuclease analysis was performed. Figure 2A shows the location of the S1 probes and the expected digestion products on the $\alpha 2$ and $\alpha 1$ globin genes. The S1 probes

Figure 2. A. Diagram of the human α globin genes showing the location and sizes of the probes used for S1 analysis (thick lines) and the expected digestion products (thin lines). B. S1 nuclease analysis of cytoplasmic RNA obtained from induced MELCs which have been stably transfected with the constructs pGW-nPRE, pGM-nPRE and pΔG-nPRE. Lane 1: Undigested probe, Lane 2: tRNA control, Lane 3: SP6 control, Lane 4: pGW-nPRE, Lane 5: pGM-nPRE, Lane 6: pΔG-nPRE. Molecular weight of markers (M) is given in base pairs. The S1 digestion products corresponding to the 3' ends of $\alpha 2$ globin and $\alpha 1$ globin are indicated. This figure is further described in the text.

used to analyse cytoplasmic RNA obtained from induced MELCs transfected with the above constructs were made by purifying and end-labelling with $\alpha^{32}\text{P}$ -dGTP a 690 bp BstEII-BstEII (Band A) and a 605 bp BstEII-BstXI (Band B) fragment from the plasmid p $\alpha 2\text{W3'PS}$ (4). Figure 2B shows the results of the S1 analysis. Lane 1 shows undigested probe. Yeast tRNA was used as a negative control (lane 2). Lane 4 shows the S1 digestion products obtained from cytoplasmic RNA from induced MELCs transfected with the construct pGW-nPRE. Two major products can be observed. The band at 217 bp corresponds to the $\alpha 2$ 3' end. The $\alpha 2$ BstEII-BstEII probe can also detect 3' ends which derive from $\alpha 1$ globin as the two regions are very similar but the resulting band is a different size (124 bp) due to a region of mismatch. Band E is an artefact which is the result of AT 'breathing'. The products obtained from S1 digestion of

Figure 3. Diagram of the region of the human α globin locus showing the $\alpha 2$ globin gene, part of the intergenic region (line) and the $\alpha 1$ globin gene. The α globin genes are depicted as boxes. The shaded areas within the genes indicate coding sequences and the blank areas are non-coding sequences. The canonical initiation sites for both genes are indicated by arrows. The position of the 13 bp 'tag' in the first exon of $\alpha 2$ globin is also indicated. The 'pause site' as defined by Enriquez-Harris *et al.* (9) is indicated by a small black box. The positions and assigned names of the PCR primer pairs (—) are shown below.

cytoplasmic RNA from induced MEL cells transfected with the construct p Δ G-nPRE (Lane 6), which does not have an $\alpha 2$ promoter, shows an intense band corresponding to the $\alpha 1$ 3' end and a very faint band corresponding to the $\alpha 2$ 3' end. This faint $\alpha 2$ band may derive from low level upstream transcripts reading into the $\alpha 2$ gene.

S1 nuclease analysis was also carried out on cytoplasmic RNA obtained from induced MELCs which had been stably transfected with the construct pGM-nPRE which contains a mutated $\alpha 2$ poly(A) signal. An *in vitro* SP6 transcribed RNA made from the template pGEM α was used as a positive control to detect total protection of both probes (Lane 3). As shown in Figure 2B, RNA derived from the construct pGM-nPRE (Lane 5) gave products corresponding to $\alpha 1$ globin 3' ends (124 bp) and a faint signal corresponding to $\alpha 2$ 3' ends. This indicates that the mutant $\alpha 2$ poly(A) signal is working at a very low efficiency or that this band is a partial S1 digestion product. Products corresponding to total protection of the BstEII-BstEII probe (Band A, 690 bp) and to RNA forming stable 3' ends at a cryptic poly(A) signal (Band C, 580 bp) respectively were also observed but at a much lower intensity relative to the $\alpha 1$ S1 product. This confirms that the mutated poly(A) signal reduces 3' end formation at the $\alpha 2$ poly(A) site and leaves a small proportion of the RNA products to be processed at the cryptic poly(A) site further downstream. This cryptic poly(A) signal has been previously described and has been shown to be relatively inefficient (4). Total protection of the BstEII-BstXI (Band B, 605 bp) probe was not observed in RNA derived from the construct pGM-nPRE. This indicates that transcripts which derive from the $\alpha 2$ promoter and read through past the cryptic poly(A) signal are not processed at the $\alpha 1$ poly(A) site and are hence unstable. A doublet, (D), is presumably another 'breathing' artefact since it disappears with higher concentrations of S1 nuclease (data not shown).

Development of a quantitative RT/PCR assay in order to analyse human α globin gene expression in stably transfected MELCs

As illustrated in Figure 3, we have designed PCR primers that will detect transcripts originating from the first exon of $\alpha 1$ globin, the second intron of $\alpha 2$ and $\alpha 1$ globin and at intervals in the 3' flanking region of the $\alpha 2$ globin gene. The sequences of the

Figure 4. Detection of primary transcripts in 100 ng of nuclear RNA isolated from induced MELCs stably transfected with the construct pGW-nPRE. The nuclear RNA was subjected to reverse transcription (+) with specific 3' primers and then amplified for 28 cycles with the appropriate primer pairs. The primer pairs used are indicated below each lane. Each reaction was done in duplicate. A negative control, where reverse transcriptase was omitted (—) in the reverse transcription step was included for each primer pair. The resulting PCR products were electrophoresed on a 5% polyacrylamide gel, stained with ethidium bromide and photographed under U.V. Molecular weight of markers (M) are given in bp. In the second panel, the arrows on the left indicate the PCR products which detect transcripts initiating upstream of the canonical initiation sites of the $\alpha 2$ and $\alpha 1$ globin genes. The arrows on the right indicate the PCR products which detect transcripts resulting from the unspliced introns of the $\alpha 2$ and $\alpha 1$ globin genes. The PCR products shown in this Figure are further described in the text.

primers used for RT/PCR and the expected sizes of the resulting PCR products are listed in Table 1. Using a computer, primers were checked for sequences that could cross react with the endogenous mouse α globin gene. No significant matches were found. We also performed 30 cycles of PCR using 100 ng of induced but untransfected MELC nuclear RNA. With some primer pairs faint bands were observed but they were not the correct size (data not shown).

We reverse transcribed 100 ng of nuclear RNA (obtained from MELCs which have been stably transfected with the construct pGW-nPRE) with the specific 3' primers listed in Table 1 and amplified the resulting cDNAs with the appropriate primer pairs for 28 cycles of PCR. This is substantially more cycles than we would normally carry out but it allows visualization of the resulting PCR products. Furthermore, the experiment was not performed under quantitative conditions so the intensities of each band cannot be considered to be a measure of the relative RNA levels. Each reaction was carried out in duplicate. A negative control, that contained all the reaction components except reverse transcriptase, was carried out for each primer pair. Figure 4 shows that transcripts can be detected in all of the regions delineated by the primer pairs and that the presence of a PCR

Figure 5. Nuclear RNA, in the amounts indicated, from induced MELCs stably transfected with the construct pGW-nPRE was reverse transcribed with the primer 672 and then amplified for 20 cycles with 32 P-labelled primer pair 'Exon 1' (see Table 1). Each reaction was carried out in duplicate (as indicated by the open and solid squares). The resulting PCR products were electrophoresed on a 5% polyacrylamide gel, stained with ethidium bromide, visualised under U.V. and then quantitated by counting the radioactivity of the excised gel bands. The resulting values were corrected for background amplification, which was obtained from negative controls where no reverse transcriptase was added, and plotted against the amount of input nuclear RNA.

product is dependent on the addition of reverse transcriptase. The PCR products obtained were the expected sizes and were also verified by restriction enzyme analysis (data not shown). The PCR product resulting from amplification with primer pair 'Exon 1' only detects transcripts originating from $\alpha 1$ globin as the corresponding primer binding site for primer 670 in exon 1 of $\alpha 2$ globin has been disrupted by the 13 bp 'tag'. The PCR products resulting from amplification with primer pairs 'Intron 2' and 'Upstream Initiation' will be discussed further. All other primer pairs amplify PCR products that originate from sequences downstream of the $\alpha 2$ globin gene. Several of the duplicate lanes in Figure 4 have extra bands that do not correspond to an expected product (Upstream initiation: ~230 bp, Intron 2: ~180 bp and Pre-Alu: ~75 bp). We think that these are simply artefact bands resulting from the relatively low temperature at which reverse transcription takes place as these bands do not arise when an identical DNA template is amplified (data not shown).

As we are interested in comparing the levels of the relatively low abundance transcripts that are present downstream of the $\alpha 2$ globin gene and in assessing their significance it was necessary to make the assay quantitative. Robinson and Simon (17) have utilised the theoretical linearity of input molecules of RNA vs PCR product to measure mRNA levels of Pgk-2 during spermatogenesis. They, and others (14, 15, 16), have shown that reverse transcription with subsequent PCR can be performed in a quantitative fashion provided that the RNA template is limiting and number of cycles are within the exponential phase of the amplification curve. We therefore made serial dilutions of nuclear RNA from stably transfected induced MELCs and subjected them to a fixed amount of cycles of PCR in order to determine the optimum amount of RNA for meaningful assessment of RNA

Table 2. Human $\alpha 2$ and $\alpha 1$ Globin gene expression in stably transfected MEL cells.

$\alpha 2:\alpha 1$ RNA		
Introns	63:37	
Upstream initiation sites	61:39	
[Normal adult reticulocytes	60:40	(20)].
$\alpha 1$ Upstream initiation site		
	% of total $\alpha 1$ transcription	
Nuclear	3	
Cytoplasmic	2	

levels. In the experiment illustrated in Figure 5 we used RNA obtained from MELCs transfected with clone pGW-nPRE and performed RT/PCR assays with the primer pair that detects transcripts from exon 1 of $\alpha 1$ globin. The primer pair was end-labelled with [γ - 32 P]ATP to allow quantitation. The reactions were carried out in duplicate and the signal obtained was corrected for background amplification. Figure 5 shows that at 20 cycles of PCR the amount of input RNA was approximately linearly related to the amount of PCR product obtained between 1 and 50 ng of input RNA. The graph begins to plateau at 100 ng RNA and then plateaus dramatically after 250 ng RNA. We also subjected these samples to different cycles of PCR to determine at which cycle number the amount of PCR product relative to input RNA started to decline. With 100 ng RNA the amount of PCR product was linear up to 18 cycles and plateaued from 20 to 30 cycles (data not shown). This procedure was carried out for each RNA sample assayed.

Analysis of human α globin gene expression in stable MEL cell transformants by quantitative RT/PCR

As shown in Figure 4, the PCR products for the $\alpha 2$ intron 2 can be resolved from the $\alpha 1$ intron 2 on an acrylamide gel as they differ in size by 7 bp. The extra band at approximately 180 bp does not correspond to an expected product. In order to quantitatively analyse the relative expression of the two transfected α globin genes we reverse transcribed 100 ng of nuclear RNA (obtained from MELCs transfected with the construct pGW-nPRE) with the 3' intron 2 primer and subsequently amplified the resulting cDNAs for 18 cycles of PCR with the intron 2 specific primers. We then used the ratios of the two α globin introns to determine the average ratio of $\alpha 2$ to $\alpha 1$ expression as approximately 63:37 (Table 2) which is similar to the $\alpha 2:\alpha 1$ mRNA levels which have been reported for normal erythroid cells (28).

Hess *et al.* (28) have reported the presence of an upstream transcription initiation site in the $\alpha 1$ globin gene which is present in human bone marrow and also in Cos cells transiently transfected with human $\alpha 1$ globin DNA. As illustrated in Figure 3 the $\alpha 2$ globin gene in the construct pGW-nPRE has a 13 bp tag which allowed us to differentiate between the PCR products representing transcripts initiating upstream of the canonical transcriptional initiation site of $\alpha 2$ globin and those initiating upstream of the canonical initiation site of $\alpha 1$ globin. Using a 3' primer positioned in the first exon of each α globin gene (primer 672) and a 5' primer (primer 671), (the 5' end of which

Figure 6. Levels of primary transcripts in the 3' flanking region of the human $\alpha 2$ globin gene. 100 ng of nuclear RNA was obtained from induced MELCs stably transfected with the construct pGW-nPRE (●) reverse transcribed and amplified with the specific primer pairs (which were end-labelled with ^{32}P) indicated on the x axis. The conditions of reverse transcription and amplification are described in the text. Each experiment was carried out at least twice and the results are shown as the mean $\pm$ S.E. The resulting PCR products were electrophoresed on a 5% polyacrylamide gel, stained with ethidium bromide and the excised bands were counted for incorporated radioactivity. For primer pairs where the amount of PCR product was very low, a small amount of the corresponding non-radioactive PCR product was added to aid visualisation. The amount of radioactivity in each PCR product was normalized relative to the levels obtained from an *in vitro* transcribed SP6 RNA from the template pGEM α . The normalized levels of transcription are shown as a percentage of the level obtained from unspliced $\alpha 2$ intron detected by the primer pair 'intron 2'. Shown on the same graph are the levels of primary transcripts in the 3' flanking region of a human $\alpha 2$ globin gene with a non functional poly(A) signal. 100 ng of nuclear RNA was obtained from induced MELCs stably transfected with the construct pGM-nPRE (■) and was reverse transcribed and amplified with the specific primer pairs (which were end-labelled with ^{32}P) indicated on the x axis. The conditions of reverse transcription and amplification are described in the text. The results were obtained as described above.

is located at -41 relative to the canonical initiation site of each α globin gene), we performed an RT/PCR assay on nuclear and cytoplasmic RNA obtained from MELCs transfected with the construct pGW-nPRE. We found that upstream transcripts were present in both α globin genes in nuclear RNA (as shown in Figure 4) and cytoplasmic RNA (data not shown). An extra band at approximately 230 bp is present which does not correspond to an expected product. Using quantitative conditions we found that the ratio of $\alpha 2$ to $\alpha 1$ globin upstream initiated transcripts in nuclear RNA was similar to the ratio of $\alpha 2$ to $\alpha 1$ gene expression which was obtained by comparing unspliced $\alpha 2$ intron 2 to $\alpha 1$ intron 2. We compared the level of transcription from the $\alpha 1$ upstream initiation site to the amount of signal obtained from a primer pair in which the 3' primer is identical (primer 672) but where the 5' primer (primer 670) is located 3' to the canonical initiation site ('Exon 1'). An example of this PCR product is shown in Figure 4. The amount of PCR product from this primer pair is a measure of the total amount of $\alpha 1$ globin transcription. The number of transcripts originating from the upstream initiation site comprised approximately 3% of the total number of $\alpha 1$ globin transcripts present in the nucleus (Table 2). A similar level (2%) of transcripts which have initiated upstream of the canonical $\alpha 1$ initiation site was detected in

cytoplasmic RNA (Table 2) indicating that these transcripts are efficiently transported from the nucleus to the cytoplasm.

Determination of the region of transcriptional termination in the human α globin genes using quantitative RT/PCR

In order to assess where $\alpha 2$ globin transcripts cease to be detectable in our assay, a condition which we judge to indicate termination of transcription, we subjected 100 ng of RNA obtained from MELCs transfected with the construct pGW-nPRE to reverse transcription with specific 3' primers detecting transcripts from unspliced $\alpha 2$ intron 2 and in the 3' flanking region of $\alpha 2$ globin. The resulting cDNAs were then amplified with the corresponding ^{32}P end-labelled primer pairs for 18 cycles of PCR. The amount of radioactivity present in the respective PCR products obtained for each primer pair was corrected for background amplification, as shown by the minus reverse transcriptase controls, and normalized relative to the values obtained from the SP6 control RNA which was reverse transcribed and amplified in a similar fashion to the test RNA. The values shown in Figure 6 are the mean [$\pm$ Standard Error (S.E.)] of at least two experiments where, in each experiment, the reactions for each primer pair were carried out in duplicate. Figure 6 shows the levels of transcripts that can be detected in the intergenic region downstream of the $\alpha 2$ poly(A) site expressed as a percentage of the unspliced message originating from second intron of the $\alpha 2$ globin gene. Our results demonstrate that the transcript level detected by the pair of primers which span the $\alpha 2$ poly(A) site, primer pair 'Poly(A)', is approximately 40% of the $\alpha 2$ unspliced message. We then show that the signal obtained from primer pairs located over the next 300 bp remains relatively constant, drops to a substantially lower level at the primer pair that is located immediately after the pause site ('Post-Pause') and then declines further at the primer pair located just before the Alu sequence ('Pre-Alu'). In order to be sure that these transcripts were linked to the $\alpha 2$ promoter we carried out a quantitative RT/PCR assay with nuclear RNA from MELCs transfected with the construct pGAP-nPRE, in which the $\alpha 2$ promoter is deleted. We found that there was a low level of background transcription when an $\alpha 2$ promoter was not present although the overall pattern obtained, (as shown in Figure 6 for pGW-nPRE), was similar. This may not reflect the true situation when the normal promoter is present since these transcripts, which are presumably reading either from random upstream initiation sites or as a consequence of tandem insertion of the α globin genes, may be blocked by the normal $\alpha 2$ promoter.

To demonstrate what effect a mutated poly(A) signal would have on the levels of $\alpha 2$ globin RNA, as shown by our quantitative RT/PCR assay, we transfected MELCs containing a construct which was identical to pGW-nPRE but where the canonical poly(A) signal AATAAA in the upstream α globin gene is mutated to AATAAG (pGM-nPRE), a situation analogous to a naturally occurring α -thalassaemic mutant (30). This mutated poly(A) signal has been previously shown to polyadenylate and cleave inefficiently. Whitelaw and Proudfoot (4) have demonstrated that when RNA polymerase II encounters the mutant $\alpha 2$ poly(A) signal it fails to terminate transcribing until a cryptic polyadenylation signal is encountered in an Alu sequence downstream of $\alpha 2$ globin.

100 ng of nuclear RNA, obtained from induced MELCs stably transfected with the construct pGM-nPRE, was subjected to the RT/PCR assay under quantitative conditions. As shown in Figure 6 the transcript levels detected by primer pairs 'Poly(A)'

and 'Pre-Alu' are higher than the level of $\alpha 2$ unspliced message and declines after the so-called cryptic poly(A) signal in the Alu sequence which is detected by primer pair 'Post-Alu'. These results are entirely consistent with previous studies (4) and indicate that much higher levels of transcript are detectable past the mutated poly(A) signal and extend up to the cryptic poly(A) site. Clearly the normal termination process of the $\alpha 2$ globin gene is disrupted by the poly(A) signal mutation.

DISCUSSION

Mapping the primary transcript of a gene has proved to be a technically difficult procedure. Although steady state mRNA can be readily characterised, unprotected transcripts are in general highly unstable and therefore hard to detect. To circumvent these difficulties nuclear run-off analysis (11), an *in vitro* procedure, is invariably employed which effectively maps the polymerase density over a gene and thus by inference the extent of the gene's primary transcripts. However it must always remain a possibility that this *in vitro* approach may not reflect the true physiological situation in a gene. Furthermore, using this method it is notoriously hard to obtain reproducible data, especially when analyzing single copy Pol II genes where levels of transcription are relatively low. We have therefore developed a more direct approach to measure the level and position of primary transcripts from the duplicated human α globin genes in steady state nuclear RNA. To do this we have employed a strategy based on reverse transcription and amplification by the polymerase chain reaction. We found that the most accessible way to evaluate the varying levels of transcripts detected by different primer pairs was to normalize them to the corresponding levels obtained by an exogenous synthetic message that was reverse-transcribed and amplified in parallel with the test RNA. It was not necessary to construct standard curves for each of the primer pairs in order to quantitate the level obtained from the test RNA as we were interested in determining the relative, rather than absolute, number of RNA molecules present at a given region. Furthermore, to study human α globin gene transcription we have recreated an *in vivo* system which is amenable to genetic manipulation by stably transfecting MELCs with α globin constructs that contain a region that has been shown to be important in regulating α globin gene expression.

These results have revealed a number of new features associated with human α globin gene transcription. We show that the 5' positioned $\alpha 2$ globin gene produces more transcript than the 3' $\alpha 1$ globin gene in a ratio of 63:37 and that approximately 3% of these transcripts initiate from upstream promoters. These results are in close agreement with previous studies (28, 29), which effectively validates our experimental approach. Our results then demonstrate that significant levels of unprocessed nuclear RNA exist in the 3' flanking region of the $\alpha 2$ globin gene at approximately 40% of the level of unspliced intron transcripts. This figure indicates that transcripts past the $\alpha 2$ globin gene's poly(A) site are readily detectable by PCR but are somewhat less stable than intron RNA. The principle result of these studies is to show that the 3' flanking region transcripts are reduced to low levels about 500 bp past the poly(A) signal, defining this part of the $\alpha 2$ globin gene as the 3' extremity of the primary transcript. We feel that this RT/PCR methodology provides a useful complementary approach to nuclear run-off analysis.

ACKNOWLEDGEMENTS

We are grateful to the other members of our laboratory for critical review of the manuscript. We thank the Wellcome Trust for financial support (program grant no. 03277/Z/90/Z). P. E.-H. was funded by an MRC Student Training Fellowship.

REFERENCES

- Manley, J.L. (1988). *Biochim. Biophys. Acta*, **950** 1–12.
- Proudfoot, N.J. and Whitelaw, E. (1988). In Hames, B.D. and Glover, D.M. (eds), *Transcription and Splicing* (Frontiers in Molecular Biology). Oxford University Press, Oxford, pp.97–129.
- Humphrey, T. and Proudfoot, N.J. (1988). *Trends. Genet.*, **4** 243–245.
- Whitelaw, E. and Proudfoot, N. (1986). *EMBO J.*, **5** 2915–2922.
- Connelly, S. and Manley, J.L. (1988). *Genes. Dev.*, **2** 440–452.
- Logan, J., Falck-Pederson, E., Darnell, J.E. Jr. and Shenk, T. (1987). *Proc. Natl. Acad. Sci. U.S.A.*, **84** 8306–8310.
- Proudfoot, N.J. (1989). *Trends Biochem Sci.*, **14** 105–110.
- Brennan, C.A., Dombroski, A.J. and Platt, T. (1987). *Cell*, **48** 945–952.
- Enriquez-Harris, P., Levitt, N., Briggs, D. and Proudfoot, N.J. (1991). *EMBO J.*, **10** 1833–1842.
- Vandenberg, D.J., James-Pederson, M. and Hardison, R.C. (1991). *J. Mol. Biol.*, **220** 255–270.
- Maa, M.-C., Chinsky, J., Ramamurthy, V., Martin, B.D., Kellems, R.E. (1990). *J. Biol. Chem.* **265** 12513–12519.
- Groudine, M., Peretz, M. and Weintraub, H. (1981). *Mol. Cell. Biol.*, **1** 281–288.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988). *Science*, **239** 487–491.
- Rappolee, D.A., Brenner, C.A., Schultz, R., Mark, D. and Werb, Z. (1988). *Science*, **241** 1823–1825.
- Singer-Sam, J., Robinson, M.O., Bellve, A.R., Simon, M.I. and Riggs, A.D. (1990). *Nucl. Acids Res.*, **18** 1255–1259.
- Murphy, L.D., Herzog, C.E., Rudick, J.B., Fojo, A.T. and Bates, S.E. (1990). *Biochemistry*, **29** 10351–10356.
- Robinson, M.O. and Simon, M.I. (1991). *Nucl. Acids Res.*, **19** 1557–1562.
- Higgs, D.R., Wood, W.G., Jarman, A.P., Sharpe, J., Lida, J., Pretorius, I.M. and Ayyub, H. (1990). *Genes. Dev.*, **4** 1588–1601.
- Jarman, A.P., Wood, W.G., Sharpe, J.A., Gourdon, G., Ayyub, H. and Higgs, D.R. (1991). *Mol. Cell. Biol.*, **11** 4679–4689.
- Lauer, J., Shen, C.-K. J. and Maniatis, T. (1980). *Cell*, **20** 119–130.
- Proudfoot, N.J. (1986). *Nature*, **322** 562–565.
- Liebhauer, S.A., Goossens, M. and Kan, Y.W. (1981). *Nature*, **290** 26–29.
- Southern, P.J. and Berg, P. (1982). *J. Mol. Appl. Genet.*, **1** 327–341.
- Norman, C., Raymond, V. and Proudfoot, N.J. (1987). In Stamatoyannopoulos, G. and Nienhuis, A.W. (eds) *Developmental control of globin gene expression*. Alan R. Liss Inc. New York. pp 235–252.
- Melton, D.A., Krieg, P.A., Rabagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984). *Nucl. Acids Res.*, **12** 7035–7056.
- Levitt, N., Briggs, D., Gil, A. and Proudfoot, N.J. (1989). *Genes Dev.*, **3** 1019–1025.
- Kwok, S. (1990). In Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (eds), *PCR Protocols: A guide to methods and applications*. Academic Press Inc. pp 142–145.
- Orkin, S.H. and Goff, S.C. (1981). *Cell*, **24** 345–351.
- Hess, J., Perez-Stable, C., Deisseroth, A. and Shen, C.-K.J. (1985). *Nucl. Acids Res.*, **13** 6059–6074.
- Higgs, D.R., Goodbourn, S.E.Y., Lamb, J., Clegg, J.B., Weatherall, D.J. and Proudfoot, N.J. (1983). *Nature*, **306** 398–400.