The EMBO Journal vol.4 no.2 pp.453–456, 1985

Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human β-globin gene

Stuart H.Orkin, Tu-Chen Cheng[1], Stylianos E.Antonarakis[1] and Haig H.Kazazian,Jr.[1]

Division of Hematology-Oncology, Childrens Hospital and the Dana Farber Cancer Institute, Department of Pediatrics, Harvard Medical School, Boston, MA 02115, and [1]Division of Pediatric Genetics, Johns Hopkins Hospital, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

Communicated by N.J.Proudfoot

A β-globin gene cloned from a person with β-thalassemia contained a T→C substitution within the conserved sequence AATAAA that forms a portion of the recognition signal for endonucleolytic cleavage and polyadenylation of primary mRNA transcripts. By Northern blot analysis a novel β-globin RNA species, 1500 nucleotides in length, was detected in erythroid RNA. Nuclease protection studies of erythroid RNA, as well as RNA generated upon transient expression of the cloned mutant gene in HeLa cells, located the 3′ terminus of this novel, polyadenylated RNA 900 nucleotides downstream of the normal poly(A) addition site, within 15 nucleotides of the first AATAAA in the 3′-flanking region of the β-globin gene. These findings define the *in vivo* terminus of an elongated RNA and establish that human β-globin transcription may extend at least 900 nucleotides 3′ of the normal polyadenylation site.

Key words: RNA processing/transcription termination/transient gene expression

Introduction

The majority of nonhistone mRNAs in higher eukaryotes are polyadenylated at their 3′ ends (Proudfoot, 1982). The highly conserved sequence AAUAAA found 10–30 nucleotides upstream of most polyadenylation sites in mRNA forms a portion of a recognition signal for endonucleolytic cleavage of primary transcripts (Fitzgerald and Shenk, 1981; Proudfoot, 1984). Recent evidence indicates that sequences located immediately downstream of the poly(A) addition site also contribute to the selection of cleavage sites (Simonsen and Levinson, 1983; McDevitt *et al.*, 1984; Woychik *et al.*, 1984). Polyadenylation of cleaved RNAs may involve small nuclear ribonucleoprotein particles, perhaps of the U4 snRNP class (Moore and Sharp, 1984; Berget, 1984).

When the DNA sequence corresponding to the AAUAAA signal, AATAAA is altered to AAGAAA or AATAAG in adenovirus (Montell *et al.*, 1983) and human α-globin genes (Higgs *et al.*, 1983), respectively, elongated RNA transcripts have been observed. These findings are consistent with transcription and subsequent polyadenylation at positions beyond the normal polyadenylation site. The manner in which transcription terminates 3′ to eukaryotic genes is uncertain. Although transcription proceeds downstream from the mouse β-globin and other genes (Nevins and Darnell, 1978; Nevins *et al.*, 1980; Hofer and Darnell, 1981; Hofer *et al.*, 1982; Salditt-Georgieff and Darnell, 1983, 1984), discrete termination sites have yet to be defined. Rather transcription appears to diminish over a stretch of DNA >500 bp downstream from the poly(A) addition site (Hofer and Darnell, 1981; Salditt-Georgieff and Darnell, 1983, 1984; Rohrbaug *et al.*, 1984).

Here we describe a human β-globin gene isolated from a patient with β-thalassemia in which a T→C substitution within the AATAAA sequence was observed. In stable erythroid RNA we observed β-globin RNA which was polyadenylated 900 bp downstream from the normal position and just beyond the first AAUAAA in the flanking sequence. These data locate the *in vivo* 3′ terminus of an elongated RNA and demonstrate that a fraction of primary β-RNA transcripts may extend at least 900 bp 3′ of the normal polyadenylation site.

Results

Identification of a novel β-thalassemia gene

As part of a comprehensive study of β-thalassemia in black Americans (Antonarakis *et al.*, 1984) we identified a family in which an abnormally large β-globin RNA (~1500 nucleotides in length) was present in Northern blot analysis of erythroid RNA (Figure 1). Three family members had this novel RNA species in association with a β-thalassemia gene within a chromosome of polymorphism haplotype 2 (Antonarakis *et al.*, 1984; Orkin and Kazazian, 1984). Approximately 25% of β-thalassemia genes among black Americans are found in this haplotype (Antonarakis *et al.*, 1984). In individuals heterozygous for both the sickle β-globin gene (β^S) and the β-thalassemia gene described here, Hb A was 22–24%, Hb S

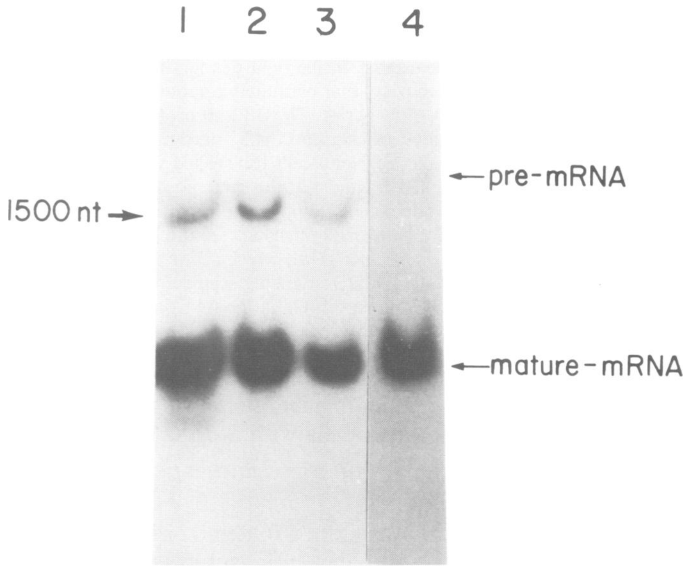

Fig. 1. Northern blot analysis of erythroid RNA. **Lanes: 1, 2, 3** = β^{thal}, β^S heterozygotes. **4** = normal individual. The precursor for β-RNA (pre-mRNA) is ~1600 nucleotides in length.

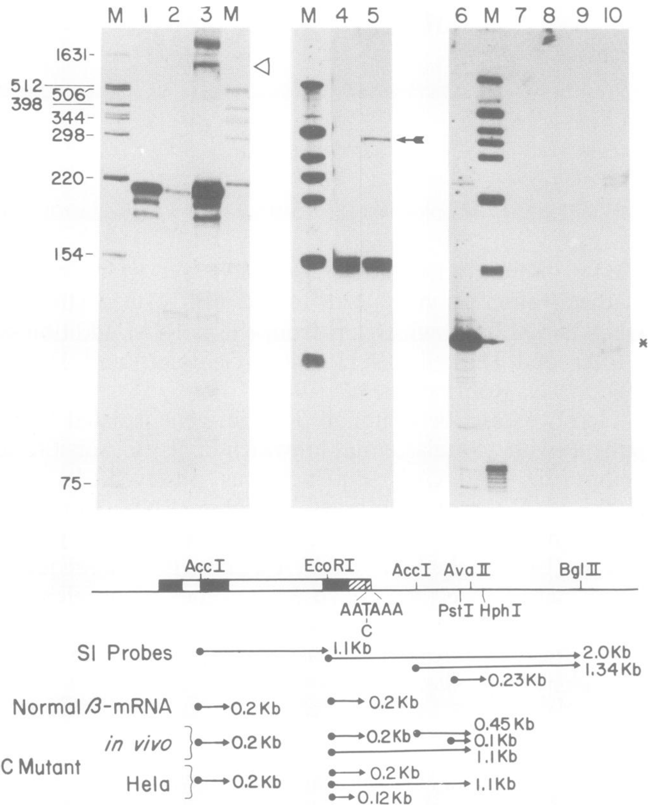

Fig. 2. S1 nuclease mapping of normal and mutant RNA transcripts. S1 nuclease mapping probes were 3′ end-labeled at the indicated restriction sites below the map of the β-gene and its downstream region. Probes used: left panel, 2.0-kb *Eco*RI*-*Bgl*II fragment. Center, mixed probe consisting of 1.1-kb *Acc*I*-*Eco*RI and 1.34-kb *Acc*I*-*Bgl*II. Right, 0.23-kb *Pst*I-*Hph*I. The asterisk denotes the site of the 3′ end label. **Lanes: 1** = RNA prepared from HeLa cells transfected with the normal β-globin gene. **2** = RNA prepared from HeLa cells transfected with the mutant gene. **3,5,10** = erythroid RNA from patients with novel elongated β-RNA. **4,8,9** = normal bone marrow RNA. **7** = no RNA added control. **6** = 3′ end-labeled *Pst*I*-*Hph*I S1 probe digested with *Ava*II. **M** = pBR322 DNA digested with *Hin*fI. The sizes of the markers are provided only for the left panel. The lowest band in the other marker lanes is 75 nucleotides. Comparison of **lanes 6** and **10** demonstrates that the 3′ end of the elongated RNA is located ~3–5 nucleotides upstream of the 3′-flanking *Ava*II site. The various fragments protected in normal and mutant RNA samples are shown schematically below. The open arrow in the left panel, the closed arrow in the center, and the asterisk in the right denote the 1.1-kb, 0.45-kb and 0.1-kb fragments, respectively, that permitted accurate mapping of the 3′ end of the elongated RNA transcript. The 0.12-kb protected fragment in **lane 2** derived from alternative splicing into the 3′-untranslated region was routinely seen in RNA samples of HeLa cells transfected with the mutant gene. Although some protected fragment at the same position is evident in **lane 3**, it was not seen reproducibly and appears to represent minor nicked input probe (as exemplified by other minor bands seen in **lanes 1** and **3**). Use of an additional S1 probe *Eco*RI*-*Pst*I, which spans the 3′-untranslated region and the initial 3′-flanking segment of the gene, led to protection of the 0.12-kb fragment in HeLa RNAs from cells transfected with the mutant gene, but not from *in vivo* erythroid RNA (not shown).

70–74% and Hb F <1% with a total hemoglobin of ~10 g/100 ml. These patients have a clinical course similar to that of other black Americans with S-β-thalassemia.

A mutation (T→C) lies within the cleavage-polyadenylation signal (AATAAA)

The β-thalassemia gene was cloned (Orkin *et al.*, 1982) from an individual with sickle-β-thalassemia, subjected to DNA sequence analysis (Maxam and Gilbert, 1980), and expressed transiently in HeLa cells (Treisman *et al.*, 1982, 1983). Within the DNA sequence coding for the conserved 3′-non-coding sequence AAUAAA in mRNA, a T→C substitution was found. The DNA sequences of the 5′-flanking region to position −300 and of all splice junctions were normal. As is typical of β-genes of the framework 1 variety (Orkin *et al.*, 1982; Antonarakis *et al.*, 1984; Orkin and Kazazian, 1984), no DNA sequence polymorphisms were observed within the gene.

RNA cleavage occurs 900 nucleotides downstream

Upon transient expression in cultured HeLa cells this gene directed about one-fifth to one-tenth as much stable β-RNA as a normal β-gene (Figure 2, lanes 1 and 2). S1 nuclease mapping of RNAs generated in the transient expression assay and of *in vivo* erythroid RNA was performed. Using an end-labeled probe extending from the exon-3 *Eco*RI site 2 kb in the 3′ direction, we detected three protected fragments (1.1, 0.2, and 0.12 kb) upon hybridization with RNA from HeLa cells transfected with the mutant gene (Figure 2, lane 2) and two prominent fragments (1.1 and 0.2 kb) with *in vivo* RNA (Figure 2, lane 3). Fractionation of RNA samples on oligo(dT)-cellulose demonstrated protection only with the poly(A)$^+$ fraction (not shown). In addition to the normal protected fragment of 0.2 kb (lane 1), therefore, a novel 1.1-kb nucleotide species was evident in both HeLa and *in vivo* RNAs (lanes 2 and 3). This species, which was absent in normal blood and marrow RNA samples, could be generated either by an RNA cleavage or splicing event ~900 bp downstream from the β-globin gene. The 0.12-kb protected fragment, seen clearly and reproducibly only in HeLa cells transfected with the mutant gene (lane 2), apparently arises from a splicing event within the 3′-untranslated region at a cryptic donor-like sequence (AAGGTTCCT, nucleotides 34–42 of the untranslated region).

To delineate the nature of the RNA leading to the novel 1.1-kb protected fragment, additional S1 nuclease mapping of the *in vivo* RNA was performed with various probes (Figure 2, lanes 4–10). Protected fragments were consistent only with either an RNA cleavage or splicing event 900 bp downstream from the normal polyadenylation site. Given the size of normal β-globin RNA (~600 nucleotides), the extent of the 3′ extension (900 nucleotides), and the size of the novel RNA detected by Northern blot analysis (1500 nucleotides), we conclude that this position marks the 3′ terminus of the RNA. Using S1 nuclease probes located further downstream (up to 1.8 kb), no protected fragments were observed upon hybridization to erythroid RNA of one of the patients (not shown). Due to the low abundance of β-globin RNA in HeLa cells transfected with the mutant gene, we have not been successful in defining the acceptor sequence spliced to the donor-like sequence in the 3′-untranslated region during transient expression.

The S1 nuclease mapping experiments indicate that the site of RNA cleavage and polyadenylation downstream from the mutant gene lies within five nucleotides 5′ to an *Ava*II site [GG(A or T)CC] (Figure 2, lanes 6 and 10) situated 900 nucleotides downstream of the normal poly(A) addition site. The DNA sequence of this region is AATAAAATAT GAGTCTCAAG TGGTCCTTGT (Poncz *et al.*, 1983). We infer from these findings that cleavage and subsequent polyadenylation occurred within 10–15 nucleotides downstream of the underlined AATAAA sequence, the first encountered in the 3′-flanking region (Poncz *et al.*, 1983). The patterns of RNA cleavage and polyadenylation deduced from the nuclease protection studies are depicted in Figure 3.

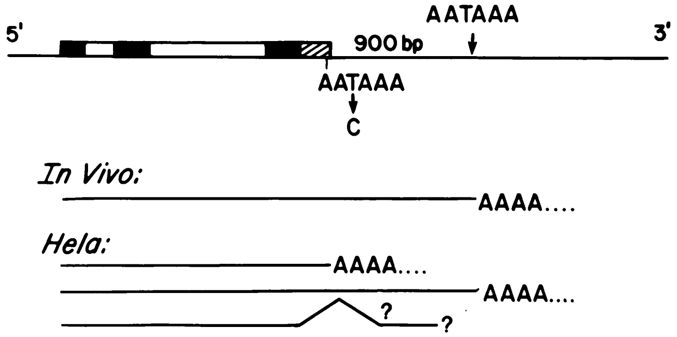

Fig. 3. Schematic representation of the RNAs detected by S1 nuclease mapping. The mutant gene contained the T→C substitution indicated in the 3'-untranslated region of the β-globin gene. *In vivo* erythroid RNA contained a species with an additional 900 nucleotides at the 3' end. This RNA was polyadenylated and had its 3' terminus just beyond the first AATAAA in the downstream sequence. In HeLa cells transfected with the mutant gene this novel RNA was found at a low level in addition to a small amount of normally cleaved and polyadenylated RNA and another species in which splicing occurred at a cryptic donor sequence within the untranslated region (bottom pattern). The downstream acceptor was not identified. Transcripts from the mutant gene within erythroid cells may have contained some normally cleaved and polyadenylated RNA but this could not be assessed due to the presence of a β^S-gene in *trans* in the patient.

Discussion

The formation of 3' ends of eukaryotic polymerase II-generated mRNAs involves endonucleolytic cleavage and polyadenylation (Montell *et al.*, 1983; Proudfoot, 1984). What role, if any, specific transcription termination signals play in this process is uncertain. Considerable previous work has demonstrated that the transcriptional unit of polymerase II-transcribed genes may extend considerably past the poly(A) addition site, perhaps as far as 1000–2500 nucleotides (Hofer and Darnell, 1981; Salditt-Georgieff and Darnell, 1983, 1984; Rohrbaug *et al.*, 1984). Specifically, data on the mouse β^{maj} and rabbit $\beta 1$ genes indicates that high level transcription may proceed for ~600 nucleotides beyond the poly(A) site (Salditt-Georgieff and Darnell, 1983; Rohrbaug *et al.*, 1984). Further downstream transcription appears to decrease over a region of ~1000 additional nucleotides. It has recently been suggested that an inverted repeat located ~600 nucleotides beyond the poly(A) addition site of the rabbit $\beta 1$ gene may be involved in this attenuation of transcription (Rohrbaug *et al.*, 1984). The sequences of the 3'-flanking regions of the human β- and rabbit $\beta 1$-globin genes share considerable homology in this region (Rohrbaug *et al.*, 1984).

Our results with the β-thalassemia gene carrying a naturally occurring mutation (AATAAA→AACAAA) within the polyadenylation-cleavage signal support and extend these general findings. First, we infer from the transient expression of this gene that cleavage and polyadenylation at the normal poly(A) addition site is markedly inefficient but not totally extinguished (Figure 2, lane 2) due to the single base substitution within the conserved sequence AATAAA. A T→G substitution in the same position of an adenovirus gene (Montell *et al.*, 1983) had a similar effect on the efficiency of formation of wild-type 3' ends. Second, our findings demonstrate that transcription may proceed at least 900 nucleotides beyond the poly(A) addition site of the human β-globin gene when normal cleavage and polyadenylation is largely prevented (Figures 2 and 3). Since the level of the extended β-mRNA from the mutant allele in erythroid cells is lower than that of mature β-mRNA from the β^S allele (Figure 1), we cannot estimate the actual efficiency of transcription through the 900 nucleotide segment downstream from the normal poly(A) addition site. Several additional factors may contribute to a reduced level of this species. For one, some extended mRNAs may be processed aberrantly using the cryptic donor site detected in the transient expression assay (Figure 2, lane 2 and Figure 3). In addition, the cleavage-polyadenylation signal that is utilized 900 nucleotides downstream may function inefficiently and lead to the further loss of transcripts, especially since recent evidence favors the role of sequences 3' to the poly(A) addition site in determining the efficiency with which an AATAAA sequence is utilized (Simonsen and Levinson, 1983; McDevitt *et al.*, 1984; Woychik *et al.*, 1984). Although we cannot accurately assess the contributions of these various factors, we can confidently conclude that transcription may proceed at least a fraction of the time through the 900 nucleotides 3' to the β-globin gene and that no absolute transcription termination signals appear to exist in this vicinity. Third, our results establish the *in vivo* 3' terminus of an extended mRNA just beyond the first AATAAA sequence downstream from the β-gene. In a previously reported instance of α-thalassemia in which AATAAA was altered to AATAAG, elongated transcripts were seen in transient assays but not *in vivo*. We surmise that the extended β-mRNA is more stable within erythroid cells than the extended α-mRNA. In part this may relate to the proximity of the 'cryptic' cleavage-polyadenylation signals to the respective normal poly(A) addition sites of these genes. In the study of the α-thalassemia gene, transient assay was performed using a construct with a short downstream segment (Higgs *et al.*, 1983). Therefore, the potential site(s) of cleavage and poly(A) addition could not be assessed.

The 'cryptic' polyadenylation signal 900 bp downstream from the β-globin gene appears, in this instance, to be the predominant site utilized in generation of stable erythroid cell transcripts. RNAs from erythroid and transfected HeLa cells did not protect DNA probes situated further downstream (not shown). Either few transcripts escape cleavage and polyadenylation at this AATAAA signal or those that do cannot be appropriately cleaved at the numerous AATAAA sequences (Poncz *et al.*, 1983) located in the ensuing 1000 nucleotides.

The phenotype of individuals carrying this β-thalassemia gene is somewhat milder than anticipated from the above results and the findings of others. Specifically, ~20–25% of the total hemoglobin is normal (HbA, $\alpha_2\beta_2$) in individuals heterozygous for both β-thalassemia and sickle cell anemia. In view of the inefficient cleavage and polyadenylation at the normal poly(A) addition site upon transient expression of the mutant gene in HeLa cells (Figure 2, lane 2), less normal hemoglobin might be anticipated. As no homozygotes for this mutation have been identified, the percent of transcripts actually cleaved and polyadenylated at the normal position *in vivo* is unknown. It is possible that cleavage and polyadenylation at the normal position within erythroid cells may exceed that evident in the HeLa cell transient expression assay. Whether the 1500 nucleotide β-RNA seen in the patients is translated has not been assessed. However, the appreciable level of HbA in individuals heterozygous for this form of β-thalassemia and sickle cell anemia suggests that it is likely to contribute substantially to globin production.

Materials and methods

Haplotype analysis, gene cloning and DNA sequencing

Determination of DNA polymorphisms within the β-globin gene cluster in the propositus and family members was performed as described previously (Orkin *et al.*, 1982; Antonarakis *et al.*, 1984; Orkin and Kazazian, 1984). In this family the chromosomes bearing the normal (β^A), sickle (β^S) and thalassemic β-globin genes could be distinguished by their patterns of polymorphic restriction sites (haplotypes). The β-thalassemia gene was present on a chromosome of haplotype 2 in blacks as defined in Orkin and Kazazian (1984) and Antonarakis *et al.* (1984). The β-thalassemia gene of an individual heterozygous for both β-thalassemia and sickle cell anemia was cloned as a 7.5-kb *Hind*III fragment in bacteriophage Charon 28 as described (Orkin *et al.*, 1982).

Transient gene expression and S1 nuclease mapping

The mutant β-globin gene was subcloned as a 5.2-kb *Bgl*II fragment in the expression vector πSVplac (Triesman *et al.*, 1983) and introduced into HeLa cells by calcium phosphate precipitation. This construct contains ~2 kb of downstream flanking sequences. The normal β-globin gene was introduced into HeLa cells as a 3.7-kb *Bgl*II/*Pst*I fragment in this vector as before (Treisman *et al.* 1983). 48 h after transfection, HeLa cells were harvested and total cellular RNA was isolated (Antonarakis *et al.*, 1984). S1 nuclease mapping probes were prepared as described (Favaloro *et al.*, 1980), labeled singly at either the 5′ or 3′ end (Maxam and Gilbert, 1980). 3′-Flanking restriction sites were derived from the sequence reported by Poncz *et al.* (1983). Probes were hybridized to RNAs as described (Favaloro *et al.*, 1980) and then digested with S1 nuclease (Sigma) at 1000 units/ml for 45 min at 37°C. Protected DNA fragments were electrophoresed in urea-acrylamide gels (Maxam and Gilbert, 1980).

Northern blot analysis

Peripheral blood RNA samples were prepared as described previously (Antonarakis *et al.*, 1984), formaldehyde-treated, electrophoresed and then transferred to nitrocellulose (Maniatis *et al.*, 1982). The filter was hybridized with a 0.9-kb *Bam*HI-*Eco*RI fragment of the β-globin gene (Antonarakis *et al.*, 1984).

Acknowledgements

The skilled technical assistance of Julianne P.Sexton was greatly appreciated. We thank Ross Hardison for the opportunity to see recent transcriptional data on the rabbit β1-globin gene prior to publication (Rohrbaug *et al.*, 1984). This work was supported by grants from the National Institutes of Health and the National Foundation-March of Dimes.

References

Antonarakis,S.E., Orkin,S.H., Cheng,T.C., Scott,A.F., Sexton,J.P., Trusko, S.P., Charache,S. and Kazazian,H.H.,Jr. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1154-1158.

Berget,S.M. (1984) *Nature*, **309**, 179-182.

Favaloro,J., Treisman,R. and Kamen,R. (1980) *Methods Enzymol.*, **65**, 718-749.

Fitzgerald,M. and Shenk,T. (1981) *Cell*, **24**, 251-260.

Higgs,D.R., Goodbourn,S.E.Y., Lamb,J., Clegg,J.B., Weatherall,D.J. and Proudfoot,N.J. (1983) *Nature*, **306**, 398-400.

Hofer,E. and Darnell,J.E. (1981) *Cell*, **23**, 585-593.

Hofer,E., Hofer-Warbinek,R. and Darnell,J.E. (1982) *Cell*, **29**, 887-893.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.

Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499.

McDevitt,M.A., Imperiale,M.J., Ali,H. and Nevins,J.R. (1984) *Cell*, **37**, 993-999.

Montell,C., Fisher,E.F., Caruthers,M.H. and Berk,A.J. (1983) *Nature*, **305**, 600-605.

Moore,C.L. and Sharp,P.A. (1984) *Cell*, **36**, 581-591.

Nevins,J.R. and Darnell,J.E. (1978) *Cell*, **15**, 1477-1493.

Nevins,J.R., Blanchard,J.M. and Darnell,J.E. (1980) *J. Mol. Biol.*, **144**, 377-386.

Orkin,S.H. and Kazazian,H.H.,Jr. (1984) *Annu. Rev. Genet.*, **18**, 131-171.

Orkin,S.H., Kazazian,H.H.,Jr., Antonarakis,S.E., Goff,S.C., Boehm,C.D., Sexton,J.P., Waber,P.G. and Giardina,P.J.V. (1982) *Nature*, **296**, 627-631.

Poncz,M., Schwartz,E., Ballantine,M. and Surrey,S. (1983) *J. Biol. Chem.*, **258**, 11599-11609.

Proudfoot,N.J. (1982) *Nature*, **298**, 516.

Proudfoot,N.J. (1984) *Nature*, **307**, 412-213.

Rohrbaug,M.L., Jonson,J.E.,III, James,M.D. and Hardison,R.C. (1984) *Mol. Cell. Biol.*, in press.

Salditt-Georgieff,M. and Darnell,J.E. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4694-4698.

Salditt-Georgieff,M. and Darnell,J.E. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2274.

Simonsen,C.C. and Levinson,A.D. (1983) *Mol. Cell. Biol.*, **3**, 2250-2258.

Treisman,R., Proudfoot,N.J., Shander,M. and Maniatis,T. (1982) *Cell*, **29**, 903-911.

Treisman,R., Orkin,S.H. and Maniatis,T. (1983) *Nature*, **302**, 591-596.

Woychik,R.P., Lyons,R.H., Post,L. and Rothman,F.M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3944-3948.

Received on 25 October 1984; revised on 4 December 1984