25. Komar, A. *J. Reprod. Fert.* **35,** 433–443 (1973).
26. Dyban, A. P. & Khozhai, L. I. *Bull. exp. Biol. Med.* **89,** 528–530 (1980).
27. Kaufman, M. H. *Nature* **302,** 258–260 (1983).
28. Whittingham, D. G. *J. Reprod. Fert.* Suppl. **14,** 7–22 (1971).
29. Whitten, W. K. in *Advances in the Biosciences* Vol. 6, (ed. Raspe, G.) 129–139 (Pergamon, Oxford, 1971).
30. O'Farrell, P. H. *J. biol. Chem.* **250,** 4007–4021 (1975).
31. Ansorge, W. *Analyt. Biochem.* (in the press).

Regulated expression of the human β-globin gene family in murine erythroleukaemia cells

S. Wright, E. deBoer, F. G. Grosveld* & R. A. Flavell†

Laboratory of Gene Structure and Expression, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

Chemically induced differentiation of cultured murine erythroleukaemia (MEL) cells results in a several hundred-fold increase in transcription of the adult mouse globin genes[1–3] and thus serves as a model for gene activation during erythropoiesis. One approach to study gene regulation in this system has been to analyse the expression of foreign globin genes introduced into MEL cells[4–8]. By introducing cosmid DNA containing the human adult(β), fetal(γ) and embryonic(ε)-globin genes, we have shown here that expression of the β, but not the γ or ε genes, is regulated during MEL differentiation. Regulated expression of the human β-globin gene was observed when it was introduced either as part of the intact globin gene cluster or as an individual gene with 1.5 kilobases (kb) of 5′ flanking DNA. Transcription from a herpes simplex virus (HSV) promoter adjacent to the thymidine kinase (*tk*) gene is also inducible in MEL cells.

Cosmids containing regions of the human β-globin gene cluster (see Fig. 1) were introduced into thymidine kinase-negative (tk^-) MEL cells[6] by calcium phosphate transformation[9]. Before transformation, DNA was linearized within the vector sequences by cleavage with *Pvu*I. Stable tk^+ transformants were obtained at a frequency of 1–10 clones per 10^6 cells per μg DNA. Transformants were shown by Southern blot analysis[10] to contain 1–15 copies of non-rearranged introduced DNA per cell (data not shown); 10–15 transformants containing each cosmid were induced into erythroid differentiation by culturing for 3 days in the presence of 3 mM hexamethylene bisacetamide (HMBA). Levels of human and mouse globin transcripts were quantitated by S_1 nuclease analysis[11,12] using probes which mapped the 5′ and 3′ ends of the mRNAs.

Endogenous mouse β-major globin mRNA levels in transformants as measured by S_1 nuclease analysis of the 3′ end of the mRNA are shown in Table 1 and Fig. 2*a*; 52 of 54 clones tested showed a >100-fold increase in mouse β-globin mRNA levels on differentiation, corresponding to >10,000 copies of mRNA per induced cell. Six of the transformants which showed induction had a high level of mouse β-major globin mRNA before HMBA treatment (>1,000 copies per cell).

Of 40 MEL transformants that contained a foreign human β-globin gene, 30 showed a 4–100-fold increase in the level of human β-globin mRNA on differentiation as measured by S_1 nuclease analysis of the 3′ end of the mRNA (Fig. 2*b*, *c*; Table 1). Of the 10 transformants that showed no induction, two were non-inducible for endogenous mouse β-globin mRNA (one CosHG6N and one β-pRT) and four showed a high level of mouse globin mRNA before addition of the inducing agent (two CosHG6N and two CosHG28TK). Correct initiation of the human globin transcripts was shown by S_1 nuclease mapping of the 5′ end (Fig. 2*d*). Induced human β-globin mRNA levels corresponded to 120–2,500 copies per cell, and induction levels did not correlate with the copy number of human β-globin genes present in the transformant.

Of 19 transformants containing the complete human γ-globin genes (CosHG25 and CosHG28TK), 17 showed the same or a smaller amount of γ-globin mRNA after differentiation (~50 copies per cell; Fig. 2*e*, Table 1). Two transformants showed a less than sixfold inducibility of γ-globin mRNA. Similarly, transformants which contained the complete ε-globin gene (CosHG25) showed the same or reduced levels of ε-globin mRNA after differentiation (data not shown).

CosHG29TK and CosHGΔ17 contain the 3′ ends of the A_γ- and ε-globin genes adjoined to vector sequences as shown in Figs 1 and 3. Surprisingly, transformants containing these cosmids showed a 10–30-fold inducibility of RNAs which were detected by S_1 nuclease analysis using probes which hybridized to the 3′ ends of A_γ- and ε-globin mRNAs. Northern blot analysis[13] showed that the inducible RNA was 700 base pairs (bp) long and that it hybridized with both vector and globin DNA probes (not shown). If we assume the only RNA splicing in this vector–globin hybrid mRNA to be removal of the globin large intervening sequence, the observed RNA length places the transcription initiation site close to that of the HSV *tk* gene[14], but on the opposite DNA strand. This was confirmed

* To whom correspondence should be addressed.
† Present address: Biogen Research Corporation, 14 Cambridge Center, Cambridge, Massachusetts 02142, USA.

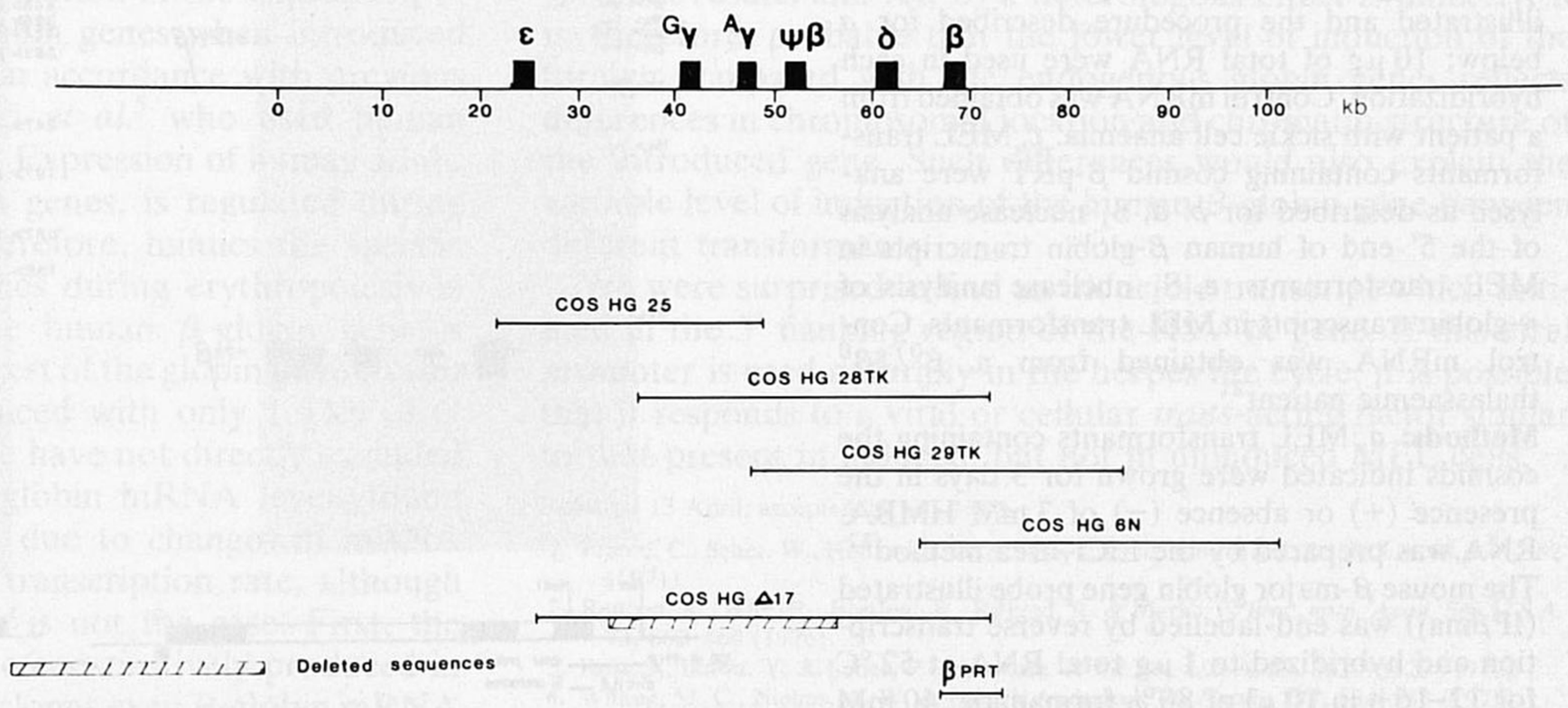

Fig. 1 Structure of human globin gene cosmids introduced into MEL cells. Regions of the human globin cluster as indicated were cloned in cosmid vectors containing the HSV *tk* gene[17,18] (vector sequences not shown). CosHG25 and CosHG6N were isolated directly from a tk vector library. CosHG28TK and CosHG29TK were constructed from previously isolated cosmids, CosHG28 and CosHG29 (ref. 18) by the exchange of vector sequences containing the *tk* gene. CosHGΔ17 was constructed by packaging and transduction of a ligation mixture containing a *Cla*I–*Bam*HI fragment from cosmid pRT, a *Bam*HI–*Xma*I fragment containing the 3′ side of the ε-gene, and *Xma*I–*Pvu*I fragment containing the Gγ-, Aγ-, δ- and β-globin genes and part of the cosmid vector pJB8 (ref. 19). Ligation yielded a DNA concatemer which deleted during the *in vitro* packaging to give deletion cosmids including CosHGΔ17. β-pRT is a subclone of the 4.7-kb β-globin *Bga*I fragment inserted in the *Bam*HI site of vector pRT; the *tk* and β-globin genes in this construction are transcribed in the same direction. Cosmid DNA was prepared by the method of Birnboim and Doly[20] followed by purification on CsCl gradients. *Pvu*I-linearized cosmid DNA was introduced into tk^- MEL cells[6] by calcium phosphate transformation[8] and selection in hypoxanthine–aminopterin–thymidine medium.

Table 1 Induction of human and mouse globin mRNAs in cosmid-containing MEL transformants

Cosmid	Mouse β-major globin mRNA			Human γ-globin mRNA			Human β-globin mRNA		
	No. of clones tested	No. showing induction	No. with high non-induced levels	No. of clones tested	No. showing induction	No. showing no induction	No. of clones tested	No. showing induction	No. showing no induction
CosHG25	9	9	2	9	0	9	—	—	—
CosHGΔ17	8	8	0	—	—	—	8	6	2
CosHG28TK	10	10	2	10	2	8	10	7	3
CosHG29TK	9	9	0	9*	6*	3*	4	4	0
CosHG6N	8	7	2	—	—	—	8	5	3
β-pRT	10	9	0	—	—	—	10	8	2
Total	54	52	6	19	2	17	40	30	10

* Transcription of the γ-globin gene in MEL clones containing CosHG29TK initiates at a vector promoter (see text) and is not included in the totals column.

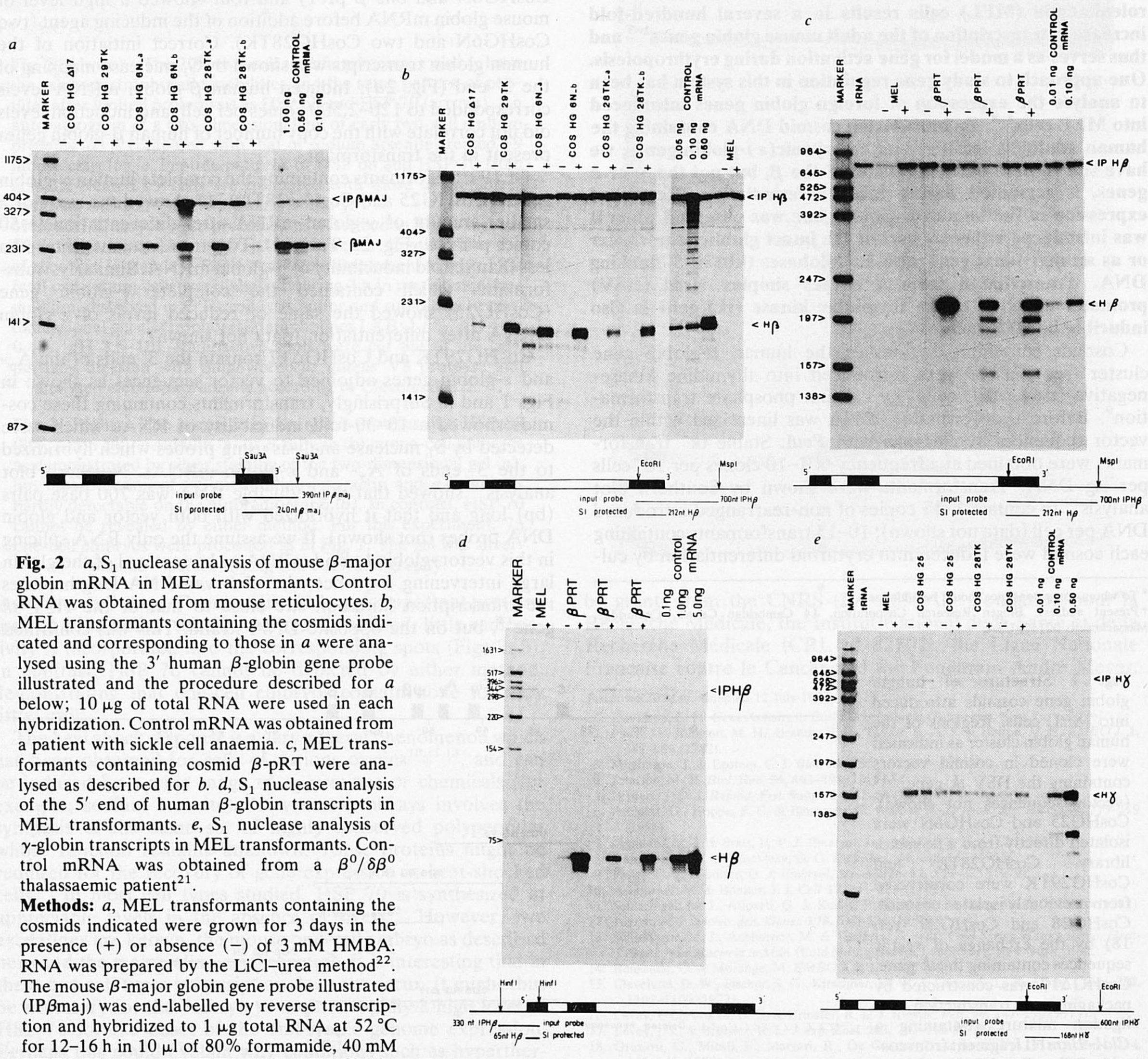

Fig. 2 *a*, S_1 nuclease analysis of mouse β-major globin mRNA in MEL transformants. Control RNA was obtained from mouse reticulocytes. *b*, MEL transformants containing the cosmids indicated and corresponding to those in *a* were analysed using the 3′ human β-globin gene probe illustrated and the procedure described for *a* below; 10 μg of total RNA were used in each hybridization. Control mRNA was obtained from a patient with sickle cell anaemia. *c*, MEL transformants containing cosmid β-pRT were analysed as described for *b*. *d*, S_1 nuclease analysis of the 5′ end of human β-globin transcripts in MEL transformants. *e*, S_1 nuclease analysis of γ-globin transcripts in MEL transformants. Control mRNA was obtained from a $\beta^0/\delta\beta^0$ thalassaemic patient[21].

Methods: *a*, MEL transformants containing the cosmids indicated were grown for 3 days in the presence (+) or absence (−) of 3 mM HMBA. RNA was prepared by the LiCl–urea method[22]. The mouse β-major globin gene probe illustrated (IPβmaj) was end-labelled by reverse transcription and hybridized to 1 μg total RNA at 52 °C for 12–16 h in 10 μl of 80% formamide, 40 mM PIPES *p*H 6.4, 1 mM EDTA, 400 mM NaCl. The mixture was digested with 3,000 units of S_1 nuclease (Boehringer) in 300 μl of 300 mM NaAc *p*H 4.8, 200 mM NaCl, 2 mM $ZnSO_4$ for 2 h at 20 °C. S_1-protected DNA was ethanol-precipitated and electrophoresed on a 7 M urea, 7% acrylamide gel. *d*, MEL transformants containing cosmid β-pRT and corresponding to those used in *c* were analysed. The procedure was as for *a* except that the 5′ human β-globin probe was end-labelled by kinase, strand-separated, and hybridized to 120 μg total RNA in a 50 μl hybridization volume. *e*, RNA from MEL transformants containing the cosmids indicated was analysed using the γ-globin gene probe illustrated and the procedure described for *a*. 10 μg of total RNA were used in each hybridization.

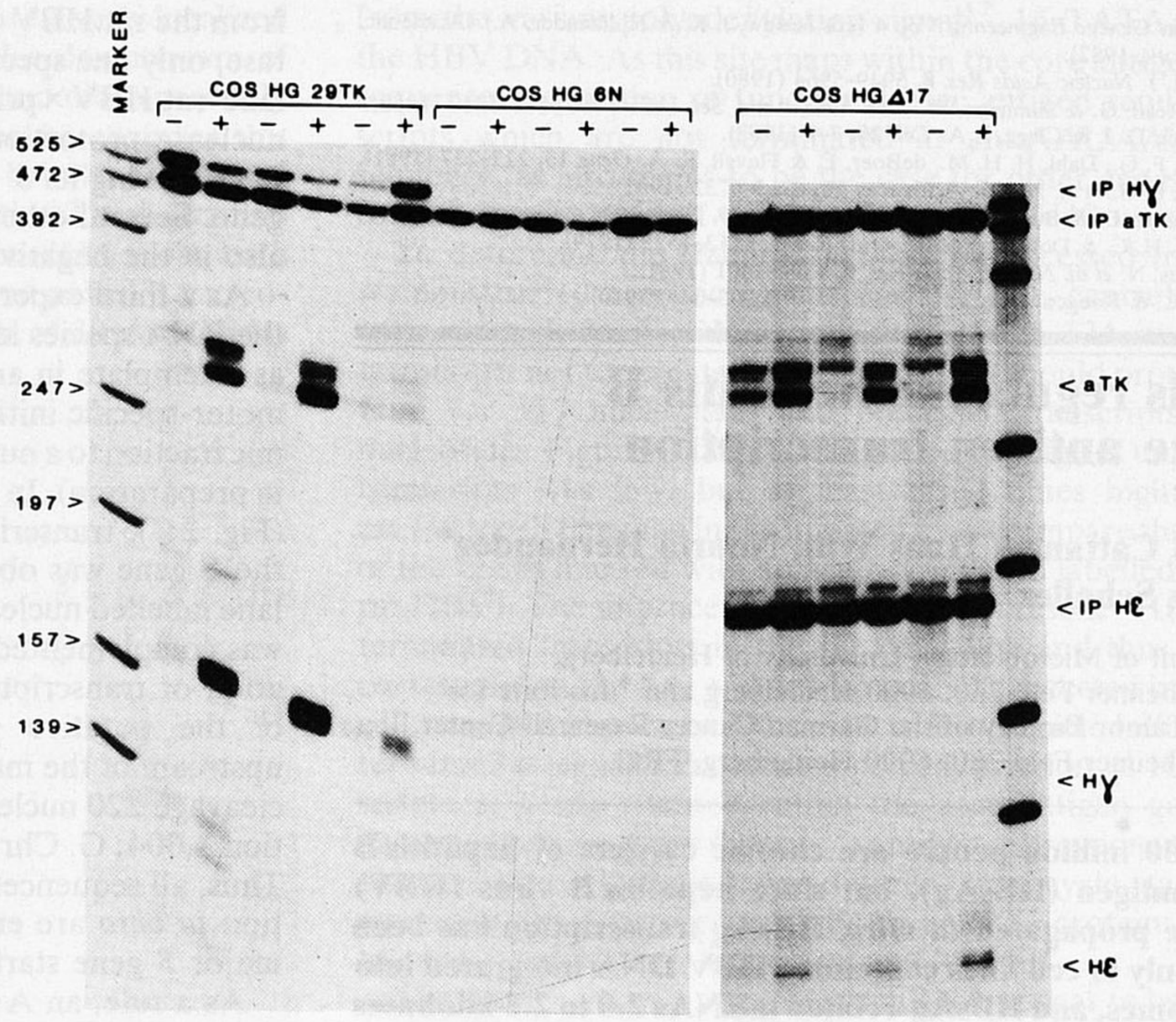

Fig. 3 S_1 nuclease analysis of anti-*tk* transcripts in MEL transformants. MEL transformants containing cosHG29TK, cosHG6N and cosHGΔ17 were grown for 3 days in the presence (+) or absence (−) of 3 mM HMBA. CosHG29TK and cosHGΔ17 contain the 3′ ends of the Aγ- and ε-globin genes adjoined to vector sequences near the HSV *tk* gene as shown. The 5′ ends of *a-tk* transcripts initiating near the *tk* gene were mapped with a 510 bp *Xma*I–*Bgl*II probe (IP aTK) using the procedure described for Fig. 2*a*. This probe covers the expected 5′ end of the *a-tk* transcript and gives an S_1 nuclease-protected fragment of 280–300 nucleotides (aTK). RNA from cosHG29TK and cosHGΔ17 transformants was simultaneously hybridized to 3′ γ-globin and 3′ ε-globin gene probes, respectively.

by S_1 nuclease mapping using a 510-bp *Xma*I–*Bgl*II probe which covers the predicted 5′ end and gives a 280–300-bp protected band. The 5′ end of this 'anti-tk' (*a-tk*) transcript thus maps ~160 bp away from the *tk* gene cap site (see Fig. 3). A transcript initiating at this site has been previously reported from *in vitro* transcription of HSV DNA fragments[15]. ATA and CAAT box sequences required for transcription initiation are located at the appropriate distances 5′ to the *a-tk* transcript. When this promoter is not linked to known coding sequences (in CosHG25, CosHG28TK and CosHG6N), no transcripts originate from it (Fig. 3, centre panel), suggesting that either an RNA splice or transcription terminator are required to yield a stable transcript.

We have shown independent regulation in the expression of individual human ε-, γ- and β-globin genes when introduced into MEL cells as a gene cluster, in accordance with previous work by Willing *et al.*[4] and Pyati *et al.*[5] who used human chromosome 11–MEL cell hybrids. Expression of human adult, but not fetal or embryonic globin genes, is regulated during MEL cell differentiation. This, therefore, mimics the specific activation of adult-type globin genes during erythropoiesis in man. Regulated expression of the human β-globin gene is independent of the presence of the rest of the globin gene cluster and is observed when it is introduced with only 1.5 kb of 5′ flanking DNA (βpRT; Fig. 2*c*). We have not directly excluded the possibility that the increased globin mRNA levels found after MEL cell differentiation are due to changes in mRNA stability as opposed to changes in transcription rate, although strong arguments suggest that this is not the case. First, the human ε- and γ-globin mRNAs are constitutively produced in MEL cells and in some exceptional clones even β-globin mRNA is constitutive. This suggests that ε- and γ-globin mRNAs are stable in uninduced MEL cells and therefore makes it unlikely that β-globin mRNA is unstable. Note that γ- and β-globin mRNAs stably coexist in human erythroid cells, for example, in heterozygotes for hereditary persistence of fetal haemoglobin[16]. Moreover, the 3′ regions of the same ε- and γ-globin mRNAs become inducible when linked to the HSV promoter, which suggests that the induction is not related to the stability of the mRNA.

The level of induction of the introduced gene is about one order of magnitude lower than that observed for the endogenous mouse β-major globin gene. This lower level probably cannot be explained by the fact that the human β-globin gene and mouse MEL cells represent a heterologous system; Chao *et al.*[8] have recently shown that the transcription of a mouse–human hybrid gene is also inducible in this system. As the level of hybrid mouse–human mRNA derived from the mouse promoter is about the same as that derived from the human promoter (present results and ref. 8) a heterologous effect is unlikely. It is, therefore, probable that the lower level of induction of the foreign compared with the endogenous globin genes reflects differences in chromosomal location and chromatin structure of the introduced gene. Such differences would also explain the variable level of induction of the human β-globin gene between different transformants.

We were surprised to find an inducible transcript which initiated in the 5′ flanking region of the HSV *tk* gene. If this viral promoter is used naturally in the herpes life cycle, it is possible that it responds to a viral or cellular *trans*-acting factor similar to that present in induced, but not in uninduced MEL cells.

Received 13 April; accepted 25 July 1983.

1. Friend, C., Scher, W., Holland, J. G. & Sato, T. *Proc. natn. Acad. Sci. U.S.A.* **68,** 378–382 (1971).
2. Reuben, R., Wife, R., Breslow, R., Rifkind, R. & Marks, P. *Proc. natn. Acad. Sci. U.S.A.* **73,** 862–866 (1976).
3. Ross, J., Ikawa, Y. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **69,** 3620–3623 (1972).
4. Willing, M. C., Niehus, A. W. & French Anderson, W. *Nature* **277,** 534–538 (1979).
5. Pyati, J., Kucherlapati, R. S. & Skoultchi, A. J. *Proc. natn. Acad. Sci. U.S.A.* **77,** 3435–3439 (1980).
6. Spandidos, D. & Paul, J. *EMBO J.* **1,** 15–20 (1982).
7. Flavell, R. A. *et al.* in *UCLA Symp. molec. cell. Biol.* **26,** 17–33 (1982).
8. Chao, M. V., Mellon, P., Charnay, P., Maniatis, T. & Axel, R. *Cell* **32,** 483–493 (1983).
9. Wigler, M. *et al. Cell* **16,** 777–785 (1979).
10. Southern, E. M. *J. molec. Biol.* **98,** 503–517 (1975).
11. Berk, A. J. & Sharp, P. A. *Cell* **12,** 721–732 (1977).
12. Weaver, R. F. & Weissmann, C. *Nucleic Acids Res.* **6,** 1175–1192 (1979).

13. Seed, B. in *Genetic Engineering* Vol. 4 (eds Setlow, J. K. & Hollaender, A.) (Academic, New York, 1982).
14. McKnight, J. *Nucleic Acids Res.* **8,** 5949–5964 (1980).
15. Sullivan Read, G. & Summers, W. *Proc. natn. Acad. Sci. U.S.A.* **79,** 5212–5219 (1982).
16. Weatherall, D. J. & Clegg, J. A. *Cell* **29,** 7–9 (1982).
17. Grosveld, F. G., Dahl, H. H. M., deBoer, E. & Flavell, R. A. *Gene* **13,** 227–237 (1981).
18. Grosveld, F. G. *et al. Nucleic Acids Res.* **10,** 6715–6732 (1982).
19. Ish-Horowicz, D. & Burke, J. F. *Nucleic Acids Res.* **9,** 2989–2998 (1981).
20. Birnboim, H. C. & Doly, J. *Nucleic Acids Res.* **7,** 1513–1523 (1979).
21. Moschonas, N. *et al. Nucleic Acids Res.* **9,** 4391–4401 (1981).
22. Auffray, C. & Rougeon, F. *Eur. J. Biochem.* **107,** 303–314 (1980).

Signals regulating hepatitis B surface antigen transcription

Roberto Cattaneo, Hans Will, Nouria Hernandez* & Heinz Schaller

Department of Microbiology, University of Heidelberg, Im Neuenheimer Feld 230, 6900 Heidelberg and * Institute for Cell and Tumor Biology of the German Cancer Research Center, Im Neuenheimer Feld 280, 6900 Heidelberg, FRG

About 200 million people are chronic carriers of hepatitis B surface antigen (HBsAg), but since hepatitis B virus (HBV) cannot be propagated *in vitro*, HBsAg transcription has been studied only in cell lines containing HBV DNA integrated into chromosomes, and HBsAg-related mRNAs 2.0 to 2.5 kilobases (kb) long have been described[1–4]. We have analysed the transcripts produced in an infected chimpanzee liver and in a rat cell line containing HBV DNA[5]. In contrast to previous suppositions[1,2] we report here that the major *S* gene transcript initiates close to the *S* gene, that is, within the 'pre-*S*' region[6] and is processed/polyadenylated at a site situated within the core gene. The efficiency of processing/polyadenylation at this site varies between the chimpanzee liver and the rat cell line studied. The *S* gene promoter does not contain a TATA box but instead has a sequence homologous to that which positions the 5′ ends of the major simian virus 40 (SV40) late transcript[7].

We extracted RNA from a liver specimen obtained from a chimpanzee infected with HBV of known sequence (refs 8, 9 and H.W. *et al.*, manuscript in preparation). However, this material was limited and contained only about 20 copies of HBV *S* gene transcripts per cell. Therefore we also prepared RNA from a cell line whose HBV-specific transcripts had been previously analysed by Northern hybridization[2] (rat 2/130.4/TK4; ref. 5), designated here as rat.HBV. As the concentration of *S* gene transcripts in this line is about 100 molecules per cell, this RNA was used to establish optimal conditions for the analysis of the liver RNA.

The *S* gene transcription unit had been defined as covering ~2.7 kb, with ~900 bp upstream of the *S* gene coding region[1,10–12]. Therefore, we used a probe covering this region (probe *Ava*I; Fig. 1, centre) in an S_1 protection experiment to map initiation of transcription. Figure 2*A* shows that only one major DNA species was protected by RNA from both sources (lanes liver × S_1 and rat.HBV × S_1, bands marked with an asterisk). The corresponding RNA has its 5′ ends at about HBV nucleotide 1,255 (Fig. 1, centre), that is, 185 nucleotides upstream of the *S* gene coding region. Essentially the same fragment of probe *Ava*I was protected when the RNA–DNA hybrids were digested with exonuclease VII (Fig. 2*A*, lanes liver × ExoVII and rat.HBV × ExoVII), which does not attack the loop formed by a genomic DNA probe when hybridizing with a spliced RNA[13]. These results argue against the existence of an intron in the first part of the *S* gene transcription unit. The 5-nucleotide difference in length between the S_1 and exonuclease VII digestion products is due to the fact that exonuclease VII does not digest to completion the ends of single-stranded DNA[14] probes.

Further evidence that the 5′ end detected 185 nucleotides upstream of the *S* gene does correspond to a transcription initiation site comes from primer extension experiments: when primer S (Fig. 1, centre) was hybridized with RNA extracted from the rat.HBV cell line and elongated with reverse transcriptase, only one specific band (marked with an asterisk in Fig. 2*B*, lane rat.HBV × primer S) was identified. As was the case for nuclease protection experiments, this band corresponds to an RNA having its 5′ ends at ~185 nucleotides upstream of the *S* gene. Several other bands resulting from self-priming occurred also in the negative control.

As a third experimental approach to investigate the origin of the RNA species identified above, we used cloned HBV DNA[15] as a template in an *in vitro* transcription system in which promoter-specific initiation depends on the addition of a cytoplasmic fraction to a nuclear extract (N.H. and W. Keller, manuscript in preparation). In an S_1 protection experiment with probe *Ava*I (Fig. 2*C*), transcription initiation 185 nucleotides upstream of the *S* gene was observed (band marked with an asterisk in the lane labelled nuclear XT + cyt.fr.) only when the nuclear extract was complemented with the cytoplasmic fraction. Specific initiation of transcription at this site was abolished after cleavage of the putative promoter sequence about 30 nucleotides upstream of the major *S* start (Fig. 3, site *Fnu*4H) but not after cleavage 220 nucleotides further upstream (site *Bam*HI at position 1,004; G. Christofori, N.H. and R.C., unpublished results). Thus, all sequences necessary for initiation of *S* gene transcription *in vitro* are encoded in the 250 nucleotides preceding the major *S* gene start site.

As a rule, an A + T-rich region (TATA box) responsible for correct positioning of transcription initiation[16] is encoded 30 nucleotides upstream of the start site of eukaryotic mRNAs. Exceptions to this rule are the SV40 and polyoma late promoters and the adenovirus 2 early region 2 promoter. Recently, Brady *et al.*[7] concluded that the sequence GGTACCTAACC (underlined in Fig. 3) located 30 nucleotides upstream of the SV40 major late start is important in the control of SV40 late expression. We find a very similar sequence, which is part of a region of extensive structural homology between HBV and SV40 (box in Fig. 3), in the same position relative to the major HBV *S* mRNA start. A second element of the SV40 late promoter was recently identified within the SV40 origin of replication[14,15]. Again, there is a stretch of strong sequence homology between the SV40 origin and sequences about 60 nucleotides upstream of the major *S* mRNA start site. These structural homologies and the fact that both SV40 late and HBV *S* promoters control the expression of the major viral proteins suggest that these promoters could define a novel class of regulating sequences.

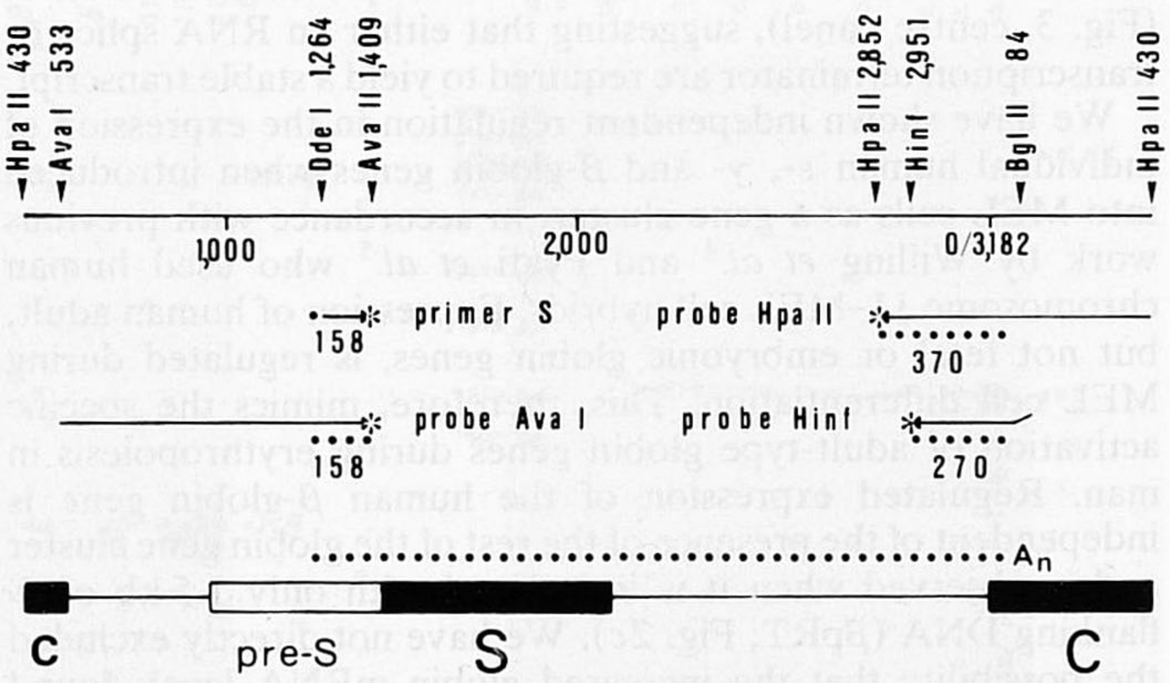

Fig. 1 Map of the HBV genome, the probes and primers used, and the transcripts detected. Top line, restriction sites used for the construction of the probes. The circular HBV genome is opened at the *Hpa*II site at nucleotide 430 (convention of Pasek *et al.*[24]), and the position of the first nucleotide of the recognition sequence of each enzyme is indicated. Centre, probes and primers used and transcripts detected. The probes and primer S (for details see Fig. 2) are drawn as lines, the position of the label indicated with an asterisk and the RNA species protected or elongated depicted as dotted lines. Bottom, map of the HBV genome. The surface antigen (S) and the core antigen (C) genes are shown as solid boxes, the pre-*S* region as an open box, and the *S* mRNA as a dotted line.