MOLECULAR AND CELLULAR BIOLOGY, Aug. 1985, p. 1894–1900
0270-7306/85/081894-07$02.00/0

Vol. 5, No. 8

Synthesis of Predominantly Unspliced Cytoplasmic RNAs by Chimeric Herpes Simplex Virus Type 1 Thymidine Kinase–Human β-Globin Genes

DANIEL S. GREENSPAN* AND SHERMAN M. WEISSMAN

Department of Human Genetics, Yale University School of Medicine, New Haven, Connecticut 06510

Received 20 February 1985/Accepted 16 May 1985

The herpes simplex virus type 1 thymidine kinase (*tk*) gene lacks introns and produces stable mRNA in the absence of splicing. We have prepared a hybrid gene by placing the first exon, first intron (first intervening sequence, designated IVS1), and most of the second exon of the normal human β-globin gene into the 3′ untranslated region of the *tk* gene. Although this hybrid gene contains all globin sequences presumed necessary for the splicing of IVS1, predominantly, unspliced stable cytoplasmic RNA is produced in both long- and short-term expression assays. Moreover, stable unspliced cytoplasmic RNA is detected whether the intron is situated in a sense or an antisense orientation. Efficient splicing of IVS1 is obtained either by deleting the majority of *tk* coding sequences or by relocating the globin sequences from the 3′ to the 5′ untranslated region of the *tk* gene.

Several lines of evidence have indicated splicing to be an essential function in the biogenesis of certain mRNAs (13, 15). The globin genes have been studied extensively in this regard, and it has been noted that globin transcripts in the sense orientation driven from viral promoters are spliced and form stable mRNA, whereas transcripts containing the intron in the antisense orientation do not form stable cytoplasmic or nuclear (14) RNA species. This difference was considered evidence for the requirement of splicing in the expression of globin mRNAs. Further evidence was derived from the study of altered globin genes produced both by in vitro mutagenesis (38) and by spontaneous mutations as in the human thalassemias (8). A number of mutations that affect the splice junction sequences in either the large or small intron have been found to decrease the amount of stable RNA formed. Cytoplasmic RNAs that were formed used cryptic alternative splice sites, and in no case was a stable unspliced globin mRNA formed. Other studies have indicated that unspliced globin RNA (5) and unspliced RNA of other genes (24) may be selectively retained in the nucleus.

In contrast to the above results, a number of genes including those for herpes simplex virus (HSV) type 1 (HSV-1) thymidine kinase (25, 36), histone (17), interferon (28), and a variety of polymerase III transcription products such as 5S RNA (18) lack intervening sequences and yet produce stable mRNAs which are exported from the nucleus. Other unspliced mRNAs have been noted. For example, cDNA corresponding to an early gene of polyoma is expressed as stable RNA in the absence of splicing (34). Genes which contain intervening sequences may also produce stable cytoplasmic RNA retaining intact introns, as is the case for the unspliced late 19S mRNA of simian virus 40 (SV40; 11, 12). A deletion in adenovirus early region E1a which removes a splice acceptor site also results in the synthesis of stable unspliced mRNA (7).

The role of splicing events in the production and export of stable mRNA remains puzzling. To examine this role in a new way, we have introduced an intron and splice junctions into the HSV-1 *tk* gene, which is ordinarily not subject to splicing mechanisms. The intron is from the human β-globin gene, which has been implicated by a number of studies as being dependent on splicing for the expression of stable cytoplasmic RNA. This chimeric gene is capable of directing the synthesis of large amounts of unspliced transcripts, which are exported to the cytoplasm. This effect was reproducible in both long- and short-term expression systems and was obtained by using two different RNA extraction procedures in two different cell types. *tk* sequences inhibited splicing when located 5′ but not 3′ of the globin sequences.

MATERIALS AND METHODS

Cell culture and transfection. Mouse Ltk$^-$ aprt$^-$ and human 143B tk$^-$ (1) cell lines were provided by W. Summers, Yale University, and K. Subramanian, University of Illinois, respectively. Each cell line was maintained in Dulbecco modified essential medium with 10% fetal calf serum. The 143B maintenance medium also contained 100 μg of 5-bromo-2′-deoxyuridine per ml. All 143B cells were passaged twice in 5-bromo-2′-deoxyuridine-free medium before transfection. For long-term transformation of cells to a tk$^+$ phenotype, L cells and 143 B cells were transfected with 50 ng and 10 μg, respectively, of recombinant DNA per 10-cm dish (10^6 cells) as described by Wigler et al. (39). For L-cell transfections, 10 μg of high-molecular-weight Ltk$^-$ DNA or salmon sperm DNA (Worthington Diagnostics) per dish was included as carrier. Carrier DNA was found to be unnecessary in the transformation of 143B cells. Hypoxanthine-aminopterin-thymidine (HAT)-resistant clonal lines of L or 143B cells were obtained by picking individual colonies of cells with cloning cylinders. For transient expression studies, five 10-cm dishes of Ltk$^-$ cells were transfected for each recombinant with 50 μg of recombinant DNA and 100 μg of salmon sperm DNA per dish. DNA-calcium phosphate coprecipitates were left on cells for 12 h, after which the medium was replaced with fresh, nonselective medium. Cells were harvested 48 h later.

Construction of recombinant DNAs. Plasmid tk-pBR322 was obtained from W. Summers, and normal human β-globin

* Corresponding author.

sequences, clone HβG1 (21), were obtained from T. Maniatis, Harvard University.

pDG001 and pDG002. Plasmid tk-pBR322 contains a 3.5-kilobase *Bam*HI fragment of HSV-1 strain CL101 DNA containing the entire *tk* gene plus flanking sequences (26). To construct pDG001 and pDG002 (Fig. 1), a 71-base-pair (bp) fragment containing the *tk* gene translation termination codon was isolated from a *Hha*I digest of tk-pBR322, blunt ended by being trimmed with S1 nuclease, and ligated to *Hin*dIII linkers. Subsequent cleavage with *Xma*I and *Hin*dIII resulted in a 64-bp fragment containing the translation terminator. Digestion of tk-pBR322 with *Xma*I and *Hin*dIII produced a 4,300-bp vector with a *Hin*dIII end made up of pBR322 minus its small *Bam*HI-*Hin*dIII fragment and an *Xma*I end made up of 285 bp of the 5'-most HSV-1 flanking sequences. Cloning of the 64-bp *Xma*I-*Hin*dIII translation terminator fragment into this vector produced recombinant pDG001. Subsequent cloning of a 1,653-bp *tk* *Xma*I DNA fragment into the unique *Xma*I site of pDG001 reconstituted the *tk* gene with all 5' flanking sequences intact but with all HSV-1 sequences downstream of the *tk* translation terminator removed (pDG002).

pDG003 and pDG004. To provide the *tk* gene of recombinant pDG002 with a polyadenylation signal, pDG002 was cleaved at the unique pBR322 *Cla*I site just 6 bases downstream of the truncated *tk* gene (Fig. 1). The SV40 *Hin*dII–*Hin*dIII-G fragment (364 bp) which contains both early and late polyadenylation signals (10, 31) was then blunt-end ligated into the *Cla*I site which had been 3' filled with Klenow fragment. Recombinants pDG003 and pDG004 thus contain *tk* genes designed to utilize either the late (pDG003) or early (pDG004) polyadenylation signals of SV40.

pDG005 and pDG006. The 433-bp *Nco*I-*Bam*HI fragment of human β-globin DNA extends from (and includes) the AUG translation initiation codon through the small first intervening sequence (IVS1) and ends just 14 bp upstream of the 3' end of exon 2. This fragment was blunt ended by 3' filling with Klenow fragment and ligated to *Hin*dIII linkers. It was then cloned into the unique *Hin*dIII site of pDG004 between the *tk* translation terminator and the SV40 early polyadenylation signal in either an antisense (pDG005) or a sense (pDG006) orientation (Fig. 1).

pDG007. To form pDG007 (Fig. 2), the 296-bp *Nco*I fragment of SV40 containing the 72-bp repeats was blunt ended with Klenow fragment and ligated to *Kpn*I linkers. *Kpn*I digestion cleaved linkers and an internal *Kpn*I site to generate a 256-bp fragment, containing both repeats, which was subsequently cloned into a unique pDG006 *Kpn*I site in HSV sequences 500 bases upstream from the *tk* cap site. The orientation of the 72-bp repeats was such that their early region sides face the *tk* gene.

pDG033, pDG034, pDG036, pDG037, and pDG038. To form pDG033, pDG034, pDG036, pDG037, and pDG038 (Fig. 2), recombinant pDG007 was cleaved with *Bgl*II, blunt ended by 3' filling with Klenow fragment, and then cut at the unique *Sal*I site in its pBR322 sequences. This *Sal*I-*Bgl*II piece was then ligated to the appropriate *Sma*I-*Sal*I pDG007 piece to derive recombinant pDG033. Similarly, the same *Sma*I-*Sal*I piece ligated to an *Eco*RV-*Sal*I piece or to an *Apa*I-*Sal*I piece, in which *Apa*I site had been blunt ended by trimming with T4 polymerase I (23), resulted in recombinant pDG034 or pDG038, respectively. The same blunt-ended *Bgl*II-*Sal*I piece used in the construction of pDG033 produced pDG036 when ligated to an *Sph*I-*Sal*I fragment in which the *Sph*I site had been blunt ended with T4 polymerase I. Utilizing the unique *Bgl*II site in the 5' leader sequences of the *tk* gene, *Bgl*II-*Sph*I and *Bgl*II-*Apa*I pieces in which both *Apa*I and *Sph*I sites had been blunt ended with T4 polymerase I were ligated to form recombinant pDG0037.

pDG016. To form pDG016, the 256-bp *Kpn*I fragment containing the 72-bp repeats of SV40 was prepared as described above for pDG007 and inserted into a unique *Kpn*I site in HSV sequences 500 bases upstream of the *tk* cap site in tk-pBR322.

pDG042. To form pDG042, the 433-bp *Nco*I-*Bam*HI globin fragment containing IVS1 was 3' filled with Klenow fragment, ligated to *Bgl*II linkers, and inserted into the unique *Bgl*II site in the 5' untranslated region of the *tk* gene in pDG004, which then received the SV40 72-bp repeat sequences as described above for pDG007.

JEM 13β. JEM 13β was provided by J. Metherall, Yale University, and is made up of a 4.7-kilobase *Bgl*II fragment of human genomic DNA containing the β-globin gene and extensive 5' and 3' flanking sequences cloned into expression vector pLTN3 (16).

RNA extraction. Cytoplasmic RNA was isolated from cloned tk^+ L-cell lines by Dounce homogenization of cells in hypotonic RSB buffer (10 mM Tris hydrochloride [pH 7.4]–10 mM NaCl–3 mM $MgCl_2$) as previously reported by this laboratory (29). Nuclei were separated from cytoplasm by low-speed centrifugation through a 0.25 M sucrose pad, with recentrifugation of the cytoplasmic fraction to remove residual nuclei, and polyadenylated RNA was selected from cytoplasmic RNA by chromatography on oligo$(dT)_{12-15}$ cellulose (type 2; Collaborative Research, Inc.) as previously reported (29). Cytoplasmic and nuclear RNA fractions were isolated from transient expression assay L cells and cloned tk^+ 143B cell lines by disruption of cells in isotonic lysis buffer in the presence of Nonidet P-40 (Shell Chemicals) as described previously (23), except that lysis buffer was pH 8.0 and cytoplasmic fractions were centrifuged twice to remove residual nuclei. Nuclear RNA samples were treated with 10 μg of DNase (Worthington) made RNase free by chromatography on UDP-cellulose (40) per ml.

S1 nuclease mapping. The S1 nuclease protection experiments were carried out as described previously (4, 37) with previously described modifications (9).

RESULTS

Long-term expression of recombinant *tk* clones. It was not known, a priori, whether an intron introduced into a gene which ordinarily has none would (i) be spliced or (ii) preclude the formation of stable RNA species if retained in transcripts. It was also not known if splicing, if it occurred, would enhance or decrease expression of a gene not normally subject to splicing mechanisms. Two *tk* recombinants containing the small intron of normal human β-globin (IVS1) either in the sense (pDG006) or antisense (pDG005) orientation were transfected into mouse Ltk^- cells. To judge the effect of the globin sequences on expression, the number of colonies obtained after selection in HAT medium were compared with the number of colonies obtained by transfection with the original tk-pBR322 recombinant or with recombinants pDG002, pDG003, and pDG004. Transformation rates in such long-term transformation assays have been found to reflect initial rates of *tk* gene expression (33). A derivative of tk-pBR322 containing the SV40 72-bp repeat enhancing sequences (pDG016) was included in the study as a positive control for the enhancement of *tk* expression.

Unexpectedly, intermediate recombinants in which the *tk* gene utilizes either the late (pDG003) or early (pDG004)

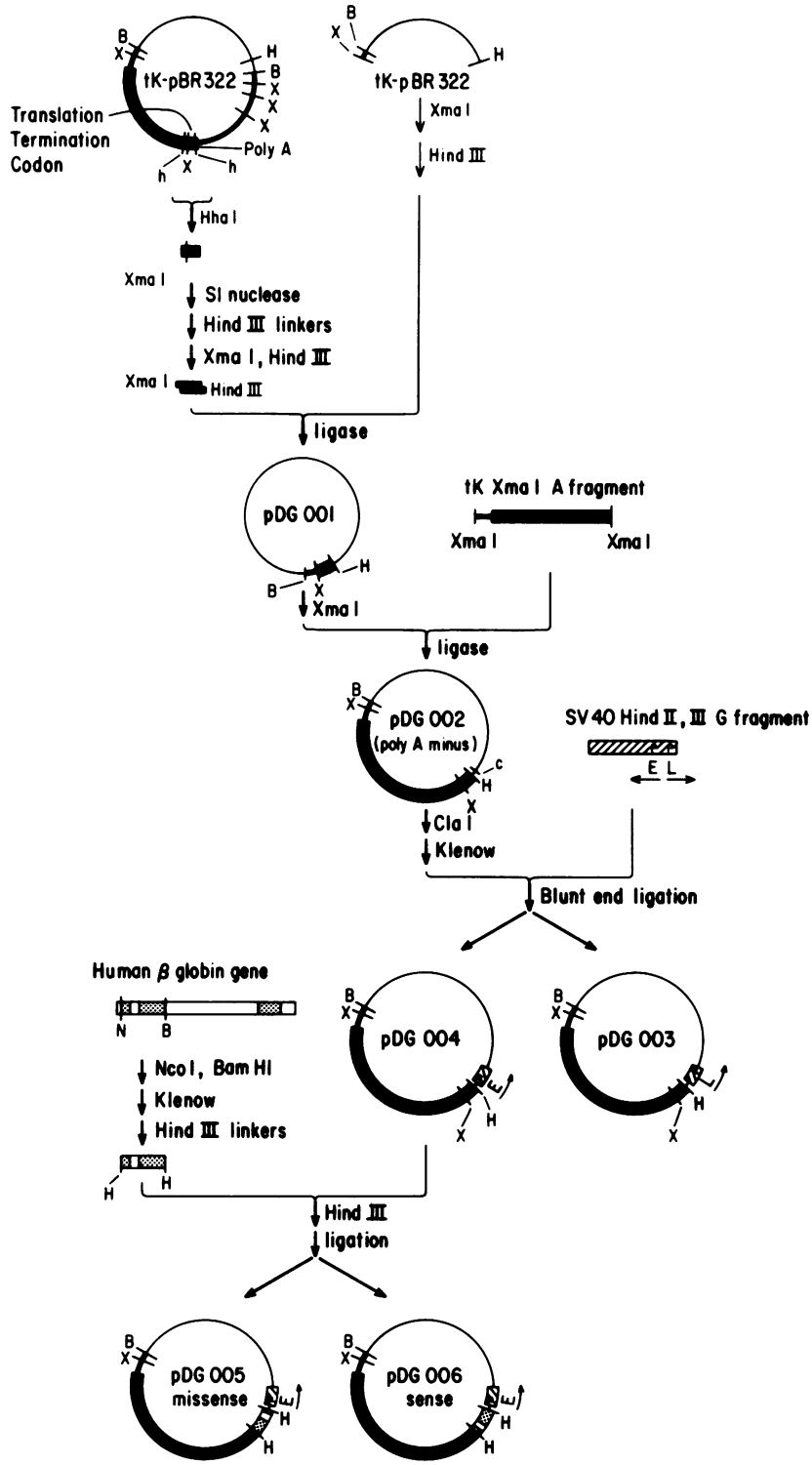

FIG. 1. Construction of *tk*–β-globin recombinants. Details of the construction of *tk*–β-globin recombinants and the origin of starting materials are described in the text. Symbols: dark box, HSV-1 *tk* gene and flanking sequences; hatched box, SV40 sequences; stippled box, globin coding sequences; open box, globin intron and flanking sequences; single line, pBR322. Abbreviations: B, *Bam*HI; C, *Cla*I; h, *Hha*I; H, *Hin*dIII; N, *Nco*I; X, *Xma*I; P, polyadenylation signal.

SV40 polyadenylation site yielded numbers of colonies of HAT-resistant L cells similar to those obtained for pDG016. In 143B cells in which the SV40 72-bp repeats have been found to produce a 20-fold enhancement in gene expression (6), pDG004 is again found to produce numbers of colonies similar to those obtained for pDG016 (Table 1). However, pDG003 does not produce a large number of colonies. This apparent stimulation of *tk* expression by 3′ SV40 sequences

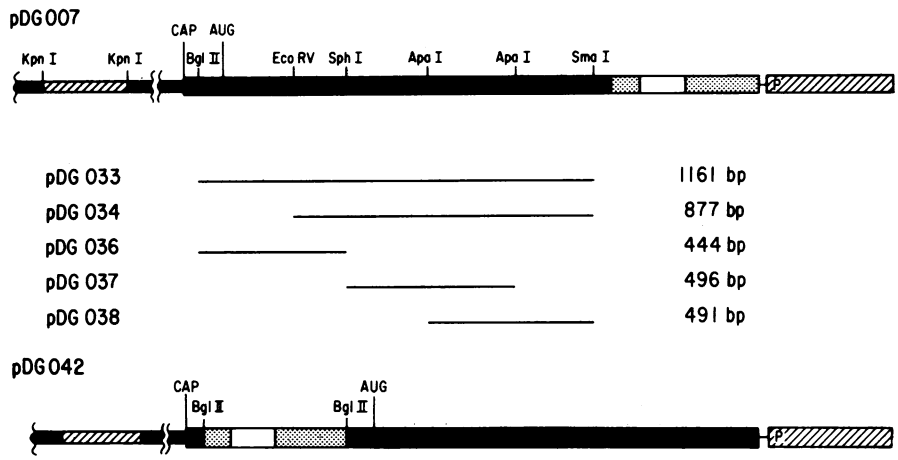

FIG. 2. Schematic representation of recombinants pDG007 and pDG042 and recombinants derived from pDG007 by partial deletion *tk* sequences. Recombinants pDG033, pDG034, pDG036, pDG037, and pDG038 were derived by deleting *tk* sequences from pDG007 as described in the text. The extent of each deletion is represented by a bar aligned with the pDG007 map. The number of bp encompassed by each deletion is to the right of the bar. Symbols are the same as those as in Fig. 1. CAP and AUG denote *tk* mRNA cap site and translational start codon, respectively.

is currently under study. In either L cells or 143B cells, however, the insertion of IVS1 in either a sense or antisense orientation reduces the number of colonies obtained to a level comparable with that of tk-pBR322 (Table 1, pDG006 and pDG005). This reduction in the level of *tk* transformation may be due to a distancing of *tk* sequences from those of SV40 by a spacer effect of the globin DNA or to the production of less stable RNA species due to the retention of intervening sequences. In either case, if splicing had occurred, its effect on expression would appear to be minimal, since pDG005, whose globin sequences are in an antisense orientation, gives similar numbers of colonies as pDG006, whose splice junctions are oriented correctly.

Globin-specific sequences in pDG006 transcripts are predominantly unspliced. Northern blot analysis of cytoplasmic polyadenylated RNA from cloned lines of tk^+ L cells transformed by the various recombinants found that *tk*-specific RNA forms from each line corresponded in size to faithful transcripts of the appropriate recombinant gene (data not shown). Transcripts of cells transformed by pDG005 or pDG006 were also found to hybridize a probe specific for IVS1, suggesting the retention of at least some intronic sequences. To further analyze the extent to which introns are retained in the cytoplasmic RNA of pDG006-transformed L cells, cytoplasmic polyadenylated RNA was hybridized to a globin-specific probe and subjected to S1 nuclease digestion. Hybridization of unspliced and spliced RNA transcripts to the 5'-end-labeled *Bam*HI-*Sph*I globin probe should generate protected DNA fragments of 433 and 209 bp, respectively, with S1 digestion. Under conditions of probe excess, the amount of a protected DNA fragment of given size is proportional to the amount of the corresponding RNA species in a sample. With such analysis, cytoplasmic RNA produced by pDG006 was found to be made up predominantly of unspliced transcripts (Fig. 3, lane 2). A small amount of RNA which had been spliced at the normal globin splice junctions was also detected (Fig. 3, lane 2). Forms utilizing alternative globin splice sites were not detected, and

TABLE 1. Number of tk^+ colonies transformed by constructs

Amt of construct DNA per dish[a] (μg)	Colonies per dish for[b]:						
	tk-pBR322	pDG002	pDG003	pDG004	pDG005	pDG006	pDG016
Expt 1[c]							
0.5	550	114		>1,000	564	527	
Expt 2[d]							
0.25	50	16	204	226	93	105	195
0.05	33	1	111	173	34	58	126
Expt 3[e]							
0.25	53	15	150	177	64	84	173
0.05	37	0	103	123	28	31	105
Expt 4[f]							
10.0	3		6	327	4	1	370
10.0	2		3	113	2	2	102

[a] Transfections were performed 1 day after initial plating of the cells.
[b] Average number of colonies per each of four replicate dishes.
[c] Initial plating of LTK^- cells was at 1.4×10^6 cells per dish; selection was in 1× HAT.
[d] Initial plating of LTK^- cells was at 0.7×10^6 cells per dish; selection was in 1× HAT.
[e] Initial plating of LTK^- cells was at 0.7×10^6 cells per dish; selection was in HAT medium containing 1/30 the usual concentration of thymidine.
[f] Initial plating of 143B cells was at 1.4×10^6 cells per dish; selection was in 1× HAT.

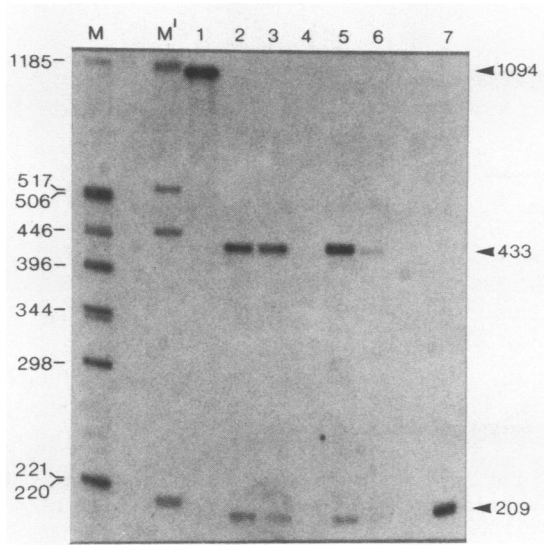

FIG. 3. S1 nuclease analysis of splicing for globin-specific RNA sequences in pDG005- or pDG006-transformed mouse and human cell lines. Polyadenylated cytoplasmic RNA from cloned, transformed lines of mouse L or human 143B cells was hybridized to a 1,094-bp double-stranded fragment of β-globin DNA. This probe extends from the *Bam*HI site in exon 2 to an *Sph*I site 660 bases upstream of the globin AUG and was 5′ end labeled at the *Bam*HI site. RNA-DNA hybrids were digested with S1 nuclease at various temperatures and analyzed as described in the text. The 5′-end-labeled DNA products were protected from S1 digestion by the following: lane 2, 3.5 μg of pDG006 L-cell RNA at 23°C; lane 3, 10 μg of pDG006 143B cell RNA at 23°C; lane 4, 10 μg of pDG005 143B cell RNA at 23°C; lane 5, 10 μg of pDG006 143B cell RNA at 17°C; lane 6, 10 μg of pDG006 143B cell RNA at 30°C; lane 7, 15 μg of whole-cell human reticulocyte RNA at 23°C. Lane 1, undigested probe; M, ^{32}P-labeled markers of *Hin*fI-*Eco*RV-digested pBR322; M′, additional markers of 1169, 1118, 526, 447, and 215 bp.

all DNA fragments protected by spliced RNA comigrated with DNA protected by authentic spliced β-globin from human reticulocytes (Fig. 3, lane 7). Analyses with *tk*-specific probes found no splicing involving the *tk* gene (data not shown).

RNA of pDG006-transformed human 143B cells gave the same distribution of spliced and unspliced forms as the RNA from the mouse L cells (Fig. 3, lane 3). Thus, the pattern of spliced and unspliced cytoplasmic RNA is neither cell line- nor species-specific and is, instead, apparently intrinsic to the pDG006 recombinant. S1 digestion of DNA-RNA hybrids at various temperatures (Fig. 3, lanes 3, 5, and 6) demonstrated a stronger signal upon digestion at 17°C than at higher temperatures, without the appearance of spurious bands due to incomplete digestion. Therefore, 17°C was used in all subsequent S1 nuclease digestions.

Transient expression analysis of *tk*-globin recombinants. Selection in HAT medium and the establishment of long-term tk$^+$ cell cultures might have affected the expression of transfected recombinants in unforeseen ways. To avoid this possibility, a short-term transfection assay was established for the transient expression of recombinants in mouse L cells. To ensure that transcription in such an assay was at detectable levels, the 72-bp repeat enhancer sequences of SV40 (2, 27) were inserted upstream of the *tk* gene in pDG006 to yield recombinant pDG007 (Fig. 2). At 48 h after transfection with pDG007, L cells were harvested, and their RNA was subjected to analysis by S1 nuclease digestion. Such analysis (Fig. 4A, lane 1) shows that the cytoplasm of these cells contains a distribution of spliced and unspliced globin-specific RNA species very similar to the distribution found in the cytoplasm of long-term pDG006-transformed cell lines.

To examine further the role of *tk* sequences in the accumulation of unspliced cytoplasmic RNA, two large deletions were made in the *tk* gene of pDG007 before transient expression and S1 analysis. One deletion yielded recombinant pDG033 (Fig. 2) with an internal deletion of 1,161 bp that removed almost the entire *tk* gene and left only 57 bp of *tk* 5′ leader sequence fused to 56 bp from the 3′ end of the gene. Recombinant pDG034 contains a smaller deletion in which an additional 53 bp of 5′ leader sequence and 231 bp of coding sequence from the 5′ end of the *tk* gene have been retained (Fig. 2). Recombinant pDG033 produced a dramatic shift in the distribution of spliced and unspliced globin-specific RNAs accumulated in the cytoplasm of transfected L cells (Fig. 4A, lane 3), with a marked decrease in the amount of unspliced RNA and a concomitant increase in the amount of accurately spliced forms.

The hypothetical objection could be raised that unspliced RNA found in the cytoplasm of transfected cells arose by leakage from the nuclei during cell fractionation. There are important reasons to suppose that this leakage is not the case. (i) Recombinant JEM 13β (see above), which contains the entire human β-globin gene and is known to produce normally processed globin RNA (J. Metherall, personal communication), was transfected into mouse cells at the same time as pDG007, pDG034, and pDG033 and was found to result in the detection of only accurately spliced forms of globin RNA in the cytoplasm (Fig. 4A, lane 4). In addition, the nuclear fractions of the cells transfected with recombinants pDG007, pDG033, and JEM 13β (Fig. 4) were isolated, and nuclear RNA samples were analyzed by S1 nuclease digestion (Fig. 4A, lanes 5, 6, and 7). Equal quantities of RNA were used for S1 analysis of cytoplasmic and nuclear fractions under conditions of probe excess. Lower levels of both spliced and unspliced RNA species were found in the nuclei than in the cytoplasm of cells transfected by pDG007, pDG033, or JEM 13β. There is relatively little unspliced nuclear RNA, therefore, which might serve as a source for hypothetical leakage. (ii) Similar quantities of unspliced cytoplasmic RNA have been detected whether RNA extractions were performed under hypotonic or isotonic conditions. Since extraction under isotonic conditions greatly reduces leakage of nuclear RNAs into cytoplasm (22, 35), this finding also indicates that cytoplasmic RNAs are not nuclear contaminants. Isotonic conditions were employed in all but the initial RNA extractions in this study.

A 14-fold-longer exposure of lanes 5, 6, and 7 (Fig. 4B) reveals features of the nuclear RNA species. Recombinant JEM 13β is seen to produce four forms of unspliced RNA. The 483-bp fragment represents an unspliced globin transcript initiated at the correct transcriptional start site (21). Three larger fragments are seen, representing unspliced transcripts having aberrant start sites in the human 5′ flanking sequences upstream from the globin gene (J. Metherall, personal communication).

The effect that *tk* sequences have on the accumulation of unspliced RNA might have been localized to one small part of the *tk* gene. To examine this possibility, recombinants pDG036, pDG037, and pDG038 were derived from pDG007 by the deletion of 400- to 500-bp stretches of *tk* sequence from the 5′ end, middle, and 3′ end of the gene, respectively

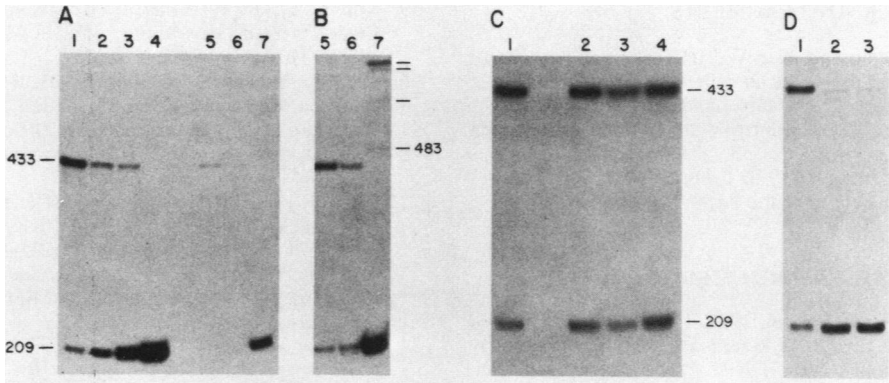

FIG. 4. Analysis of splicing in RNA from transient expression assays. Transfections of L cells for transient expression assays and preparation of total cytoplasmic and nuclear RNA fractions 48 h later are described in the text. Of each RNA sample (cytoplasmic or nuclear), 50 μg was hybridized to the 5'-end-labeled globin *Bam*HI-*Sph*I probe (described in the legend to Fig. 2) and subjected to S1 digestion at 17°C. (A and B) Autoradiographs of portions of the same gel after exposures of 12 h and 7 days, respectively. Lanes 1 through 4, DNA products protected from S1 digestion by cytoplasmic RNA from L cells transfected with pDG007 (lane 1), pDG034 (lane 2), pDG033 (lane 3), and JEM 13β (lane 4). Lanes 5 through 7, DNA protected from S1 digestion by nuclear RNA from L cells transfected with pDG007 (lane 5), pDG033 (lane 6), and JEM 13β (lane 7). (C) DNA protected from S1 digestion by cytoplasmic RNA from L cells transfected with pDG007 (lane 1), pDG036 (lane 2), pDG037 (lane 3), and pDG038 (lane 4). (D) DNA protected from S1 digestion by cytoplasmic RNA from L cells transfected with pDG007 (lane 1), pDG042 (lane 2), and pDG033 (lane 30). Details of the construction of recombinants are described in the text.

(Fig. 2). However, no dramatic shift from unspliced to spliced RNA forms was found for any of the three deletion constructs (Fig. 4).

To determine whether the *tk* gene would inhibit the splicing of IVS1 if the intervening sequence was located upstream of the tk sequences, recombinant pDG042 was constructed (Fig. 2). Splicing of IVS1 was found to be as efficient for pDG042 as for recombinant pDG033, from which the majority of *tk* sequences had been removed (Fig. 4D, lanes 2 and 3).

DISCUSSION

The data presented here indicate that splice junctional sequences and recently described sequences within introns employed for lariat formation (32) are not always sufficient for splicing to occur. Sequences in the transcript external to the intron and even far removed from splice junctions can markedly affect the efficiency of splicing.

Analysis of nuclear transcripts provides limits to possible mechanisms leading to the export of unspliced RNA. We did not detect an accumulation of excess nuclear RNA in cells expressing constructs containing the intron compared with cells containing the authentic globin gene alone. This finding argues that primary transcripts of *tk*-globin recombinants do not accumulate in the nucleus and spill out into the cytoplasm because of a high nuclear concentration. Conversely, we did not detect a decrease in the amount of nuclear transcripts in cells transfected with *tk*-globin hybrids compared with the amount of transcripts in cells transfected with the intact globin gene, which suggests that the hybrid transcripts are not exported more rapidly from the nucleus before splicing can occur. Observations are more consistent with the hypothesis that whatever is involved in transport from the nucleus to the cytoplasm proceeds with similar efficiency in cells having the globin gene alone or the fused *tk*-globin constructs. Alteration of primary transcripts due to insertion of the *tk* gene segments upstream of globin sequences, therefore, appears to affect the availability of globin sequences for splicing without altering transport to the cytoplasm.

To further explore the role of *tk* sequences in the accumulation of unspliced cytoplasmic RNA, a number of *tk*-globin recombinants containing different deletions of *tk* sequence were examined. Deletion of the majority of *tk* sequences caused a shift from predominantly unspliced to predominantly spliced cytoplasmic RNA. Several smaller deletions, however, could not localize this effect. Possibly, a gene which has adapted to expressing unspliced RNA may have evolved multiple regions within it that would favor nuclear export without splicing of its transcripts. A second possibility is that the length of 5' RNA per se, before the first splice signal in these constructs, was sufficient to inhibit splicing. Recombinants producing predominantly unspliced RNA contained from 900 to 1,400 bases upstream of the first splice junction. Although the latter possibility has not been precluded by the present study, it should be noted that naturally occurring RNAs with long 5' unspliced sequences of 550 (3) or 600 (30) bases before the first splicing event are spliced quite efficiently. Thus, a combination of length and specific features may be involved in the observed inhibition of splicing.

Whatever the mode of inhibition, a strong positional effect on the splicing efficiency of a given intron has been demonstrated. The identical intron and junctional sequences which spliced with low efficiency when located downstream of *tk* sequences were spliced with high efficiency when located near the 5' end of the chimeric gene (Fig. 4D, lanes 1 and 2). One possible explanation of the observed difference in efficiencies is conformational, i.e., that the configuration of the RNA precursor is such that its 5' end is more "open" to splicing than the 3' end, which is relatively masked. In such a case, however, the 3' site would have to be masked by a variety of different conformations, since splicing is inefficient in four recombinants, three of which contain extensive but different deletions of *tk* sequence (Fig. 4C). An alternative explanation is that there is a 5'→3' directional component, perhaps involving localized scanning or processivity, involved in the splicing process such that splicing is blocked by *tk* sequences only when the sequences lie upstream of the globin intron. Although hypothetical, such 5'→3' directionality is consistent with previous reports of apparent 5'→3' directionality in donor splice site selection (19, 20) and in the mechanism of intron excision (32).

ACKNOWLEDGMENTS

We thank Daniel DiMaio and Alan Weiner for critical reading of the manuscript, Robert Conroy for excellent technical assistance, and Ann M. Mulvey for preparing this manuscript.

This work was supported by grant MV-10-0 from the American Cancer Society. D.S.G. was a U.S. Public Health Services postdoctoral trainee (grant CA-0959-09 from the National Institutes of Health) and is now a Fellow of the Arthritis Foundation.

LITERATURE CITED

1. **Bacchetti, S., and F. L. Graham.** 1977. Transfer of the gene for thymidine kinase to thymidine kinase-deficient human cells by purified herpes simplex viral DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:1590–1594.
2. **Banerji, J., S. Rusconi, and W. Schaffner.** 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell **27**:299–308.
3. **Battey, J., C. Moulding, R. Taub, W. Murphy, T. Stewart, H. Potter, G. Lenoir, and P. Leder.** 1983. The human c-myc oncogene: structural consequences of translocation into the IgH locus in Burkitt lymphoma. Cell **34**:779–787.
4. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell **12**:721–732.
5. **Busslinger, M., N. Moschonas, and R. A. Flavell.** 1981. β^+ thalassemia: aberrant splicing results from a single point mutation in an intron. Cell **27**:289–298.
6. **Byrne, B. J., M. S. Davis, J. Yamaguchi, D. J. Bergsma, and K. N. Subramanian.** 1983. Definition of the SV40 early promoter region and demonstration of a host range bias in the enhancement effect of the SV40 72-bp repeat. Proc. Natl. Acad. Sci. U.S.A. **80**:721–725.
7. **Carlock, L., and N. C. Jones.** 1981. Synthesis of an unspliced cytoplasmic message by an adenovirus 5 deletion mutant. Nature (London) **294**:572–574.
8. **Collins, F., and S. M. Weissman.** 1984. The molecular genetics of human hemoglobin, p. 315–462. *In* K. Moldave and W. E. Cohn (ed.), Progress in nucleic acid research and molecular biology, vol. 31. Academic Press, Inc., New York.
9. **Favaloro, T., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA: analysis by two-dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
10. **Fiers, W., R. Contreras, G. Haegeman, R. Rogiers, A. Van de Voorde, H. Van Heuverswyn, J. Van Herreweghe, G. Volckaert, and M. Ysebaert.** 1978. Complete nucleotide sequence of SV40 DNA. Nature (London) **273**:113–120.
11. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissman.** 1978. Heterogeneity and 5'-terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. **126**:813–846.
12. **Ghosh, P. K., P. Roy, A. Barkan, J. E. Mertz, S. M. Weissman, and P. Lebowitz.** 1981. Unspliced functional late 19S mRNAs containing intervening sequences are produced by a late leader mutant of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **78**:1386–1390.
13. **Gruss, P., C. J. Lai, R. Dhar, and G. Khoury.** 1979. Splicing as a requirement for biogenesis of functional 16S mRNA of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **76**:4317–4321.
14. **Hamer, D. H., and P. Leder.** 1979. Splicing and the formation of stable RNA. Cell **18**:1299–1302.
15. **Hamer, D. H., K. D. Smith, S. H. Boyer, and P. Leder.** 1979. SV40 recombinants carrying rabbit β-globin gene coding sequences. Cell **17**:725–735.
16. **Humphries, R. K., T. Ley, P. Turner, A. D. Moulton, and A. W. Nienhuis.** 1982. Differences in human α, β- and δ-globin gene expression in monkey kidney cells. Cell **30**:173–183.
17. **Kedes, L. H.** 1978. Histone gene and histone messengers. Annu. Rev. Biochem. **48**:837–870.
18. **Korn, L. J., and D. D. Brown.** 1978. Nucleotide sequence of Xenopus borealis oocyte 5S DNA: comparison of sequences that flank several related eukaryotic genes. Cell **15**:1145–1156.
19. **Kuhne, T., B. Weiringa, J. Reiser, and C. Weissman.** 1983. Evidence against a scanning model of RNA splicing. EMBO J. **2**:727–733.
20. **Lang, K. M., and R. A. Spritz.** 1983. RNA splice site selection: evidence for a 5'→3' scanning model. Science **220**:1351–1355.
21. **Lawn, R. M., A. Efstratiadis, C. O'Connell, and T. Maniatis.** 1980. The nucleotide sequence of the human β-globin gene. Cell **21**:647–651.
22. **Levis, R., and S. Penman.** 1977. The metabolism of poly(A)$^+$ and poly(A)$^-$ hnRNA in cultured drosophila cells studied with a rapid uridine pulse-chase. Cell **11**:105–113.
23. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **May, E., M. Kress, and P. May.** 1978. Characterization of two SV40 early mRNAs and evidence for a nuclear "prespliced" RNA species. Nucleic Acids Res. **5**:3083–3099.
25. **McKnight, S. L.** 1980. The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. Nucleic Acids Res. **8**:5949–5964.
26. **McKnight, S. L., and R. Kingsbury.** 1982. Transcriptional control signals of a eukaryotic protein-coding gene. Science **217**:316–324.
27. **Moreau, P., R. Hen, B. Wasylyk, R. Everett, M. Gaub, and P. Chambon.** 1981. The SV40 72-base repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res. **9**:6047–6068.
28. **Nagata, S., N. Mantei, and C. Weissman.** 1980. The structure of one of the eight or more distinct chromosomal genes for human interferon-alpha. Nature (London) **287**:401–408.
29. **Piatak, M., K. N. Subramanian, P. Roy, and S. M. Weissman.** 1981. Late messenger RNA production by viable simian virus 40 mutants with deletions in the leader region. J. Mol. Biol. **153**:589–618.
30. **Reddy, V. B., P. K. Ghosh, P. Lebowitz, M. Piatak, and S. M. Weissman.** 1979. Simian virus 40 early mRNA's. I. Genomic localization of 3' and 5' termini and two major splices in mRNA from transformed and lytically infected cells. J. Virol. **30**:279–296.
31. **Reddy, V. B., B. Thimmappaya, R. Dhar, K. N. Subramanian, B. S. Zain, J. Pan, P. K. Ghosh, M. L. Celma, and S. M. Weissman.** 1978. The genome of simian virus 40. Science **200**:494–502.
32. **Ruskin, B., A. R. Krainer, T. Maniatis, and M. R. Green.** 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing *in vitro*. Cell **38**:317–331.
33. **Spandidos, D. A., and N. M. Wilkie.** 1983. Host-specificities of papillomavirus, Maloney murine sarcoma virus and simian virus 40 enhancer sequences. EMBO J. **2**:1193–1199.
34. **Treisman, R., V. Novak, J. Favaloro, and R. Kamen.** 1981. Transformation of rat cells by an altered polyoma virus genome expressing only the middle-T protein. Nature (London) **292**:595–600.
35. **Villarreal, L. P., and S. Jovanovich.** 1980. Leakage of nuclear transcripts late in simian virus 40-infected CV1 cells: quantification of spliced and unspliced late 19S RNAs. J. Virol. **36**: 595–600.
36. **Wagner, M. J., J. A. Sharp, and W. C. Summers.** 1981. Nucleotide sequence of the thymidine kinase gene of herpes simplex virus type 1. Proc. Natl. Acad. Sci. U.S.A. **78**:1441–1445.
37. **Weaver, R. F., and C. Weissman.** 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map coordinates. Nucleic Acids Res. **7**:1175–1193.
38. **Wieringa, B., F. Meyer, J. Reiser, and C. Weissmann.** 1983. Unusual splice sites revealed by mutagenic inactivation of an authentic splice site of the rabbit β-globin gene. Nature (London) **301**:38–43.
39. **Wigler, M., A. Pellicer, S. Silverstein, R. Axel, G. Urlaub, and L. Chasin.** 1979. DNA-mediated transfer of the adenine phosphoribosyl transferase locus into mammalian cells. Proc. Natl. Acad. Sci. U.S.A. **76**:1373–1376.
40. **Wilchek, M., and M. Gorecki.** 1969. Affinity chromatography of bovine pancreatic ribonuclease A. Eur. J. Biochem. **11**:491–494.