THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 258, No. 10, Issue of May 25, pp. 6043-6050, 1983
Printed in U.S.A.

Expression of the Human Insulin Gene and cDNA in a Heterologous Mammalian System*

(Received for publication, October 18, 1982)

Orgad Laub‡ and William J. Rutter

From the Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143

The human insulin gene or the corresponding cDNA has been inserted into the early region of a simian virus 40 vector in which all SV40 splice junctions were deleted while the early promoter and polyadenylation regions remained intact. The expression of insulin-coding sequences was tested in permissive monkey COS cells.

The insulin cDNA was transcribed from the early promoter to produce a stable polyadenylated RNA which was translated, and immunoreactive human proinsulin accumulated in the medium. Thus RNA splicing is not obligatory for insulin expression in this system.

The genomic insulin transcript was also initiated from the SV40 promoter and terminated at the SV40 polyadenylation site. S1 endonuclease mapping revealed that the transcript is processed via two alternative splicing pathways within the insulin gene. About one-third of the total transcripts are processed normally with removal of the two insulin-specific introns. This transcript is apparently translated normally since immunoreactive proinsulin accumulates in the medium.

About two-thirds of the transcripts are processed via an alternative splicing pathway involving a new splice acceptor site located within the coding region of the insulin gene. This results in a codon frameshift such that translation would produce a novel chimeric peptide containing the insulin NH_2-terminal B chain, but a different COOH terminus containing human and SV40 sequences. A peptide of the predicted size is detected in the COS cell extract.

Most eukaryotic genes are mosaic structures in which the coding regions (exons) represented in the mRNA are interrupted by intervening sequences (introns) which are subsequently removed from the primary transcript by RNA splicing (Sharp, 1982). In most cellular genes thus far studied, the primary transcript gives rise to a single mature mRNA species. There are however notable exceptions in which alternative splicing is employed: in the early gene of SV40, two donor sites are spliced to one common acceptor site (Berk and Sharp, 1978). In the adenovirus 2 late genes, one donor site is spliced to several acceptor sites (Chow *et al.*, 1977). Alternative RNA splicing pathways also contribute to the diversity of the expression of immunoglobulin genes (Marcu, 1982), calcitonin (Rosenfeld *et al.*, 1982; Amara *et al.*, 1982), and to the tissue-specific expression of salivary and liver amylase (Young *et al.*, 1981).

The nucleotide sequence of the human insulin gene and its flanking regions has been determined (Bell *et al.*, 1979, 1980). A comparison of the human insulin gene with the insulin cDNA and with other insulin genes indicates that this gene has two intervening sequences. Although the DNA sequence provides crucial structural information, it does not decisively locate the boundary region for the RNA splicing junction nor the regions regulating gene expression. To elucidate these features biological systems must be employed.

We have used SV40 as a vector to express the human insulin gene in permissive monkey cells. This system is particularly useful for these and other studies because during infection the virus reaches high titer within the cells and consequently high levels of transcription are achieved. For these experiments our construction employed the SV40 early promoter but the SV40 splice sites were eliminated. A full length cDNA was inserted to test for the effect of introns on the expression of insulin-coding sequences. Introns are required for effective expression of the mouse β^{maj} globin gene in another SV40 vector-host system (Hamer and Leder, 1979).[1] We show that cultured transformed monkey cells infected with this SV40-insulin recombinant express high levels of insulin-coding mRNA without splicing and secrete immunoreactive proinsulin into the medium. The transcripts from the genomic human insulin DNA were processed by two splicing pathways. In the first, the precursor RNA is processed normally; in the second, a new splice acceptor site is recruited from within the insulin-coding sequences.

MATERIALS AND METHODS

Enzymes and Radioisotopes—Restriction enzymes were purchased from Bethesda Research Laboratories or New England Biolabs. *Escherichia coli* polymerase I, T4 polynucleotide kinase, and T4 DNA ligase were from New England Biolabs. S1 nuclease was obtained from Miles Laboratories. All radioisotopes were from Amersham.

Cell Transfection and Virus Strain—The construction of SV40-insulin recombinants is described in the results sections. Transformed African green monkey (COS-7) cells (Gluzman, 1981) were maintained in Dulbecco's modified Eagle's medium containing penicillin, streptomycin, and 10% fetal calf serum. Twenty-four hours after seeding (1.5×10^6 cells/10-cm plate), the cultures were transfected by the $CaPO_4$ procedure (Graham and Van Der Eb, 1973; Parker and Stark, 1979). 5–10 μg of circularized SV40-insulin recombinant DNA were used for each plate and virus stocks were prepared from cell lysates 14–21 days after transfection. The titer of the recombinant viruses was estimated by comparison with wild type virus stock of known titer, assayed under the same conditions. The comparisons were performed by comparing cytopathic effects of the two stocks and alternatively by comparing amounts of free viral DNA (Hirt, 1967) 48-h post-infection.

* This research was supported by National Institutes of Health Grants GM 28520 and AM 21344 and by Grant 1-745 from the March of Dimes. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Recipient of the Weizmann Fellowship and a fellowship from the Cystic Fibrosis Foundation.

[1] A. Buchman and P. Berg, personal communication.

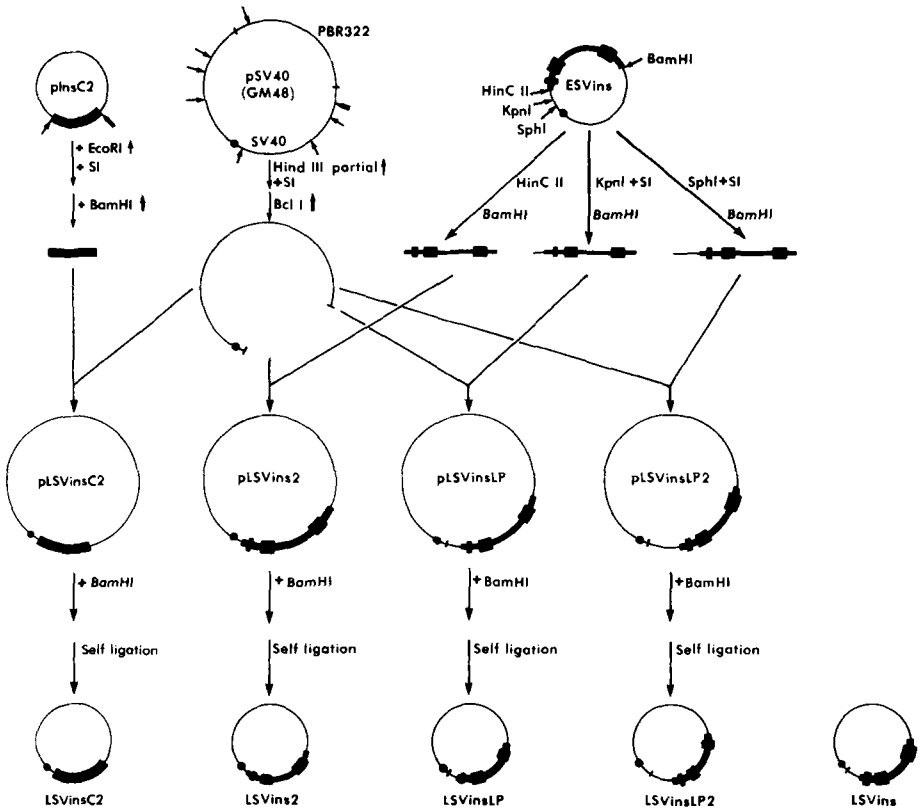

FIG. 1. **SV40-Human insulin recombinant designed to express the human insulin sequences.** The insulin sequences are indicated by the *dark solid line*, and the coding sequences are indicated by *three blocks*. The details for these constructions are summarized in the text.

DNA Preparations—SV40 strain 777 DNA was used for the construction of the SV40 vectors. The *Hinc*II/*Bam*HI human insulin DNA fragment was purified from pIns96 which is a subclone of pHi300 (Bell *et al.*, 1980). All SV40 and human DNA fragments were purified by agarose gel electrophoresis. DNA fragments were eluted by shaking the gel in 0.2 M NaCl, 1 mM EDTA, 10 mM Tris, pH 7.5. Eluted DNA was filtered through GF/C filters and concentrated with butanol-1. Fragments were ligated as described by Maniatis *et al.* (1978) and SV40-insulin recombinants were cloned in pBR322 and amplified in *E. coli* (Clewell and Helinski, 1969). SV40-insulin recombinant DNA was extracted from infected COS cells as described (Hirt, 1967; Randloff *et al.*, 1967). Preparation of uniformly labeled viral DNA has been previously described (Zasloff *et al.*, 1982). End-labeled DNA probes were prepared by using T4 DNA polymerase for 3′ end labeling (O'Farrell *et al.*, 1980) and T4 polynucleotide kinase for 5′ end labeling (Maxam and Gilbert, 1980).

Analysis of RNA—Polyadenylated and nonpolyadenylated cytoplasmic mRNAs from COS cells infected at a multiplicity of 10–100 plaque-forming units/cell with the LSV[2]-insulin recombinants were isolated as described previously (Laub and Aloni, 1975). Mapping of RNAs by the S1 method of Berk and Sharp (1977) was done with labeled insulin probes prepared from LSV_{ins} DNA. After hybridization at 50–52 °C for 4 h, the DNA:RNA hybrids were digested with 1000 units of S1 nuclease, denatured with formamide, and analyzed by electrophoresis on a 5% polyacrylamide, 8 M urea gel (Maxam and Gilbert, 1980). For Northern analysis polyadenylated RNA was electrophoresed on a methyl mercury gel (Alwine *et al.*, 1977) and blotted onto nitrocellulose paper. The resulting blot was hybridized to a nick-translated insulin probe (Maniatis *et al.*, 1978), washed with 0.1 × SSC at 50 °C, and autoradiographed.

Protein Analysis—COS cells were infected with the $LSV_{ins}2$ or $LSV_{ins}C2$ virus stocks at 10–100 plaque-forming units/cell. Cells were maintained for 24 h in cysteine-depleted medium followed by 6-h labeling with 10 μCi/ml of L-[^{35}S]cysteine. Cell lysate or culture medium was immunoprecipitated with guinea pig anti-bovine insulin serum and analyzed on a 12.5% acrylamide/sodium dodecyl sulfate gel (Laemmli, 1970). Quantitative radioimmunoassays were performed as described (Rall *et al.*, 1973).

[2] The abbreviations used are: LSV_{ins}, late simian virus 40-insulin recombinant; ESV_{ins}, early simian virus 40-insulin recombinant; $LSV_{ins}C2$, late simian virus 40-insulin cDNA recombinant; $LSV_{ins}2$, late simian virus 40-insulin genomic recombinant; Pipes, 1,4-piperazinediethanesulfonic acid; pLSV, late simian virus 40 vector.

RESULTS

Construction of SV40-Human Insulin DNA Recombinants—Fig. 1 summarizes the procedure for constructing the SV40-insulin recombinants. The LSV vector was generated from an SV40 genome inserted in the *Bam*HI site of pBR322 (pSV40) and amplified in *E. coli* GM48 cells. pSV40 was linearized by partial digestion with *Hind*III restriction endonuclease (SV40 nucleotide 5107) followed by S1 treatment to produce blunt ends. The linear pSV40 DNA was digested with *Bcl*1 restriction endonuclease (SV40 nucleotide 2706) and the 7.2-kilobase pair pLSV vector was purified by preparative agarose gel electrophoresis. The resulting pLSV vector contains the SV40 origin of replication and the coding information for the SV40 late genes. Most of the coding region for the SV40 early genes (nucleotides 5107 to 2706) was deleted from the pLSV vector. The vector retains the early SV40 promoters and 105 noncoding nucleotides downstream from the 5′ end as well as 136 nucleotides prior to the polyadenylation site at the 3′ end of the early gene region. In $pLSV_{ins}C2$ a 545-bp *Bam*HI/*Eco*RI blunt-ended insulin cDNA fragment was inserted in this site. The human insulin gene including 69 nucleotides in the 5′ flanking and 119 nucleotides in the 3′ flanking region was present in the 1603-bp *Hinc*II/*Bam*HI fragment that was used in all the insulin genomic recombinants. In $LSV_{ins}2$ the 1603-bp insulin fragment was inserted 105 bp downstream from the early SV40 major cap site. In $LSV_{ins}LP$ the late SV40 leader sequences (SV40 nucleotides 234–437) are placed between the SV40 early promoter and the human insulin gene. In $LSV_{ins}LP2$, the late SV40 promoter and leader (SV40 nucleotides 140–437) were inserted between the early SV40 promoter and insulin gene. The constructions of the late SV40 replacement recombinant, ESV_{ins}, and the nonexpressing

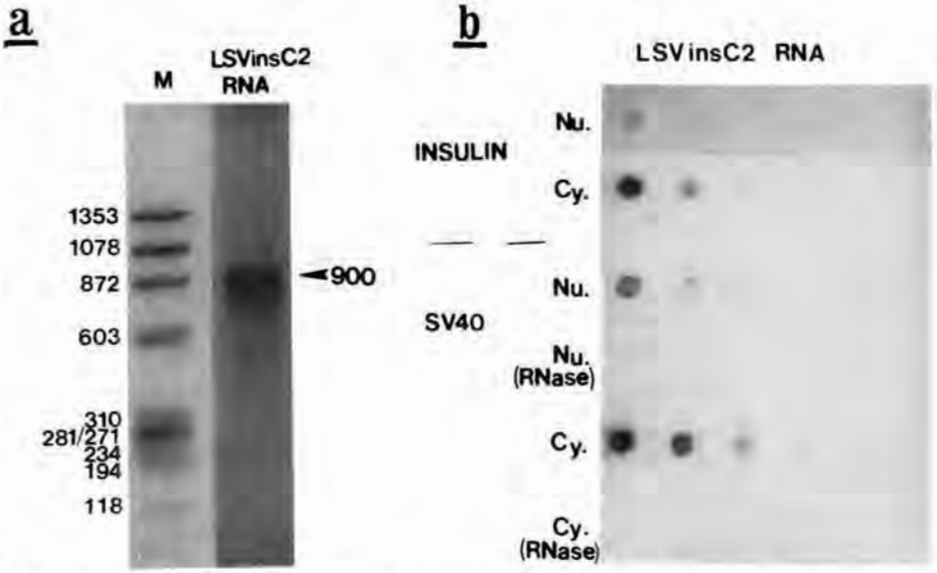

FIG. 2. **Transcription of the LSV$_{ins}$C2 recombinant.** COS cells were infected with 10–100 plaque-forming units/cell of the LSV$_{ins}$C2 virus stock and incubated for 48 h at 37 °C. Nuclear (*Nu.*) and cytoplasmic (*Cy.*) RNA was prepared as described previously (Laub and Aloni, 1975). *a*, polyadenylated cytoplasmic RNA was denatured in 10 mM methyl mercury, electrophoresed through a 5 mM methyl mercury, 1.5% agarose gel, and transferred to nitrocellulose. The resulting blots were preincubated in 50% formamide, 4 × SSC, 3 × Denhardt's, and hybridized at 42 °C to an insulin ^{32}P-labeled DNA probe. The blot was washed in 0.1 × SSC at 50 °C and autoradiographed. *Lane M* contains DNA size markers labeled with [γ-^{32}P]ATP and polynucleotide kinase (Maxam and Gilbert, 1980). *b*, 5-fold dilutions of the nuclear and cytoplasmic RNAs and controls of RNase-treated samples were spotted on nitrocellulose and hybridized to nick-translated SV40 and insulin [^{32}P]DNA probes.

LSV$_{ins}$ hybrid have been described elsewhere.[3] All insulin DNA fragments had a blunt 5′ end and a *Bam*HI 3′ end. These fragments were ligated between the *Hind*III/S1 blunt end and the *Bcl*1 end within the pLSV vector.

Ampicillin-resistant colonies were screened by hybridization to ^{32}P-labeled insulin and SV40 probes, and the positive colonies were analyzed by restriction endonuclease mapping. The resulting plasmids contain a *Bam*HI insert which includes the SV40 origin of replication, a functional set of late SV40 genes and human insulin sequences inserted in the sense direction relative to the early SV40 promoter and polyadenylation sites. These *Bam*HI fragments containing the LSV-insulin hybrids were self-ligated to form circular DNA and subsequently infected into permissive monkey COS cells (Gluzman, 1981).

Transcription of Insulin Sequences from the SV40-Insulin cDNA Recombinant (LSV$_{ins}$C2)—The intronless LSV$_{ins}$C2 recombinant produces high levels of cytoplasmic polyadenylated RNA containing insulin coding sequences. A Northern blot analysis of this RNA (Fig. 2*a*) revealed one insulin-specific band which corresponds in size (≃900 bases) to an insulin mRNA which is initiated at the SV40 early promoter and is polyadenylated at the early SV40 termination signal. This RNA contains the 748 nucleotides of nonspliced insulin coding RNA and about 150 poly(A) residues.

Because of the previous reports suggesting a role of intron removal in mRNA stability (Hamer and Leder, 1979) and transport from the nucleus to the cytoplasm (Lai and Khoury, 1979), we have assayed the steady state levels of insulin-coding sequences in the nucleus and the cytoplasm. The results (Fig. 2*b*) show that the partitioning of nonspliced insulin mRNA between the nucleus and the cytoplasm is indistinguishable from the partition of normally spliced late SV40 mRNA. Thus, in this system, there is no apparent barrier to the transport of intronless mRNA from the nucleus to the cytoplasm.

[3] Laub, O., Rall, L., Bell, G. I., and Rutter, W. J. (1983) *J. Biol.*

Transcription of Insulin-coding Sequences from SV40-Insulin Gene Recombinants—The RNA present in the cytoplasm of COS cells infected with each of the LSV-insulin recombinants was analyzed by the method of Berk and Sharp (1978). A uniformly labeled insulin-specific probe was prepared from COS cells infected with the LSV$_{ins}$ virus stock and labeled with (^{32}P)orthophosphate for 12–16 h. Viral supercoiled [^{32}P]DNA was purified (Hirt, 1967; Randloff *et al.*, 1967) and the insulin probe was isolated as a 1.7-kilobase pair *Hinc*II fragment. This labeled probe contains the 1603 nucleotides of the human insulin insert and an additional 103 bases derived from the 3′ end of the early SV40 gene. The ^{32}P probe was denatured and hybridized to poly(A$^+$) and poly(A$^-$) RNA isolated from infected COS cells under conditions favoring RNA-DNA duplexes (Casey and Davidson, 1977). The resulting hybrids were digested with S1 nuclease and analyzed on a 5% acrylamide, 8 M urea sequencing gel (Maxam and Gilbert, 1980). As shown in Fig. 3, *lane A*, poly(A) RNA from the ESV$_{ins}$ recombinant produced six bands; the band 206- and 218-nucleotides long correspond to the two insulin-coding exons and co-migrate with the S1-protected fragments obtained with human insulinoma RNA treated in a similar manner.[3] LSV$_{ins}$ is not transcribed (Fig. 3, *lane B*). This is due to the strong inhibitory effect of an SV40 sequence (map units 0.76 to 0.86) interposed between the promoter and downstream sequences. If this inhibitory fragment is removed (LSV$_{ins}$2) or replaced (LSV$_{ins}$LP and LSV$_{ins}$LP2), the insulin sequences are transcribed and expressed at high levels.

LSV$_{ins}$2, LSV$_{ins}$LP, and LSV$_{ins}$LP2 (Fig. 3, *lanes C*, *D*, and *E*, respectively) are expressed efficiently, and produce 5- to 10-fold more RNA than the late SV40 replacement recombinant. The second exon of the human insulin gene, 206 nucleotides in length, is present in the poly(A$^+$) fraction of all LSV-insulin constructions; however, exon 1 (42 bp) and exon 3 (218 bp) are not detected in any of the LSV recombinants. This result indicates that most RNA transcripts initiate and ter-

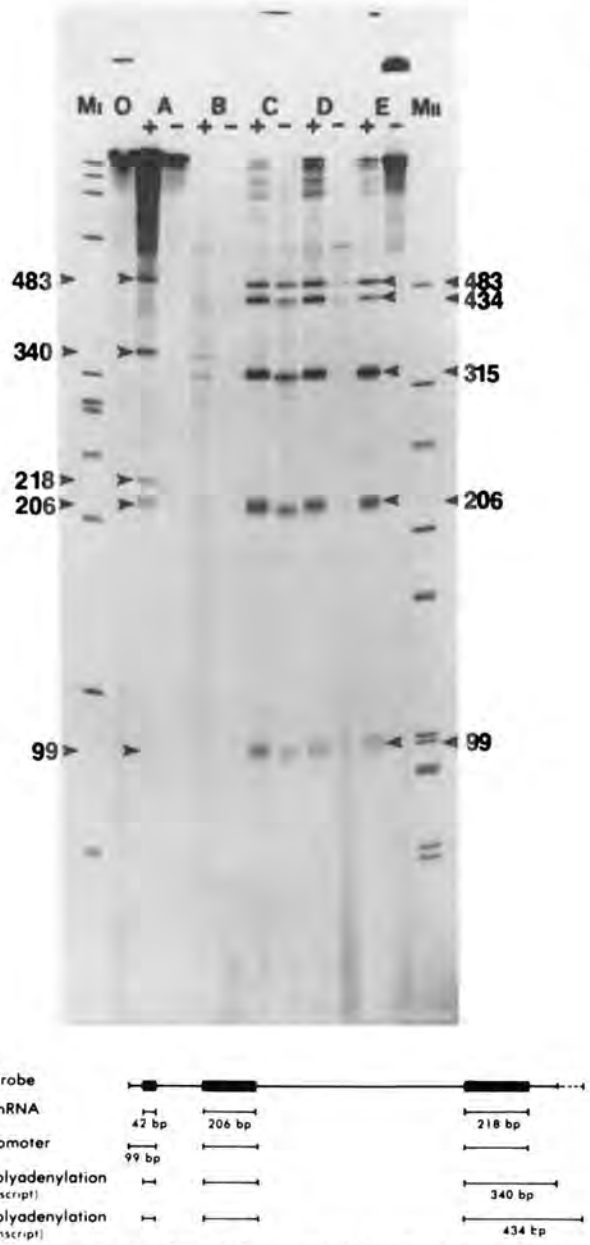

FIG. 3. **S1 analysis of RNA from SV40-insulin-infected cells.** Purified non-poly(A) (−)- and poly(A)-containing (+) cytoplasmic RNAs from infected COS cells were used. RNAs were hybridized to an insulin/*Hinc*II DNA fragment purified from uniformly labeled LSV_{ins} [^{32}P]DNA. Hybridization mixture in 80% formamide, 0.4 M NaCl, 0.01 M Pipes, pH 6.4 was denatured 2 min at 80 °C, annealed 3 h at 52 °C, digested 1 h with 1000 units of S1 nuclease, and subjected to a 5% acrylamide, 8 M urea sequencing gel. *Tracks MI* and *MII* contain size markers; *Lane A*, ESV_{ins} RNA; *B*, LSV_{ins} RNA; *C*, $LSV_{ins}2$ RNA; *D*, $LSV_{ins}LP$ RNA; *E*, $LSV_{ins}LP2$ RNA; *O*, control of probe without RNA. The *diagram* represents the predicted S1-protected fragments. The band 315 nucleotides in length was unpredicted.

length, corresponds to RNA that is initiated at the SV40 cap site and spliced at the donor site of insulin exon 1. The band, 443 ± 8 nucleotides long, corresponds to the RNA extending from the splice acceptor site of insulin exon 3 to the 3′ end of the insulin probe and therefore results from transcripts polyadenylated at the early SV40 site. The largest protected band, 483 ± 8 nucleotides in length, was also observed with the ESV_{ins} recombinant (Fig. 3, *lane A*). This fragment is derived from a partially spliced RNA in which only the second intron is removed, thus this particular RNA class contains exon 1, nucleotides in length, was not predicted; the next set of experiments was aimed at the mapping and characterization of this transcript.

*Mapping the Novel Transcript—Hae*III restriction endonuclease is one of the few restriction enzymes which cuts single-stranded DNA at its specific recognition site (CCGG). The DNA 315 nucleotides in length derived from the S1 endonuclease protection experiment was eluted from the sequencing gel and digested with an excess of *Hae*III. The cleavage products were analyzed on a 5% acrylamide, 8 M urea sequencing gel. As shown in Fig. 4, two new bands, 192 ± 3 and 123 ± 3 bases long, were detected. Only two locations within the insulin gene insert could produce a 315-nucleotide

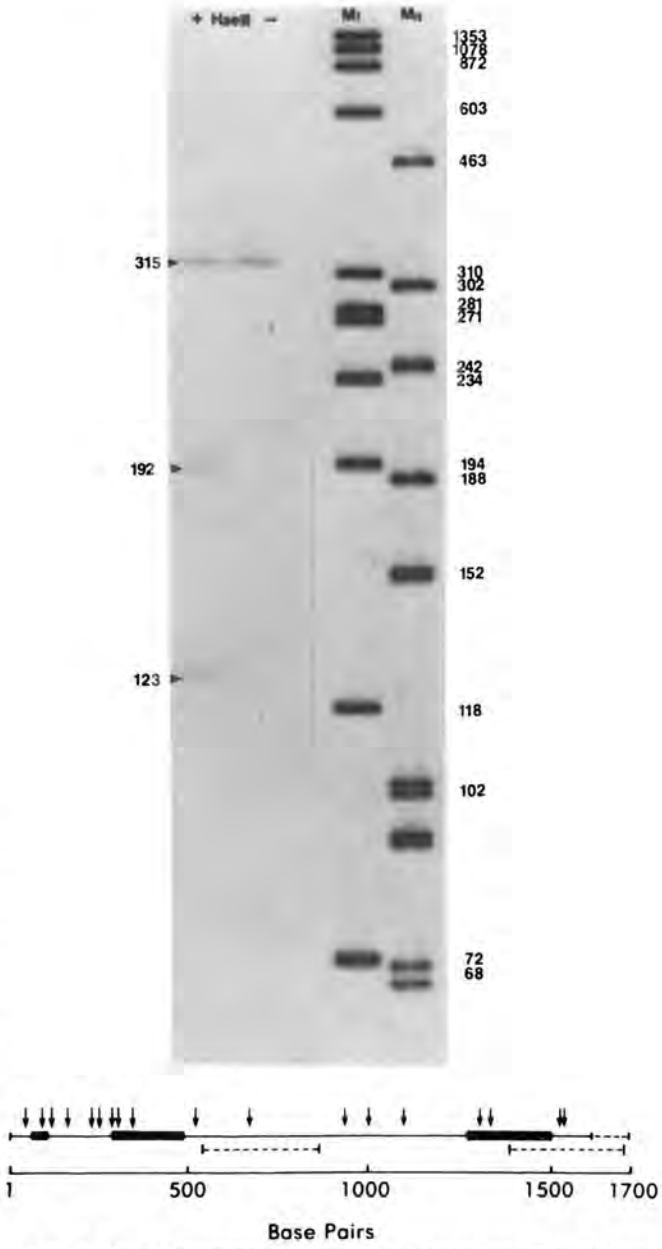

FIG. 4. ***Hae*III mapping of the 315 transcript.** The 315-base [^{32}P]DNA derived from the S1 nuclease protection experiment (Fig. 3) was eluted from the sequencing gel by passive shaking of the gel as described under "Materials and Methods." The eluted [^{32}P]DNA was digested with 3 units of *Hae*III, and the cleavage products were analyzed on a 5% acrylamide, 8 M urea sequencing gel. *Lane MI* and *MII* are end-labeled [^{32}P]DNA size markers. The *diagram* represents the *Hae*III restriction map of the insulin DNA. The *dotted lines*

fragment which, when cleaved with *Hae*III, would yield 123- and 192-bp fragments (Fig. 4). Because the *Hae*III products appear as doublets, we reasoned that the fragment must be located at the 3′ end of the insulin insert. In this region there are two *Hae*III sites separated by 6 bases and located on opposite DNA strands, thus the single-stranded DNA can form a loop that contains a double-stranded *Hae*III site, which is presumed to be the substrate for this enzyme (Bron and Murray, 1975).

Mapping the 3′ End of the New Exon—A probe for mapping the new exon was constructed from LSV_{ins} DNA digested with *Pvu*II labeled with ^{32}P using T4 DNA polymerase (O'Farrell *et al.*, 1980) and subsequently cleaved with *Hinc*II. The 445-bp *Pvu*II/*Hinc*II-labeled fragment was purified by gel electrophoresis. This DNA probe contains 98 bp derived from the third insulin exon, 117 noncoding nucleotides from the 3′ end of the insulin insert, and 227 bp derived from the 3′ end of the SV40 early genes. This [^{32}P]DNA probe was hybridized to cytoplasmic polyadenylated RNA extracted from COS cells infected with the ESV_{ins} or the $LSV_{ins}2$ virus stocks. As shown in Fig. 5, *lane A*, the ESV_{ins} transcripts protected two DNA bands; the 98 ± 2-base fragment corresponds to insulin transcripts polyadenylated at the insulin poly(A) site and the 215 ± 4-base fragment extends beyond the end of the insulin insert and reflects polyadenylation at the late SV40 poly(A) site. Hybridizing the probe with $LSV_{ins}2$ RNA (Fig. 5,

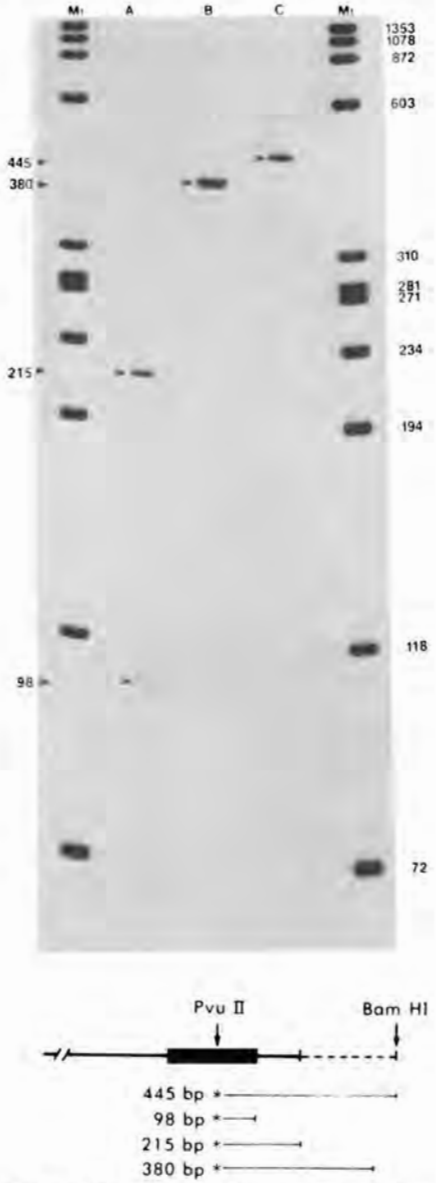

FIG. 5. **Mapping the 3′ end of the new exon.** LSV_{ins} DNA was cut with *Pvu*II restriction endonuclease and labeled with T4 DNA polymerase (O'Farrell *et al.*, 1980) using [γ-^{32}P]dCTP. The labeled DNA fragments were cut with *Hinc*II, and a 445-bp *Pvu*II/*Hinc*II-labeled fragment was purified by gel electrophoresis. This DNA probe was annealed with $LSV_{ins}2$ RNA and analyzed by S1 nuclease as described for Fig. 3. *Lane MI*, end-labeled [^{32}P]DNA size markers; *A*, S1 mapping of RNA from a late SV40-replacement recombinant $(ESV_{ins})^3$, *B*, RNA from COS cells infected with the $LSV_{ins}2$ recombinant; *C*, untreated labeled probe. The *diagram* represents the predicted S1-protected fragment, terminating at the insulin poly(A)

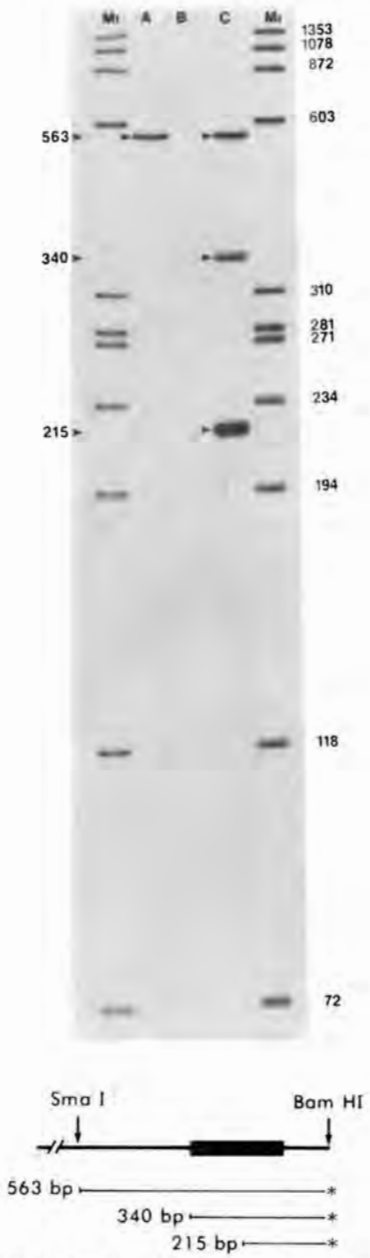

FIG. 6. **Mapping the 5′ end of the new exon.** The 1603-bp *Hinc*II/*Bam*HI insulin-coding DNA fragment was end-labeled with [γ-^{32}P]ATP and polynucleotide kinase. The labeled fragment was digested with restriction endonuclease *Sma*I and the resulting 563-bp labeled fragment was purified by gel electrophoresis. This DNA probe was annealed with $LSV_{ins}2$ RNA and analyzed by S1 nuclease as described for Fig. 3. *Lane MI*, DNA size markers; *A*, untreated probe; *B*, no RNA; *C*, S1 mapping of RNA extracted from COS cells infected

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.