Reports

Rat Insulin Genes: Construction of Plasmids Containing the Coding Sequences

Abstract. *Recombinant bacterial plasmids have been constructed that contain complementary DNA prepared from rat islets of Langerhans messenger RNA. Three plasmids contain cloned sequences representing the complete coding region of rat proinsulin I, part of the preproinsulin I prepeptide, and the untranslated 3′ terminal region of the mRNA. A fourth plasmid contains sequences derived from the A chain region of rat preproinsulin II.*

In the 55 years since the first isolation of insulin (*1*), enormous progress has been made in understanding the role of this hormone in normal glucose homeostasis and in diabetes. Little is known, however, about the control of insulin gene expression in normal and pathological states. Although insulin is composed of two polypeptide chains (A and B), it is the product of a single gene. The immediate precursor of insulin is a single polypeptide, termed proinsulin, that contains the two insulin chains A and B connected by another peptide, C (*2*). Recently it has been reported that the initial translation product of insulin messenger RNA (mRNA) is not proinsulin itself, but (another precursor preproinsulin) that contains more than 20 additional amino acids on the amino terminus of proinsulin (*3*). Thus, the structure of the preproinsulin molecule can be represented schematically as NH_2–(prepeptide)–B chain–(C peptide)–A chain–COOH.

To determine the structure of the insulin gene and to study its regulation in normal and pathological states, as well as to investigate the possibility of the synthesis of insulin in an alternate biological system such as bacteria, we have isolated the coding region of the insulin gene by cloning in bacterial plasmids the complementary DNA (cDNA) synthesized in vitro from rat insulin mRNA. We describe here the construction of these plasmids, and the nucleotide sequence of the cloned DNA. The data show that the plasmids contain the coding region for most of the translated portion of the gene for rat insulin I and the segment coding for the A chain of rat insulin II (*4*).

The general scheme used for the isolation of bacterial plasmids containing nucleotide sequences coding for rat preproinsulin is illustrated in Fig. 1, and is described below.

Isolation of cDNA complementary to rat islet mRNA. The isolation of insulin mRNA is complicated by the low proportion of the endocrine B cells (which produce insulin) in the pancreas and the high levels of ribonuclease in the dominant acinar cells. These problems have been circumvented by adapting procedures for the relatively large scale isolation of islets of Langerhans (*5*) (in which the majority of the cells are B cells) from the rat pancreas, and by using a method that allows the extraction of intact, translatable mRNA from sources rich in ribonuclease (*6*).

Polyadenylated RNA was isolated by chromatography of the total RNA preparation on oligodoxythymidylate-cellulose and transcribed into cDNA with the use of avian myeloblastosis virus (AMV) reverse transcriptase and dT_{12-18} as primer (*7*), and the RNA removed from the mRNA-cDNA hybrid by alkali treat-

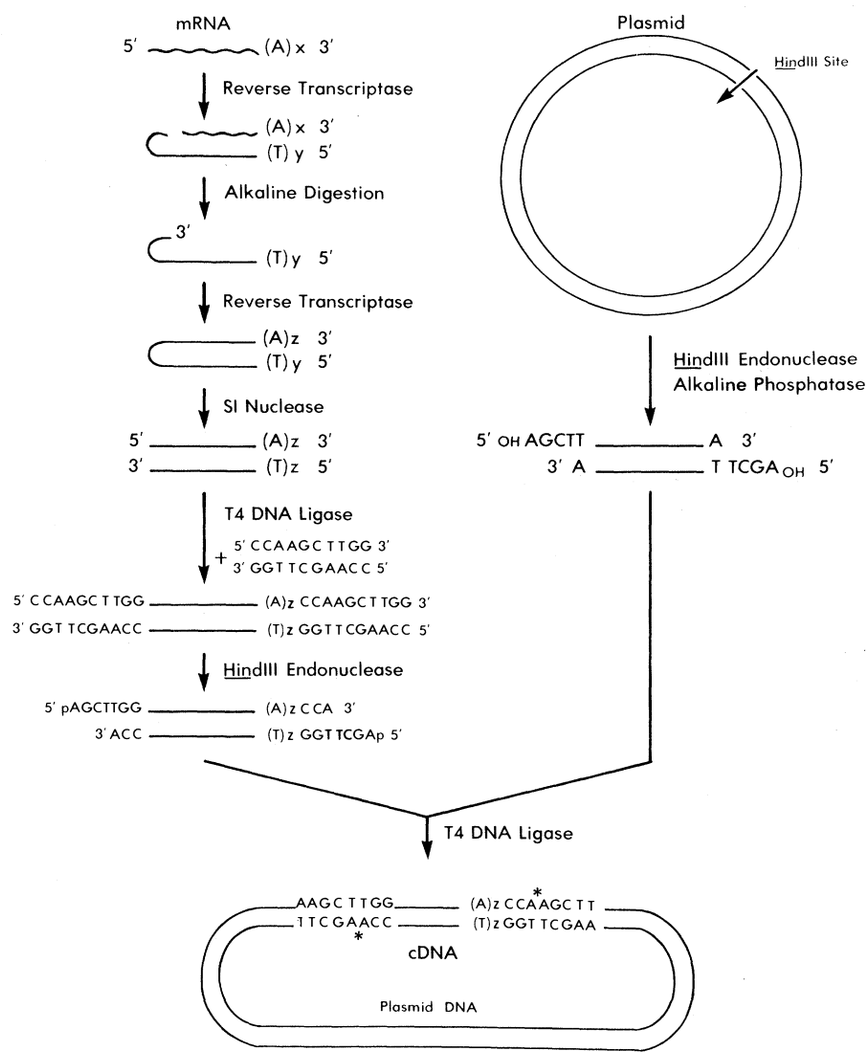

Fig. 1. Schematic diagram for insertion of cDNA into bacterial plasmids with the use of chemically synthesized restriction site linkers. The asterisks in the recombinant plasmid indicate the position where a phosphodiester bond has not formed because of the absence of a 5′ terminal phosphate. The example shown here is for the Hind III decanucleotide. In certain cases the Eco RI octanucleotide (5′)TGAATTCA(3′) was used; in these cases insertion was made into the Eco RI site of the plasmid.

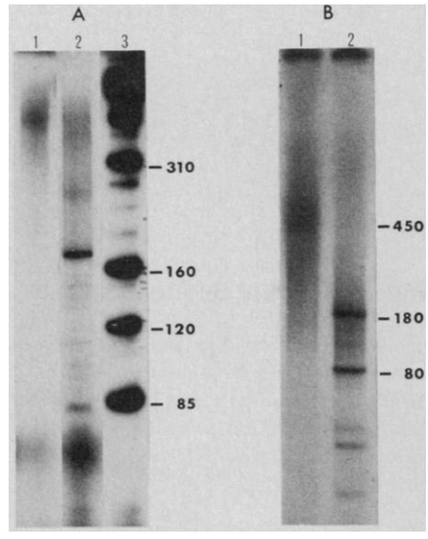

Fig. 2. Analysis of complementary DNA (cDNA) derived from total rat islet of Langerhans mRNA. Avian myeloblastosis virus (AMV) reverse transcriptase (provided by D. J. Beard, Life Sciences, Inc., St. Petersburg, Florida) was used to transcribe total polyadenylated RNA from rat islets of Langerhans into cDNA (*7*). The reactions were carried out in 50 m*M* tris-HCl, *p*H 8.3, 9 m*M* $MgCl_2$, 30 m*M* NaCl, 20 m*M* β-mercaptoethanol, 1 m*M* unlabeled deoxyribonucleoside triphosphates, 250 μ*M* α-^{32}P-labeled nucleoside triphosphate (specific activity: 50 to 200 c/mole) (each reaction contains three different unlabeled and one labeled triphosphates), 20 μg of oligo(dT_{12-18}) (Collaborative Research) per milliliter, 100 μg of polyadenylated RNA per milliter, and 220 units of reverse transcriptase per milliliter. The mixture was incubated at 45°C for 15 minutes. After addition of the disodium salt of EDTA (EDTA-Na_2) to a concentration of 25 mmole/liter, the solution was extracted with an equal volume of water-saturated phenol, followed by chromatography of the aqueous phase on a Sephadex G-100 column (0.3 by 10 cm) in 10 m*M* tris-HCl, *p*H 9, 100 m*M* NaCl, 2 m*M* EDTA. Nucleic acid eluted in the void volume was precipitated with ethanol after addition of ammonium acetate, *p*H 6.0, to a concentration of 0.25*M*. After centrifugation, the pellet was dissolved in 50 μl of 0.1*M* NaOH (freshly prepared) and the RNA was hydrolyzed at 70°C for 20 minutes. Sodium acetate (1*M*, *p*H 4.5) was added for neutralization, and the ^{32}P-labeled cDNA was precipitated with ethanol and dissolved in water. This cDNA was 5 to 8 percent resistant to digestion by S1 nuclease at 45°C (*13*). Synthesis of the second strand was performed in 50 m*M* tris-HCl, *p*H 8.3, 9 m*M* $MgCl_2$, 10 m*M* dithiothreitol, 500 μ*M* each of the three unlabeled deoxyribonucleoside triphosphates, 100 μ*M* α-^{32}P-labeled deoxynucleoside triphosphate (specific activity: 1 to 10 c/mmole), 50 μg of cDNA per milliliter, and 220 units of reverse transcriptase per milliliter at 45°C for 120 minutes. The reaction was stopped by addition of EDTA-Na_2 to a concentration of 25 m*M*; the mixture was extracted with phenol and subjected to chromatography on Sephadex G-100, and the nucleic acids were precipitated by ethanol. Portions (5,000 to 10,000 count/min) of single- or double-stranded cDNA were digested with an excess of restriction endonuclease Hae III (*9*) and analyzed by electrophoresis on polyacrylamide slab gels (*8*). The gels were dried, and the ^{32}P-labeled DNA was detected by autoradiography with the use of Kodak No-Screen (NS-2T) film. (A) Lane 1, total single-stranded cDNA transcribed from mRNA isolated from islets of Langerhans from the rat pancreas; lane 2, Hae III cleavage pattern of this cDNA; and lane 3, single-stranded Hae III fragments of M13 DNA. (B) Lane 1, double-stranded cDNA from mRNA isolated from islets of Langerhans from the rat pancreas; and lane 2, Hae III fragments of this cDNA. Sizes were determined with unlabeled Hae III fragments of SV40 DNA as markers (*28*).

ment. This cDNA was heterodisperse, as judged by polyacrylamide gel electrophoresis (*8*). It contained at least one predominant cDNA species of about 450 nucleotides (Fig. 2A, lane 1), which gave rise to two prominent bands on cleavage with the restriction endonuclease Hae III (Fig. 2A, lane 2) (*9*). The single-stranded cDNA was converted to the double-stranded form with AMV reverse transcriptase, labeled deoxynucleoside triphosphates, and the self-priming ability of single-stranded cDNA (see Fig. 1) (*7*). The product of this reaction is known to consist of a long hairpin loop structure (*10*). Digestion of this DNA (Fig. 2B, lane 1) by Hae III restriction endonuclease gave fragments (Fig. 2B, lane 2) corresponding to those obtained by digestion of the single-stranded cDNA. It seemed likely that the Hae III fragments were derived from insulin cDNA because of their prominence in the total cDNA preparation. Therefore, these fragments as well as the complete cDNA preparations were used in the cloning experiments.

Molecular cloning of cDNA fragments. We have developed an improved method of cloning cDNA into bacterial plasmids that involves the ligation of chemically synthesized restriction site linkers to cDNA and then cleavage with the appropriate restriction endonuclease to produce cDNA molecules with cohesive termini for ligation to similarly cleaved plasmid DNA. This approach, illustrated in Fig. 1, was first used for the

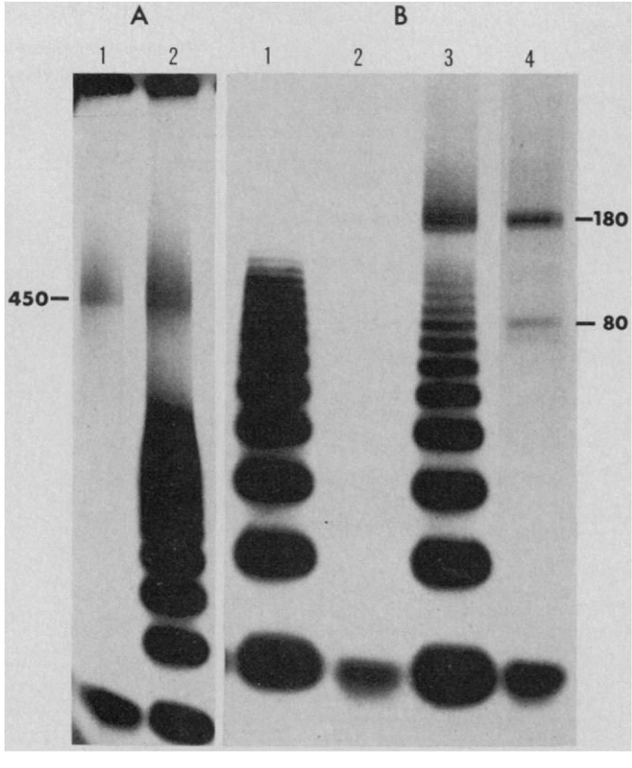

Fig. 3. Blunt-end ligation of Hind III decanucleotide linkers to rat islet cDNA. Double-stranded cDNA (2 to 5 μg/ml) was treated with 30 units of S1 nuclease (1200 units per milliliter; Miles) in 0.03*M* sodium acetate, *p*H 4.6, 0.3*M* NaCl, 4.5 m*M* $ZnCl_2$ at 22°C for 30 minutes followed by 15 minuters at 10°C. Addition of tris base to 0.1*M*, EDTA to 25 m*M*, and *Escherichia coli* transfer RNA to 50 μg/ml was used to stop the digestion. After phenol extraction of the reaction mixture and Sephadex G-100 chromatography, the excluded ^{32}P-labeled cDNA was precipitated with ethanol. This treatment results in a high yield of cDNA molecules with base-paired ends necessary for the blunt-end ligation to chemically synthesized decanucleotides. Double-stranded cDNA that was not treated with S1 nuclease but fragmented by Hae III endonuclease to generate blunt-ended fragments (*9*) was also ligated as follows. Ligation of Hind III decamers to cDNA was carried out by incubation at 14°C in 66 m*M* tris-HCl, *p*H 7.6, 6.6 m*M* $MgCl_2$, 1 m*M* ATP, 10 m*M* dithiothreitol m*M*, 3 μ*M* 5′-^{32}P-labeled Hind III decamers (10^6 count/min per picomole), and T4 DNA ligase (~ 500 units per milliliter) for 1 hour. The reaction mixture was then heated to 65°C for 5 minutes to inactivate the ligase; KCl (to 50 m*M*), β-mercaptoethanol (to 1 m*M*), and EDTA (to 0.1 m*M*) were then added, and the mixture was digested with 150 units of Hind III endonuclease per milliliter for 2 hours at 37°C. (A) Lanes 1 and 2, Hind III decamer ligation to total cDNA derived from rat islets of Langerhans RNA before (lane 2) and after (lane1) cleavage with Hind III endonuclease. (B) Lanes 1 and 2, Hind III decamer ligation without added cDNA before (lane 1) and after (lane 2) Hind III endonuclease digestion. Lanes 3 and 4, Hind III decamer ligation to Hae III fragments of total cDNA derived from rat islets of Langerhans RNA before (lane 3) and after (lane 4) Hind III endonuclease digestion. Autoradiograph B has been enlarged in order to show ligation of multiple Hind III decamers to the 180 base pair Hae III fragment. SV40 DNA digested by Hae III has been used to determine the sizes of the cDNA Hae III fragments (*28*).

cloning of synthetic DNA fragments (*11*) and makes use of the ability of T4 DNA ligase to catalyze the joining of blunt-ended DNA molecules (*12*).

Since the double-stranded cDNA molecules generated by self-priming of the single strand are in the form of long hairpin structures, it was first necessary to remove the hairpin and any non-base-paired regions, so that the final product would be a perfectly base-paired duplex DNA. This was achieved by digestion of the double-stranded cDNA with the single-strand specific nuclease S1 (*13*); it was not necessary for the cloning of the cDNA–Hae III fragments since Hae III cleavage itself produces molecules with base-paired ends (*9*). The ligation of blunt-ended cDNA to an excess of the Hind III restriction site decamer (5′)CCAAGCTTGG(3′) (*14*) is shown in Fig. 3 (C, cytosine; A, adenine; G, guanine; T, thymine). This decamer was chosen since neither the major com-

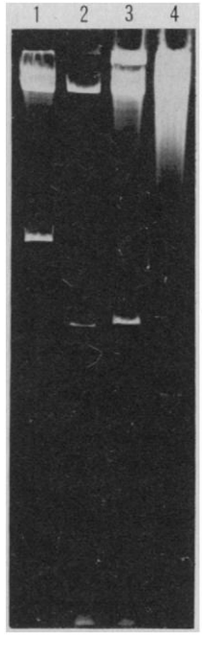

Fig. 4. Restriction endonuclease analysis of the isolated recombinant plasmids. Crude plasmid preparations (2 to 5 μg) were digested with an excess of the appropriate restriction endonuclease (*19a*). EDTA-Na_2 (10 m*M*) and sucrose (10 percent final concentration) were then added, and the mixture was subjected to electrophoresis on an 8 percent polyacrylamide gel. Hind III digest of pAU-1 (lane 1), Eco RI digests of pAU-2 (lane 2), pAU-3 (lane 3), and pAU-4 (lane 4), respectively. The stained material at the bottom of lanes 2 and 3 is RNA that is still present in the crude extract.

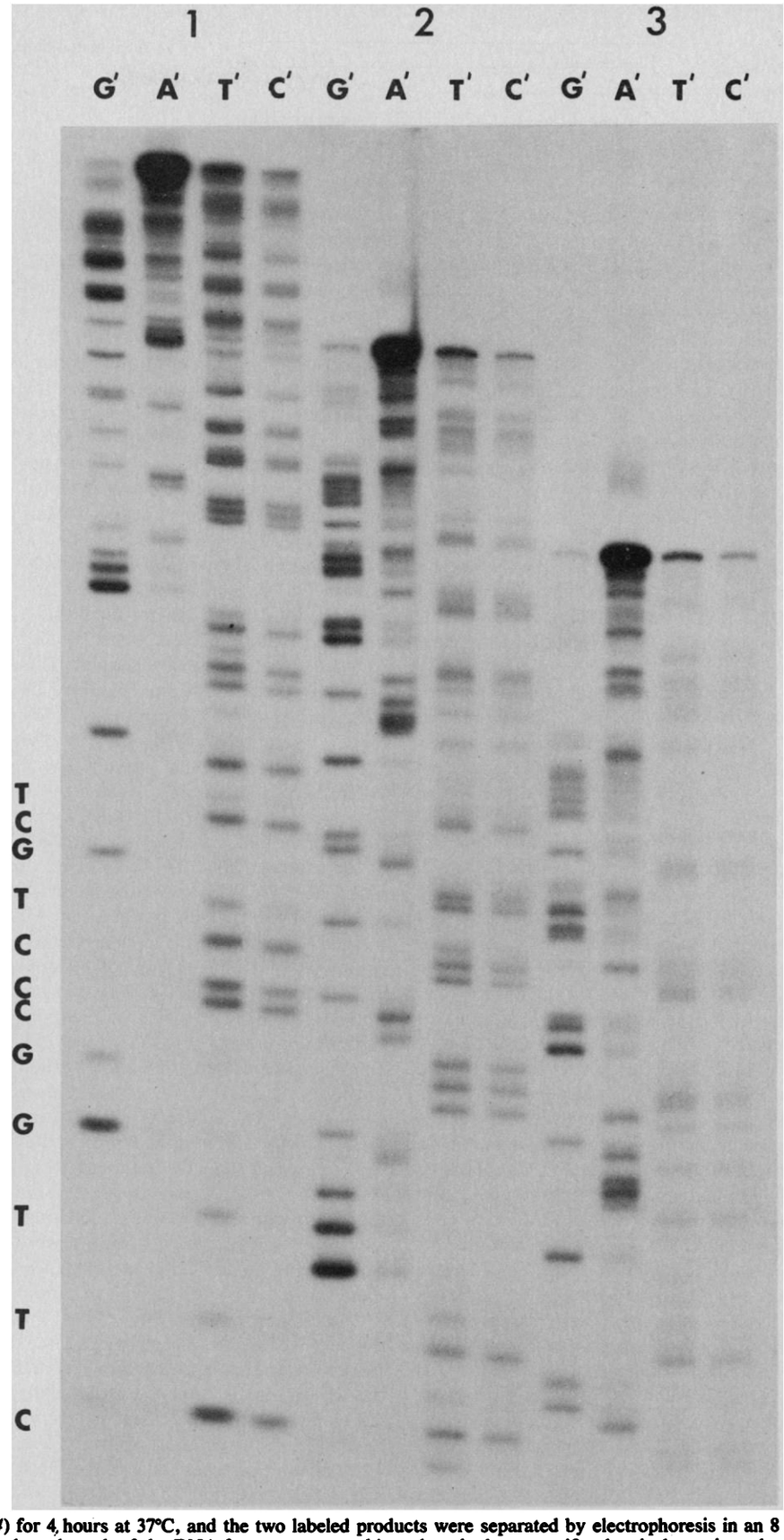

Fig. 5. Autoradiogram of a sequence gel for pAU-2. The 180 base pair DNA fragment released by Hind III digestion of pAU-2 was isolated by electrophoresis in an 8 percent polyacrylamide gel. After being stained with ethidium bromide, the cloned DNA fragment was excised and eluted electrophoretically. After elution, the 5′ termini of the fragment were labeled with [γ-^{32}P]ATP and polynucleotide kinase (*21*). The terminally labeled DNA was then digested with Alu I endonuclease (*14*) for 4 hours at 37°C, and the two labeled products were separated by electrophoresis in an 8 percent polyacrylamide gel. After elution from the gel, each of the DNA fragments was subjected to the base-specific chemical reactions described by Maxam and Gilbert (*21*). The products of each of the four reactions (G′, A′, T′, C′) were separated by electrophoresis in a 20 percent polyacrylamide gel containing 7*M* urea, and fragments labeled at the 5′ terminus were visualized by autoradiography. Identical samples from each reaction were placed on the gels and run for 36, 24, and 12 hours respectively, in the order 3 to 1. The nucleotide sequence of the larger fragment generated by Alu I digestion can be read directly from the autoradiogram as shown on the left figure. The 5′ terminal pentanucleotide sequence (5′)CTTGG is derived from the 3′-end of the Hind III linker. The 5′-terminal dinucleotide (pAG) has been run off the gel.

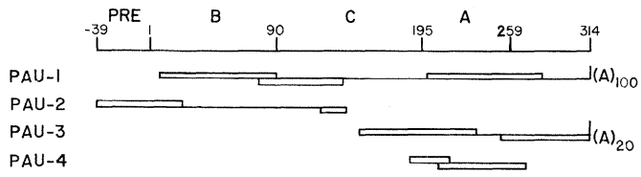

Fig. 6. Schematic diagram of rat preproinsulin mRNA and the corresponding cloned DNA fragments. The boxed areas on each DNA fragment represent the regions directly sequenced and correspond to the italic type in Fig. 7.

ponent of the total islet cDNA nor the two Hae III fragments were themselves cleaved by Hind III endonuclease (data not shown). The self-ligation of the decamers is shown in Fig. 3B, lane 1, and the ligation of multiple decamers to full-length cDNA and to cDNA cleaved by Hae III endonuclease are shown in Fig. 3A, lane 2, and Fig. 3B, lane 3, respectively. Treatment of the ligation mixtures with Hind III endonuclease yields duplex cDNA with Hind III single-stranded cohesive termini (Fig. 3A, lane 1; Fig. 3B, lane 4) as well as the cleaved decanucleotide (Fig. 3B, lane 2). Since the cleaved decamers also contained Hind III termini, and hence would compete with cDNA for ligation to the plasmid, the cDNA was purified by polyacrylamide gel electrophoresis (*8*) before ligation to the plasmid.

We used the bacterial plasmid pMB9, a 3.5-million-dalton molecule containing single Hind III and Eco RI sites (*15*). Infection of *Escherichia coli* with pMB9 confers tetracycline resistance and colicin immunity. Since the Hind III site is localized within the promoter for the gene responsible for tetracycline resistance, not all bacteria containing a plasmid formed by insertion of DNA in the Hind III site are tetracycline-sensitive, but some retain variable levels of tetracycline resistance (*16*).

To ensure ligation of most of the cDNA molecules to plasmid DNA, it is necessary to add a molar excess of the plasmid DNA. However, this results in the majority of plasmids circularizing without an inserted cDNA fragment, and thus the subsequently transformed cells contain mainly pMB9 and not the recombinant plasmids. In order to reduce the number of colonies to be screened for recombinant plasmids, we first treated the Hind III–cut pMB9 DNA with alkaline phosphatase (see Fig. 1) (*17*). This removes the 5′ terminal phosphates from the Hind III endonuclease-generated ends of the plasmid and prevents self-ligation of the plasmid DNA, ensuring that circle formation (and hence transformation) is dependent on the insertion of a DNA fragment containing 5′-phosphorylated termini.

The mixtures containing recombinant cDNA–pMB9 were used to transform the EK2 host *E. coli* χ1776 (*18*). Transformants were selected by growth on medium containing tetracycline (*16, 19*) and screened for recombinant plasmids (*19a*). One combinant plasmid (pAU-1) obtained by transformation with the total rat islet cDNA contained an inserted DNA fragment approximately 410 nucleotides in length, which was released from the plasmid by Hind III endonuclease digestion (Fig. 4, lane 1). This cloned DNA fragment, which hybridized to rat islet cDNA (data not shown), was isolated and subjected to DNA sequence analysis (see below). Transformation with the electrophoretically purified Hae III fragments (Fig. 3b, lane 4) did not yield any clones in the Hind III site of pMB9 (*16*). The Hind III termini were therefore converted to Eco RI cohesive termini. This was achieved by "filling in" the Hind III ends using reverse transcriptase, followed by blunt-end ligation to the Eco RI restriction site octanucleotide (5′)TGAATTCA(3′) (*20*). Cleavage of this duplex cDNA with Eco RI yields cohesive termini, thus allowing insertion of the fragment into the Eco RI site of pMB9 (*20*). Transformations by means of the electrophoretically purified Hae III fragments treated in this manner yielded a number of clones, all of which contained insertions of foreign DNA at the Eco RI site (*16*). The size of these insertions corresponded to that of the original Hae III fragments used for ligation to the plasmid DNA. Two recombinant plasmids, pAU-2 and pAU-3, obtained by transformation with the approximately 80-base-pair Hae III fragments (Fig. 3B, lane 4) and one, pAU-4, from transformation with the smaller Hae III fragment (80 base pairs; Fig. 3B, lane 4) were isolated (Fig. 4, lanes 2, 3, and 4, respectively).

Sequence analysis of cloned DNA. Purified plasmid DNA from pAU-1 was cleaved with endonuclease Hind III and the 410-base-pair insertion isolated by electrophoresis on a 6 percent polyacrylamide gel (similar to those shown in Fig. 4). After elution from the gel, the DNA was labeled at the 5′ termini by incubation with [γ-^{32}P]ATP and polynucleotide kinase (*21*). The labeled DNA was cleaved with Hae III endonuclease and the two labeled fragments (about 265 and 135 base pairs) separated on a polyacrylamide gel. The isolated fragments were subjected to the base specific cleavage reactions developed by Maxam and Gilbert (*21*). The cleaved DNA was separated on a 20 percent polyacrylamide gel containing 7*M* urea and analyzed (*21*). By these methods, the nucleotide sequence of the DNA can be read directly from the gel (for example, Fig. 5).

The nucleotide sequences of the smaller cloned DNA fragments from plasmids pAU-2 (Fig. 5), pAU-3, and pAU-4 were similarly determined after isolation of the inserted DNA fragments from a polyacrylamide gel.

Rat insulin gene sequences are contained in the DNA clones. The determined nucleotide sequences of cloned DNA's from plasmids pAU-1, -2, -3 are clearly overlapping (Figs. 6 and 7) and thus represent regions of the same molecule. The reconstructed sequence is a continuous stretch of 354 nucleotides, terminated at one end by varying lengths of poly(dA · dT). The 5′ portion of the mRNA is not established by these experiments (*22*). Since the mRNA contains a terminal poly(A) sequence, the DNA strand containing 3′ terminal poly(dA) is of the same sense. This strand determines an amino acid sequence which exactly corresponds to the entire coding region for rat proinsulin I and 13 out of 23 amino acids of the prepeptide sequence. The nucleotide sequences of the mRNA's therefore confirm the previously determined amino acid sequence for rat proinsulin I (*4*), and in addition show that the C peptide is connected to the B chain by the sequence -Arg-Arg-(arginine-arginine) and to the A chain by a -Lys-Arg-(lysine-arginine) sequence. Previous data did not provide assignments for the basic arginine or lysine residues connecting the B-C-A peptides, although they had been inferred by analogy with the bovine, porcine, and human proinsulin sequences (*4*).

The portion of the amino acid sequence of the prepeptide determined from the available mRNA sequence (Fig. 8) is consistent with the partial amino acid sequence reported by Chan *et al.* (*3*). In agreement with these workers we find leucine at positions −9, −11, and −12, but we find alanine at −13, where they had made a tentative assignment of lysine on the basis of a single sequencer run. The mRNA sequence also estab-

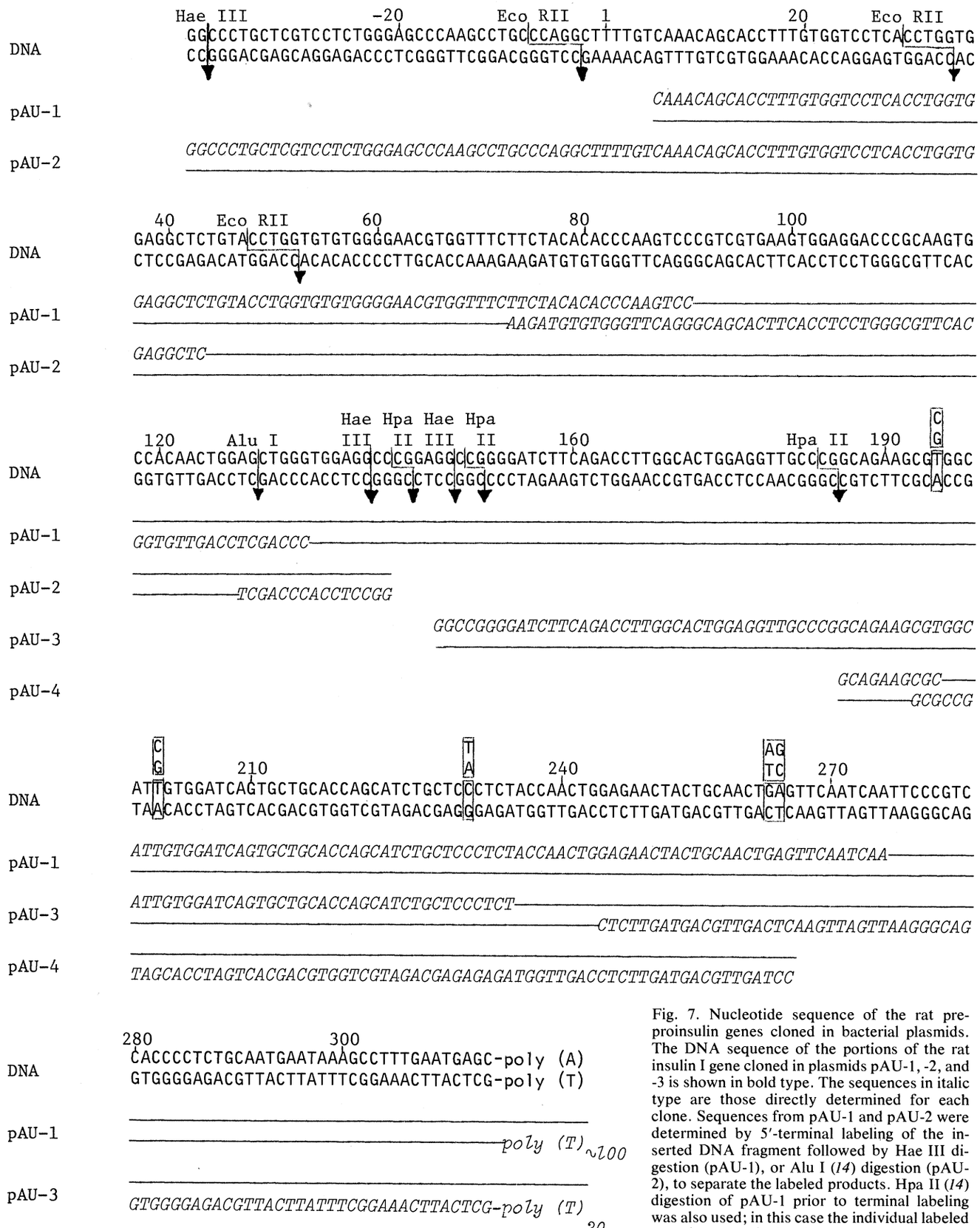

Fig. 7. Nucleotide sequence of the rat preproinsulin genes cloned in bacterial plasmids. The DNA sequence of the portions of the rat insulin I gene cloned in plasmids pAU-1, -2, and -3 is shown in bold type. The sequences in italic type are those directly determined for each clone. Sequences from pAU-1 and pAU-2 were determined by 5′-terminal labeling of the inserted DNA fragment followed by Hae III digestion (pAU-1), or Alu I (*14*) digestion (pAU-2), to separate the labeled products. Hpa II (*14*) digestion of pAU-1 prior to terminal labeling was also used; in this case the individual labeled strands were isolated by denaturation and polyacrylamide gel electrophoresis (*21*). Sequences from pAU-3 were also determined by 5′-labeling of the cloned DNA fragment followed by strand separation. pAU-4 was labeled at the 5′ terminus and treated with Hha I; the labeled fragments were then resolved by polyacrylamide gel electrophoresis prior to sequencing. The exact number of dT residues at the end of the pAU-3 insert was determined from the sequencing gel, whereas it was not possible to determine precisely the corresponding number in the pAU-1 clone. The boxed base pairs indicate the differences between the nucleotide sequence of proinsulin I and proinsulin II (above the line) in the A chain region. The restriction sites that were used for the DNA sequence analysis are shown together with the Eco RII sites (*14*). The latter presumably are methylated in the χ1776 strain at the starred position in the sequences CC*TGG or CC*AGG, thereby blocking the dimethyl sulfate cleavage reaction at this point during DNA sequencing. Only those restriction sites confirmed by direct digestion are shown.

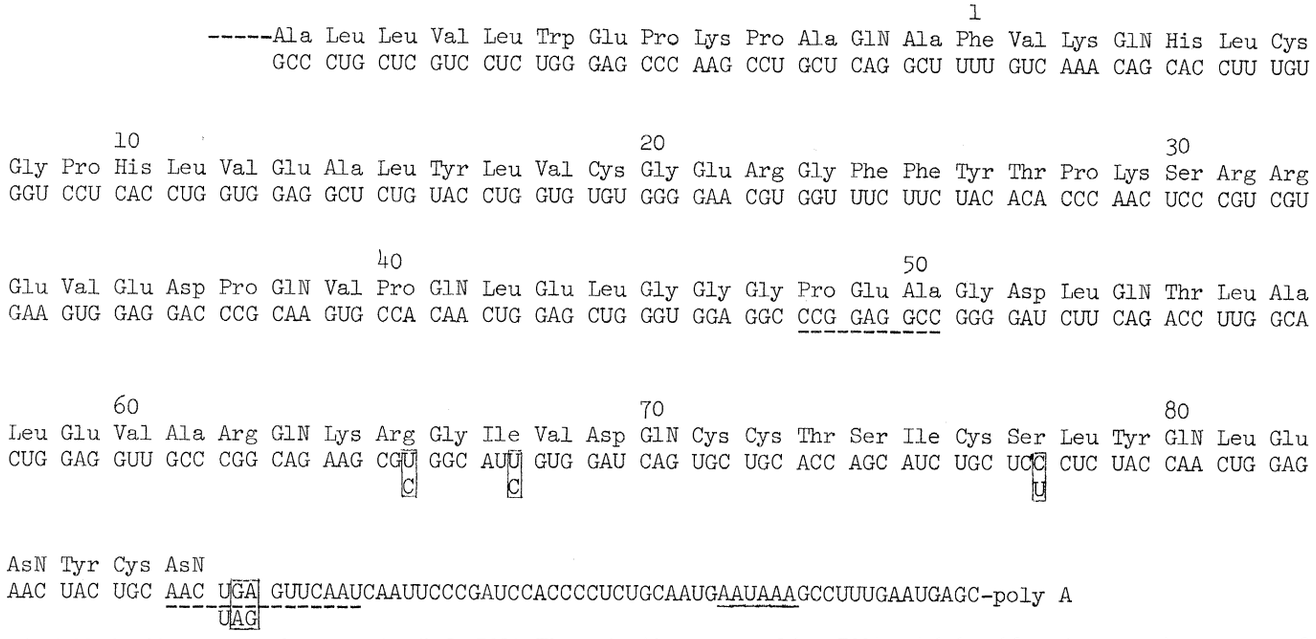

Fig. 8. Nucleotide sequence of rat preproinsulin I mRNA. The nucleotide sequence of the mRNA was deduced from the composite sequences of the cloned DNA fragments as shown in Fig. 7. The amino acid sequence was predicted from the mRNA sequence and agrees with that previously published for rat proinsulin I (*4*). The amino acid sequence of the prepeptide (−1 to −13) determined from the nucleotide sequence of pAU-2 was previously unknown except for amino acids −1 to −13, which are identical to previously published data except for −10 [see text and (*3*)]. The underlined sequence (AAUAAA) is that previously found in the 3′ untranslated region of eukaryotic mRNA's (*27*). The sequences underlined with a broken line have not been fully confirmed. The sequence shown is for preproinsulin I mRNA; nucleotide changes observed in the A chain region of proinsulin II are boxed below the line.

lishes the amino acid sequence at residue −10 and −8 to −1, which have not been previously determined, thereby demonstrating the utility of determination of the amino acid sequence via the sequence of mRNA or its cDNA.

The junction of the prepeptide with rat proinsulin I is at an alanine-phenylalanine (Ala-Phe) bond (Fig. 8). This bond must be enzymatically cleaved to form proinsulin. A protease specific for the Ala-Phe bond has not yet been identified, but it is noteworthy that all of the exocrine secretory proteins (that is, digestive enzymes) whose prezymogens have been partially sequenced from dog pancreas have an Ala-Phe sequence near the end of the prepeptide (*23*). In addition, 15 percent of all of the dipeptide sequences in the precursor portions are either Ala-Phe or Ala-Tyr; thus a protease specific for these sequences could rapidly remove and degrade the prepeptides from the secretory proteins (*24*). This processing mechanism is consistent with the proposed common origin of exocrine and endocrine cells (*25*).

Previous studies have demonstrated the presence of two types of proinsulin (I and II) in the rat (*4*). Rat proinsulin II has four amino acids (two in the B chain and two in the C chain) that differ from the amino acid sequence of proinsulin I. No clones have thus far been obtained that encompass these regions of proinsulin II. The cloned cDNA fragment in pAU-4, however, may represent the A chain region of proinsulin II mRNA. The original cDNA fragment which gave rise to this clone was clearly transcribed from the A chain region of insulin mRNA. Since the amino acid sequence of the A chain of insulins I and II are identical but the nucleotide sequence of this fragment is slightly different from the corresponding region of proinsulin I mRNA, we infer that it is derived from proinsulin II mRNA. There are three base changes in the amino acid coding portion of the gene in the proinsulin II clone; all of these involve interconversion of C and T residues in the third position of the codons (Fig. 8). In one case such a change has produced an Hha I restriction site (nucleotide 195, Fig. 7). A further difference is found in the termination codons: for rat proinsulin II it is UAG, whereas for proinsulin I, it is UGA (*26*) (U, uracil).

An examination of the coding regions of the mRNA's (Fig. 8) suggests there may be some bias in the use of particular codons; for example, for glutamic acid (two possible codons), GAG is used in 8 out of 11 occurrences; for leucine (six codons), CUG is used 9 out of 16 times, and UUA and CUA are not used; for tyrosine (two codons), UAC is used 6 out of 6 times; and for valine (four codons), GUG is used 6 out of 9 times.

The sequence AAUAAA (nucleotides 295 to 300, Fig. 8), which has been reported to occur in many eukaryotic mRNA's in a similar position in the 3′ untranslated region (*27*), is also present in the corresponding region of rat preproinsulin I mRNA.

Our results demonstrate the utility of molecular cloning in analyzing gene structures. These clones containing insulin gene segments should be very useful for studying regulation of insulin mRNA biosynthesis and its expression in bacteria.

AXEL ULLRICH, JOHN SHINE
JOHN CHIRGWIN, RAYMOND PICTET
EDMUND TISCHER, WILLIAM J. RUTTER
HOWARD M. GOODMAN
Department of Biochemistry and Biophysics, University of California, San Francisco, 94143

References and Notes

1. F. G. Banting and C. H. Best, *J. Lab. Clin. Med.* 7, 251 (1922).
2. D. F. Steiner, D. Cunningham, L. Spigelman, B. Aten, *Science* **157**, 697 (1967).
3. S. J. Chan, P. Keim, D. F. Steiner, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 1964 (1976); P. T. Lomedico and G. F. Saunders, *Nucleic Acids Res.* **3**, 381 (1976).
4. There are two nonallelic genes, I and II, for insulin in the rat. L. F. Smith, *Am. J. Med.* **40**, 662 (1966); J. L. Clark and D. F. Steiner, *Proc. Natl. Acad. Sci. U.S.A.* **62**, 278 (1969); J. Markussen and F. Sundby, *Eur. J. Biochem.* **25**, 153 (1972).
5. The procedure of P. E. Lacy and M. Kostianovsky, [*Diabetes* **16**, 35 (1967)], was modified by digesting each pancreas with 12 mg of collagenase (Worthington CLS IV) plus 1 mg of soybean trypsin inhibitor. Islets were isolated on Ficoll gradients as described by D. W. Scharp, C.

B. Kemp, M. J. Knight, W. F. Ballinger, P. E. Lacy [*Transplantation* **16**, 686 (1973)].
6. Islets were homogenized in 4*M* guanidinium thiocyanate (Tridom/Fluka), 1*M* mercaptoethanol, buffered to *p*H 5.0 (J. M. Chirgwin, A. Pryzybyla, W. J. Rutter, in preparation). The homogenate was layered over 5.7*M* cesium chloride, and the RNA was sedimented as described by V. Glisin, R. Crkvenjakov, C. Byus, [*Biochemistry* **13**, 2633 (1974)].
7. H. Aviv and P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1408 (1972); A. Efstratiadis, F. C. Kafatos, A. M. Maxam, T. Maniatis, *Cell* **7**, 279 (1976).
8. C. W. Dingman and A. C. Peacock, *Biochemistry* **7**, 659 (1968).
9. K. Horiuchi and N. D. Zinder, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2555 (1975); R. W. Blakesley and R. D. Wells, *Nature (London)* **257**, 421 (1975).
10. W. A. Salser, *Ann. Rev. Biochem.* **43**, 923 (1974).
11. H. L. Heyneker, J. Shine, H. M. Goodman, H. W. Boyer, J. Rosenberg, R. E. Dickerson, S. A. Narang, K. Itakura, S. Lin, A. D. Riggs, *Nature (London)* **263**, 748 (1976).
12. V. Sgaramella, J. H. Van de Sande, H. G. Khorana, *Proc. Natl. Acad. Sci. U.S.A.* **67**, 1468 (1970); A. Sugina, N. Cozzarelli, H. Heyneker, J. Shine, H. Boyer, H. M. Goodman, *J. Biol. Chem.*, in press.
13. V. M. Vogt, *Eur. J. Biochem.* **33**, 192 (1973).
14. R. H. Scheller, R. E. Dickerson, H. W. Boyer, A. D. Riggs, K. Itakura, *Science* **196**, 177 (1977); Hsu I endonuclease was used interchangeably with Hind III; R. J. Roberts, *Crit. Rev. Biochem.* **3**, 123 (1976).
15. R. L. Rodriguez, F. Bolivar, H. M. Goodman, H. W. Boyer, M. Betlach, in *ICN/UCLA Symposium on Molecular Mechanisms in the Control of Gene Expression*, D. P. Nierlich, W. J. Rutter, C. F. Fox, Eds. (Academic Press, New York, 1976), pp. 471–477.
16. Insertion of DNA fragments into the Hind III site of pMB9 reduces the level of tetracycline resistance of cells carrying such recombinant plasmids to varying degrees dependent on the sequences cloned into this site [R. L. Rodriguez, R. Tait, J. Shine, F. Bolivar, H. Heyneker, M. Betlach, H. W. Boyer in *Tenth Annual Miami Winter Symposium* (Academic Press, New York, in press)]. We have previously observed that the insertion of DNA molecules containing poly dA · dT regions allows the expression of tetracycline resistance at reduced levels (5 to 10 μg/ml). Screening for recombinant plasmids was therefore carried out at 5 μg of tetracycline per milliliter for transformation into the Hind III site, and 20 μg of tetracycline per milliliter for transformation into Eco R1 site.
17. Plasmid DNA cleaved by Hind III or Eco RI was treated with bacterial alkaline phosphatase (Worthington, BAPF, 0.1 unit per microgram of DNA) at 65°C in 25 m*M* tris-HCl, *p*H 8, for 30 minutes, followed by phenol extraction to remove the phosphatase. After precipitation by ethanol, the phosphatase-treated plasmid DNA was added to cDNA (containing Hind III or Eco RI cohesive terminals) at a molar ratio of 3 : 1 (plasmid : cDNA). The mixture was incubated in 66 m*M* tris, *p*H 7.6, 6.6 m*M* $MgCl_2$, 10 m*M* dithiothreitol, 1 m*M* adenosine triphosphate (ATP) for 1 hour at 14°C in the presence of 50 units of T4 DNA ligase per milliliter. The ligation mixture was added directly to the χ1776 cells for transformation.
18. A modification (W. Salser, personal communication) of the transformation procedure originally provided to all recipients of χ1776 by R. Curtiss III was used. The work with this strain was done in a P3 physical containment facility in the EK-2 host-vector system *E. coli* χ1776–pMB9 in compliance with the NIH guidelines for recombinant DNA research.
19. F. Bolivar, R. L. Rodriguez, M. C. Betlach, H. W. Boyer, *Gene*, in press.
19a. R. B. Meagher, R. C. Tait, H. Betlock, H. W. Boyer, *Cell* **10**, 521 (1977).
20. Complementary DNA containing Hind III cohesive terminals was incubated at 37°C for 30 minutes with AMV reverse transcriptase under the conditions described in the legend to Fig. 2 for synthesis of single-stranded cDNA. The reaction was followed by using α-^{32}P-labeled dTTP (deoxythymidine triphosphate). Ligation of the Eco RI site octanucleotide (5′)TGAATTCA(3′) to the "filled in" cDNA was performed as described with the Hind III decamer in the legend to Fig. 3. Digestion with an excess of Eco RI endonuclease produced cDNA with Eco RI cohesive terminals. Since a Hind III site is recreated by this process, this procedure has the additional advantage that the cloned fragment also can be released from the plasmid by Hind III restriction endonuclease.
21. A. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
22. J. R. Duguid, D. F. Steiner, W. L. Chick [*ibid.* **73**, 3539 (1976)] have reported that insulin mRNA isolated from a rat islet cell tumor is approximately 600 nucleotides in length. Assuming a poly(A) length of 100 residues, the sequence of approximately 150 nucleotide residues remains to be established. Of these, approximately 30 residues are required to define the remaining portion of the known prepeptide sequence, leaving 120 residues of 5′-noncoding sequence. However, Duguid *et al.* used for part of their analyses ribosomal RNA markers in formamide polyacrylamide gels to arrive at their estimate of molecular weight. This method is believed to result in an overestimate of mRNA sizes [S. L. C. Woo, J. M. Rosen, C. D. Liarakos, Y. C. Choi, H. Busch, A. R. Means, B. N. O'Malley, D. L. Robberson, *J. Biol. Chem.* **250**, 7027 (1975)]. Our single-stranded cDNA forms a band on polyacrylamide gels with an estimated size of 450 nucleotides (Fig. 3A, lane 1). If this band comes from molecules initiated by dT_{12-18} very close to the poly(A)-nontranslated region junction, then there remain only about 50 nucleotides left to be determined. This latter number is more consistent with the small 5′ noncoding regions that have been observed to date for eukaryotic mRNA's.
23. A. Devillers-Thiery, T. Kindt, G. Scheele, G. Blobel, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 5016 (1975).
24. P. N. Campbell and G. Blobel, *FEBS Lett.* **72**, 215 (1976).
25. R. L. Pictet and W. J. Rutter, in *Handbook of Physiology*, section 7, *Endocrinology*, vol. 1, *Endocrine Pancreas* (American Physiological Society, Washington, D.C., 1972), p. 25; C. de Haen, E. Swanson, D. C. Teller, *J. Mol. Biol.* **106**, 639 (1976).
26. The sequences in this region of proinsulin I are not completely confirmed.
27. N. J. Proudfoot, *J. Mol. Biol.* **107**, 491 (1976).
28. K. N. Subramanian, J. Pan, S. Zain, S. M. Weissman, *Nucleic Acids Res.* **1**, 727 (1974).
29. Supported in part by NIH grants GM 21830 (to W.J. R.), CA 14026 (to H.M.G.), and GM 05385 (to J.C.).; by a postdoctoral fellowship from Deutsche Forschungs Gemeinschaft (to A.U.); and by an NIH career development award (to R.P.). We are indebted to W. Imagawa for aid in the isolation of rat islets. The Hind III decamer and T4 ligase were provided by R. H. Scheller, California Institute of Technology, and H. Heyneker, University of California, San Francisco, respectively.

12 May 1977

Environmental Asbestos Pollution Related to Use of Quarried Serpentine Rock

Abstract. *Crushed serpentinite quarried in Montgomery County, Maryland, has been extensively used for paving roads and other surfaces. The mineral assemblage includes antigorite or lizardite as well as chrysotile and tremolite. Air samples taken in the vicinity of serpentine-paved roads show that chrysotile concentrations are about 10^3 times greater than those typically found in urban ambient air in the United States.*

A rock quarry located near Rockville, Maryland, is the major source of crushed stone in the area north of Washington, D.C., including Montgomery and Prince Georges counties in Maryland. An estimated 28 million cubic meters or more of rock has been quarried during some 20 years of operation. The crushed stone is used as road metal, base course, and for resurfacing of highways, parking lots, and driveways. It has also been used as concrete aggregate and for other purposes in the construction industry, and as filler-binder for asphalt in blacktop paving.

The quarry is located in the Hunting Hill pluton, which is about 1½ km wide and 6½ km long and consists of serpentinized dunite cut by gabbro dikes. The structure, petrography, and mineralogy were described by Larrabee (*1*), who concluded that the bulk of the serpentinite consisted of antigorite in both platy and fibrous forms. Veins or lenticular bodies of chrysotile, tremolite, deweylite, talc, anthophyllite, clinozoisite, penninite, and other silicate minerals were reported to occur in the serpentinite. The chrysotile veins are generally less than 1 mm wide. Since the fiber is the harsh variety and occurs in short lengths it is not a commercial source of asbestos. Chlorite is common in the serpentinite, occurring chiefly as a replacement of antigorite.

The widespread and large-scale use of such crushed serpentinite raised the possibility of environmental contamination by the asbestos and asbestiform minerals present—chrysotile, tremolite, and anthophyllite. These minerals have a known human disease potential (*2*). The biological potential of platy and fibrous antigorite has not been determined.

We have investigated the possibility of environmental contamination by asbestos mineral dusts from quarry operations such as blasting, crushing, and truck loading and by the dissemination of dust along the quarry truck delivery routes as well as during unloading and installation of the crushed rock. We have been mindful that, after emplacement, considerable dust may be continually created by abrasion and erosion of pavement surfaces, particularly where the crushed rock is not bound by asphalt base or topping. Braking, turning, and acceleration of vehicles on these surfaces produce visible dust clouds. An example is seen in Fig. 1. Vegetation along roads paved with crushed stone is often heavily coated with road dust, particularly in the vicinity of intersections.