Proc. Natl. Acad. Sci. USA
Vol. 75, No. 8, pp. 3727–3731, August 1978
Biochemistry

A bacterial clone synthesizing proinsulin

(rat preproinsulin/cDNA cloning/solid-phase radioimmunoassay/DNA sequence/fused proteins)

LYDIA VILLA-KOMAROFF*, ARGIRIS EFSTRATIADIS*, STEPHANIE BROOME*, PETER LOMEDICO*, RICHARD TIZARD*, STEPHEN P. NABER†, WILLIAM L. CHICK†, AND WALTER GILBERT*

* Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138; and † Elliot P. Joslin Research Laboratory, Harvard Medical School, and the Peter Bent Brigham Hospital, Boston, Massachusetts 02215

Contributed by Walter Gilbert, June 9, 1978

ABSTRACT **We have cloned double-stranded cDNA copies of a rat preproinsulin messenger RNA in *Escherichia coli* χ1776, using the unique *Pst* endonuclease site of plasmid pBR322 that lies in the region encoding amino acids 181–182 of penicillinase. This site was reconstructed by inserting the cDNA with an oligo(dG)-oligo(dC) joining procedure. One of the clones expresses a fused protein bearing both insulin and penicillinase antigenic determinants. The DNA sequence of this plasmid shows that the insulin region is read in phase; a stretch of six glycine residues connects the alanine at position 182 of penicillinase to the fourth amino acid, glutamine, of rat proinsulin.**

Can the structural information for the production of a higher cell protein be inserted into a plasmid in such a way as to be expressed in a transformed bacterium? To attack this problem, we used as a model rat insulin, an interesting protein that can be identified by immunological and biological means.

Although mature insulin contains two chains, A and B, it is the product of a single longer polypeptide chain. The hormone is initially synthesized as a preproinsulin structure (1, 2). A hydrophobic leader sequence of 23 amino acids at the amino terminus of the nascent chain is cleaved off, presumably as the polypeptide chain moves through the endoplasmic reticulum (2–4), producing a proinsulin molecule. The proinsulin chain folds up and then the C peptide is cleaved from its middle (5). Thus each of the two (nonallelic) insulin genes in the rat (6–8) encodes a polypeptide 109 amino acids long, whose initial structure is NH_2—leader sequence—B chain—C peptide—A chain.

Ullrich *et al.* (9) have cloned double-stranded cDNA copies of rat preproinsulin mRNA isolated from pancreatic islets and determined sequences covering much of those two genes. We have made double-stranded cDNA copies of mRNA from a rat insulinoma (10) and cloned these in the *Pst* (*Providencia stuartii* endonuclease) site of pBR322 (11), which lies within the penicillinase gene.

The *Escherichia coli* penicillinase is a periplasmic protein, the gene for which was recently sequenced (12). Penicillinase is synthesized as a preprotein with a 23 amino acid leader sequence (12, 13), which presumably serves as a signal to direct the secretion of the protein to the periplasmic space, and is removed as the protein traverses the membrane. Insertion of the structural information for insulin into the penicillinase gene should cause expression of the insulin sequence as a fusion product transported outside the cell.

MATERIALS AND METHODS

Bacterial Strains. *E. coli* K-12, strain HB101 [*hsm*$^-$, *hrs*$^-$, *recA*$^-$, *gal*$^-$, *pro*$^-$, *str*^{r} (14)] was initially obtained from H. Boyer. *E. coli* K-12 strain χ1776 (15) (F$^-$, *tonA53*, *dapD8*, *minA1*, *supE42*, Δ40[*gal-uvrB*], λ$^-$, *minB2*, *rfb-2*, *nalA25*, *oms-2*, *thyA57*, *metC65*, *oms-1*, Δ29[*bioH-asd*], *cycB2*, *cycA1*, *hsdR2*) was provided by R. Curtiss.

DNA and Enzymes. pBR322 DNA, a gift from A. Poteete, was used to transform *E. coli* HB101. Plasmid DNA was purified according to the procedure of Clewell (16). Avian myeloblastosis virus reverse transcriptase (RNA-dependent DNA polymerase), *E. coli* DNA polymerase I, and terminal transferase were gifts from T. Papas, M. Goldberg, and J. Wilson, respectively. Restriction enzymes were purchased from Bethesda Research Labs and New England BioLabs.

RNA Purification. An x-ray-induced, transplantable rat beta cell tumor (10) was used as source of preproinsulin mRNA. Tumor slices (20 g per preparation) were homogenized, and a cytoplasmic RNA (about 2 mg/g of tissue) was purified from a postnuclear supernatant by Mg^{2+} precipitation (17), followed by extraction with phenol and chloroform, and enriched for poly(A)-containing RNA by oligo(dT)-cellulose chromatography (18). About 4% of the material binds to the column (data from eight preparations). Further purification of the oligo(dT)-cellulose-bound material by sucrose gradient centrifugation and/or polyacrylamide gel electrophoresis showed that the preproinsulin mRNA was a minor component of the preparation.

Double-Stranded cDNA Synthesis. Oligo(dT)-cellulose-bound RNA was used directly as template for double-stranded cDNA synthesis (19), except that a specific p(dT)$_8$dG-dC primer (Collaborative Research) was utilized for reverse transcription. The concentrations of RNA and primer were 7 mg/ml and 1 mg/ml, respectively. All four [α-^{32}P]dNTPs were at 1.25 mM (final specific activity 0.85 Ci/mmol). The reverse transcript was 2% of the input RNA, and 25% of it was finally recovered in the double-stranded DNA product.

Construction of Hybrid DNA Molecules. pBR322 DNA (5.0 μg) was linearized with *Pst*, and approximately 15 dG residues were added per 3′ end by terminal transferase at 15° in the presence of 1 mM Co^{2+} (20) and autoclaved gelatin at 100 μg/ml. Similarly, dC residues were added to 2.0 μg of double-stranded cDNA, which was then electrophoresed in a 6% polyacrylamide gel. Following autoradiography, molecules in the size range of 300 to 600 base pairs (0.5 μg) were eluted from the gel (21). Size selection was done after tailing rather than before because previous experience had indicated that occasionally impurities contaminating DNA extracted from gels inhibits terminal transferase. The eluted double-stranded cDNA was concentrated by ethanol precipitation, redissolved in 10 mM Tris·HCl at pH 8, mixed with 4 μg of dG-tailed pBR322, and dialyzed versus 0.1 M NaCl/10 mM EDTA/10 mM Tris, pH 8. The mixture (4 ml) was then heated at 56° for 2 min, and annealing was performed at 42° for 2 hr. The hybrid DNA was used to transform *E. coli* χ1776.

Transformation and Identification of Clones. Transformation of *E. coli* χ1776 (an EK2 host) with pBR322 (an EK2

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. §1734 solely to indicate this fact.

vector) was performed in a biological safety cabinet in a P3 physical containment facility in compliance with NIH guidelines for recombinant DNA research published in the *Federal Register* [(1976) **41**, 27902–27943].

χ1776 was transformed by a transfection procedure (22) adapted to χ1776 by A. Bothwell (personal communication) and slightly modified as follows: χ1776 was grown in L broth (23) supplemented with diaminopimelic acid at 10 μg/ml and thymidine (Sigma) at 40 μg/ml to OD_{590} of 0.5. Cells (200 ml) were sedimented at 500 × *g* and resuspended by swirling in 1/10th vol of cold buffer containing 70 mM $MnCl_2$, 40 mM Na acetate at pH 5.6, 30 mM $CaCl_2$, and kept on ice for 20 min. The cells were repelleted and resuspended in 1/30th of the original volume in the same buffer. Two milliliters of the annealed DNA preparation was added to the cells. Aliquots of this mixture (0.3 ml) were placed in sterile tubes and incubated on ice for 60 min. The cells were then placed at 37° for 2 min. Broth was added to each tube (0.7 ml) and the tubes were incubated at 37° for 15 min; 200 μl of the cells was spread on sterile nitrocellulose filters (Millipore) overlaying agar plates containing tetracycline at 15 μg/ml. (The filters were boiled to remove detergents before use.) The plates were incubated at 37° for 48 hr. Replicas of the filters were made by a procedure developed by D. Hanahan (personal communication): The nitrocellulose filters containing the transformants were removed from the agar and placed on a layer of sterile Whatman filter paper. A new sterile filter was placed on top of the filter containing the colonies and pressure was applied with a sterile velvet cloth and a replica block. A sterile needle was used to key the filters. The second filter was placed on a new agar plate and incubated at 37° for 48 hr. The colonies on the first filter were screened by the Grunstein–Hogness technique (24), using as probe an 80-nucleotide-long fragment produced by *Hae* III digestion of high specific activity cDNA (9). Positive colonies were rescreened by hybrid-arrested translation (25) as described in the legend of Table 1.

Radioimmunoassays. Two-site solid-phase radioimmunoassays were performed (28). Cells from colonies to be tested were transferred with an applicator stick onto 1.5% agarose containing 30 mM Tris·HCl, pH 8, lysozyme at 0.5 mg/ml, and 10 mM EDTA; released antigen was adsorbed to an IgG-coated polyvinyl disk during a 1-hr incubation at 4°. The wash buffer contained streptomycin sulfate at 300 μg/ml and normal guinea pig serum (Grand Island Biological Co.) instead of normal rabbit serum. Guinea pig antiserum to bovine insulin was purchased from Miles Laboratories.

Standard (liquid) radioimmunoassays were performed using the back titration procedure employing alcohol precipitation of insulin–antibody complexes (29).

DNA Sequencing. DNA sequencing was performed as described by Maxam and Gilbert (30).

RESULTS

Construction and Identification of cDNA Clones. We isolated poly(A)-containing RNA from a transplantable rat insulinoma. This preparation contained preproinsulin mRNA, because it directed the synthesis in a cell-free system of a product precipitable with anti-insulin antibody (data not shown). However, the mRNA yield after further purification was not sufficient for cloning, and therefore we decided to clone cDNA synthesized from the total preparation. In an attempt to enrich the reverse transcript for insulin sequences, we utilized the DNA sequence reported by Ullrich *et al.* (9) to choose a specific primer, $(dT)_8dG$-dC. The product of double-stranded cDNA synthesis (19) was extended by a short oligo(dC) tail about 15 nucleotides in length, and sized on a polyacrylamide gel. A broad size cut averaging 500 base pairs was selected in order to enrich for full-length sequences. We inserted these molecules into the *Pst* site of pBR322 after elongating the 3′-terminal extension of the cleavage site with oligo(dG). We used this oligo(dG)·oligo(dC) joining procedure in order to reconstruct the *Pst* recognition sequence (ref. 31; W. Rowenkamp and R. Firtel, personal communication); approximately 40% of the inserts were excisable with *Pst* after cloning. From about 0.25 μg of tailed cDNA we obtained 2355 transformants in *E. coli* strain χ1776. To identify clones containing insulin sequences, we first screened one-third of the transformants, using as a probe an 80-nucleotide-long *Hae* III fragment of cDNA synthesized from oligo(dT)-bound RNA because the results of Ullrich *et al.* (9) suggested that such a fragment should be insulin specific. About 20% of the clones were positive, but restriction analysis of plasmid DNA from a few candidates showed that the inserts were not insulin sequences. We concluded that our probe was not pure and rescreened some of the positive clones, using hybrid-arrested translation (25). This method is based on the principle that mRNA in the form of an RNA·DNA hybrid does not direct cell-free protein synthesis. We incubated aliquots of oligo(dT)-bound RNA with linearized

Table 1. Hybrid-arrested translation and immunoprecipitation of the cell-free products

Source of arresting DNA	Radioactivity, cpm/20 μl: Acid insoluble	Immunoprecipitable: − Insulin	Immunoprecipitable: + Insulin	% Immunoprecipitable*
Control I (−DNA, −RNA)†	2,570			
Control II (−DNA, +RNA)‡	35,700	12,300	310	36.2
pBR322	28,800	7,850	245	29.0
Clone 3	15,100	3,630	264	26.9
Clone 13	19,600	5,190	350	28.4
Clone 15	18,600	4,850	252	28.7
Clone 16	29,200	8,830	247	32.2
Clone 17	24,000	6,700	316	30.0
Clone 18	15,900	3,690	251	25.8
Clone 19	8,650	587	277	5.0
Clone 20	15,100	4,070	231	30.6
Clone 21	21,100	5,170	223	26.7

Plasmid DNA (about 3 μg) was digested with *Pst*, precipitated with ethanol, and dissolved directly in 20 μl of deionized formamide. After heating for one minute at 95° each sample was placed on ice. Following the addition of 1.5 μg of oligo(dT)-cellulose-bound RNA, piperazine-*N,N′*-bis(2-ethanesulfonic acid) (Pipes) at pH 6.4 to 10 mM, and NaCl to 0.4 M, the mixtures were incubated for 2 hr at 50°. They were then diluted by the addition of 75 μl of H_2O and ethanol precipitated in the presence of 10 μg of wheat germ tRNA, washed with 70% (vol/vol) ethanol, dissolved in H_2O, and added to a wheat germ cell-free translation mixture (26) containing 10 μCi of [^{3}H]leucine (60 Ci/mmol). Fifty-microliter reaction mixtures were incubated at 23° for 3 hr and then duplicate 2-μl aliquots were removed for trichloroacetic acid precipitation. From the remainder two 20-μl aliquots were treated with ribonuclease, diluted with immunoassay buffer, and analyzed for the synthesis of immunoreactive preproinsulin by means of a double antibody immunoprecipitation (27) in the absence or presence of 10 μg of bovine insulin. The washed immunoprecipitates were dissolved in 1 ml of NCS (Amersham) and assayed in 10 μl of Omnifluor (New England Nuclear) by liquid scintillation counting.

* Calculated using the formula [(immunoprecipitable radioactivity in the absence of insulin) − (immunoprecipitable radioactivity in the presence of insulin)]/[(acid-insoluble radioactivity) − (acid-insoluble radioactivity of control I)].

† Reaction mixture incubated in the absence of added RNA.

‡ Cell-free translation by the direct addition of oligo(dT)-cellulose-bound RNA into the reaction mixture.

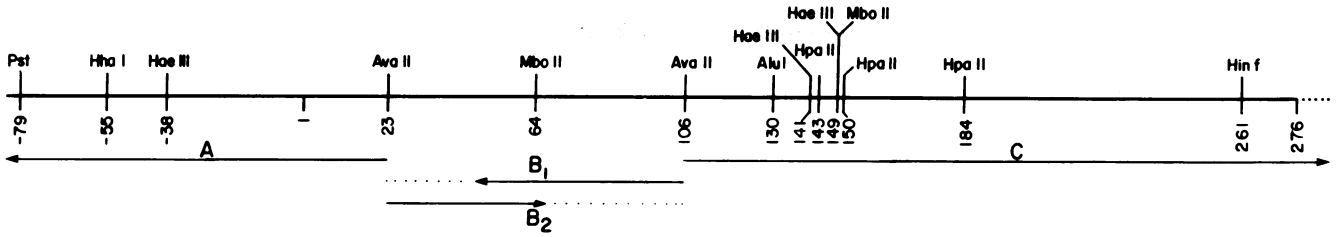

FIG. 1. Restriction map of the insertion in clone pI19. Each restriction site is identified by a number indicating the 5′-terminal nucleotide generated by cleavage at the message strand. Nucleotides are numbered beginning with the first base of the sequence encoding proinsulin. Nucleotides in the 5′ direction from position 1 in the message strand are identified by negative numbers, beginning with −1. Arrows indicate the sequenced fragments; those pointing to the left indicate sequences derived from the antimessage strand, and those pointing to the right indicate sequences derived from the message strand. The uniquely labeled restriction fragments were generated as follows: Following excision with *Pst*, DNA of the insertion was digested with *Ava* II and end labeled. Fragments A and C purified from a polyacrylamide gel were sequenced directly because the *Pst* ends do not label significantly. Fragment B was strand separated on a polyacrylamide gel and sequenced in both directions. The exact number of C·G pairs in the right-hand tail before the *Pst* site could not be counted.

DNA from nine clones under conditions favoring DNA·RNA hybridization (32), added them to cell-free translation systems, and assayed for a specific inhibition of insulin synthesis. Table 1 shows that one of the plasmids, pI19, inhibited the synthesis of immunoprecipitable material. Restriction endonuclease digestions of the *Pst*-excised insert of pI19 with several enzymes generated fragments whose sizes were consistent with the sequence of Ullrich *et al.* (9). We confirmed the presence of insulin DNA in pI19 by direct DNA sequence analysis and screened the rest of the clones with purified pI19 insert labeled by nick translation. About 2.5% (48/1745) of the clones hybridized strongly to this probe. There must have been enrichment for insulin sequence at some step of our procedure, because hybridization analysis using cloned insulin DNA as probe showed the presence of only 0.3% insulin mRNA in the original oligo(dT)-bound RNA.

Sequence Information. Fig. 1 shows the restriction map of the insertion in clone pI19 and Fig. 2 shows the sequence of the insert. It corresponds to rat insulin I (5, 33) and encodes the entire preproinsulin chain with the exception of the first two amino acid residues of the reported preregion (1). It therefore extends the sequence determined by Ullrich *et al.* (9) by twenty-five 5′-terminal nucleotides. It also verifies the reported amino acid residues for positions −14, −17, −18, and −20; it identifies the previously uncertain residue −15; and it identifies the unknown residue −19. However, the residues at positions −16 and −21 differ from those reported (1).

The sequence deviates from that determined by Ullrich *et al.* (9) at the region immediately after the UGA terminator, where a GAGTC sequence occurs, predicting a *Hin*f cleavage site that we have experimentally verified. Furthermore, only moderate agreement exists between the two sequences for the next 15 nucleotides of the 3′ untranslated region.

Expression. Almost two-thirds of the clones carrying inserts were ampicillin resistant; thus the active site of penicillinase must lie between amino acid residues 23 and 182 (12). The degree of resistance was variable, suggesting the expression of different sequences from the inserts in the form of fused translation products, probably differing in length and stability.

We therefore screened colonies of the 48 clones containing insulin sequence for the presence of insulin antigenic determinants, using a solid-phase radioimmunoassay (28). Polyvinyl sheets coated with antibody molecules will bind specific antigens released from bacteria. The immobilized antigen can then be detected by autoradiography following exposure of the sheets to ^{125}I-labeled antibody. This method permits detection of as little as 10 pg of insulin in a colony. We coated plastic disks with anti-insulin antibody and used ^{125}I-labeled anti-insulin to detect solely insulin antigenic determinants. Disks coated with anti-penicillinase antibody and exposed to ^{125}I-anti-insulin detect the presence of a fused protein, as do disks coated with anti-insulin and exposed to radioiodinated anti-penicillinase.

One clone, pI47, gave positive responses with all of the combinations described above; this indicates the presence of a penicillinase–insulin hybrid polypeptide. Fig. 3 shows some of the results. To determine whether this fused protein is secreted, we grew clone pI47 in liquid culture and extracted the proteins in the periplasmic space by osmotic shock, a method that does not lyse bacteria (34). Fig. 4 shows that the insulin

FIG. 2. DNA sequence of the insertion in clone pI19. Nucleotides are numbered using the convention described in Fig. 1. Accordingly, amino acids are numbered beginning with the first amino acid of proinsulin, while the last amino acid of the leader sequence (pre region) is numbered as −1. Restriction endonuclease cleavage sites experimentally verified are underlined and identified. The arrows indicate, in order, the ends of the leader sequence and the peptides B, C, and A. Two nucleotides indicated by double underlining are uncertain.

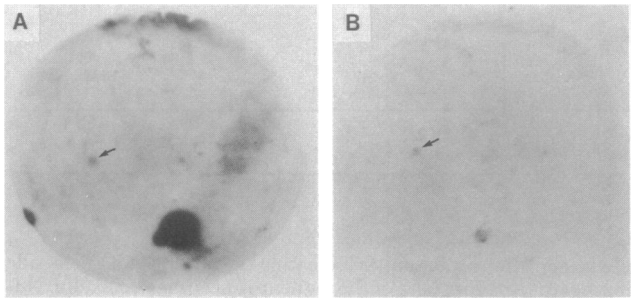

FIG. 3. Initial detection of penicillinase-insulin hybrid polypeptides in an insulin cDNA clone. Cells from colonies of the 48 insulin cDNA clones and from control colonies, χ1776 and χ1776-pBR322, were applied to an agarose/lysozyme/EDTA plate. Positive controls, 5 ng of insulin and 5 ng of penicillinase, each in 1 μl of wash buffer, also were spotted on plate. Antigen was adsorbed to an IgG-coated polyvinyl disk during a 1-hr incubation at 4°. Immobilized antigen was labeled by setting the plastic disk on a solution containing radioiodinated anti-insulin IgG. The autoradiographs are of disks precoated with anti-insulin IgG (*A*) or anti-pencillinase IgG (*B*), exposed on Kodak X-Omat R film using a Du Pont Cronex Lightning Plus intensifying screen for 12 hr at −70°. The arrows indicate the signal generated by clone pI47. The large exposed area in the lower right of (*A*) is the positive control for insulin detection.

antigen was recovered in the distilled water wash of the shock procedure. Table 2 shows that the insulin antigen in the wash is also detectable and quantifiable by a standard radioimmunoassay. The yield of antigen depended on the growth medium; antigen was released by cells grown in M9/glucose/amino acids medium but not by cells grown in brain/heart infusion. We estimate a recovery of about 100 molecules per cell.

Structure of the Fused Protein. We sequenced pI47 to determine the sequence around the junctions. Fig. 5 shows that a proinsulin I cDNA lies in the *Pst* site in the correct orientation and in phase, so that a fused protein can be synthesized. In pI19, the insert is in the correct orientation, but not in phase. In pI47 the oligo(dG)-oligo(dC) region encodes six glycines that connect the penicillinase sequence, ending at amino acid 182 (alanine), to the fourth amino acid (glutamine), of the proinsulin sequence. The cDNA sequence in pI47 extends 26 base pairs past the UGA terminator. Thus, we infer the structure of the fused protein to be penicillinase(24–182)-$(Gly)_6$-proinsulin(4–86).

DISCUSSION

The coding regions of eukaryotic structural genes are often interrupted by introns (35–38), whose transcripts are spliced out of the mature mRNA. Because prokaryotes do not appear to process their messengers, double-stranded cDNA made from a mature messenger is the material of choice to carry eukaryotic structural information into bacteria.

By using cDNA cloning technology and an extremely sensitive method to assay expression, we were able to construct a derivative of *E. coli* strain χ1776 carrying an insulin gene sequence and to detect the synthesis and secretion into the periplasmic space of a fused protein carrying antigenic determinants of both insulin and penicillinase. This was accomplished simply by inserting double-stranded cDNA carrying the structural information for insulin into a restriction site within the structural gene for penicillinase. Not only is the fused DNA sequence expressed as a chain of amino acids, but also the polypeptide folds so as to reveal insulin antigenic shapes. Thus we expect soon to be able to demonstrate biological function for this, or for a similar, fused protein.

We anticipate that the joining of cDNA sequences to nucleotides that lie ahead of the *Pst* site in the penicillinase gene will also produce fused and secreted molecules. Moreover, if the fusion replaces the preproinsulin leader with that of penicillinase it is likely that the new protein will also be secreted by the *E. coli* cell and may even be correctly matured by cleavage of the leader sequence.

Clearly, we have exploited a general method that should lead to the expression and secretion of any eukaryotic protein provided another protein, such as penicillinase, will serve as a carrier, by virtue of its leader sequence. Moreover, the secretion of the eukaryotic protein sequence to the periplasm or extracellular space will both permit its harvest in a purified form and probably eliminate intracellular sources of instability.

Often just an expression of antigens is the goal. In a "shotgun" screening, the existence of a fused protein antigen could be used to identify transformants carrying desired eukaryotic gene fragments. On the other hand, the insertion of a DNA fragment coding for surface antigenic determinants of a virus into a carrier protein should lead to the secretion of a fused protein that could serve as a vaccine, even though no entirely correct virus product is ever produced.

Table 2. Immunoreactive insulin concentration in distilled water wash of osmotic shock procedure

Exp.	Insulin, μunits/ml	Cells/ml
1	318	1.5×10^{10}
2	166	6.0×10^{9}
3	386	4.2×10^{10}

Duplicate 0.1-ml aliquots of each sample prepared as described in the legend to Fig. 4 were assayed (29) in a final volume of 0.4 ml using rat insulin standard, a gift from J. Schlichtkrull. One unit = 48 μg. The NaCl/Tris wash, the 20% sucrose wash, and the media of χ1776-pI47 as well as the water wash from osmotic shock of χ1776-pBR322 gave values below the sensitivity of the assay (25 μunits/ml).

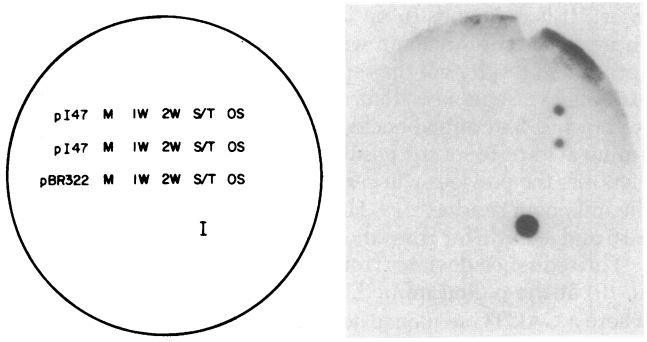

FIG. 4. Release of insulin antigen from χ1776-pI47 cells by osmotic shock. One liter of χ1776-pI47 cells growing at 37° in M9 medium supplemented with 1 g of tryptone, 0.5 g of yeast extract, and 0.5% glucose was harvested at a density of 5×10^7 cells per ml and washed two times in 10 ml of cold 10 mM Tris·HCl, pH 8/30 mM NaCl. The cells were then osmotically shocked (34) in the following manner: The final wash pellet was resuspended in 10 ml of 20% sucrose per 30 mM Tris·HCl, pH 8, at room temperature, made 1 mM in EDTA, shaken at room temperature for 10 min, centrifuged out, resuspended in 10 ml of cold distilled water, shaken in an ice bath for 10 min, and again pelleted. The resulting supernatant was termed the "water wash." As a control, 1 liter of χ1776-pBR322 was grown and treated in a similar manner. Aliquots (1 μl) of each fraction to be assayed for the presence of insulin antigen were applied to the surface of a 1.5% agar plate. (*A*) Positions of each fraction on the plate. M, medium; 1W, first wash supernatant; 2W, second wash supernatant; S/T, sucrose/Tris supernatant; OS, distilled water wash; I, insulin. (*B*) Autoradiograph showing results of a two-site radioimmunoassay of these fractions. Antigen was adsorbed to a polyvinyl disk and labeled by using anti-insulin IgG. The labeled areas correspond to the water washes and the positive control (1 ng insulin). A spectrophotometric assay for β-galactosidase (23) indicated that no more than 4% of cells lyse during this procedure.

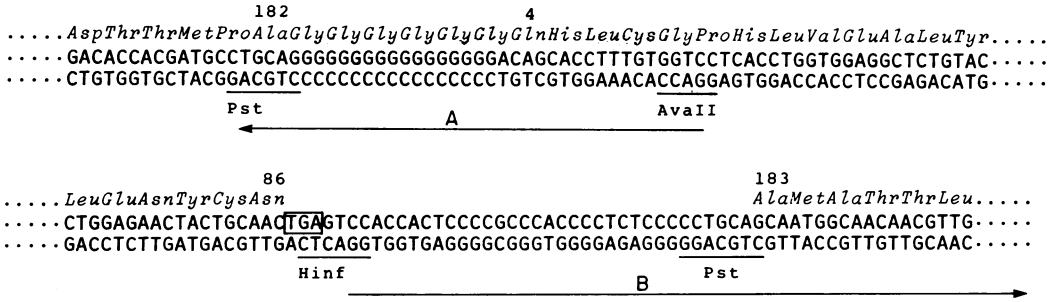

FIG. 5. Partial DNA sequence of the insertion in clone pI47. Clone pI47 DNA was digested with *Hinf* and two fragments, H1 and H2 (ca. 1700 and 280 base pairs long, respectively) were isolated. H1 contains the amino-terminal portion of the penicillinase gene and the bulk of the cDNA insert. H1 was digested with *Ava* II, end labeled, and digested again with *Pst*. A fragment 39 nucleotides long (fragment A, arrow) was isolated and sequenced. Fragment H2 was end labeled and digested with *Alu* I (which cuts at the region corresponding to amino acid 200 of penicillinase). A fragment 88 base pairs long (fragment B, arrow) was isolated and sequenced. The termination sequence TGA is boxed.

We thank David Baltimore, Philip Sharp, and Salvador Luria for the use of the Massachusetts Institute of Technology P3 laboratory. We thank Macy Koehler for help with the figures; Fotis Kafatos for use of facilities; Philip Sharp, Al Bothwell, Shirley Tilghman, Doug Hanahan, and Richard Firtel for discussions. W.G. is an American Cancer Society Professor of Molecular Biology. W.L.C. is an Established Investigator of the American Diabetes Association. This work was supported by National Institutes of Health Grants AM 21240 and GM 09541-17 to W.G. and AM 15398 to W.L.C.

1. Chan, S. J., Keim, P. & Steiner, D. F. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 1964–1968.
2. Chan, S. J. & Steiner, D. F. (1977) *Trends Biochem. Sci.* **2,** 254–256.
3. Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67,** 835–851.
4. Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67,** 852–862.
5. Steiner, D. F., Kemmler, W., Tager, H. S. & Peterson, J. D. (1974) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **33,** 2105–2115.
6. Smith, L. F. (1966) *Am. J. Med.* **40,** 662–666.
7. Clark, J. L. & Steiner, D. F. (1969) *Proc. Natl. Acad. Sci. USA* **62,** 278–285.
8. Markussen, J. & Sundby, F. (1972) *Eur. J. Biochem.* **25,** 153–162.
9. Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W. J. & Goodman, H. M. (1977) *Science* **196,** 1313–1319.
10. Chick, W. L., Warren, S., Chute, R. N., Like, A. A., Lauris, V. & Kitchen, K. C. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 628–632.
11. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crossa, J. H. & Falkow, S. (1977) *Gene* **2,** 95–113.
12. Sutcliffe, J. G. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3737–3741.
13. Ambler, R. P. & Scott, G. K. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3732–3736.
14. Boyer, H. W. & Rouland-Dussoix, D. (1969) *J. Mol. Biol.* **41,** 459–472.
15. Curtiss, R., III, Pereira, D. A., Hsu, J. C., Hull, S. C., Clarke, J. E., Maturin, L. J., Sr., Goldschmidt, R., Moody, R., Inoue, M. & Alexander, L. (1977) in *Recombinant Molecules: Impact on Science and Society. Proceedings of the 10th Miles International Symposium,* eds. Beers, R. F., Jr., & Bassett, E. G. (Raven, New York), pp. 45–56.
16. Clewell, D. B. (1972) *J. Bacteriol.* **110,** 667–676.
17. Palmiter, R. (1974) *Biochemistry* **13,** 3603–3615.
18. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
19. Efstratiadis, A., Kafatos, F. C., Maxam, A. M. & Maniatis, T. (1976) *Cell* **7,** 279–288.
20. Roychoudhury, R., Jay, E. & Wu, R. (1976) *Nucleic Acid Res.* **3,** 101–116.
21. Gilbert, W. & Maxam, A. M. (1973) *Proc. Natl. Acad. Sci. USA* **70,** 3581–3584.
22. Enea, V., Vovis, G. F. & Zinder, N. D. (1975) *J. Mol. Biol.* **96,** 495–509.
23. Miller, J. M. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York), pp. 431–435.
24. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 3961–3965.
25. Paterson, B. M., Roberts, B. E. & Kuff, E. L. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 4370–4374.
26. Roberts, B. E. & Paterson, B. M. (1973) *Proc. Natl. Acad. Sci. USA* **70,** 2330–2334.
27. Lomedico, P. T. & Saunders, G. F. (1976) *Nucleic Acids Res.* **3,** 381–391.
28. Broome, S. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2746–2749.
29. Makula, D. R., Vichnuk, D., Wright, P. H., Sussman, K. E. & Yu, P. L. (1969) *Diabetes* **18,** 660–689.
30. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
31. Boyer, H. W., Betlach, M., Bolivar, F., Rodriguez, R. L., Heyneker, H. L., Shine, J. & Goodman, H. M. (1977) in *Recombinant Molecules: Impact on Science and Society. Proceedings of the 10th Miles International Symposium,* eds. Beers, R. F., Jr. & Bassett, E. G. (Raven, New York), pp. 9–20.
32. Casey, J. & Davidson, N. (1977) *Nucleic Acids Res.* **4,** 1539–1552.
33. Humbel, R. E., Bosshard, H. R. & Zahn, H. (1972) in *Handbook of Physiology, Section 7 (Endocrinology),* eds. Steiner, D. F. & Freinkel, N., (American Physiological Society, Washington, DC), Vol. 1, pp. 111–132.
34. Neu, H. C. & Heppel, L. A. (1965) *J. Biol. Chem.* **240,** 3685–3692.
35. Tilghman, S. M., Tiemeier, D. C., Seidman, J. G., Peterlin, B. M., Sullivan, M., Maizel, J. V. & Leder, P. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 725–729.
36. Jeffreys, A. J. & Flavell, R. A. (1977) *Cell* **12,** 1097–1108.
37. Breathnach, R., Mandel, J. L. & Chambon, P. (1977) *Nature* **270,** 314–319.
38. Gilbert, W. (1978) *Nature* **271,** 501.