Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 6838–6842, November 1983
Biochemistry

Isolation of a genomic clone for bovine pancreatic trypsin inhibitor by using a unique-sequence synthetic DNA probe

(genomic libraries/synthetic oligonucleotides/hybridization with single-stranded probes)

STEPHEN ANDERSON* AND I. BARRY KINGSTON

Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH England

Communicated by F. Sanger, August 10, 1983

ABSTRACT **Unique-sequence synthetic DNA probes, based on the known amino acid sequence of bovine pancreatic trypsin inhibitor, were constructed from oligodeoxynucleotides. In genomic Southern blot experiments, these probes were shown to hybridize specifically to discrete restriction fragments. A synthetic probe also was used to isolate a cloned BPTI gene from a bovine genomic library. DNA sequence analysis of this clone indicated that the BPTI coding region was neither preceded by a start codon nor immediately followed by a termination codon. This suggests that the mature form of BPTI may be produced through proteolytic processing from a larger polypeptide precursor.**

Bovine pancreatic trypsin inhibitor (BPTI) is a small 58-amino-acid basic protein found in a variety of bovine tissues, including lung, parotid gland, spleen, liver, and pancreas (1, 2). Although its exact physiological role and site of synthesis are unclear, its interactions with trypsin and other serine proteases such as kallikrein, chymotrypsin, and plasmin have been well characterized (1, 2). The refolding pathway, involving disulfide bond rearrangements, is better understood for BPTI than for any other protein (3). Moreover, because of its small size, high solubility, and stability, BPTI has been an object of intense physical and theoretical study as a model globular polypeptide (4–6).

The availability of a cloned gene for BPTI would enable us to learn more about the properties of the protein *in vivo*, including the site and mode of its synthesis, and its evolutionary relationship to other homologous proteins. Moreover, by coupling the BPTI gene to a suitable expression system, it would be possible to engineer BPTI variants genetically and, thus, to test the structure–function relationships of an extremely well-studied polypeptide. For these reasons we undertook the isolation of the gene for BPTI.

Our initial approach was to isolate mRNA with the intention of synthesizing a cDNA copy. However, we were unable to demonstrate the presence of BPTI in *in vitro* translations of mRNA prepared from bovine parotid gland or lung despite the fact that these tissues are good sources of the native protein. In view of the uncertainty regarding its site of synthesis and the abundance of its mRNA, we decided to adopt an approach using synthetic probes to isolate a genomic clone.

Unique-sequence synthetic oligodeoxynucleotide probes have been used previously to isolate genomic clones for protein genes from lower eukaryotes (7, 8). However, in these cases prior data from frameshift mutants allowed the exact sequences of the genes to be inferred. More generally, mixed-sequence oligodeoxynucleotide probes have been used to isolate protein genes of unknown sequence from cDNA libraries (9). Such probes are typically mixtures of 8–32 oligonucleotides, 14–17 nucleotides in length, representing every possible codon combination for a small stretch (5–6 residues) of amino acid sequence. Under stringent hybridization conditions that discriminate against incorrectly base-paired probes, these mixtures are capable of locating specific gene sequences in clone libraries of low-to-moderate complexity (9–12). Nevertheless, because of their short length and heterogeneity, mixed probes often lack the specificity required for probing sequences as complex as a mammalian genome. This makes such a method impractical for the isolation of mammalian protein genes when the corresponding mRNAs are unavailable.

We describe here a novel method of gene isolation that uses unique-sequence synthetic gene fragments as specific probes. We have used this method to isolate a gene for BPTI from a bovine genomic library. A similar approach also has been used recently by Jaye *et al.* (13) to isolate a cDNA clone for human factor IX.

MATERIALS AND METHODS

Materials. Protected nucleotide monomers were purchased from Cruachem (Bend, OR); Klenow DNA polymerase, from Boehringer Mannheim; proteinase K, from British Drug House; restriction enzymes, from New England BioLabs; T4 polynucleotide kinase and deoxynucleoside and dideoxynucleoside triphosphates, from P-L Biochemicals; r[γ-^{32}P]ATP and d[α-^{32}P]-ATP (10 mCi/ml, aqueous; 1 Ci = 37 GBq), from Amersham; and *Escherichia coli* DNA, from Sigma.

Oligonucleotide Synthesis. DNA oligonucleotides were synthesized by using the phosphotriester method with monomer addition on a solid-phase support (14, 15). Oligonucleotides were purified by HPLC or by polyacrylamide gel electrophoresis in 7 M urea. The oligonucleotides synthesized and their sequences were as follows:

P1	C-T-A-G-G-C-A-C-C-A-C-C-G-C
P2	C-G-T-A-C-C-T-G-C-G-G-T-G-G
P3	A-G-G-T-A-C-G-C-A-T-A-C-A-G
P4	C-C-G-A-G-G-A-C-T-G-T-A-T-G
P5	T-C-C-T-C-G-G-C-G-C-T-C-T-T
P6	A-A-C-T-T-C-A-A-G-A-G-C-G
P7	G-A-A-G-T-T-G-T-T-G-C-G-C-T-T-G-G-C
P8	T-G-T-C-G-C-G-C-C-A-A-G-C-G-C-A-A-C
P9	G-C-G-A-C-A-G-C-C-A-C-C-G-T-A-C-A-C-G
P10	A-A-A-C-C-T-T-C-G-T-G-T-A-C-G-G-T-G-G-C

Probe Construction. Probes were assembled by annealing together the oligonucleotides to form a duplex structure, then joining these in a ligation reaction (16). Full-length strands were purified by polyacrylamide gel electrophoresis under denatur-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: BPTI, bovine pancreatic trypsin inhibitor; bp, base pair(s); kbp, kilobase pair(s); AMV, avian myeloblastosis virus.

* Present address: Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080.

ing conditions (17), then reannealed and treated with 40 units of avian myeloblastosis virus (AMV) reverse transcriptase in a buffer containing 50 mM Tris·HCl (pH 7.8), 50 mM KCl, 7 mM $MgCl_2$, 2 mM dithiothreitol, and 200 μM dNTP (dATP, dCTP, dGTP, and dTTP) (final volume, 50 μl) for 60 min at 23°C. The flush-ended DNA produced by this treatment was cloned into the *Sma* I site of the vector M13mp8 (18), and the sequence was verified by the dideoxynucleotide chain termination method (19).

Probe Labeling. Probes were labeled by annealing 20 μg of single-stranded DNA from each M13 recombinant with 1.5 pmol of "sequencing" primer (14), then carrying out a DNA synthesis reaction in the presence of radioactive deoxynucleoside triphosphates. Reactions were carried out in 20 mM Tris·HCl, pH 7.8/7 mM $MgCl_2$/50 μM dNTP (dCTP, dGTP, and dTTP) containing 100 μCi of d[α-^{32}P]ATP (at a specific radioactivity of 800 Ci/mmol) and 10 units of Klenow DNA polymerase in a volume of 120 μl for 30 min at 23°C. Reactions were "chased" for a further 15 min by the addition of 12 μl of 500 μM dATP; then 14 μl of 1.0 M Tris·HCl (pH 7.8) and 6 μl of 1.0 M NaCl were added, and the polymerase was inactivated by heating the reaction mixture for 15 min at 70°C. The DNA in the cooled reaction mixture was digested with 100 units of *Eco*RI, and then the solution was adjusted to contain 0.2% Sarkosyl, 12 mM EDTA, and 250 μg of proteinase K per ml and was incubated for 30 min at 50°C. The DNA was extracted with phenol/chloroform, precipitated with ethanol, and redissolved in 50 μl of 10 mM Tris·HCl, pH 7.8/1 mM EDTA, then heat-denatured in 98% formamide/10 mM EDTA, and electrophoresed for 2 hr at 1.6 kV on a 40 cm × 20 cm × 0.4 mm 6% polyacrylamide/7 M urea gel. The region of the gel containing the radioactive probe was detected by a 3-min autoradiographic exposure, excised, and soaked overnight in 3 ml of 0.2 M sodium phosphate, pH 6.2/0.67 M NaCl/6.7 mM EDTA, to elute the probe.

DNA Preparation. High molecular weight DNA was isolated from bovine kidney nuclei (20) with a homogenization buffer of 15 mM Tris·HCl, pH 7.8/15 mM NaCl/60 mM KCl/5 mM EDTA/0.5 mM spermidine/0.34 M sucrose. DNA from $NaDodSO_4$-lysed nuclei was deproteinized by proteinase K treatment and phenol/chloroform extraction and then purified by equilibrium density gradient centrifugation in CsCl; the size of the DNA was >100 kilobase pairs (kbp) as analyzed by agarose gel electrophoresis with intact phage λ DNA as a marker.

Blot Hybridizations. Bovine DNA was digested with restriction enzymes, subjected to agarose gel electrophoresis, and blotted onto nitrocellulose filters by a modification of the method of Southern (21, 22). Filters were pretreated with 0.2 M sodium phosphate, pH 6.2/0.67 M NaCl/6.7 mM EDTA containing 100 μg of sonicated denatured *E. coli* DNA and 20 μg of sonicated denatured M13mp8 DNA per ml, 10× Denhardt's solution (1× Denhardt's solution is 0.02% Ficoll, 0.02% polyvinylpyrollidone, and 0.02% bovine serum albumin), and 1% Sarkosyl for 1 hr at 65°C. Hybridizations were performed at 65°C for 48 hr in a solution identical to that used for prehybridization except that dextran sulfate was also present at a concentration of 10% and the probe was present at a concentration of 10^6 cpm/ml. After hybridization, filters were washed twice at room temperature in 400 ml of 0.2 M sodium phosphate, pH 6.2/0.67 M NaCl/6.7 mM EDTA/1% Sarkosyl and then four times at 65°C in 250 ml of the same solution. Autoradiographic exposures of the filters were performed with preflashed film and an intensifying screen (23) for 2 days at −70°C.

Genomic Library Construction. A bovine genomic library was constructed by the method of Karn *et al.* (24). Briefly, bovine genomic DNA was partially digested with *Sau*3A I, fractionated by agarose gel electrophoresis, and a size fraction corresponding to fragments of 11–24 kbp was recovered from the gel by electroelution. This was ligated to *Bam*HI-digested λ1059, and the resultant chains were packaged by using extracts from the phage λ lysogens NS428 and NS433 (25). The library, containing ≈2×10^6 independent recombinants, was amplified on strain Q359 [r^-_k, m^+_k, su^+_{II}, 80^R, P2] (24) and stored over $CHCl_3$ at 4°C.

Library Probing. Recombinant phage were plated on strain D91 [r^+_k, m^+_k, Δ lac-pro] (24) at a density of 2×10^4 plaque-forming units per 90-mm dish; a total of 10^6 plaque-forming units (50 dishes) were screened by the plaque hybridization method of Benton and Davis (26). Duplicate nitrocellulose filters were lifted from each dish, and these were hybridized to the probe exactly as described above for the blot hybridization.

DNA Sequence Determination. After plaque purification of the positive clones, phage stocks were prepared, and the DNA from one of these was subjected to sequence analysis. Fifteen micrograms of DNA from the recombinant phage was fragmented with DNase I in the presence of Mn^{2+} (27) or by sonication (28), size-fractionated by agarose gel electrophoresis, electroeluted, and cloned by blunt-end ligation with T4 DNA ligase into *Sma* I-cut, phosphatase-treated M13mp8 (18); ≈900 independent subclones were obtained. These were picked into a fresh lawn of JM101 (29), transferred to nitrocellulose filters, and hybridized with the probe by a modification of the method of Hu and Messing (30, 31). The sequences of positively hybridizing subclones were determined by the dideoxynucleotide chain termination method (19).

RESULTS

Probe Design. Synthetic gene fragments were designed by using the known amino acid sequence for BPTI (32) and consensus codon usage data compiled from the sequences of 26 mammalian genes (33). Initially, a synthetic gene sequence (including a termination codon) was written by using, for each amino acid, the most likely codon from the mammalian consensus data. This sequence was then compared with its complement by using the computer program DIAGON (34) to detect any highly self-complementary regions that might interfere with the construction of the probe. Each such region greater than four nucleotides in length and within 50 nucleotides of its complement was removed from the sequence by substituting, where appropriate, the next most likely codon. Only a few substitutions were required to reduce the probe self-complementarity to within the stated limits.

Probe Construction. Ten DNA oligonucleotides (P1 through P10), corresponding to overlapping portions of both the synthetic BPTI gene sequence and its complement, were made by the solid-phase phosphotriester method. These oligonucleotides were annealed and ligated to form a longer double-stranded DNA fragment and then were converted to the fully duplex form by "filling-in" the ends. An 86-base-pair (bp) synthetic BPTI gene fragment was made in this manner and cloned in both orientations by using the vector M13mp8 (Fig. 1*A*). A single-stranded probe was prepared by primed DNA synthesis across the insert in the M13 recombinant, followed by *Eco*RI digestion of the partially duplex DNA and isolation of the probe fragment by polyacrylamide gel electrophoresis under denaturing conditions (Fig. 1*B*). Shorter probes, made from fewer oligonucleotides, also were constructed in a similar manner.

Genomic Southern Blots. High molecular weight DNA prepared from bovine kidney nuclei was digested with various restriction enzymes, and then the resulting fragments were separated by agarose gel electrophoresis and probed by Southern blot hybridization (Fig. 2). The hybridization and washing temperatures were adjusted until several specific bands were clearly

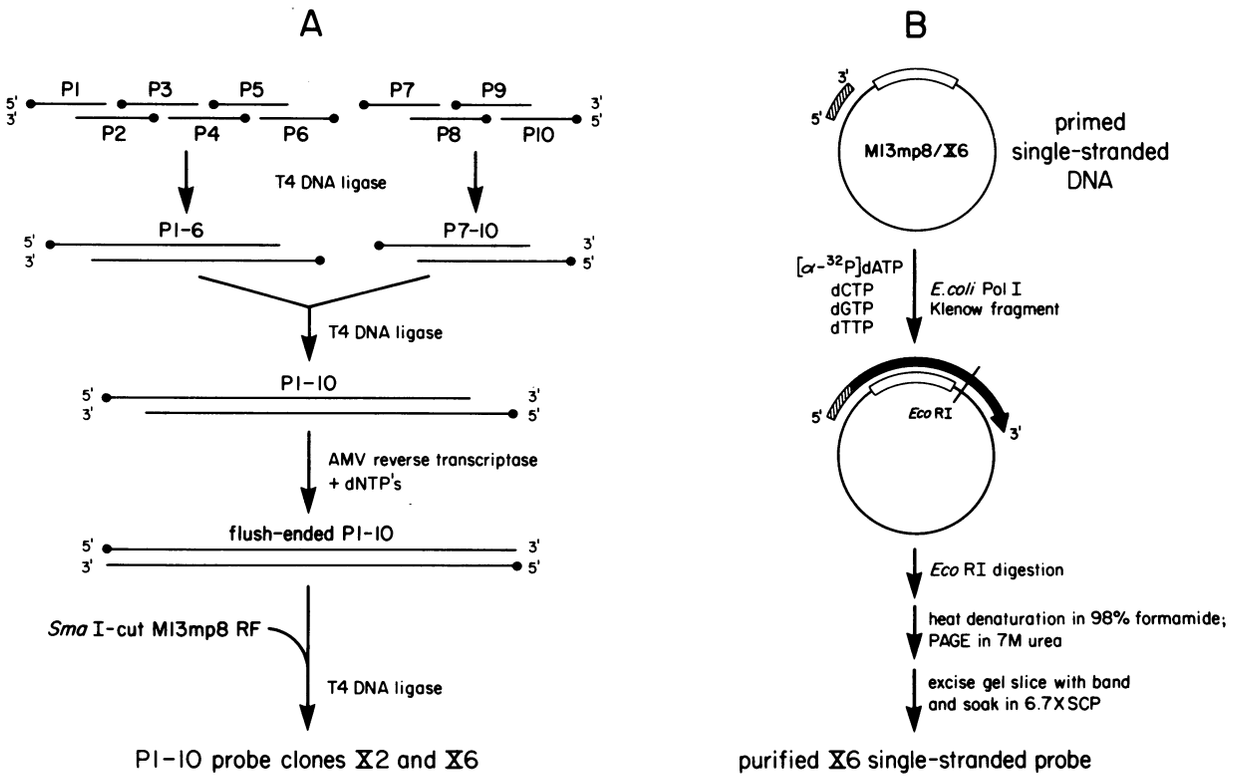

FIG. 1. Construction of the synthetic probes. (*A*) Two mixtures of oligonucleotides, P1–P6 and P7–P10, were separately annealed and ligated, and the resultant products were purified by polyacrylamide gel electrophoresis (PAGE) under denaturing conditions. These products, P1–P6 and P7–P10, were mixed, annealed, and ligated again. The result of this ligation, P1–P10, was purified by gel electrophoresis, "filled-in" by treatment with AMV reverse transcriptase to produce an 86-bp fragment, and cloned into the *Sma* I site of M13mp8. M13 recombinants with the P1–P10 sequence in each orientation, X2 and X6, were obtained. A 66-bp synthetic gene fragment composed of oligonucleotides P1–P8 also was assembled and cloned using the same procedures. (*B*) A standard DNA sequence assay primer (14) was annealed to single-stranded DNA from the M13 recombinant, and a DNA synthesis reaction was carried out in the presence of ^{32}P-labeled deoxynucleoside triphosphates. The semiduplex product of this reaction was cut at the unique *Eco*RI site distal to the insert (18), and the probe was isolated by polyacrylamide gel electrophoresis under denaturing conditions. SCP buffer (1×), 30 mM sodium phosphate, pH 6.2/100 mM NaCl/1 mM EDTA (22); □, synthetic BPTI gene fragment insert; ▨, DNA sequence assay primer; ➡, ^{32}P-labeled DNA; Pol I, DNA polymerase I.

visible over the nonspecific background hybridization. With the 86-mer probes X2 and X6, only two bands were visible in the *Hin*dIII and *Bam*HI digests, and a single band of double intensity was visible in the *Eco*RI digest. The same bands also were present in a blot hybridization with a shorter 66-mer probe, but with this probe the nonspecific background hybridization was higher and some additional faint bands were evident. No bands were visible in the control hybridization with a single-stranded probe prepared from the vector M13mp8 without an insert (Fig. 2).

Probing a Bovine Genomic Library. A bovine genomic library containing approximately 2×10^6 independent recombinants was prepared by using the vector λ1059 (24). Approximately 10^6 recombinant plaques were screened by hybridization with the 86-mer probe, and five positively hybridizing clones were detected. During the plaque hybridizations, which were carried out under the same conditions as used for the Southern blot hybridizations, the level of nonspecific binding of the probe to the filter was extremely low (Fig. 3).

DNA Sequence Analysis. Four of the five positively hybridizing clones were plaque-purified and analyzed further. Preliminary restriction mapping data (not shown) indicated that two of the clones represented partially overlapping segments of one region of the genome, and the other two represented partially overlapping segments of a second region. The restriction digests contained 3.8- and 4.0-kbp *Eco*RI fragments and 3.5- and 7.3-kbp *Bam*HI fragments that hybridized to the probe; these correspond to the sizes of the *Eco*RI and *Bam*HI fragments that were detected in the genomic Southern blot hybridization with the same probe (Fig. 2). However, for each clone and for any given digest, there was only a single positively hybridizing fragment. The simplest interpretation of these data is that the haploid bovine genome contains only two genes that hybridize to the 86-mer probe under the conditions used and that these genes are separated in the genome by at least 11 kbp (the minimum insert size).

One of the clones (λX6-6B) was chosen for sequence analysis. Total DNA from the phage λ recombinant was randomly fragmented and "shotgun" cloned into the *Sma* I site of the vector M13mp8 (18, 27). Approximately 900 independent M13 subclones were probed with the positive strands of the 86-mer M13 recombinants X2 and X6 (30, 31). Six positively hybridizing subclones were isolated and their sequences determined by the dideoxynucleotide chain termination method. A contiguous 677 nucleotides of bovine DNA sequence were obtained from the overlapping subclone sequences; contained within this was an open reading frame, the translation of which perfectly matched the amino acid sequence of BPTI (Fig. 4). However, the BPTI coding region was neither preceded by a start codon nor followed by a termination codon but rather was embedded in a longer open reading frame.

Probe Homology. Fig. 5 shows an alignment of the 86-mer and 66-mer probes with the BPTI coding region. In each case the homology between the probe and the actual genomic sequence was 74%. Three arginine AG$\frac{A}{G}$ codons were found in the region of the genomic sequence overlapped by the probe,

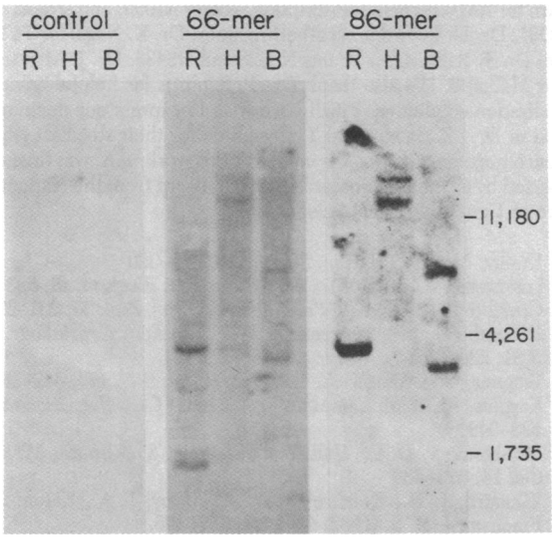

FIG. 2. Genomic Southern blots of bovine DNA. *Eco*RI (lanes R), *Hin*dIII (lanes H), and *Bam*HI (lanes B) restriction digests of bovine DNA were fractionated by electrophoresis on a 0.5% agarose gel and blotted onto nitrocellulose. (*Center* and *Right*) Results of hybridizations with probes made from cloned 66-bp and 86-bp synthetic BPTI gene fragments, respectively (see Fig. 1). (*Left*) Control, representing a hybridization performed with a probe prepared from the M13mp8 vector with no insert. Hybridizations and washings were performed as described in *Materials and Methods* except that the temperature of hybridization for the 66-mer probe was 60°C. The faint signal present in the *Hin*dIII lane of the control blot is a nonreproducible artifact of the hybridization procedure. Numbers in the margin are the lengths (in bp) of several restriction fragments of bovine mitochondrial DNA that were coelectrophoresed as size standards.

whereas the probe codons used to specify arginine were all $CG\frac{T}{C}$. This indicates that occasional grossly incorrect codon choices do not affect seriously the specificity of the probe.

DISCUSSION

In this paper we describe the isolation of a gene for bovine pancreatic trypsin inhibitor with a unique-sequence synthetic DNA probe. The translated DNA sequence of the BPTI genomic clone contains a region that exactly matches the amino acid sequence of BPTI (Fig. 4), but definitive proof that this is an active gene will require the isolation of a BPTI cDNA clone. The BPTI gene is located on a 4.0-kbp *Eco*RI fragment and a 3.5-kbp *Bam*HI fragment, and we have also cloned a positively hybridizing gene located on a 3.8-kbp *Eco*RI fragment and a 7.3-kbp *Bam*HI fragment. Together, these two cloned genes can account for all of the bands detected by the 86-mer probe on the genomic Southern blot (Fig. 2). DNA sequence analysis of the second gene (unpublished data) indicates that it codes for a protein similar but not identical to bovine serum trypsin inhibitor (36).

The DNA sequence upstream of the BPTI coding region contains two in-frame termination codons but no in-frame initiation codons (Fig. 4). In between the two termination codons and the start of the BPTI coding region, there are two long pyrimidine stretches, each of which is followed by the dinucleotide A-G; both of these closely match a consensus intron/exon boundary sequence (35). Downstream, an in-frame termination codon exists 52 nucleotides from the end of the BPTI coding region. This stretch is very purine-rich and also contains a possible exon/intron junction (Fig. 4). Therefore, it is likely that

FIG. 4. BPTI gene sequence. Shown is a partial sequence of the bovine genomic clone λX6-6B, representing the region of the cloned DNA that hybridizes to the synthetic 86-mer probe. Overlined is the portion of the translated DNA sequence that matches the amino acid sequence of BPTI. *, In-frame termination codons; ▲, potential intron/exon ("*A*ceptor") and exon/intron ("*D*onor") splice junctions (35).

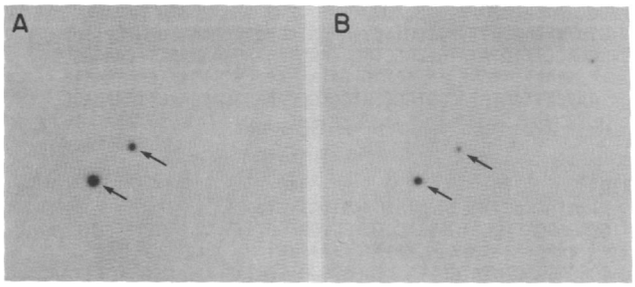

FIG. 3. Plaque hybridizations. Duplicate filters (*A* and *B*) were lifted from a 90-mm Petri dish (containing 20,000 confluent plaques from the bovine genomic library), hybridized with the synthetic 86-mer probe under the same conditions used for the Southern blots shown in Fig. 2, and autoradiographed. Arrows indicate signals from two positively hybridizing clones. Note that the circular outline of the filter, due to nonspecific background hybridization, is virtually invisible.

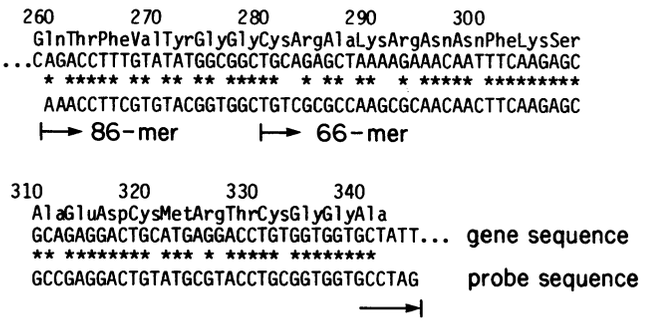

FIG. 5. Probe homology. The sequences of the synthetic BPTI gene fragment probes and the actual BPTI gene sequence (Fig. 4) are shown aligned. Homology is indicated by asterisks. The 86-mer probe and the 66-mer probe are homologous to nucleotides 260–345 and 280–345 (Fig. 4), respectively.

the BPTI coding region exists as part of an exon in the gene for a larger polypeptide and that BPTI is released through proteolytic processing at both its NH_2 and COOH termini from this primary translation product. The nature of such a precursor is unknown, however.

The synthetic gene fragment probing procedure given here is a general one that can be used to isolate the gene for any protein, as long as sufficient amino acid sequence information is available and it is possible to prepare a genomic DNA library from the organism making that protein. This cloning approach should prove especially useful in those cases where the site of synthesis of a particular protein is unknown or when it is difficult to isolate mRNA from the relevant tissue. The minimum probe length that will work at the expected 70–80% homology level has not been established and must vary somewhat depending on the amino acid sequence, codon choice, and whether or not an intron interrupts the coding region being sought. Nevertheless, 20–25 residues of reliable, contiguous amino acid sequence for a given protein should render its gene a candidate for cloning by this method. In general, the proper hybridization/washing conditions can be found and the effectiveness of a given probe tested by performing a simple genomic Southern blot hybridization (Fig. 2).

At present, the only reliable way to determine splice junctions is to compare the cDNA sequence for a gene with its corresponding genomic sequence. Hence, in order to fully characterize the information content of a gene, it is necessary to isolate both cDNA and genomic clones. The unique-sequence synthetic DNA probes discussed here and elsewhere (13) should be capable of isolating cDNA clones of even extremely low-abundance mRNAs from any cDNA library, as long as it is large enough to contain at least one copy of the desired clone. Alternatively, a genomic clone for the gene of interest could be isolated and subcloned into a retroviral or simian virus 40 expression vector; subsequent transfection of an appropriate host cell should then produce correctly spliced transcripts of the cloned gene (37–39), from which cDNA could be prepared.

We gratefully acknowledge the advice and guidance concerning oligonucleotide synthesis that we received from Drs. M. J. Gait, H. W. D. Matthes, B. Sproat, M. Singh, and D. Brown; we especially thank Drs. M. Singh, B. Sproat, and H. W. D. Matthes for synthesizing oligodeoxynucleotides P2, P4, and P9, respectively. We are indebted to many colleagues for gifts of materials used in this work: Drs. M. J. Gait, H. W. D. Matthes, and B. Sproat for derivatized support resins; Dr. J. Karn for AMV reverse transcriptase, vector λ1059, and strains Q359 and D91; Dr. D. Brown for synthetic primer; Dr. K. Nagai for T4 DNA ligase; Dr. T. Rabbitts for strains NS428 and NS433; Dr. J. Messing for vector M13mp8. We also thank Dr. T. Rabbitts for helpful advice on hybridization conditions. Finally, we wish to express our deep appreciation to Drs. F. Sanger and T. Creighton for their steadfast support and encouragement during the course of this work. S.A. was financially supported by a National Research Service Award from the National Institute of General Medical Sciences.

1. Kunitz, M. (1947) *J. Gen. Physiol.* **30**, 311–320.
2. Laskowski, M. & Kato, I. (1980) *Annu. Rev. Biochem.* **49**, 593–626.
3. Creighton, T. E. (1978) *Prog. Biophys. Mol. Biol.* **33**, 231–297.
4. Deisenhofer, J. & Steigemann, W. (1975) *Acta. Crystallogr. Sect. B* **31**, 238–250.
5. Wagner, G. & Wuthrich, K. (1982) *J. Mol. Biol.* **155**, 347–366.
6. Karplus, M. & McCammon, J. A. (1981) *Crit. Rev. Biochem.* **9**, 293–349.
7. Montgomery, D. L., Hall, B. D., Gillam, S. & Smith, M. (1978) *Cell* **14**, 673–680.
8. Kinnaird, J. H., Keighren, M. A., Kinsey, J. A., Eaton, M. & Fincham, J. R. S. (1982) *Gene* **20**, 387–396.
9. Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. H. & Itakura, K. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6613–6617.
10. Gillam, S., Waterman, K. & Smith, M. (1975) *Nucleic Acids Res.* **2**, 625–633.
11. Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T. & Itakura, K. (1979) *Nucleic Acids Res.* **6**, 3543–3557.
12. Wallace, R. B., Johnson, M. J., Hirose, T., Miyake, T., Kawashima, E. H. & Itakura, K. (1981) *Nucleic Acids Res.* **9**, 879–894.
13. Jaye, M., de la Salle, H., Schamber, F., Balland, A., Kohli, V., Findeli, A., Tolstoshev, P. & Lecocq, J.-P. (1983) *Nucleic Acids Res.* **11**, 2325–2335.
14. Duckworth, M. L., Gait, M. J., Goelet, P., Hong, G.-F., Singh, M. & Titmas, R. C. (1981) *Nucleic Acids Res.* **9**, 1691–1706.
15. Gait, M. J., Matthes, H. W. D., Singh, M., Sproat, B. S. & Titmas, R. C. (1982) *Nucleic Acids Res.* **10**, 6243–6254.
16. Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F. & Boyer, H. W. (1977) *Science* **198**, 1056–1063.
17. Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87**, 107–110.
18. Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276.
19. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
20. Burgoyne, L. A., Wagar, M. A. & Atkinson, M. R. (1970) *Biochem. Biophys. Res. Commun.* **39**, 254–259.
21. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
22. Bentley, D. L. & Rabbitts, T. H. (1983) *Cell* **32**, 181–189.
23. Laskey, R. A. & Mills, A. D. (1977) *FEBS Lett.* **82**, 314–316.
24. Karn, J., Brenner, S., Barnett, L. & Cesareni, G. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5172–5176.
25. Sternberg, N., Tiemeier, D. & Enquist, L. (1977) *Gene* **1**, 255–280.
26. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
27. Anderson, S. (1981) *Nucleic Acids Res.* **9**, 3015–3027.
28. Deininger, P. L. (1983) *Anal. Biochem.* **129**, 216–223.
29. Messing, J. (1979) *Recomb. DNA Techn. Bull.* **2**, 43–48.
30. Hu, N.-T. & Messing, J. (1982) *Gene* **17**, 271–277.
31. Brown, D. M., Frampton, J., Goelet, P. & Karn, J. (1982) *Gene* **20**, 139–144.
32. Kassell, B., Radicevic, M., Ansfield, M. J. & Laskowski, M., Sr. (1965) *Biochem. Biophys. Res. Commun.* **18**, 255–258.
33. Grantham, R., Gautier, C., Gong, M., Jacobzone, M. & Mercier, R. (1981) *Nucleic Acids Res.* **9**, r43–r74.
34. Staden, R. (1982) *Nucleic Acids Res.* **10**, 2951–2961.
35. Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472.
36. Wachter, E., Deppner, K., Hochstrasser, K., Lempart, K. & Geiger, R. (1980) *FEBS Lett.* **119**, 58–62.
37. Shimotohno, K. & Temin, H. M. (1982) *Nature (London)* **299**, 265–268.
38. Hamer, D. H. & Leder, P. (1979) *Nature (London)* **281**, 35–40.
39. Mulligan, R. C. & Berg, P. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2072–2076.