Method for polypeptide production involving expression of a heterologous gene, recombinant microbial cloning vehicle containing said gene and bacterial culture transformed by said cloning vehicle.

Priority: 08.11.77 US 849591

Date of publication of application: 16.05.79 Bulletin 79/10

Publication of the grant of the patent: 24.05.89 Bulletin 89/21

Designated Contracting States: BE CH DE FR GB LU NL SE

References cited:
CHEMICAL & ENGINEERING NEWS, vol. 50, no. 45, 7 Nov 1977; “Human gene in E. coli: it works!”, p. 4
CHEMICAL ABSTRACTS, vol. 89, 1978, Columbus, OH (US); L.VILLA-KOMAROFF et al.: “A bacterial clone synthesizing proinsulin”, no. 159988n

Proprietor: GENENTECH, INC.
460 Point San Bruno Boulevard
South San Francisco California 94080 (US)

Inventor: Riggs, Arthur Dale
4852 Saint Andrews
La Vern, CA (US)

Representative: Goldin, Douglas Michael et al
J.A. KEMP & CO. 14, South Square Gray’s Inn
London WC1R 5EU (GB)

References cited:
CHEMICAL ABSTRACTS, vol. 86, 1977, Columbus, OH (US); L.A.McREYNOLDS et al.: “The ovalbumin gene. Ill. Insertion of ovalbumin gene sequences in chimeric bacterial plasmids”, no. 184866s
CHEMICAL ABSTRACTS, vol. 87, 1977, Columbus, OH (US); P.HUMPHRIES et al.: "Molecular cloning of extensive sequences of the in vitro synthesized chicken ovalbumin structural gene", no. 114458g

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European patent convention).
References cited:

Columbus, OH (US); F.DIVAR et al.: "Construction and characterization of new cloning vehicles. II. A multipurpose cloning system", no. 117616p
Description

Technical field

This invention relates to microbial polypeptide expression.

Background

Genetic information is encoded on double-stranded deoxyribonucleic acid ("DNA" or "genes") according to the order in which the DNA coding strand presents the characteristic bases of its repeating nucleotide components. "Expression" of the encoded information to form polypeptides involves a two-part process. According to the dictates of certain control regions ("regulons") in the gene, RNA polymerase may be caused to move along the coding strand, forming messenger RNA (ribonucleic acid) in a process called "transcription". In a subsequent "translation" step the cell’s ribosomes in conjunction with transfer RNA convert the mRNA "message" into polypeptides. Included in the information mRNA transcribes from DNA are signals for the start and termination of ribosomal translation, as well as the identity and sequence of the amino acids which make up the polypeptide. The DNA coding strand comprises long sequences of nucleotide triplets called "codons" because the characteristic bases of the nucleotides in each triplet or codon encode specific bits of information. For example, 3 nucleotides read as ATG (adenine-thymine-guanine) result in an mRNA signal interpreted as "start translation", while termination codons TAG, TAA and TGA are interpreted "stop translation". Between the start and stop codons lie the so-called structural gene, whose codons define the amino acid sequence ultimately translated. That definition proceeds according to the well-established "genetic code" (e.g., J. D. Watson, Molecular Biology of the Gene W. A. Benjamin Inc., N. Y., 3rd ed. 1976) which describes the codons for the various amino acids. The genetic code is degenerate in the sense that different codons may yield the same amino acid, but precise in that for each amino acid there are one or more codons for it and no other. Thus, for example, all of the codons TCT, TCC, TCA, TCG, AGT and AGC when read as such, encode for serine and no other amino acid. During translation the proper reading phase or reading frame must be maintained. Consider for example what happens when the ribosome reads different bases as the beginning of a codon (underlined) in the sequence

\[
\text{GCT GGT TGT AAG} \quad \rightarrow \quad \text{Ala-Gly-Cys-Lys} \\
\text{G CTT GGT GTA AG} \quad \rightarrow \quad \text{Leu-Val-Val} \\
\text{GC TGG TGG TAA G} \quad \rightarrow \quad \text{Trp-Leu-(STOP)}
\]

The polypeptide ultimately produced, then, depends vitally upon the spatial relationship of the structural gene with respect to the region.

A clearer understanding of the process of genetic expression will emerge once certain components of genes are defined:

- **Operon**—A gene comprising structural gene(s) for polypeptide expression and the control region ("regulon") which regulates that expression.
- **Promoter**—A gene within the regulon to which RNA polymerase must bind for initiation of transcription.
- **Operator**—A gene to which repressor protein may bind, thus preventing RNA polymerase binding on the adjacent promoter.
- **Inducer**—A substance which deactivates repressor protein, freeing the operator and permitting RNA polymerase to bind to promoter and commence transcription.

**Catabolite Activator Protein ("CAP") binding site**—A gene which binds cyclic adenosine monophosphate ("cAMP")-mediated CAP, also commonly required for initiation of transcription. The CAP binding site may in particular cases be unnecessary. For example, a promoter mutation in the lactose operon of the phage λ plac UV5 eliminates the requirement for cAMP and CAP expression. J. Beckwith et al., *J. Mol. Biol. 69*, 155–160 (1972).

**Operon-operator system**—As used herein, an operable control region of an operon, with or without respect to its inclusion of a CAP binding site or capacity to code for repressor protein expression.

Further by way of definition, and for use in the discussion of recombinant DNA which follows, we define the following:

- **Cloning vehicle**—Non-chromosomal double stranded DNA comprising an intact "replicon" such that the vehicle is replicated, when placed within a unicellular organism ("microbe") by a process of "transformation". An organism so transformed is called a "transformant".
- **Plasmid**—For present purposes, a cloning vehicle derived from viruses or bacteria, the latter being "bacterial plasmids".
- **Complementary**—A property conferred by the base sequences of single strand DNA which permits the formation of double stranded DNA through hydrogen bonding between complementary bases on the respective strands. Adenine (A) complements thymine (T), while guanine (G) complements cytosine (C).
EP 0 001 930 B1

Advances in biochemistry in recent years have led to the construction of “recombinant” cloning vehicles in which, for example, plasmids are made to contain exogenous DNA. In particular instances the recombinant may include “heterologous” DNA, by which is meant DNA that codes for polypeptides ordinarily not produced by the organism susceptible to transformation by the recombinant vehicle. Thus, plasmids are cleaved to provide linear DNA having ligatable termini. These are bound to an exogenous gene having ligatable termini to provide a biologically functional moiety with an intact replicon and a desired phenotypical property. The recombinant moiety is inserted into a microorganism by transformation and transformants are isolated and cloned, with the object of obtaining large populations capable of expressing the new genetic information. Methods and means of forming recombinant cloning vehicles and transforming organisms with them have been widely reported in the literature. See, e.g., H. L. Heynecker et al., *Nature* 263, 748–752 (1976); Cohen et al., *Proc. Nat. Acad. Sci. USA* 69, 2110 (1972); *ibid.*, 70, 1293 (1973); *ibid.*, 70, 3240 (1973); *ibid.*, 71, 1030 (1974); Morrow et al., *Proc. Nat. Acad. Sci. U.S.A.* 71, 1743 (1974); Novick, *Bacteriological Rev.*, 33, 210 (1969); Hershfield et al., *Proc. Nat'l. Acad. Sci. U.S.A.* 71, 3455 (1974) and Jackson et al., *ibid.* 69, 2904 (1972). A generalized discussion of the subject appears in S. Cohen, *Scientific American* 233, 24 (1975). These and other publications alluded to herein are incorporated by reference.

A variety of techniques are available for DNA recombination, according to which adjoining ends of separate DNA fragments are tailored in a number of ways to facilitate ligation. The latter term refers to the formation of phosphodiester bonds between adjoining nucleotides, most often through the agency of the enzyme T4 DNA ligase. Thus, blunt ends may be directly ligated. Alternatively, fragments containing complementary single strands at their adjoining ends are advantage by hydrogen bonding which positions the respective ends for subsequent ligation. Such single strands, referred to as cohesive termini, may be formed by the addition of nucleotides to blunt ends using terminal transferase, and sometimes simply by chewing back one strand of a blunt end with an enzyme such λ-exonuclease. Again, and most commonly, resort may be had to restriction endonucleases, which cleave phosphodiester bonds in and around unique sequences of nucleotides of about 4–6 base pairs in length. Many restriction endonucleases and their recognition sites are known, the so-called Eco RI endonuclease being most widely employed. Restriction endonucleases which cleave double-stranded DNA at rotationally symmetric “palindromes” leave cohesive termini. Thus, a plasmid or other cloning vehicle may be cleaved, leaving termini comprising half the restriction endonuclease recognition site. A cleavage product of exogenous DNA obtained with the same restriction endonuclease will have ends complementary to those of the plasmid termini. Alternatively, as disclosed *infra*, synthetic DNA comprising cohesive termini may be provided for insertion into the cleaved vehicle. To discourage rejoiner of the vehicles' cohesive termini pending insertion of exogenous DNA, the termini can be digested with alkaline phosphatase, providing molecular selection for closures incorporating the exogeneous fragment. Incorporation of a fragment having the proper orientation relative to other aspects of the vehicle may be enhanced when the fragment supplants vehicle DNA excised by two different restriction endonucleases, and itself comprises termini respectively constituting half the recognition sequence of the different endonucleases.

Despite wide-ranging work in recent years in recombinant DNA research, few results susceptible to immediate and practical application have emerged. This has proven especially so in the case of failed attempts to express polypeptides and the like coded for by “synthetic DNA”, whether constructed nucleotide by nucleotide in the conventional fashion or obtained by reverse transcription from isolated in RNA (complementary or “cDNA”). In this application we describe what appears to represent the first expression of a functional polypeptide product from a synthetic gene, together with related developments which promise wide-spread application. The product referred to is somatostatin (Guillemin, U.S.P. 3,904,694), an inhibitor of the secretion of growth hormone, insulin and glucagon whose effects suggest its application in the treatment of acromegaly, acute pancreatitis and insulin-dependent diabetes. See R. Guillemin et al, *Annual Rev. Med.* 27 379 (1976). The somatostatin model clearly demonstrates the applicability of the new developments described here on numerous and beneficial fronts, as will appear from the accompanying drawings and more clearly from the detailed description which follows.

Summary of invention

According to the invention there is provided a process for producing a desired heterologous polypeptide which comprises growing a bacterium transformed with a recombinant plasmid comprising a homologous regulon, a DNA sequence in proper reading frame therewith, and one or more termination codons, wherein said DNA sequence is positioned between said regulon and the termination codon(s) and encodes said heterologous polypeptide, additional protein, and a selective cleavage site between the desired heterologous polypeptide and additional protein, wherein the DNA sequence is expressed and the resulting conjugate polypeptide is sufficiently large not to be degraded by endogenous proteolytic enzymes and is specifically cleaved to give desired heterologous polypeptide.

Also according to the invention there is provided a recombinant plasmid comprising a homologous regulon, a DNA sequence encoding a desired specific heterologous polypeptide, a selective cleavage site and additional protein, and one or more termination codons, wherein the said DNA sequence is interposed in proper reading frame between said regulon and termination codon(s) such that a conjugate polypeptide comprising both the desired heterologous polypeptide and additional protein results from expression in a
bacterium transformed with the plasmid, there being a selective cleavage site between the desired heterologous polypeptide and the additional protein.

Brief description of the drawings

The accompanying drawings illustrate one context in which preferred embodiments of the invention find application, i.e., expression of the hormone somatostatin by bacterial transformants containing recombinant plasmids.

Figure 1. Schematic outline of the process: the gene for somatostatin, made by chemical DNA synthesis, is fused to the E. coli β-galactosidase gene on the plasmid pBR322. After transformation into E. coli, the recombinant plasmid directs the synthesis of a precursor protein which can be specifically cleaved in vitro at methionine residues by cyanogen bromide to yield active mammalian polypeptide hormone. A, T, C and G denote the characteristic bases (respectively adenine, thymine, cytosine and guanine) of the deoxyribonucleotides in the coding strand of the somatostatin gene.

Figure 2. Schematic structure of a synthetic gene whose coding strand (i.e., the "upper" strand) comprises codons for the amino acid sequence of somatostatin (given).

Figure 3. Schematic illustration of a preferred method for construction of nucleotide trimers used in constructing synthetic genes. In the conventional notation employed to depict nucleotides in Figure 3, the 5' OH is to the left and the 3' OH to the right, e.g.

\[
\begin{align*}
5' & \quad B \\
OH & \quad 3' \\
\end{align*}
\]

Figure 4. Flow chart for the construction of a recombinant plasmid (e.g., pSOM11-3) capable of expressing a somatostatin ("SOM")-containing protein, beginning with the parental plasmid pBR322. In Figure 4 the approximate molecular weight of each plasmid is stated in daltons ("d"). Ap' and Tc' respectively denote genes for ampicillin and tetracycline resistance, while Tc" denotes tetracycline susceptibility resulting from excision of a portion of the Tc' gene. The relative positions of various restriction endonuclease specific cleavage sites on the plasmids are depicted (e.g., Eco RI, Bam I, etc.).

Figures 5A and 5B. The nucleotide sequences of key portions of two plasmids are depicted, as is the direction of messenger RNA ("mRNA") transcription, which invariably proceeds from the 5' end of the coding strand. Restriction endonuclease substrate sites are as shown. Each depicted sequence contains both the control elements of the lac (lactose) operon, and codons for expression of the amino acid sequence of somatostatin (italics). The amino acid sequence numbers for β-galactosidase ("β-gal") are in brackets.

Figures 6—8. As more particularly described in the "Experimental" discussion, infra, these depict the results of comparative radioimmune assay experiments which demonstrate the somatostatin activity of the product expressed by the recombinant plasmids.

Figure 9. Schematic structure of synthetic genes whose coding strands comprise codons for the amino acid sequences of the A and B strands of human insulin.

Figure 10. Flow chart for construction of a recombinant plasmid capable of expressing the B chain of human insulin.

Detailed description

1. Preparation of genes coding for heterologous polypeptide

DNA coding for any polypeptide of known amino acid sequence may be prepared by choosing codons according to the genetic code. For ease in purification, etc., oligodeoxyribonucleotide fragments of, for example, from about 11 to about 16 nucleotides are prepared separately, then assembled in the desired sequence. Thus, one prepares first and second series of oligodeoxyribonucleotide fragments of convenient size. The first series, when joined in proper sequence, yield a DNA coding strand for polypeptide expression (see, e.g., Figure 2, fragments A, B, C and D). The second series, when likewise joined in proper sequence, yield a strand complementary to the coding strand (e.g., Figure 2, fragments E, F, G and H). The fragments of the respective strands preferably overlap such that complementarity promotes their self assembly through hydrogen bonding of the cohesive termini of fragment blocks. Following assembly, the structural gene is completed by ligation in the conventional manner.

The degeneracy of the genetic code permits substantial freedom in the choice of codons for any given amino acid sequence. For present purposes, however, codon choice was advantageously guided by three considerations. First, codons and fragments were selected, and fragment assembly was staged, so as to avoid undue complementarity of the fragments, one with another, save for fragments adjacent one another in the intended gene. Secondly, sequences rich in AT base pairs (e.g., about five or more) are avoided, particularly when preceded by a sequence rich in GC base pairs, to avoid premature termination of transcription. Thirdly, at least a majority of the codons chosen are those preferred in the expression of microbial genomes (see, e.g., W. Fiers, et al, Nature 260, 500 (1976). For purposes of the appended claims, we define the following as codons "preferred for the expression of microbial genomes":

5
### Table I

**Preferred assignment of codons**

<table>
<thead>
<tr>
<th>First position (5' end)</th>
<th>Second position (read across)</th>
<th>Third position (3' end)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(read down)</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>phe</td>
<td>ser</td>
</tr>
<tr>
<td></td>
<td>phe</td>
<td>tyr</td>
</tr>
<tr>
<td></td>
<td>leu</td>
<td>Stop</td>
</tr>
<tr>
<td></td>
<td>leu</td>
<td>pro</td>
</tr>
<tr>
<td></td>
<td>leu</td>
<td>pro</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>pro</td>
</tr>
<tr>
<td></td>
<td>il e</td>
<td>thr</td>
</tr>
<tr>
<td></td>
<td>thr</td>
<td>as n</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>met (start)</td>
</tr>
<tr>
<td></td>
<td>val</td>
<td>ala</td>
</tr>
<tr>
<td></td>
<td>val</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>val</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Most preferably in the case of somatostatin, the amino acid (codon) relationships of the structural gene are: gyl (GGT); cys (TGT); lys (AAG); trp (TGG); ala (GCT, GCG); asn (AAT, AAC); phe (TTC, TTT); thr (ACT, ACG); and ser (TCC, TCG).

Where the structural gene of a desired polypeptide is to be inserted in a cloning vehicle for expression as such, the gene is preceded by a "start" codon (e.g., ATG) and immediately followed by one or more termination or stop codons (see Fig. 2). However, as described *infra*, the amino acid sequence of a particular polypeptide may be expressed with additional protein preceding and/or following it. If the intended use of the polypeptide requires cleavage of the additional protein, appropriate cleavage sites are coded for adjacent the polypeptide—additional protein codon junction. Thus, in Figure 1 as an example, the expression product is a precursor protein comprising both somatostatin and the greatest part of the β-galactosidase polypeptide. Here ATG is not required to code for the start of translation because ribosomal translation of the additional β-gal polypeptide reads through into the somatostatin structural gene. Incorporation of the ATG signal, however, codes for the production of methionine, an amino acid specifically cleaved by cyanogen bromide, affording a facile method for converting precursor protein into the desired polypeptide.

Figure 2 also exemplifies a further feature preferred in heterologous DNA intended for recombinant employment, i.e., the provision of cohesive termini, preferably comprising one of the two strands of a restriction endonuclease recognition site. For reasons previously discussed, the termini are preferably designed to create respectively different recognition sites upon recombination.

While the developments described here have been demonstrated as successful with the somatostatin model, it will be appreciated that heterologous DNA coding for virtually any known amino acid sequence may be employed, *mutatis mutandis*. Thus, the techniques previously and hereafter described are applicable, *mutatis mutandis*, to the production of poly(amino)acids, such as polyleucine and polyalanine; enzymes; serum proteins; analgesic polypeptides, such as β-endorphins, which modulate thresholds of pain, etc. Most preferably, the polypeptides produced as such will be mammalian hormones or intermediates therefor. Among such hormones may be mentioned, e.g., somatostatin, human insulin, human and bovine growth hormone, luteinizing hormone, ACTH, pancreatic polypeptide, etc. Intermediates include, for example, human proinsulin, human proinsulin, the A and B chains of human insulin and so on. In addition to DNA made in vitro, the heterologous DNA may comprise cDNA resulting from reverse transcription from mRNA. See, e.g., Ulrich et al, *Science* 196, 1313 (1977).

2. Recombinants coding for the expression of precursor protein

In the process schematically depicted in Figure 1, expression yields a precursor protein comprising both a polypeptide coded for by a specific heterologous structural gene (somatostatin) and additional protein (comprising a portion of the β-galactosidase enzyme). A selective cleavage site adjacent the somatostatin amino acid sequence permits subsequent separation of the desired polypeptide from
superfluous protein. The case illustrated is representative of a large class of procedures made available by the techniques described herein.

Most commonly, cleavage will be effected outside the replicative environment of the Plasmid or other vehicle as, for example, following harvest of the microbial culture. In this fashion temporary conjugation of small polypeptides with superfluous protein may preserve the former against, e.g., in vivo degradation by endogenous enzymes. At the same time, the additional protein will ordinarily rob the desired polypeptide of bioactivity pending extra-cellular cleavage, with the effect of enhancing the biosafety of the procedure. In particular instances, of course, it may prove desirable to effect cleavage within the cell. For example, cloning vehicles could be provided with DNA coding for enzymes which convert insulin precursors to the active form, operating in tandem with other DNA coding for expression of the precursor form.

In the preferred case, the particular polypeptide desired lacks internal cleavage sites corresponding to that employed to shed superfluous protein, although it will be appreciated that where the condition is not satisfied competition reactions with will yet give the desired product, albeit in lower yield. Where the desired product is methionine-free, cyanogen bromide cleavage at methionine adjacent the desired sequence has proven highly effective. Likewise, arginine- and lysine-free products may be enzymatically cleaved with, e.g., trypsin at arg-arg, lys-lys or like cleavage sites adjacent the desired sequence. In the case where cleavage leaves, e.g., unpaired arginine attached to desired product, it may be removed by carboxypeptidase digestion. When trypsin is employed to cleave at arg-arg, lysine sites within the desired polypeptide may first be protected, as with maleic or citraconic anhydrides. The cleavage techniques discussed here by way of example are but representative of the many variants which will occur to the art-skilled in light of the specification.

Cleavable protein may be expressed adjacent either the C- or N-terminals of a specific polypeptide, or even within the polypeptide itself, as in the case of the included sequence which distinguishes proinsulin and insulin. Again, the vehicle employed may code for expression of protein comprising repeated sequences of the desired polypeptide, each separated by selective cleavage sites. Most preferably, however, codons for superfluous protein will be translated in advance of the structural gene of the desired product, as in the case illustrated in the Figures. In every case care should be taken to maintain the proper reading frame relative to the regulon.

3. Expression of immunogens

The ability to express both a specific polypeptide and superfluous protein provides useful tools for the production of immunogenic substances. Polypeptide “haptens” (i.e. substances containing determinants specifically bound by antibodies and the like but ordinarily too small to elicit an immune response) can be expressed as conjugates with additional protein sufficient in size to confer immunogenicity. Indeed, the β-gal-somatostatin conjugate produced here by way of example is of immunogenic size and may be expected to raise antibodies which bind the somatostatin hapten. Proteins comprising in excess of 100 amino acids, most commonly in excess of 200 such, exhibit immunogenic character.

Conjugates prepared in the foregoing fashion may be employed to raise antibodies useful in radioimmune or other assays for the hapten, and alternatively in the production of vaccines. We next describe an example of the latter application. Cyanogen bromide—or other cleavage products of viral coat protein will yield oligopeptides which bind to antibody raised to the protein itself. Given the amino acid sequence of such an oligopeptide hapten, heterologous DNA therefore may be expressed as a conjugate with additional protein which confers immunogenicity. Use of such conjugates as vaccines could be expected to diminish side reactions which accompany use of coat protein itself to confer immunity.

4. The control elements

Figure 1 depicts a process wherein a transformant organism expresses polypeptide product from heterologous DNA brought under the control of a regulon, “homologous” to the organism in its untransformed state. Thus, lactose-dependent E. Coli. chromosomal DNA comprises a lactose or “lac” operon which mediates lactose digestion by, inter alia, elaborating the enzyme β-galactosidase. In the particular instance illustrated, the lac control elements are obtained from a bacteriophage, λ phage, λ and in turn, was derived by transduction from the same bacterial species, hence the “homology”. Homologous regulons suitable for use in the disclosed process may alternatively derive from plasmidic DNA native to the organism.

The simplicity and efficiency of the lac promoter-operator system commend its use in the systems we describe, as does its ability to be induced by IPTG (isopropylthio-β-D galactoside). of course, other operons or portions thereof could be employed as well, e.g., lambda promoter-operator, arabinoose operon (phi 80 dera), or the colicine E1, galactose, alkaline phosphatase or tryptophan operons. Promoter-operators derived from the latter (i.e., “tryp operon”) would be expected to confer 100% repression pending induction (with indoleacrylic acid) and harvest.

5. Plasmid construction generally

The details of the process schematically illustrated in Figure 4 appear from the Experimental section, infra. At this point, however, it is useful to briefly discuss various of the techniques employed in constructing the recombinant plasmid of the preferred embodiment.
The cloning and expression of the synthetic somatostatin gene employed two plasmids. Each plasmid has an EcoRI substrate site at a different region of the β-galactosidase structural gene (see Figures 4 and 5). The insertion of the synthetic somatostatin DNA fragment into the EcoRI sites of these plasmids brings the expression of the genetic information in that fragment under control of the lac operon controlling elements. Following the insertion of the somatostatin fragment into these plasmids, translation should result in a somatostatin polypeptide preceded either by 10 amino acid (pSOM1) or by virtually the whole β-galactosidase subunit structure (pSOM1-3).

The plasmid construction scheme initiates with plasmid pBR322, a well-characterized cloning vehicle. Introduction of the lac elements to this plasmid was accomplished by insertion of a HaelII restriction endonuclease fragment (203 nucleotides) carrying the lac promoter, CAP binding site, operator, ribosome binding site, and the first 7 amino acid codons of the β-galactosidase structural gene. The HaelII fragment was derived from λ plac5 DNA. The EcoRI-cleaved pBR322 plasmid, which had its termini repaired with T4 DNA polymerase and deoxyribonucleotide triphosphates, was blunt-end ligated to the HaelII fragment to create EcoRI termini at the insertion points. Joining of these HaelII and repaired EcoRI termini generate the EcoRI restriction site (see Fig. 4 and 5) at each terminus. Transformants of E. coli RR1 with this DNA were selected for resistance to tetracycline (Tc) and ampicillin (Ap) on 5-bromo-4-chloro-indolylgalactoside (X-gal) medium. On this indicator medium, colonies constitutive for the synthesis of β-galactosidase, by virtue of the increased number of lac operators titrating repressor, are identified by their blue color. Two orientations of the HaelII fragment are possible but these were distinguished by the asymmetric location of an HhaI restriction site in the fragment. Plasmid pH10 was further modified to eliminate the EcoRI endonuclease site distal to the lac operator (pBH20).

The eight chemically synthesized oligodeoxyribonucleotides (Fig. 2) were labeled at the 5' termini with [32P]-ATP by polynucleotide kinase and joined with T4 DNA ligase. Through hydrogen bonding between the overlapping fragments, the somatostatin gene self-assembles and eventually polymerizes into larger molecules because of the cohesive restriction site termini. The ligated products were treated with EcoRI and BamHI restriction endonucleases to generate the somatostatin gene as depicted in Figure 2.

The synthetic somatostatin gene fragment with EcoRI and BamHI termini was ligated to the pH20 plasmid, previously treated with the EcoRI and BamHI restriction endonucleases and alkaline phosphatase. The treatment with alkaline phosphatase provides a molecular selection for plasmids carrying the inserted fragment. Ampicillin-resistant transformants obtained with this ligated DNA were screened for tetracycline sensitivity and several were examined for the insertion of an EcoRI-BamHI fragment of the appropriate size.

Both strands of the EcoRI-BamHI fragments of plasmids from two clones were analyzed by nucleotide sequence analysis starting from the BamHI and EcoRI sites. The sequence analysis was extended into the lac controlling elements; the lac fragment sequence was intact, and in one case, pSOM1, the nucleotide sequence of both strands was independently determined each giving the sequence depicted in Figure 5A.

The EcoRI-Pst fragment of the pSOM1 plasmid, with the lac-controlling element, was removed and replaced with the EcoRI-Pst fragment of pBR322 to produce the plasmid pSOM11. The EcoRI fragment of λ plac 5, carrying the lac operon control region and most of the β-galactosidase structural gene, was inserted into the EcoRI site of pSOM1. Two orientations of the EcoRI lac fragment of λ plac 5 were expected. One of these orientations would maintain the proper reading frame into the somatostatin gene, the other would not. Analysis of independently isolated clones for somatostatin activity then identified clones containing the properly oriented gene, of which the clone designated pSOM1-3 was one.

6. The microorganism

Various unicellular microorganisms have been proposed as candidates for transformation, such as bacteria, fungi and algae. That is, those unicellular organisms which are capable of being grown in cultures or fermentation. Bacteria are for the most part the most convenient organisms to work with. Bacteria which are susceptible to transformation include members of the Enterobacteriaceae, such as strains of Escherichia coli and Salmonella; Bacillaceae, such as Bacillus subtilis; Pneumococcus; Streptococcus, and Haemophilus influenzae.

The particular organism chosen for the somatostatin work next discussed was E. coli. strain RR1, genotype: Pro+Leu+ Thr+ Met+ Leu+ Serr+ Lys+ Lac+. E. coli. RR1 is derived from E. coli. HB101 (H. W. Boyer, et al., J. Mol. Biol. (1969) 47, 459-472) by mating with E. coli K12 strain KL16 as the Hfr donor. See J. H. Miller, Experiments in Molecular Genetics (Cold Spring Harbor, New York, 1972). Cultures of both E. coli RR1 and E. coli. RR1 (pBR322) have been deposited with the American Type Culture Collection without restriction as to access, respectively ATCC Nos. 31343 and 31344 both deposited 8th November 1977. The somatostatin-producing organism has likewise been deposited [ATCC No. 31447] deposited 28th October 1978.

In the case of human insulin, A and B chain genes were cloned in E coli K-12 strain 294 (end A, thi-, har-, hsm,+) ATCC No. 31446 deposited 28th October 1978, and that organism employed in expression of the A chain. (E. Coli K-12 strain 294 [pLA1], ATCC No. 31448) deposited 28th October 1978. The B chain of human insulin was first expressed in a derivative of HB101, i.e., E. coli K-12 strain D1210 a lac- (Hollb), and that B gene-containing organism has likewise been deposited [ATCC No. 31449] deposited 28th October 1978. Alternatively, the B gene may be inserted in and expressed from the organism first mentioned, i.e., strain 294.
Experimental

I Somatostatin

1. Construction of somatostatin gene fragments

Eight oligodeoxyribonucleotides respectively labeled A through H in Figure 2 were first constructed, principally by the modified triester method of K. Ikura et al., *J. Am. Chem. Soc.* 97, 7327 (1975). However, in the case of fragments C, E and H resort was had to an improved technique in which fully protected trimers are first prepared as basic units for building longer oligodeoxyribonucleotides. The improved technique is schematically depicted in Figure 3, wherein B is thymine, N-benzyolated adenine, N-benzyolated cytosine or N-isobutyrylated guanine. In brief, and with reference to Figure 3, with an excess of I (2 mmole), the coupling reaction with II (1 mmole) went almost to completion in 60 min with the aid of a powerful coupling reagent, 2,4,6-trisopropylbenzenesulfonyl tetrazolide (TPSTE, 4 mmole; 2). After removal of the 5’-protecting group with 2% benzene sulfonic acid solution, the 5’-hydroxyl dimer V could be separated from an excess of 3’-phosphodieste monomer IV by simple solvent extraction with aqueous NaHCO₃ solution in CHCl₃. The fully protected trimer block was prepared successively from the 5’-hydroxyl dimer V, I (2 mmole), and TPSTE (4 mmole) and isolated by chromatography on silica gel, as in B. T. Hunt et al., *Chem. and Ind.* 1967, 1968 (1967). The yields of trimers made according to the improved technique appear from Table II.

The eight oligodeoxyribonucleotides, after removal of all protecting groups, were purified by high-pressure liquid chromatography on Permaphase AAR (R. A. Henry et al. *J. Chrom. Sci. II*, 358 (1973)). The purity of each oligomer was checked by homochromatography on thin-layer DEAE-cellulose and also by gel electrophoresis in 20% acrylamide slab after labeling of the oligomers with [γ-³²P]-ATP in the presence of polynucleotide kinase. One major labeled product was obtained from each DNA fragment.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Yield</th>
<th>Sequence</th>
<th>Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>81%</td>
<td>ATG</td>
<td>69%</td>
</tr>
<tr>
<td>TTT</td>
<td>75%</td>
<td>GCC</td>
<td>61%</td>
</tr>
<tr>
<td>GGA</td>
<td>41%</td>
<td>CCA</td>
<td>72%</td>
</tr>
<tr>
<td>AGA</td>
<td>49%</td>
<td>CAA</td>
<td>72%</td>
</tr>
<tr>
<td>ATC</td>
<td>71%</td>
<td>TTA</td>
<td>71%</td>
</tr>
<tr>
<td>CCT</td>
<td>61%</td>
<td>CAT</td>
<td>52%</td>
</tr>
<tr>
<td>ACA</td>
<td>63%</td>
<td>CCC</td>
<td>73%</td>
</tr>
<tr>
<td>ACC</td>
<td>65%</td>
<td>AAC</td>
<td>59%</td>
</tr>
<tr>
<td>CGT</td>
<td>51%</td>
<td>GAT</td>
<td>60%</td>
</tr>
</tbody>
</table>

2. Ligation and acrylamide gel analysis of somatostatin DNA

The 5’ OH termini of the chemically synthesized fragments A through H were separately phosphorylated with T4 polynucleotide kinase. [³²P]-γ-ATP was used in phosphorylation so that reaction products could be monitored autoradiographically, although it will be appreciated that unlabelled ATP would serve as well were autoradiography dispensed with. Just prior to the kinase reaction, 25 μCi of [γ-³²P]-ATP (approx. 1500 cpm/Mol) (Maxam and Gilbert, *Proc. Nat. Acad. Sci. U.S.A.* 74, 1507 (1977) was evaporated to dryness in 0.5 ml Eppendorf tubes. Five micrograms of fragment were incubated with 2 units of T4 DNA kinase (hydroxylapatite fraction, 2500 units/ml); in 70 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 5 mM dithiothreitol in a total volume of 150 μl for 20 min at 37°C. To insure maximum phosphorylation of the fragments for ligation purposes, 10 μl of a mixture consisting of 70 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 5 mM dithiothreitol, 0.5 mM ATP and two units of DNA kinase were added and incubation continued for an additional 20 min at 7°C. The fragments (250 ng/μl) were stored at −20°C without further treatment. Kinased fragments A, B, E, and F (1.25 μg each) were ligated in a total volume of 50 μl in 20 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 10 mM dithiothreitol, 0.5 mM ATP and 2 units of T4 DNA ligase (hydroxylapatite fraction, 400 units/ml; 27), for 16 hr at 4°C. Fragments C, D, G and H were ligated under similar conditions. Samples of 2 μl were removed for analysis by electrophoresis on a 10% polyacrylamide gel followed by autoradiography (H. L. Heyneker et al., *Nature* 263, 748 (1976)) in which unreacted DNA fragments are represented by fast migrating material and wherein the monomeric form of the ligated fragments migrate with bromophenol blue dye (BPB). Some dimerization also occurs by reason of the cohesive ends of the ligated fragments A,
B, E and F, and of the ligated fragments C, D, G and H. These dimers represent the slowest migrating material, and may be cleaved by restriction endonuclease EcoRI and BamHI, respectively.

The two half molecules (ligated A+B+E+F and ligated C+D+G+H) were joined by an additional ligation step carried out in a final volume of 150 μl at 4°C for 16 hr. One microliter was removed for analysis. The reaction mixture was heated for 15 min at 65°C to inactivate the T4 DNA ligase. The heat treatment does not affect the migration pattern of the DNA mixture. Enough restriction endonuclease BamHI was added to the reaction mixture to cleave the multimeric forms of the somatostatin DNA in 30 min at 37°C. After the addition of NaCl to 100 mM, the DNA was digested with EcoRI endonuclease. The restriction endonuclease digestions were terminated by phenol-chloroform extraction of the DNA. The somatostatin DNA fragment was purified from unreacted and partially ligated DNA fragments by preparative electrophoresis on a 10% polyacrylamide gel. The band containing the somatostatin DNA fragment was excised from the gel and the DNA was eluted by slicing the gel into small pieces and extracting the DNA with elution buffer (0.5 M ammonium acetate, 10 mM MgCl₂, 0.1 mM EDTA, 0.1% SDS) overnight at 65°C. The DNA was precipitated with 2 volumes of ethanol, centrifuged, redissolved in 200 μl 10 mM Tris-HCl pH 7.6 and dialyzed against the same buffer resulting in a somatostatin DNA concentration of 4 μg/ml.

3. Construction of recombinant plasmids

Figure 4 schematically depicts the manner in which recombinant plasmids comprising the somatostatin gene were constructed, and may be referred to in connection with the following more particularized discussion.

A. The parental plasmid pBR322

The plasmid chosen for experimental somatostatin cloning was pBR322, a small (molecular wt. approx. 2.6 megadaltons) plasmid carrying resistance genes to the antibiotics ampicillin (Ap) and tetracycline (Tc). As indicated in Figure 4, the ampicillin resistance gene includes a cleavage site for the restriction endonuclease Pst I, the tetracycline resistance gene includes a similar site for restriction endonuclease BamHI, and an EcoRI site is situated between the Ap' and Tc' genes. The plasmid pBR322 is derived from pBR312, a 5.8 megadalton Ap'Te'Cot'mm plasmid (R. L. Rodríguez et al, ICN-UCLA Symposia on Molecular and Cellular Biology 6, 471—77 (1976), R. L. Rodríguez et al., Construction and Characterization of Cloning Vectors, in Molecular Mechanisms in the Control of Gene Expression, pp. 471—77, Academic Press, Inc. (1976). Plasmid pBR322 is characterized and the manner of its derivation fully described in F. Bolivar et al., "Construction and Characterization of New Cloning Vectors II. A Multipurpose Cloning System", Gene (November 1977).

B. Construction of plasmid pH10

Five micrograms of plasmid pBR322 DNA was digested with 10 units of the restriction endonuclease EcoRI in 100 mM Tris-HCl pH 7.6, 100 mM NaCl, 6 mM MgCl₂ at 37°C for 30 min. The reaction was terminated by phenol-chloroform extraction; the DNA was then precipitated with two and a half volumes of ethanol and resuspended in 50 μl of T4 DNA polymerase buffer (67 mM Tris-HCl pH 8.8, 6.7 mM MgCl₂, 16.6 mM (NH₄)₂SO₄, 167 μg/ml bovine serum albumin, 50 μM of each of the dNTP's; A. Panet et al., Biochem. 12, 5045 (1973). The reaction was started by the addition of 2 units of T4 DNA polymerase. After incubation for 30 min at 37°C the reaction was terminated by a phenol-chloroform extraction of the DNA following by precipitation with ethanol. Three micrograms of plasmid DNA (Shapiro et al Nature 224, 788 (1969)) was digested for 1 hr at 37°C with the restriction enzyme HaellI (3 units) in 6 mM Tris-HCl pH 7.6, 6 mM MgCl₂, 6 mM β-mercaptoethanol in a final volume of 20 μl. The reaction was stopped by heating for 10 min at 65°C. The pH322 treated DNA was mixed with the HaellI digested Aplac5 DNA and blunt-end ligated in a final volume of 30 μl with 1.2 units of T4 DNA ligase (hydroxylapatite fraction; A. Panet et al, supra) in 20 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 10 mM dithiothreitol, 0.5 mM ATP for 12 hrs at 12°C. The ligated DNA mixture was dialyzed against 10 mM Tris-HCl pH 7.6, and used for transformation of E. coli strain RRI conventionally. Transformants were selected for tetracycline and ampicillin resistance on minimal medium, plates containing 40 μg/ml of 5-bromo-4-chloro-3-indolylβ-D-galactoside (X-gal) medium (J. H. Miller, Experiments in Molecular Genetics (Cold Spring Harbor, New York, 1972)). Colonies constitutive for the synthesis of β-galactosidase were identified by their blue color. After screening 48 independently isolated blue colonies, three of them were found to contain plasmid DNA carrying two EcoRI sites separated by approximately 200 base pairs. The position of an asymmetrically located Hpal fragment in the 203 b.p. HaellI lac control fragment (W. Gilbert et al, in Protein-Ligand Interactions, H. Sand and G. Blauer, Eds. (De Gruyter, Berlin, 1975) pp. 193—210) allows for the determination of the orientation of the HaellI fragment, now an EcoRI fragment, in these plasmids. Plasmid pH10 was shown to carry the fragment in the desired orientation, i.e., lac transcription going into the To’ gene of the plasmid.

C. Construction of plasmid pH20

Plasmid pH10 was next modified to eliminate the EcoRI site distal to the lac operator. This was accomplished by preferential EcoRI endonuclease cleavage at the distal site involving partial protection by RNA polymerase of the other EcoRI site localized between the To’ and lac promoters, which are only about 40 base pairs apart. After binding RNA polymerase, the DNA (5 μg) was digested with EcoRI (1 unit) in a
final volume of 10 μl for 10 min at 37°C. The reaction was stopped by heating at 65°C for 10 min. The EcoRI cohesive termini were digested with S1 nuclease in 25 mM Na-acetate pH 4.5, 300 mM NaCl, 1 mM ZnCl₂, at 25°C for 5 min. The reaction mixture was stopped by the addition of EDTA (10 mM final) and Tris-Cl pH 8 (50 mM final). The DNA was phenol-chloroform extracted, ethanol precipitated and resuspended in 100 μl of T4 DNA liglation buffer. T4 DNA ligase (1 μl) was added and the mixture incubated at 12°C for 12 hr. The ligated DNA was transformed in E. coli strain RR1 conventionally and Ap'/Tc' transformants were selected on X-gal-antibiotic medium. Restriction enzyme analysis of DNA screened from 10 isolated blue colonies revealed that these clones carried plasmid DNA with one EcoRI site. Seven of these colonies had retained the EcoRI site located between the lac and Tc' promoters. The nucleotide sequence from the EcoRI site into the lac-control region of one of these plasmids, pBH20, was confirmed. This plasmid was next used to clone the somatostatin gene.

D. Construction of plasmid pSOM 1
Twenty micrograms of the plasmid pBH20 was digested to completion with restriction endonucleases EcoRI and BamHI in a final volume of 50 μl. Bacterial alkaline phosphatase was added (0.1 unit of Worthington BAPF) and incubation was continued for 10 min at 37°C. The reactions were terminated by phenol-chloroform extraction and the DNA was precipitated with 2 volumes of ethanol, centrifuged and dissolved in 50 μl 10 mM Tris-Cl pH 7.6, 1 mM EDTA. The alkaline phosphatase treatment effectively prevents self-ligation of the EcoRI, BamHI treated pH20 DNA, but circular recombinant plasmids containing somatostatin DNA can still be formed upon ligation. Since E. coli RR1 is transformed with very low efficiency by linear plasmid DNA, the majority of the transformants will contain recombinant plasmids. Fifty microliters of somatostatin DNA (4 μg/ml) were ligated with 25 μl of the BamHI, EcoRI, alkaline phosphatase-treated pBH20 DNA in a total volume of 50 μl containing 20 mM Tris-Cl pH 7.6, 10 mM MgCl₂, 10 mM dithiothreitol, 0.5 mM ATP, and 4 units of T4 DNA ligase at 22°C. After 10, 20 and 30 min, additional somatostatin DNA (40 ng) was added to the reaction mixture (the gradual addition of somatostatin DNA may favor ligation to the plasmid over self-ligation). Ligation was continued for 1 hr followed by dialysis of the mixture against 10 mM Tris-Cl pH 7.6. In a control experiment, BamHI, EcoRI, alkaline phosphatase-treated pH20 DNA was ligated in the absence of somatostatin DNA under similar conditions. Both preparations were used without further treatment to transform E. coli RR1. The transformation experiments were carried out in a P3 physical containment facility. (National Institutes of Health, U.S.A., Recombinant DNA Research Guidelines, 1976). Transformants were selected on minimal medium plates containing 20 μg/ml Ap and 40 μg/ml X-gal. Ten transformants, which were all sensitive to Tc, were isolated. For reference these were designated pSOM1, pSOM2, etc... pSOM10. In the control experiment no transformants were obtained. Four out of the ten transformants contained plasmids with both an EcoRI site and a BamHI site. The size of the small EcoRI, BamHI fragment of these recombinant plasmids was in all four instances similar to the size of the in vitro prepared somatostatin DNA. Base sequence analysis according to Maxam and Gilbert Proc. Nat. Acad. Sci. U.S.A. 74, 560 (1977), revealed that the plasmid pSOM1 had the desired somatostatin DNA fragment inserted.

The DNA sequence analysis of the clone carrying plasmid pSOM1 predicts that it should produce a peptide comprising somatostatin. However no somatostatin radioimmune activity has been detected in extracts of cell pellets or culture supernatants, nor is the presence of somatostatin detected when the growing culture is cultured directly to 70% formic acid and cyanogen bromide. E. coli RR1 extracts have been observed to degrade oligonucleotide somatostatin very rapidly. The absence of somatostatin activity in clones carrying plasmid pSOM1 could well result from intracellular degradation by endogenous proteolytic enzymes. Plasmid pSOM1 was accordingly employed to construct a plasmid coding for a precursor protein comprising somatostatin and sufficiently large as to be expected to resist proteolytic degradation.

E. The construction of plasmids pSOM11 and pSOM11-3
A plasmid was constructed in which the somatostatin gene could be located at the C-terminus of the β-galactosidase gene, keeping the translation in phase. The presence of an EcoRI site near the C-terminus of this gene and the available amino acid sequence of this protein (B. Polisky et al., Proc. Nat. Acad. Sci. U.S.A. 72, 3900 (1975); A. V. Fowler et al., ibid. at 74, 1507 (1976), A. I. Bukhari et al., Nature New Biology 243, 239 (1979) and K. E. Langley, J. Biol. Chem. 250, 2587 (1975)) permitted insertion of the EcoRI BamHI somatostatin gene into the EcoRI site while maintaining the proper reading frame. For the construction of such a plasmid, pSOM1 DNA (50 μg) was digested with the restriction enzymes EcoRI and PstI in a final volume of 100 μl. A preparative 5% polyacrylamide gel was used to separate the large PstI-EcoRI fragment that carries the somatostatin gene from the small fragment carrying the lac control elements. The large band was excised from the gel and the DNA eluted by slicing the gel into small pieces and extracting the DNA at 65°C overnight. In a similar way plasmid pBH22 DNA (50 μg) was digested with PstI and EcoRI restriction endonucleases and the two resulting DNA fragments purified by preparative electrophoresis on a 5% polyacrylamide gel. The small PstI-EcoRI fragment from pBH22 (1 μg) was ligated with the large PstI-EcoRI DNA fragment (5 μg) from pSOM1 in a final volume of 50 μl with 1 unit of T4 DNA ligase at 12°C for 12 hrs. The ligated mixture was used to transform E. coli RR1, and transformants were selected for ampicillin resistance on X-gal medium. As expected, almost all the Ap' transformants (95%) gave white colonies (no lac operator) on X-gal indicator plates. The resulting plasmid, pSOM11, was used in the
construction of plasmid pSOM11-3. A mixture of 5 µg of pSOM11 DNA and 5 µg of λ phage DNA was digested with EcoRI (10 units for 30 min at 37°C). The restriction endonuclease digestion was terminated by phenol-chloroform extraction. The DNA was then ethanol-precipitated and resuspended in T4 DNA ligase buffer (50 µl). T4 DNA ligase (1 unit) was added to the mixture and incubated at 12°C for 12 hrs. The ligated mixture was dialyzed against 10 mM Tris-HCl pH 7.6 and used to transform E. Coli strain RR1. Transformants were selected for Ap on X-gal plates containing ampicillin and screened for constitutive β-galactosidase production. Approximately 2% of the colonies were blue (pSOM11-1, 11-2 etc.).

Restriction enzyme analysis of plasmid DNA obtained from these clones revealed that all the plasmids carried a new EcoRI fragment of approximately 4.4 megadaltons, which carries the lac operon control sites and most of the β-galactosidase gene. Because two orientations of the EcoRI fragment are possible, the asymmetric location of a HindIII restriction site was used to determine which of these colonies were carrying this EcoRI fragment with lac transcription proceeding into the somatostatin gene. HindIII-BamHI double digests indicated that only the clones carrying plasmids pSOM11-3, pSOM11-5, pSOM11-6 and pSOM11-7 contained the EcoRI fragment in this orientation.

4. Radioimmuno assay for somatostatin activity

The standard radioimmune assays (RIA) for somatostatin (A. Arimura et al., Proc. Soc. Exp. Biol. Med. 148, 784 (1975)) were modified by decreasing the assay volume and using phosphate buffer. Tyr11 somatostatin was iodinated using a chloramine T procedure. (Id.) To assay for somatostatin, the sample, usually in 70% formic acid containing 5 mg/ml of cyanogen bromide was dried in a conical polypropylene tube (0.7 ml, Sarstedt) over moist KOH under vacuum. Twenty microliters of PBSA buffer (75 mM NaCl; 75 mM sodium phosphate, pH 7.2; 1 mg/ml bovine serum albumin; and 0.2 mg/ml sodium azide) was added, followed by 40 µl of a [125I] somatostatin “cocktail” and 20 µl of a 1,000-fold dilution in PBSA of rabbit antisomatostatin immune serum S39 (Vale et al., Metabolism 25, 1491 (1976). The [125I] somatostatin cocktail contained per ml of PBSA buffer: 250 µg normal rabbit gamma globulin (Antibodies, Inc.), 1,500 units protease inhibitor (“Trasylo”, Calbiochem) and about 100,000 counts of [125I] Tyr11-somatostatin. After at least 16 hour at room temperature, 0.33 ml of goat anti-rabbit gamma globulin (Antibodies, Inc., P=0.03) in PBSA buffer was added to the sample tubes. The mixture was incubated 2 hr at 37°C, cooled to 5°C, then centrifuged at 10,000 xg for 5 min. The supernatant was removed at the pellet counted in a gamma counter. With the amount of antiserum used, 20% of the counts was precipitated with no unlabeled competing somatostatin. The background with infinite somatostatin (200 ng) was usually 3%. One-half maximum competition was obtained with 10 pg of somatostatin. Initial experiments with extracts of E. Coli strain RR1 (the recipient strain) indicated that less than 10 pg of somatostatin could easily be detected in the presence of 16 µg or more of cyanogen bromide-treated bacterial protein. More than 2 µg of protein from formic acid-treated bacterial extracts interfered somewhat by increasing the background, but cyanogen bromide cleavage greatly reduced this interference. Reconstruction experiments showed that somatostatin is stable in cyanogen bromide-treated extracts.

A. Competition by bacterial extracts

Strains E. Coli RR1 (pSOM11-5) and E. Coli RR1 (pSOM11-4) were grown at 37°C to 5 x 10⁶ cells/ml in Luria broth. Then IPTG (isopropyl β-D-galactosidase, an inducer of the lac operon), was added to 1 mM and growth continued for 2 hr. One-milliliter aliquots were centrifuged for a few seconds in an Eppendorf centrifuge and the pellets were suspended in 500 µl of 70% formic acid containing 5 mg/ml cyanogen bromide. After approximately 24 hr at room temperature, aliquots were diluted tenfold in water and the volumes indicated in Figure 6 were assayed in triplicate for somatostatin. In Figure 6 “B/B,” is the ratio of [125I] somatostatin bound in the presence of sample to that bound in the absence of competing somatostatin. Each point is the average of triplicate tubes. The protein content of the undiluted samples was determined to be 2.2 mg/ml for E. Coli RR1 (pSOM11-5) and 1.5 mg/ml for E. Coli RR1 (pSOM11-4)

B. The initial screening of pSOM11 clones for somatostatin

Cyanogen bromide-treated extracts of 11 clones (pSOM11-2, pSOM11-3, etc.) were made as described above for the case of Figure 6. Thirty microliters of each extract was taken in triplicate for radioimmune assay, whose results appear from Figure 7. The range of assay points is indicated. The values for picograms somatostatin were read from a standard curve obtained as part of the same experiment.

The radioimmune assay results described thus far may be summarized as follows. In contrast to the results of experiments with pSOM1, four clones (pSOM11-3 11-5, 11-6, and 11-7) were found to have easily detectable somatostatin radioimmune activity as appears from Figures 6 and 7. Restriction fragment analysis revealed that pSOM11-3, pSOM11-5, pSOM11-6 and pSOM11-7 had the desired orientation of the lac operon, whereas pSOM11-2 and 11-4 had the opposite orientation. Thus there is a perfect correlation between the correct orientation of the lac operon and the production of somatostatin radioimmune activity.

C. Effects of IPTG induction and CNBr cleavage on positive and negative clones

The design of the somatostatin plasmid predicts that the synthesis of somatostatin would be under the control of the lac operon. The lac repressor gene is not included in the plasmid and the recipient strain (E. Coli RR1) contains the wild type chromosomal lac repressor gene which produces only 10 to 20 repressor
molecules per cell. The plasmid copy number (and therefore the number of lac operators) is approximately 20–30 per cell, so complete repression is impossible. As shown in Table III, *infra* the specific activity of somatostatin in *E. coli* RR1 (pSOM11-3) was increased by IPTG, an inducer of the lac operon. As expected, the level of induction was low, varying from 2.4 to 7 fold. In experiment 7 (Table III) α activity, a measure of the first 92 amino acids of β-galactosidase, also was induced by a factor of two. In several experiments no detectable somatostatin radioimmune activity can be detected prior to cyanogen bromide cleavage of the total cellular protein. Since the antiserum used in the radioimmune assay, S 39, requires a free N-terminal alanine, no activity was expected prior to cyanogen bromide cleavage.

**TABLE III**

<table>
<thead>
<tr>
<th>Experiment number</th>
<th>Strain</th>
<th>Medium</th>
<th>IPTG 1 mM</th>
<th>CNBr 5 mg/ml</th>
<th>pg SS/μg protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 15</td>
<td>11-2</td>
<td>LB</td>
<td>+</td>
<td>+</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>20 15</td>
<td>11-3</td>
<td>LB</td>
<td>+</td>
<td>+</td>
<td>12</td>
</tr>
<tr>
<td>20 15</td>
<td>11-4</td>
<td>LB</td>
<td>+</td>
<td>+</td>
<td>&lt;0.4</td>
</tr>
<tr>
<td>20 15</td>
<td>11-5</td>
<td>LB</td>
<td>+</td>
<td>+</td>
<td>15</td>
</tr>
<tr>
<td>20 15</td>
<td>11-3</td>
<td>LB</td>
<td>+</td>
<td>+</td>
<td>12</td>
</tr>
<tr>
<td>20 15</td>
<td>11-3</td>
<td>LB</td>
<td>+</td>
<td>−</td>
<td>61</td>
</tr>
<tr>
<td>20 15</td>
<td>11-3</td>
<td>LB</td>
<td>−</td>
<td>+</td>
<td>8</td>
</tr>
<tr>
<td>20 15</td>
<td>11-3</td>
<td>LB</td>
<td>+</td>
<td>−</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>20 15</td>
<td>11-3</td>
<td>VB+glycerol*</td>
<td>+</td>
<td>+</td>
<td>62</td>
</tr>
<tr>
<td>20 15</td>
<td>11-3</td>
<td>LB+glycerol</td>
<td>+</td>
<td>+</td>
<td>250</td>
</tr>
<tr>
<td>20 15</td>
<td>11-3</td>
<td>LB</td>
<td>+</td>
<td>+</td>
<td>320</td>
</tr>
<tr>
<td>20 15</td>
<td>11-2</td>
<td>LB</td>
<td>+</td>
<td>+</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>20 15</td>
<td>11-3</td>
<td>LB</td>
<td>+</td>
<td>+</td>
<td>24</td>
</tr>
<tr>
<td>20 15</td>
<td>11-3</td>
<td>LB</td>
<td>−</td>
<td>+</td>
<td>10</td>
</tr>
</tbody>
</table>

* Vogel-Bonner minimal medium plus glycerol.

D. Gel filtration of cyanogen bromide-treated extracts

Formic acid and cyanogen-treated extracts of the positive clones (pSOM 11-3, 11-5, 11-6, and 11-7) were pooled (Total volume 250 ml), dried, and resuspended in 0.1 ml of 50% acetic acid. [3H] leucine was added and the sample was applied to an 0.7×47 cm column of Sephadex G-50 in 50% acetic acid. Fifty-microliter aliquots of the column fractions were assayed for somatostatin. Pooled negative clone extracts (11-2, 11-4, and 11-11) were treated identically. The results appear from Figure 8. On the same column known somatostatin (Beckman Corp.) elutes as indicated (SS). In this system, somatostatin is well-separated from excluded large peptides and fully included small molecules. Only extracts of clones positive for somatostatin exhibited radioimmune activity in the column fractions and this activity elutes in the same position as chemically synthesized somatostatin.
Summary of activity information
The data establishing the synthesis of a polypeptide containing the somatostatin amino acid sequence are summarized as follows: (1) Somatostatin radioimmune activity is present in E. coli cells having the plasmid pSOM11-3, which contains a somatostatin gene of proven correct sequence and has the correct orientation of the lac EcoRI DNA fragment. Cells with the related plasmid pSOM11-2, which has the same somatostatin gene but an opposite orientation of the lac EcoRI fragment, produce no detectable somatostatin activity; (2) As predicted by the design scheme, no detectable somatostatin radioimmune activity is observed until after cyanogen bromide treatment of the cell extract; (3) The somatostatin activity is under control of the lac operon as evidenced by induction by IPTG, an inducer of the lac operon; (4) The somatostatin activity co-chromatographs with known somatostatin on Sephadex G-50; (5) The DNA sequence of the cloned somatostatin gene is correct. If translation is out of phase, a peptide will be made which is different from somatostatin at every position. Radioimmune activity is detected indicating that a peptide closely related to somatostatin is made, and translation must be in phase. Since translation occurs in phase, the genetic code dictates that a peptide with the exact sequence of somatostatin is made; (6) Finally, the above samples of E. coli RR1 (pSOM11-3) extract inhibit the release of growth hormone from rat pituitary cells, whereas samples of E. coli RR1 (pSOM11-2) prepared in parallel and with identical protein concentration have no effect on growth hormone release.

Stability, yield, and purification of somatostatin
The strains carrying the lac operon fragment (pSOM11-2, pSOM11-3, etc.) segregate with respect to the plasmid phenotype. For example, after about 15 generations, about one-half of the E. coli RR1 (pSOM11-3) culture was constitutive for β-galactosidase, i.e., carried the lac operator, and of these about half were ampicillin resistant. Strains positive (pSOM11-3) and negative (pSOM11-2) for somatostatin are unstable, and therefore, the growth disadvantage presumably comes from the overproduction of the large but incomplete and inactive galactosidase. The yield of somatostatin has varied from 0.001 to 0.03% of the total cellular protein (Table 1) probably as the result of the selection for cells in culture having plasmids with a deleted lac region. The highest yields of somatostatin have been from preparations where growth was started from a single ampicillin resistant, constitutive colony. Even in these cases, 30% of the cells at harvest had deletions of the lac region. Storage in the frozen state (lyophilization) and growth to harvest from a single such colony is accordingly indicated for the system described. Yields may be increased by, e.g., resort to bacterial strains which overproduce lac repressor such that expression of precursor protein is essentially totally repressed prior to induction and harvest. Alternatively, as previously discussed a typtophan or other operator-promoter system which ordinarily is totally repressed may be employed.

In the crude extract resulting from cell disruption in, e.g., an Eaton Press, the β-galactosidase- somatostatin precursor protein is insoluble and is found in the first low speed centrifugation pellet. The activity can be solubilized in 70% formic acid, 6 M guanidinium hydrochloride, or 2% sodium dodecyl sulfate. Most preferably, however, the crude extract from the Eaton Press is extracted with 8 M urea and the residue cleaved with cyanogen bromide. In initial experiments somatostatin activity derived from E. coli strain RR1 (pSOM11-3) has been enriched approximately 100-fold by alcohol extraction of the cleavage product and chromatography on Sephadex G-50 in 50% acetic acid. When the product is again chromatographed on Sephadex G-50 and then subjected to high pressure liquid chromatography, substantially pure somatostatin may be obtained.

II. Human insulin
The techniques previously described were next applied to the production of human insulin. Thus, the genes for insulin B chain (104 base pairs) and for insulin A chain (77 base pairs) were designed from the amino acid sequence of the human polypeptides, each with single-stranded cohesive termini for the EcoRI and BamHI restriction endonucleases and each designed for insertion separately into pBR322 plasmids. The synthetic fragments, deca- to pentadeca-nucleotides, were synthesized by the block phosphotriester method using trinucleotides as building blocks and ultimately purified with high performance liquid chromatography (HPLC). The human insulin A and B chain synthetic genes were then cloned separately in plasmid pBR322. The cloned synthetic genes were fused to an E. Coli β-galactosidase gene as before to provide efficient transcription, translation, and a stable precursor protein. Insulin peptides were cleaved from β-galactosidase precursor, detected by radioimmunoassay, and purified. Insulin radioimmunoassay activity was then generated by mixing the E. Coli products.

1. Design and synthesis of human insulin genes
The genes constructed for human insulin are depicted in Figure 9. The genes for human insulin, B chain and A chain, were designed from the amino acid sequences of the human polypeptides. The 5' ends of each gene have single stranded cohesive termini for the EcoRI and BamHI restriction endonucleases, for the correct insertion of each gene into plasmid pBR322. A HindIII endonuclease recognition site was incorporated into the middle of the B chain gene for the amino acid sequence Glu-Ala to allow amplification and verification of each half of the gene separately before the construction of the whole B chain gene. The B chain and the A chain genes were designed to be built from 29 different oligodeoxynucleotides, varying
from decamer to pentadecamers. Each arrow indicates the fragment synthesized by the improved phosphotriester method, H1 to H8 and B1 to B12 for the B chain gene and A1 to A11 for the A chain gene.

2. Chemical synthesis of oligodeoxyribonucleotides

Materials and methods for synthesis of oligodeoxyribonucleotides were essentially those described in Itakura, K. et al (1975) J. Biol. Chem. 250, 4592 and Itakura, K. et al (1975) J. Amer. Chem. Soc. 97, 7327 except for these modifications:

a) The fully protected mononucleotides, 5’-O-dimethoxytrityl-3’-p-chlorophenyl-β-cyanoethyl phosphates, were synthesized from the nucleoside derivatives using the monofunctional phosphorylating agent p-chlorophenyl-β-cyanoethyl phosphorochloridate (1.5 molar equivalent) in acetonitrile in the presence of 1-methyl imidazole Van Boom, J. H. et al (1975) Tetrahedron 31, 2953. The products were isolated in large scale (100 to 300 g) by preparative liquid chromatography (Prep 500 LC, Waters Associates).

b) By using the solvent extraction method [Hirose, T. et al (1978) Tetrahedron Letters, 2449] 32 bifunctional trimers were synthesized (see Table IV) in 5 to 10 mmole scale, and 13 trimers, 3 tetramers, and 4 dimers as the 3’ terminus blocks, in 1 mmole scale. The homogeneity of the fully protected trimers was checked by thin layer chromatography on silica gel in two methanol/chloroform solvent systems: solvent a, 5% v/v and solvent b, 10% v/v (See Table IV). Starting from this library of compounds, 39 oligodeoxyribonucleotides of defined sequence were synthesized, 18 for the B chain and 11 for the A chain gene.

The basic units used to construct polynucleotides were two types of trimer block, i.e. the bifunctional trimer blocks of Table IV and corresponding 3’-terminus trimers protected by an anisoyl group at 3’-hydroxy. The bifunctional trimer was hydrolyzed to the corresponding 3’-phosphodiester component with a mixture of pyridine-ethylamine-water (3:1:1 v/v) and also to the corresponding 5’-hydroxyl component with 2% benzenesulfonic acid. The 3’-terminus block previously referred to was treated with 2% benzenesulfonic acid to give the corresponding 5’-hydroxyl. The coupling reaction of an excess of the 3’-phosphodiester trimer (1.5 molar equivalent) with the 5’-hydroxyl component, however obtained, (1 molar equivalent) in the presence of 2,4,6-trisopropylbenzenesulfonyl tetrazolide (TPSTe, 3 to 4 equivalents) went almost to completion. To remove the excess of the 3’-phosphodiester block reactant the reaction mixture was passed through a short silica gel column set up on a sintered glass filter. The column was washed, first with CHCl3 to elute some side products and the coupling reagent, and then with CHCl3:MeOH (95:5 v/v) in which almost all of the fully protected oligomer was eluted. Under these conditions, the charged 3’-phosphodiester block reactant remained in the column. Similarly, block couplings were repeated until the desired length was constructed.

<table>
<thead>
<tr>
<th>No.</th>
<th>Compound*</th>
<th>Yield** (%)</th>
<th>Rf a.</th>
<th>Rf b.</th>
<th>Purity*** (%)</th>
<th>In Figure 9, present in:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>AAG</td>
<td>47</td>
<td>0.15</td>
<td>0.40</td>
<td>93</td>
<td>B5, B6</td>
</tr>
<tr>
<td>2.</td>
<td>AAT</td>
<td>49</td>
<td>0.25</td>
<td>0.52</td>
<td>95</td>
<td>H1, A1, A6</td>
</tr>
<tr>
<td>3.</td>
<td>AAC</td>
<td>52</td>
<td>0.28</td>
<td>0.55</td>
<td>93</td>
<td>H5, B6, A2, A8</td>
</tr>
<tr>
<td>4.</td>
<td>ACT</td>
<td>43</td>
<td>0.27</td>
<td>0.53</td>
<td>91</td>
<td>B4, B5, A6</td>
</tr>
<tr>
<td>5.</td>
<td>ACC</td>
<td>56</td>
<td>0.33</td>
<td>0.60</td>
<td>96</td>
<td>B7</td>
</tr>
<tr>
<td>6.</td>
<td>ACG</td>
<td>39</td>
<td>0.18</td>
<td>0.45</td>
<td>90</td>
<td>H5, B7</td>
</tr>
<tr>
<td>7.</td>
<td>AGG</td>
<td>45</td>
<td>0.10</td>
<td>0.26</td>
<td>89</td>
<td>H6, H7, B9</td>
</tr>
<tr>
<td>8.</td>
<td>AGT</td>
<td>33</td>
<td>0.14</td>
<td>0.40</td>
<td>98</td>
<td>B9, A2, A11</td>
</tr>
<tr>
<td>9.</td>
<td>AGC</td>
<td>50</td>
<td>0.19</td>
<td>0.48</td>
<td>92</td>
<td>H8, B1, A5, A10</td>
</tr>
<tr>
<td>10.</td>
<td>AGA</td>
<td>48</td>
<td>0.24</td>
<td>0.50</td>
<td>91</td>
<td>A9,</td>
</tr>
<tr>
<td>11.</td>
<td>TTC</td>
<td>44</td>
<td>0.26</td>
<td>0.52</td>
<td>95</td>
<td>B4, B7, A3</td>
</tr>
<tr>
<td>12.</td>
<td>TTC</td>
<td>49</td>
<td>0.11</td>
<td>0.31</td>
<td>94</td>
<td>H3, H5, A2, A3, A5</td>
</tr>
</tbody>
</table>
### TABLE IV (contd.)

Synthesis of trimer building blocks

<table>
<thead>
<tr>
<th>No.</th>
<th>Compound*</th>
<th>Yield** (%)</th>
<th>a. Rf</th>
<th>b. Rf</th>
<th>Purity*** (%)</th>
<th>In Figure 9, present in:</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.</td>
<td>TCT</td>
<td>58</td>
<td>0.24</td>
<td>0.49</td>
<td>96</td>
<td>A4</td>
</tr>
<tr>
<td>14.</td>
<td>TCA</td>
<td>45</td>
<td>0.28</td>
<td>0.53</td>
<td>92</td>
<td>H1, H2, H4, A1</td>
</tr>
<tr>
<td>15.</td>
<td>TCG</td>
<td>39</td>
<td>0.12</td>
<td>0.34</td>
<td>91</td>
<td>A2</td>
</tr>
<tr>
<td>16.</td>
<td>TGG</td>
<td>32</td>
<td>0.10</td>
<td>0.28</td>
<td>87</td>
<td>H3, A1, A10</td>
</tr>
<tr>
<td>17.</td>
<td>TGC</td>
<td>51</td>
<td>0.18</td>
<td>0.47</td>
<td>93</td>
<td>H6, B2, A4, A7, A8</td>
</tr>
<tr>
<td>18.</td>
<td>TGA</td>
<td>46</td>
<td>0.12</td>
<td>0.37</td>
<td>94</td>
<td>H7</td>
</tr>
<tr>
<td>19.</td>
<td>TAC</td>
<td>61</td>
<td>0.22</td>
<td>0.50</td>
<td>90</td>
<td>B4, A11</td>
</tr>
<tr>
<td>20.</td>
<td>TAA</td>
<td>55</td>
<td>0.17</td>
<td>0.44</td>
<td>95</td>
<td>B5, A10</td>
</tr>
<tr>
<td>21.</td>
<td>CCT</td>
<td>53</td>
<td>0.30</td>
<td>0.55</td>
<td>97</td>
<td>H3, H4, B10</td>
</tr>
<tr>
<td>22.</td>
<td>CAC</td>
<td>47</td>
<td>0.25</td>
<td>0.51</td>
<td>92</td>
<td>A3</td>
</tr>
<tr>
<td>23.</td>
<td>CAA</td>
<td>58</td>
<td>0.25</td>
<td>0.51</td>
<td>93</td>
<td>H2, H6, H8, A7</td>
</tr>
<tr>
<td>24.</td>
<td>CTT</td>
<td>41</td>
<td>0.28</td>
<td>0.54</td>
<td>92</td>
<td>B2, B9, A4</td>
</tr>
<tr>
<td>25.</td>
<td>CGA</td>
<td>40</td>
<td>0.27</td>
<td>0.52</td>
<td>93</td>
<td>A7</td>
</tr>
<tr>
<td>26.</td>
<td>CGT</td>
<td>75</td>
<td>0.25</td>
<td>0.50</td>
<td>89</td>
<td>H2, H4, B3, B1</td>
</tr>
<tr>
<td>27.</td>
<td>GGT</td>
<td>35</td>
<td>0.09</td>
<td>0.26</td>
<td>90</td>
<td>B3</td>
</tr>
<tr>
<td>28.</td>
<td>GTT</td>
<td>46</td>
<td>0.18</td>
<td>0.45</td>
<td>93</td>
<td>B2</td>
</tr>
<tr>
<td>29.</td>
<td>GTA</td>
<td>38</td>
<td>0.25</td>
<td>0.50</td>
<td>95</td>
<td>B6, B8, A6</td>
</tr>
<tr>
<td>30.</td>
<td>GAA</td>
<td>39</td>
<td>0.15</td>
<td>0.39</td>
<td>88</td>
<td>H7, B3, B8, A5</td>
</tr>
<tr>
<td>31.</td>
<td>GAT</td>
<td>52</td>
<td>0.22</td>
<td>0.49</td>
<td>89</td>
<td>B10, A9</td>
</tr>
<tr>
<td>32.</td>
<td>GCA</td>
<td>42</td>
<td>0.14</td>
<td>0.39</td>
<td>93</td>
<td>A9</td>
</tr>
</tbody>
</table>

* Fully protected trideoxynucleotides; 5'-Dimethoxytrityl-3'-p-Chlorophenyl-β-cyanoethyl phosphate.

** Yield was the overall yield calculated from the 5'-hydroxymonomers.

*** Based on HPLC analysis.

High performance liquid chromatography (HPLC) was used extensively during oligonucleotide synthesis for a) analysis of each trimer and tetramer block, b) analysis of the intermediate fragments (hexamers, nonamers, and decamers), c) analysis of the last coupling reaction, and d) purification of the final products. The HPLC was performed using a Spectra-Physics 3500B liquid chromatograph. After removal of all protecting groups by conc. NH₄OH at 50°C (6 h) and 80% AcOH at room temperature (15 min), the compounds were analyzed on a Permaphase AAX (DuPont) [Van Boom, J. et al. (1977) J. Chromatography 131, 165] column (1 m x 2 mm), using a linear gradient of solvent B (0.05 M KH₂PO₄, pH 4.0) in solvent A (0.01 M KH₂PO₄, pH 4.5). The gradient was formed starting with buffer A and applying 3% of buffer B per minute. The elution was performed at 60°C, with a flow rate of 2 ml per minute. The purification of the 29 final oligonucleotides also was performed on Permaphase AAX, under the same conditions reported above. The desired peak was pooled, desalted by dialysis, and lyophilized. After labeling the 5' termini with (γ-3P)ATP using T₄ polynucleotide kinase, the homogeneity of each oligonucleotide was checked by electrophoresis on a 20% polyacrylamide gel.
3. Assembly and cloning of B chain gene and the A chain gene

The gene for the B chain of insulin was designed to have an EcoRI restriction site on the left end, a HindIII site in the middle and BamHI site at the right end. This was done so that both halves, the left EcoRI-HindIII half (BH) and the right HindIII-BamHI half (BB), could be separately cloned in the convenient cloning vehicle pBR322 and after their sequences had been verified, joined to give the complete B gene (Figure 10). The BB half was assembled by ligation from 10 oligodeoxyribonucleotides, labeled B1 to B10 in Figure 9, made by phosphotriester chemical synthesis. B1 and B10 were not phosphorylated, thereby eliminating unwanted polymerization of these fragments through their cohesive ends (HindIII and BamHI). After purification by preparative acrylamide gel electrophoresis and elution of the largest DNA band, the BB fragment was inserted into plasmid pBR322 which had been cleaved with HindIII and BamHI. About 50% of the ampicillin resistant colonies derived from the DNA were sensitive to tetracycline, indicating that a nonplasmid HindIII-BamHI fragment had been inserted. The small HindIII-BamHI fragments from four of these colonies (pBB101 to pBB104) were sequenced and found to be correct as designed.

The BH fragment was prepared in a similar manner and inserted into pBR322 which had been cleaved with EcoRI and HindIII restriction endonucleases. Plasmids from three ampicillin resistant, tetracycline sensitive transformants (pBH1 to pBH3) were analyzed. The small EcoRI-HindIII fragments were found to have the expected nucleotide sequence.

The A chain gene was assembled in three parts. The left four middle four, and right four oligonucleotides (see Figure 9) were ligated separately, then mixed and ligated (oligonucleotides A1 and 20 0 A4). The assembled A chain gene was phosphorylated, purified by gel electrophoresis, and cloned in pBR322 at the EcoRI-BamHI sites. The EcoRI-BamHI fragments from two ampicillin resistant, tetracycline sensitive clones (pA10, pA11) contained the desired A gene sequence.

4. Construction of plasmids for expression of A and B insulin genes

Figure 10 illustrates the construction of the lac-insulin B plasmid (pB1). Plasmids pBH1 and pBB101 were digested with EcoRI and HindIII endonucleases. The small BH fragment of pBH1 and the large fragment of pBB101 (containing the BB fragment and most of pBR322) were purified by gel electrophoresis, mixed, and ligated in the presence of EcoRI-cleaved λ phagC5. The megadalton EcoRI fragment of λ phagC5 contains the lac control region and the majority of the β-galactosidase structural gene. The configuration of the restriction sites ensures correct joining of BH to BB. The lac EcoRI fragment can insert in two orientations; thus, only half of the clones obtained after transformation should have the desired orientation. The orientation of ten ampicillin resistant, β-galactosidase constitutive clones were checked by restriction analysis. Five of these colonies contained the entire B gene sequence and the correct reading frame from the β-galactosidase gene into the B chain gene. One, plB1, was chosen for subsequent experiments.

In a similar experiment, the 4.4 megadalton lac fragment from λ phagC5 was introduced into the pA11 plasmid at the EcoRI site to give pA1, pA1A is identical to pB1 except that the A gene fragment is substituted for the B gene fragment. DNA sequence analysis demonstrated that the correct A and B chain gene sequences were retained in pA1A and pB1 respectively.

5. Expression

The strains which contain the insulin genes correctly attached to β-galactosidase both produce large quantities of a protein the size of β-galactosidase. Approximately 20% of the total cellular protein was this β-galactosidase-insulin A or B chain hybrid. The hybrid proteins are insoluble and were found in the first low speed pellet where they constitute about 50% of the protein.

To detect the expression of the insulin A and B chains, we used a radioimmunoassay (RIA) based on the reconstitution of complete insulin from the separate chains. The insulin reconstitution procedure of Katsavannis et al (1967) Biochemistry 6, 2642-2655, adapted to a 27-microtiter assay volume, provides a very suitable assay. Easily detectable insulin activity is obtained after mixing and reconstituting S-Sulfonated derivatives of the insulin chains. The separate S-sulfonated chains of insulin do not react significantly, after reduction and oxidation, with the anti-insulin antibody used.

To use the reconstitution assay we partially purified the β-galactosidase-A or B chain hybrid protein, cleaved with cyanogen bromide, and formed S-sulfonated derivatives.

The evidence that we have obtained correct expression from chemically synthesized genes for human insulin can be summarized as follows: a) Radilimmunne activity has been detected for both chains. b) The DNA sequences obtained after cloning and plasmid construction have been directly verified to be correct as designed. Since radioimmune activity is obtained, translation must be in phase. Therefore, the genetic code dictates that peptides with the sequences of human insulin are being produced. c) The E. coli products, after cyanogen bromide cleavage, behave as insulin chains in three different chromatographic systems which separate on different principles (gel filtration, ion exchange, and reversed phase HPLC). d) The E coli produced A chain has been purified on a small scale by HPLC and has the correct amino acid composition.

Claims

1. A process for producing a desired heterologous polypeptide which comprises growing a bacterium transformed with a recombinant plasmid comprising a homologous regulon, a DNA sequence in proper
reading frame therewith, and one or more termination codons, wherein said DNA sequence is positioned between said regulon and the termination codon(s) and encodes said heterologous polypeptide, additional protein, and a selective cleavage site between the desired heterologous polypeptide and additional protein, wherein the DNA sequence is expressed and the resulting conjugate polypeptide is sufficiently large not to be degraded by endogenous proteolytic enzymes and is specifically cleaved to give said desired heterologous polypeptide.

2. A process according to claim 1, wherein said additional protein comprises one or more heterologous polypeptide sequences and wherein all heterologous polypeptide sequences are separated, each from the other, by selective cleavage sites.

3. A process according to either claim 1 or 2, wherein the cleavage is effected in a system outside the replicative environment of the recombinant plasmid.

4. A process according to claim 3, wherein said desired heterologous polypeptide contains no internal cleavage sites corresponding to that between the desired heterologous polypeptide and the additional protein.

5. A process according to claim 3 or claim 4, wherein the desired heterologous polypeptide lacks arginine and lysine and wherein said selective cleavage site is cleaved by trypsin.

6. A process according to claim 3 or claim 4, wherein the desired heterologous polypeptide lacks arginine, said selective cleavage site is cleaved by trypsin, and any lysine contained in said desired heterologous polypeptide is first protected against cleavage.

7. A process according to any one of claims 3 to 6, wherein said cleavage is effected with cyanogen bromide.

8. A process according to any one of the preceding claims, wherein said recombinant plasmid is a bacterial plasmid.

9. A process according to any one of the preceding claims, wherein the additional protein precedes the desired heterologous polypeptide sequence.

10. A process according to any one of the preceding claims, wherein the desired heterologous polypeptide is a mammalian hormone or an intermediate therefor.

11. A process according to claim 10, wherein the conjugate polypeptide lacks the bioactivity of the hormone.

12. A recombinant plasmid comprising a homologous regulon, a DNA sequence encoding a desired specific heterologous polypeptide, a selective cleavage site and additional protein, and one or more termination codons, wherein the said DNA sequence is interposed in proper reading frame between said regulon and termination codon(s) such that a conjugate polypeptide comprising both the desired heterologous polypeptide and additional protein results from expression in a bacterium transformed with the plasmid, there being a selective cleavage site between the desired heterologous polypeptide and the additional protein.

13. A plasmid according to claim 12, wherein the desired heterologous polypeptide sequence contains no internal cleavage sites corresponding to that between the desired heterologous polypeptide and the additional protein.

14. A plasmid according to any one of claims 12 and 13 which is a bacterial plasmid and wherein the recited elements are sequenced as follows: (regulon)—(codons for additional protein)—(codons for selective cleavage site)—(codons for heterologous polypeptide sequence)—(termination codon(s)).

15. A plasmid according to any one of claims 12 to 14, wherein the cleavage site is at methionine.

16. A plasmid according to any one of claims 12 to 15, wherein the desired heterologous polypeptide is somatostatin.

17. A plasmid according to any one of claims 12 to 15, wherein the desired heterologous polypeptide is the A chain of insulin or the B chain of insulin.

18. A bacterium transformed with a plasmid according to any one of claims 12 to 17.

19. A bacterium transformed with a plasmid selected from pSOM11-3, pIA-1 and pIB-1.

20. A bacterial culture comprising a transformed bacterium according to either claim 18 or 19.

21. A bacterial culture cloned from one or more host bacteria each comprising a plasmid according to any one of claims 12—17, the members of said culture being capable of expressing said conjugate polypeptide.

22. A bacterial culture cloned from one or more host bacteria each comprising a plasmid selected from pSOM11-3, pIA-1 and pIB-1, the members of said culture being capable of expressing said conjugate polypeptide.

23. A process according to claim 10 or 11, wherein the heterologous polypeptide is selected from somatostatin, the A chain of human insulin and the B chain of human insulin.

24. A process according to claim 23, wherein said heterologous polypeptide is somatostatin.

25. A process according claim 23, wherein said heterologous polypeptide is the A chain of human insulin.

26. A process according to claim 23, wherein said heterologous polypeptide is the B chain of human insulin.

27. A process according to any one of claims 24 to 26, wherein said cleavage site is at methionine.

28. A bacterial culture comprising E. Coli RR1 (pSOM11-3), ATCC No. 31447.
29. A bacterial culture comprising E. Coli 294 (pIA-1), ATCC 31448.
32. Plasmid pIA-1, ATCC No. 31448.
33. Plasmid pIB-1, ATCC No. 31449.

Patentansprüche

1. Verfahren zur Herstellung eines gewünschten heterologen Polypeptids, dadurch gekennzeichnet, daß man ein Bakterium züchtet, welches mit einem Rekombinationsplasmid transformiert ist, das ein homologes Regulon, eine DNA-Sequenz in geeignetem Leseraster damit und ein oder mehrere Terminationscodons enthält, wobei die DNA-Sequenz zwischen dem Regulon und dem (den) Terminationscodon(s) lokalisiert ist und für das heterologe Polypeptid, zusätzliches Protein und eine selektive Spaltungsstelle zwischen dem gewünschten heterologen Polypeptid und dem zusätzlichen Protein eng codiert, wobei die DNA-Sequenz exprimiert wird und das entstehende Konjugat-Polypeptid ausreichend groß ist, daß es durch endogene proteolytische Enzyme nicht abgebaut wird und daß es spezifisch gespalten wird, so daß das gewünschte heterologe Polypeptid erhalten wird.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß das zusätzliche Protein ein oder mehrere heterologe Polypeptideinheiten enthält, wobei alle heterologen Polypeptideinheiten voneinander durch selektive Spaltungsstellen getrennt sind.

3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß die Spaltung in einem System außerhalb der Replikationsumgebung des Rekombinationsplasmides durchgeführt wird.

4. Verfahren nach Anspruch 3, dadurch gekennzeichnet, daß das gewünschte heterologe Polypeptid keine inneren Spaltungsstellen, die denen zwischen dem gewünschten heterologen Peptid und dem zusätzlichen Protein entsprechen, enthält.

5. Verfahren nach Anspruch 3 oder 4, dadurch gekennzeichnet, daß das gewünschte heterologe Polypeptid Arginin und Lysin fehlt und daß die selektive Spaltungsstelle durch Trypsin gespalten wird.

6. Verfahren nach Anspruch 3 oder 4, dadurch gekennzeichnet, daß dem gewünschten heterologen Polypeptid Arginin fehlt, daß die selektive Spaltungsstelle durch Trypsin gespalten wird, und daß irgendwelches Lysin, welches in dem gewünschten heterologen Polypeptid enthalten ist, zuerst gegen die Spaltung geschützt wird.

7. Verfahren nach einem der Ansprüche 3 bis 6, dadurch gekennzeichnet, daß die Spaltung mit Bromcyan erfolgt.

8. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß das Rekombinationsplasmid ein bakterielles Plasmid ist.


10. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß das gewünschte heterologe Polypeptid ein Säugetierhormon oder ein Zwischenprodukt dafür ist.

11. Verfahren nach Anspruch 10, dadurch gekennzeichnet, daß dem konjugierten Polypeptid die Bioaktivität des Hormons fehlt.

12. Rekombinationsplasmid, dadurch gekennzeichnet, daß es ein homologes Regulon, eine DNA-Sequenz, die das gewünschte spezifische heterologe Polypeptid, eine selektive Spaltungsstelle und ein zusätzliches Protein codiert, und ein oder mehrere Terminationscodon(s) enthält, wobei die DNA-Sequenz in geeignetem Leseraster zwischen dem Regulon und dem (den) Terminationscodon(s) zwischengestellt ist, so daß ein Konjugatpolypeptid, das sowohl das gewünschte heterologe Polypeptid als auch zusätzliches Protein enthält, bei der Expression in einem Bakterium, das mit dem Plasmid transformiert ist, gebildet wird, wobei eine selektive Spaltungsstelle zwischen dem gewünschten heterologen Polypeptid und dem zusätzlichen Protein vorhanden ist.

13. Plasmid nach Anspruch 12, dadurch gekennzeichnet, daß die gewünschte heterologe Polypeptidsequenz keine Innenspaltungsstellen enthält, die denen zwischen dem gewünschten heterologen Polypeptid und dem zusätzlichen Protein entsprechen.

14. Plasmid nach einem der Ansprüche 12 und 13, dadurch gekennzeichnet, daß es ein bakterielles Plasmid ist und daß die aufgeführten Elemente wie folgt in Sequenzen angeordnet sind: (Regulon)—(Codons für das zusätzliche Protein)—(Codons für die selektive Spaltungsstelle)—(Codons für die heterologe Polypeptidsequenz)—(Terminationscodon(s)).

15. Plasmid nach einem der Ansprüche 12 bis 14, dadurch gekennzeichnet, daß die Spaltungsstelle beim Methionin auftritt.


17. Plasmid nach einem der Ansprüche 12 bis 15, dadurch gekennzeichnet, daß das gewünschte heterologe Polypeptid die A-Kette von Insulin oder die B-Kette von Insulin ist.

22. Bakterienkultur, dadurch gekennzeichnet, daß sie von einem oder mehreren Wirts Bakterien cloniert ist, wobei jedes ein Plasmid, ausgewählt unter pSOM11-3, plA-1 und plB-1, enthält, wobei die Mitglieder der Kultur das konjugierte Polypeptid ausdrücken können.
26. Verfahren nach Anspruch 23, dadurch gekennzeichnet, daß das heterologe Polypeptid die B-Kette von humanem Insulin ist.
27. Verfahren nach einem der Ansprüche 24 bis 26, dadurch gekennzeichnet, daß die Spaltungsstelle bei Methionin ist.
30. Bakterienkultur, dadurch gekennzeichnet, daß sie E. Coli D1210 (plB-1), ATCC Nr. 31449 enthält.
32. Plasmid plA-1, ATCC Nr. 31448.
33. Plasmid plB-1, ATCC Nr. 31449.

Revendications
30. 1. Procédé de production d’un polypeptide hétérologe désiré, qui consiste à faire croître une bactérie transformée au moyen d’un plasmide recombinant comprenant un régulon homologue, une séquence d’ADN dans un cadre de lecture approprié à celle-ci et un ou plusieurs codons de terminaison, ladite séquence d’ADN étant placée entre ledit régulon ou le ou les codons de terminaison et codant pour ledit polypeptide hétérologe, une protéine supplémentaire et un site de clivage sélectif entre le polypeptide hétérologe désiré et la protéine supplémentaire, la séquence d’ADN étant exprimée et le polypeptide conjugué résultant étant suffisamment volumineux pour ne pas être dégradé par des enzymes protéolytiques endogènes et étant clivé spécifiquement, donnant ledit polypeptide hétérologe désiré.
2. Procédé suivant la revendication 1, dans lequel la protéine supplémentaire comprend une ou plusieurs séquences polypeptidiques hétérologues et dans lequel toutes les séquences polypeptidiques hétérologues sont séparées les unes des autres par des sites de clivage sélectif.
3. Procédé suivant la revendication 1 ou 2, dans lequel le clivage est effectué dans un système à l’extérieur de l’environnement de réplication du plasmide recombinant.
4. Procédé suivant la revendication 3, dans lequel le polypeptide hétérologe désiré ne contient aucun site de clivage interne correspondant à celui entre le polypeptide hétérologe désiré et la protéine supplémentaire.
5. Procédé suivant la revendication 3 ou la revendication 4, dans lequel le polypeptide hétérologe désiré est dépourvu d’arginine et de lysine et dans lequel le site de clivage sélectif est clivé par la trypsine.
6. Procédé suivant la revendication 3 ou la revendication 4, dans lequel le polypeptide hétérologe désiré est dépourvu d’arginine, le site de clivage sélectif est clivé par la trypsine, et toute lysine présente dans ledit polypeptide hétérologe désiré est tout d’abord protégée vis-à-vis d’un clivage.
7. Procédé suivant l’une quelconque des revendications 3 à 6, dans lequel le clivage est effectué au moyen de bromure de cyanogène.
8. Procédé suivant l’une quelconque des revendications précédentes, dans lequel le plasmide recombinant est un plasmide bactérien.
10. Procédé suivant l’une quelconque des revendications précédentes, dans lequel le polypeptide hétérologe désiré est une hormone de mammifère ou un de ses intermédiaires.
11. Procédé suivant la revendication 10, dans lequel le polypeptide conjugué ne présente pas l’activité biologique de l’hormone.
12. Plasmide recombinant comprenant un régulon homologue, une séquence d’ADN codant pour un polypeptide hétérologe particulier désiré, un site de clivage sélectif et une protéine supplémentaire, et un ou plusieurs codons de terminaison, ladite séquence d’ADN étant intercalée dans le cadre de lecture convenable entre ledit régulon et le ou les codons de terminaison, de sorte qu’un polypeptide conjugué
comprenant le polypeptide hétérologue désiré et la protéine supplémentaire résulte de l'expression effectuée dans une bactérie transformée au moyen du plasmide, un site de clivage sélectif existant entre le polypeptide hétérologue désiré et la protéine supplémentaire.

13. Plasmide suivant la revendication 12, dans lequel la séquence polypeptidique hétérologue désirée ne contient aucun site de clivage interne correspondant à celui entre le polypeptide hétérologue désiré et la protéine supplémentaire.

14. Plasmide suivant l'une quelconque des revendications 12 et 13, qui est un plasmide bactérien et dans lequel les éléments mentionnés sont disposés dans l'ordre suivant : (régulon)—(codons pour la protéine supplémentaire)—(codons pour le site de clivage sélectif)—(codons pour la séquence polypeptidique hétérologue)—(le ou les codons de terminaison).

15. Plasmide suivant l'une quelconque des revendications 12 à 14, dans lequel le site de clivage est situé au niveau d'une méthionine.

16. Plasmide suivant l'une quelconque des revendications 12 à 15, dans lequel le polypeptide hétérologue désiré est la somatostatine.

17. Plasmide suivant l'une quelconque des revendications 12 à 15, dans lequel le polypeptide hétérologue désiré est la chaîne A de l'insuline ou bien la chaîne B de l'insuline.

18. Bactérie transformée au moyen d'un plasmide suivant l'une quelconque des revendications 12 à 17.


20. Culture bactérienne comprenant une bactérie transformée suivant la revendication 18 ou 19.

21. Culture bactérienne obtenue par clonage à partir d'une ou plusieurs bactéries hôtes, comprenant chacune un plasmide suivant l'une quelconque des revendications 12 à 17, les éléments de ladite culture étant capables d'effectuer l'expression du polypeptide conjugué.

22. Culture bactérienne obtenue par clonage à partir d'une ou plusieurs bactéries hôtes, comprenant chacune un plasmide choisi entre pSOM11-3, pIA-1 et pIB-1, les éléments de ladite culture étant capables d'effectuer l'expression du polypeptide conjugué.

23. Procédé suivant la revendication 10 ou 11, dans lequel le polypeptide hétérologue est choisi entre la somatostatine, la chaîne A de l'insuline humaine et la chaîne B de l'insuline humaine.

24. Procédé suivant la revendication 23, dans lequel le polypeptide hétérologue est la somatostatine.

25. Procédé suivant la revendication 23, dans lequel le polypeptide hétérologue est la chaîne A de l'insuline humaine.

26. Procédé suivant la revendication 23, dans lequel le polypeptide hétérologue est la chaîne B de l'insuline humaine.

27. Procédé suivant l'une quelconque des revendications 24 à 26, dans lequel le site de clivage est situé au niveau d'une méthionine.

28. Culture bactérienne comprenant E. Coli. RR1 (pSOM11-3), ATCC N° 31447.

29. Culture bactérienne comprenant E. Coli. 294 (pIA-1), ATCC 31448.


32. Plasmide pIA-1, ATCC N° 31448.

33. Plasmide pIB-1-ATCC N° 31449.
Fig. 1.
**Fig. 3.**

\[ \text{DMT} - \overset{\text{H}^+}{\text{O}} - \overset{\text{P} - \text{O}}{\text{O}} - \overset{\text{B} \text{or OR}}{\text{Cl}} \quad + \quad \text{HO} - \overset{\text{B}}{\text{O}} - \overset{\text{P} - \text{O}}{\text{O}} - \overset{\text{OR}}{\text{Cl}} \quad \xrightarrow{\text{TPSTe}} \quad \text{DMT} - \overset{\text{B}}{\text{O}} - \overset{\text{P} - \text{O}}{\text{O}} - \overset{\text{Cl}}{\text{O}} \]

\[ \quad \xrightarrow{\text{H}^+} \quad \text{HO} - \overset{\text{B}}{\text{O}} - \overset{\text{P} - \text{O}}{\text{O}} - \overset{\text{Cl}}{\text{O}} \quad \xrightarrow{\text{TPSTe}} \quad \text{DMT} - \overset{\text{B}}{\text{O}} - \overset{\text{P} - \text{O}}{\text{O}} - \overset{\text{R}}{\text{Cl}} \]

\[ \text{R} = \overset{\text{O}}{\text{C} - \text{O} - \text{OME or } \overset{\text{PO} - \text{O}}{\text{O}} - \overset{\text{Cl}}{\text{OCH}_2\text{CH}_2\text{CN}}} \]

- **B** =
- **DMT** =
Fig. 4.
Fig. 6.
FIG. 9.
Fig. 10.