

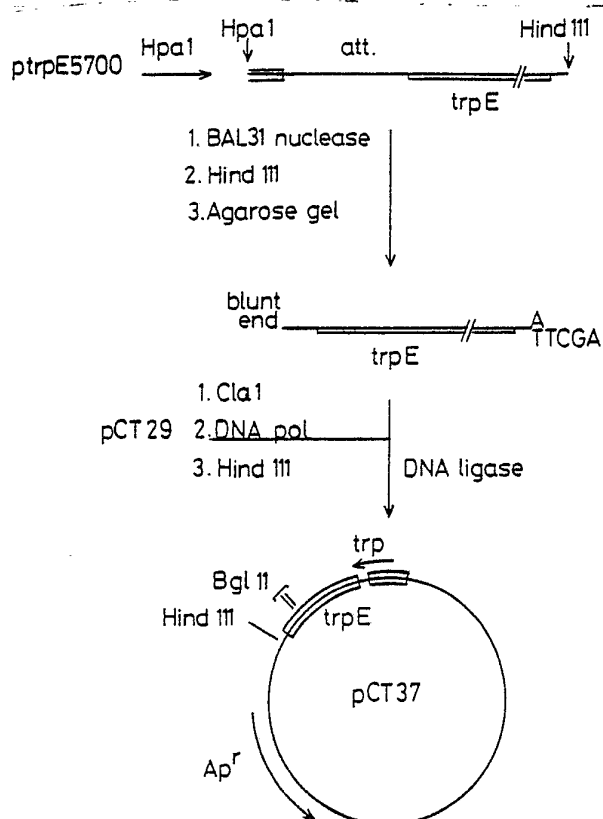


INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification³ : C12N 15/00; C12P 21/02	A1	(11) International Publication Number: WO 84/ 00380 (43) International Publication Date: 2 February 1984 (02.02.84)
(21) International Application Number: PCT/GB83/00172 (22) International Filing Date: 13 July 1983 (13.07.83) (31) Priority Application Number: 8220343 (32) Priority Date: 13 July 1982 (13.07.82) (33) Priority Country: GB (71) Applicant (for all designated States except US): CELL-TECH LIMITED [GB/GB]; 244-250 Bath Road, Slough, Berkshire SL1 4DY (GB). (72) Inventors; and (75) Inventors/Applicants (for US only) : EMTAGE, John, Spencer [GB/GB]; 12 Benjamin House, Amersham Hill, High Wycombe, Buckinghamshire (GB). HENTSCHEL, Christopher, Clive, Gabriel [GB/GB]; 48 Clifton Gardens, London W9 1AU (GB). STEPHENS, Paul, Edward [GB/GB]; 1 Selwyn Place, Cippenham, Slough, Berkshire (GB).		(74) Agent: VOTIER, Sidney, David; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB). (81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), JP, LU (European patent), NL (European patent), SE (European patent), US. Published <i>With international search report.</i>

(54) Title: VECTOR**(57) Abstract**

An expression vector suitable for the production of fusion proteins. The vector includes a nucleotide sequence comprising in phase form the 5' end; a trp promoter-operator, a trpE ribosome binding site, a structural gene comprising a trpE gene, or a 5' portion thereof and a gene coding for a heterologous polypeptide wherein the nucleotide sequence does not include a trp attenuator. The structural gene may further include a 3' portion of a trpE gene downstream of the gene coding for the heterologous polypeptide. Vectors are described which include a gene coding for human calcitonin.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	LI	Liechtenstein
AU	Australia	LK	Sri Lanka
BE	Belgium	LU	Luxembourg
BR	Brazil	MC	Monaco
CF	Central African Republic	MG	Madagascar
CG	Congo	MR	Mauritania
CH	Switzerland	MW	Malawi
CM	Cameroon	NL	Netherlands
DE	Germany, Federal Republic of	NO	Norway
DK	Denmark	RO	Romania
FI	Finland	SE	Sweden
FR	France	SN	Senegal
GA	Gabon	SU	Soviet Union
GB	United Kingdom	TD	Chad
HU	Hungary	TG	Togo
JP	Japan	US	United States of America
KP	Democratic People's Republic of Korea		

- 1 -

VECTORField of the Invention

This invention relates to the field of recombinant DNA biotechnology. In particular it relates to a vector, an organism transformed with the vector, a process for producing a polypeptide, and a polypeptide produced by the process.

Background to the Invention

The recent advances in recombinant DNA biotechnology have led to an increased demand for expression vector systems which afford particular advantages for particular desired protein products.

The expression of genes coding for relatively short polypeptides often presents problems as the expression products are in some cases unstable and may prove hard to isolate in significant quantities. It is known to overcome such problems by arranging for the production of a fusion protein comprising a sequence of amino acids homologous with the host organism in which expression is taking place, linked to the desired polypeptide which is heterologous to the host organism. The present invention is concerned with a vector capable of directing the production of such fusion proteins.

The tryptophan operon of E.coli has been the subject of detailed study in recent years. The result of this has been the elucidation of the complete nucleotide sequence of the operon (C. Yanofsky et al Nucleic Acids Research 9, 6647-6668, 1981). The tryptophan operon (hereinafter referred to as the trp operon or the trp gene) comprises, in sequence; a promoter (including an operator), a transcription start site, an attenuator, and five structural genes known as trpE, trpD, trpC, trpB and trpA (in order of their transcription). Each structural gene has an



- 2 -

associated ribosome binding site (Shine and Delgano sequence). The five structural genes code for enzymes which are used in a multi-step biosynthesis of tryptophan from chorismic acid. Expression of the genes coding for the five enzymes is controlled either by repression, resulting from excess tryptophan forming a repressor complex with another molecule, and then binding to a repression site in the promoter region, or by attenuation in the attenuator region of the trp operon.

10 The attenuator region of the trp operon possesses a gene coding for a ribosome binding site which, when transcribed, directs the production of a 14 amino acid 'leader' polypeptide. The mRNA transcript of the gene coding for the leader polypeptide assumes a secondary structure, the conformation of which determines whether the transcription of the structural genes takes place. If tryptophan is abundant, the translation of the leader mRNA is closely linked to transcription of the gene coding for the leader peptide. This causes the mRNA to assume a secondary structure which terminates transcription. If tryptophan is deficient, an alternative conformation of the leader mRNA is formed which permits transcription through the attenuator and through the structural genes (see Miozzari, G.F. and Yanofsky C.J. Bact. 133, 1457-1466, 1978; Lee, F. and Yanofsky, C. Proc. Natl. Acad. Sci. U.S.A. 74, 4365-4369, 1977; Oxender, D. Zurawski, G. and Yanofsky, C. Proc. Natl. Acad. Sci. U.S.A. 76, 5524-5527, 1979).

30 The control regions and parts of the structural genes have been used to construct vectors for the expression of heterologous polypeptides (B.E. Enger-Valk et al Gene 9, 69-85, 1980; W. Tacon, N. Carey and J.S. Emtage, Molec. Gen. Genet. 177, 427-438, 1980; R.A. Hallewell and J.S. Emtage, Gene 9, 27-47, 1980). It has been shown 35 that the control regions and the complete trpE gene can be isolated from the E.coli chromosome on a Hind III fragment having about 5,700 base pairs (A.S. Hopkins,



- 3 -

N.E. Murray and W.J. Brammar J. Mol. Biol. 107, 549-569, 1976).

The presence of the attenuator in the trp operon is disadvantageous if the control region of the operon is to be used to express heterologous genetic material. In order to express genetic material an organism requires a plentiful supply of amino acids, including tryptophan. However, the attenuator of the trp operon reduces the transcription of the structural genes in response to high levels of tryptophan. The attenuator only allows abundant expression in response to low levels of tryptophan. Such low levels of tryptophan may result in inefficient translation. This effect is not important in natural E.coli cells but is disadvantageous in a genetically manipulated system, where high expression levels are the object.

Published British patent application GB2073203A describes plasmid vectors including the trp promoter-operator system from which part of the attenuator region has been deleted. The plasmids described comprise the trp promoter-operator, nucleotides coding for the leader peptide ribosome binding site and nucleotides coding for a structural gene. The plasmids contain neither the complete trp attenuator region nor nucleotides coding for the trpE ribosome binding site. Nucleotides have been deleted from a point downstream of the ribosome binding site of the leader peptide to a point in the nucleotide sequence of trpE. The plasmids described rely upon the ribosome binding site of the leader peptide to direct production of a fusion protein comprising, for example, six amino acids of the leader polypeptide, the distal third (C-terminal) of the enzyme coded for by trpE and a heterologous polypeptide.

Summary of the Invention

According to the present invention we provide a vector including a nucleotide sequence comprising in



- 4 -

phase from the 5' end;

a trp promoter-operator,

a trpE ribosome binding site, and

a structural gene comprising a trpE gene, or a 5'

5 portion thereof and a gene coding for a heterologous polypeptide,

wherein the nucleotide sequence does not include a trp attenuator. Preferably the structural gene further includes a 3' portion of a trpE gene downstream of the part
10 of the structural gene coding for the heterologous polypeptide, without an intervening stop codon.

The vector is suitable for the production of a fusion protein comprising the heterologous polypeptide linked to the protein coded for the trpE gene or at least
15 a portion or portions thereof.

The vector is preferably a plasmid vector. The trp promoter-operator, the trpE ribosome binding site, the trpE gene or the 5' portion thereof and where appropriate the 3' portion of the trpE gene may be derived from any
20 bacterium having a trp operon. Preferably, however, they are derived from the E.coli trp operon.

The term 'heterologous polypeptide' as used herein means a polypeptide not normally produced by or found in a bacterial host organism. The term 'in phase' as
25 used herein means in the same translation reading frame.

The '5' portion' of the trpE gene is to be taken as a part of the trpE gene commencing with the 5' end of the gene. The '3' portion' of the trpE gene is here to be taken as part of the trpE gene ending with the 3' end
30 of the gene.

In another aspect of the invention we provide a process for the production of a polypeptide comprising the steps of culturing a host organism transformed with a vector of the invention, to obtain a fusion protein,
35 and cleaving the fusion protein to produce the polypeptide. Preferably the fusion protein is a protein comprising the amino acid sequence coded for by the trpE



- 5 -

gene or a 5' portion thereof and the amino acid sequence of the polypeptide. In the alternative the fusion protein may comprise, the amino acid sequence coded for by the trpE gene or a 5' portion thereof, the amino acid sequence of the polypeptide and the amino acid sequence of a 3' portion of the trpE gene.

In another aspect of the invention we provide a host organism transformed with a vector of the present invention.

10 In another aspect of the invention we provide a polypeptide produced by the process of the present invention and we further provide a fusion protein produced as an intermediate in the process of the present invention.

15 Preferably the host organisms transformed with the vector of the invention are grown in the presence of an abundance of tryptophan which has the effect of repressing expression of the heterologous genetic material. When a level of host organisms has been reached which is suitable for industrial production, the level of tryptophan in the culture medium is reduced thereby derepressing expression and allowing production of the polypeptide.

20 In yet another aspect of the invention we provide a vector including a nucleotide sequence comprising in phase, from the 5' end;

25 a trp promoter-operator,
a trpE ribosome binding site,
a trpE gene, or at least a 5' portion thereof, and
a restriction site suitable for the insertion of a
30 gene coding for a heterologous polypeptide.
wherein the nucleotide sequence does not include a trp attenuator.

Preferably the nucleotide sequence included in the vector further includes a 3' portion of a trpE gene,
35 downstream of the gene coding for the heterologous polypeptide.



- 6 -

Brief Description of the Drawings

Embodiments of the present invention are described below with reference to the accompanying drawings in which;

- 5 Figure 1 - shows a schematic process for isolating a 234 bp Taq I fragment from ptrpE₅₇₀₀ DNA, which contains the trp promoter-operator,
- Figure 2 - shows a schematic process for inserting the Taq I fragment of Figure 1 into plasmid pAT153 to form plasmids pCT12 and pCT28,
- 10 Figure 3 - shows schematically the preparation of pCT29 from pCT28 and the preparation of pCT29 for the insertion of E.coli trpE gene,
- Figure 4 - shows schematically the isolation of the trpE gene from ptrpE₅₇₀₀ DNA and its insertion into
- 15 pCT29 to form pCT37,
- Figure 5 - shows the nucleotide sequence, of part of the trp operon indicating the deletions made to the sequence to form pCT33, pCT34 and pCT37,
- 20 Figure 6 - shows a scheme for inserting a gene coding for human calcitonin, (derived from plasmid pHTB3) into plasmid pCT37,
- Figure 7 - shows schematically preparation of deletions of pCTD5,
- 25 Figure 8 - shows the nucleotide sequence around the Bgl II sites of plasmids designated 10g, 10i and 20a,
- Figure 9 - shows a scheme for the insertion of the calcitonin or calcitonin glycine gene into vectors 10g, 10i and 20a.

30



- 7 -

Description of the Invention

The invention is now described with reference to a number of embodiments.

1. Construction of vectors

5 The control region of the trp gene and the trpE gene were isolated using cosmid vectors and in vitro packaging (B. Hohn and K. Murray Proc. Natl. Acad. Sci. U.S.A. 74, 3259-3262, 1977; Collins, J and Hohn, B. Proc. Natl. Acad. Sci. U.S.A. 75, 4242-4246, 1978). E.coli chromosomal DNA
10 was partially digested with Sau 3A and the resulting fragments ligated to a 10.3 kb cosmid, 3030, that had previously been digested to completion with Bam HI. The cosmid 3030 contains a unique site for Bam HI and confers resistance to ampicillin. After ligation, the mixture
15 was packaged in vitro and used to infect a trpE⁻ strain of E.coli K12. Recombinants were selected on L-agar plates containing ampicillin from which they were replica plated onto M9 salts, glucose, casamino acids minimal agar plates supplemented with ampicillin. Recombinants were
20 thus selected that complemented the trpE⁻ strain by their ability to grow in the absence of tryptophan. Several recombinants were identified and plasmid DNA isolated and characterised by restriction enzyme analysis. One cosmid, designated 3030/trp, producing a 5700 bp fragment on
25 digestion with Hind III was selected for further use.

Cosmid 3030/trp was digested with Hind III, the fragments separated by agarose gel electrophoresis and the 5700 bp fragment recovered from a gel slice. Finally, this 5700 bp fragment was cloned into the Hind III site
30 of pAT153 to produce the plasmid p5700.

A schematic representation of a portion of the tryptophan promoter-containing fragment is given in Fig. 1. From this figure it can be seen that the complete tryptophan promoter-operator complex can be isolated on



- 8 -

a Taq I restriction fragment of 234 base pairs (bp) if the DNA is digested under conditions where about 50% of the available sites are cleaved. Thus, 10 µg of ptrp E₅₇₀₀ was incubated in 10 mM Tris-HCl pH 8.4, 100 mM NaCl, 6 mM MgCl₂ and 6 mM β-mercaptoethanol at 65°C for 20 minutes with 4 units of Taq I. After incubation the reaction mixture was extracted with phenol, precipitated with ethanol, dissolved in water and electrophoresed in a 5% polyacrylamide gel. After electrophoresis the gel was stained with ethidium bromide and the 234 bp DNA band excised from the gel and the DNA recovered (A.M. Maxam and W. Gilbert, Proc. Natl. Acad. Sci U.S.A. 74, 560-564, 1977).

This DNA fragment was then inserted into the Cla I site of the plasmid pAT153 (Twigg and Sherratt, Nature 283, 216, 1980). 10 µg of pAT153 was digested to completion with Cla I and the 5'-phosphate groups removed by incubation at 37°C for 1 hour with 0.5 units of calf intestine alkaline phosphatase in 10 mM Tris-HCl, pH 8. This procedure prevents the plasmid recircularising in the absence of an inserted DNA fragment. 100 ng of the above treated pAT153 was ligated at 15°C for 4 hours to 5 ng of the 234 bp DNA in a reaction containing 50 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 20 mM dithiothreitol, 1 mM ATP and 20 units of T4 DNA ligase. The ligated DNA was then used to transform competent E.coli K-12 strain HB101 (Boyer H.W. and Roulland Dussoix, D. J Mol. Biol. 41, 459-472, 1969) by standard techniques (Hershfield, V. et al Proc. Natl. Acad. Sci. U.S.A. 71, 3455-3459, 1974) and the bacteria plated on L-agar plates containing 100 µg/ml ampicillin. Several ampicillin resistant colonies were selected, plasmid DNA prepared and the presence of the 234 bp fragment confirmed by restriction analysis.

The resulting plasmid, designated pCT12, has the structure shown in Fig. 2. Insertion of the Taq fragment into the Cla I site reforms a Cla I site at the end



- 9 -

downstream from the trp promoter.

pCT12 was further modified to produce pCT28 and pCT29 as follows. 2 µg of pCT12 was digested to completion with Hind III and then incubated at 20°C for 30 minutes
 5 with 60 units of SI nuclease in a buffer containing 25 mM sodium acetate, pH 4.5, 0.3 M NaCl and 1 mM zinc acetate to remove the protruding Hind III ends. The reaction was terminated by raising the pH to 7.6 and extracting the mixture with phenol-chloroform (1:1). The DNA was con-
 10 centrated by ethanol precipitation and a sample incubated with T4 DNA ligase as described above and then used to transform E.coli K12 strain HB101. Several ampicillin resistant clones were selected, plasmid DNA prepared and the absence of the Hind III site confirmed by restriction
 15 analysis. The resulting plasmid, designated pCT28, has the structure shown in Fig. 2.

pCT28 was further modified. 2 µg of EcoRI-digested pCT28 was incubated with 4 units of E.coli DNA polymerase for 15 minutes at 10°C in a reaction containing 50 mM
 20 Tris-HCl pH 7.6, 10 mM MgCl₂, 10 mM β-mercaptoethanol, 0.2 mM dATP and 0.2 mM TTP. After incubation the mixture was extracted with phenol, with chloroform and then ethanol precipitated. This treatment causes the 4 nucleotides complementary to the 5' protruding ends of the EcoRI
 25 site to be filled in:-

5' A A T T C -	5' A A T T C -
3' G -	3' T T A A G -

0.6 µg of the above treated pCT28 was treated with 50 units of T4 DNA ligase in the presence of 30 picomoles
 30 of the 5'-phosphorylated synthetic oligonucleotide pCCAAGCTTGG and in 10 µl T4 DNA ligase buffer at 15°C for 16 hours. The mixture was then heated at 70°C for 10 minutes to stop the reaction and the linkers cleaved by digestion with Hind III. The linear DNA, now with Hind
 35 III ends, was separated from the linkers by electro-



- 10 -

phoresis on an agarose gel from which it was subsequently recovered by electroelution and concentrated by ethanol precipitation. Ligation, followed by transformation of E.coli K12 strain HB101, isolation of plasmid DNA and
5 identification of plasmids with Hind III and EcoRI sites, produced the plasmid designed pCT29.

pCT29, Fig. 3, contains the E.Coli tryptophan promoter-operator region and, as well, unique restriction sites for the enzymes Cla I, Hind III and EcoRI downstream
10 of the promoter region. Two of these sites, the Cla I and Hind III sites, were used to insert the trpE gene.

5 µg of pCT29 was cleaved with Cla I and the 5'-protruding ends filled in as described above using E.coli DNA polymerase in the presence of dCTP and dGTP. The
15 polymerase reaction mixture was extracted with phenol, then with chloroform and the DNA finally precipitated with ethanol. The filled-in DNA was then digested with Hind III and the resulting fragments separated by agarose gel electrophoresis. The largest fragment was isolated
20 from the gel by first staining with ethidium bromide, locating the DNA with ultraviolet light and cutting from the-gel the portion of interest. The DNA was recovered from the gel fragment by electroelution and concentrated by ethanol precipitation.

25 The trpE gene is present on the plasmid ptrp E₅₇₀₀, a portion of which is shown in Fig. 4. As described above however, ptrp E₅₇₀₀ also contains the trp attenuator. To remove the attenuator region 5 µg of ptrp E₅₇₀₀ was digested with Hpa I and then treated with nuclease BAL 31.
30 This nuclease is a highly specific nuclease that can be used to shorten DNA fragments from the ends. Further, its action produces blunt-ended molecules. 5 µg of Hpa I digested ptrp E₅₇₀₀ was treated with 1.5 units of BAL 31 nuclease in 0.6 M NaCl, 12 mM CaCl₂, 12 mM MgCl₂, 20 mM
35 Tris-HCl pH 8, 1 mM EDTA at 30°C for 1.5 minutes. The mixture was then phenol extracted, chloroform extracted



- 11 -

and ethanol precipitated. This treatment removes 150-250 bp from each end of the DNA fragments and so will remove the attenuator region from most molecules as it is 150 bp from the Hpa I site. The above treated DNA fragments were
5 digested to completion with Hind III and then separated by agarose gel electrophoresis. DNA containing the trpE gene was isolated from the gel by first staining the gel with ethidium bromide, locating the DNA with ultraviolet light and cutting from the gel DNA in the size range 1800-
10 1850 bp. This DNA was recovered from the gel fragment by electroelution and concentrated by ethanol precipitation.

The trp E fragments isolated above can be inserted into pET 29 that had been modified as described above. Fig. 4 illustrates the ligation reaction. 0.2 µg of
15 modified pCT29 and 80 µg of the trpE fragment were incubated at 20°C for 16 hours with 100 units T4 DNA ligase in 50 mM Tris pH 7.6, 10 mM MgCl₂, 1 mM ATP and 20 mM-dithiothreitol. The mixture was then used to transform E.coli K12 strain HB101 as described above and trans-
20 formants selected on L-agar plates containing ampicillin. Several ampicillin resistant colonies were selected, plasmid DNA isolated and examined by restriction analysis. This analysis confirmed the presence of the trpE gene and the loss of the attenuator region. Three plasmids were
25 isolated and were designated pCT33, pCT34 and pCT37.

Nucleotide sequence analysis was performed using the dideoxy chain termination method (Sanger et al., 1977 PNAS, 74, 5463-5467). The analysis showed that nucleotides 24 to 131, inclusive, had been lost in pCT33 and pCT37
30 and that nucleotides 24 to 151, inclusive, had been lost in pCT 34, (assuming that the start point of mRNA transcription is nucleotide 1). This further confirmed that all of the attenuator region of trpE had been deleted in these plasmids (Fig. 5).

35 Plasmid pCT37, was selected for further work involving the human clacitonin gene. (This work is described in



- 12 -

Published International Patent Applications WO 83/00327 and WO 83/00346).

2. Calcitonin and Calcitonin-Gly Fusions with trp E

The scheme outlined in Fig. 6 shows the steps carried out to combine the human calcitonin sequence derived plasmid pH-T-B3 (see above mentioned published International Patent Applications) with the expression vector pCT37 described above to produce fusion proteins. The initial step involves the modification of the calcitonin sequence so that the ultimate amino acid in the fusion protein is either the proline corresponding to the authentic terminal amino acid in calcitonin or so that a further glycine is translated. The purpose of this construction relates to the final processing envisaged in calcitonin production and is described fully below. 20 µg of pH-T-B3 were incubated in 6 mM Tris-HCl pH 7.5, 6 mM MgCl₂, 6 mM β-mercaptoethanol and 20 mM KCl with 10 units of Bst NI for 60 minutes at 60°C. After incubation the reaction was made 0.3 M in NaAcetate pH 6.0, extracted with phenol, then chloroform and concentrated by ethanol precipitation. The precipitated DNA was washed with 70% ethanol, dried under vacuum and redissolved in 20 µl of water (Step i Fig. 6). Bst NI cuts the calcitonin sequence only in the proline at position 32 and in such a way that the T residue in the anti-coding strand is removed. This T residue is added back in the next step (ii). 10 µg of the Bst NI cleaved pH-T-B3 were incubated in 60 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 10 mM β-mercaptoethanol, 0.2 mM dCTP, 0.2 mM dTTP and 6 units of Klenow enzyme (DNA-polymerase large fragment; supplied by the Boehringer Corporation (London) Ltd.). After the incubation the DNA was phenol extracted, chloroform extracted, ethanol precipitated, washed in 70% ethanol and redissolved in 10 µl of water. The resulting DNA is now flush-ended and can be ligated to



- 13 -

other flush-ended DNA molecules. The DNA was ligated (covalently joined) with two separate synthetic oligonucleotides designed A or B in Fig. 6:-

5 A = T A G G A T C C T A
 A T C C T A G G A T

 B = G G T T G A T C A A C C
 C C A A C T A G T T G G

Thus 2.5 µg of the pH-T-B3 derived DNA fragments were incubated in separate reactions with 400 ng of oligonucleotides A or B in 60 mM Tris-HCl pH 7.5, 8 mM MgCl₂, 10 mM β-mercaptoethanol, 1 mM ATP and 1.5 µl T4 DNA ligase (New England Biolabs Lot 17) for 24 hours at 16°C. Ligase activity was destroyed by raising the temperature to 70°C for 5 minutes and the DNA ethanol precipitated, washed in 15 70% ethanol, dried under vacuum and redissolved in 10 µl of water. 1 µl of both DNA samples (i.e. ligated with oligonucleotide A or B) was analysed for efficient ligation, as evidenced by its increase in size through concatemerisation and visualised by staining an agarose gel 20 with ethidium bromide following an electrophoretic size separation. Note that the two ligation reaction samples represent the start of two separate and parallel construction routes (i.e. A or B) as shown in Fig. 6. The description of the experiments below apply for both 25 series of constructions. The remaining plasmid DNA (9 µl) was separated from unreacted oligonucleotides by Sephadex G-50 chromatography and concentrated by ethanol precipitation. The DNA was redissolved in 50 µl 6 mM Tris-HCl, 50 mM NaCl, 6 mM MgCl₂, 6 mM β-mercaptoethanol and excess 30 Sau 3A enzyme added to ensure complete cutting at all Sau 3 sites after 1 hours at 37°C. The linker oligonucleotides contain the Sau 3A recognition sequence 5' GATC 3' and the Bgl II site 5' AGATCT 3' is also a Sau 3A site so that after the Sau 3A digestion, the calcitonin 35 sequence resides in DNA fragments about 110 base pairs



- 14 -

long (the exact length differing by one nucleotide depending on whether oligonucleotide A or B was used). These fragments were isolated following polyacrylamide gel electrophoresis (PAGE) as described above (Step iv Fig.6) and following ethanol precipitation, were dissolved in water (10 μ l). The vector expression plasmid pCT37 (described in detail above) was prepared for ligation into the Bgl II site as follows. 10 μ g of pCT37 was incubated in 6 mM Tris-HCl, 60 mM NaCl, 6 mM MgCl₂, 6 mM β -mercaptoethanol and 10 units Bgl II for 60 minutes at 37°C. The linearised DNA was diluted to 0.5 ml and made 0.3 M in sodium acetate pH 6.0. 5' Phosphate groups were removed by the addition of 3 units of calf intestinal alkaline phosphatase (CIAP) and incubation at 60°C for thirty minutes, after which a further 3 units of CIAP was added and a further 30 minute incubation carried out.

Phosphatase activity was destroyed by one phenol, two phenol-chloroform and three chloroform extractions, after which the DNA was ethanol precipitated and redissolved in water (10 μ l). In Step vi the calcitonin containing fragments were ligated into the Bgl II site of pCT37 as follows. 0.1 μ g of the Bgl II linearised and CIAP treated pCT37 was incubated with approximately 2 ng of the calcitonin containing fragments in 60 mM Tris-HCl pH 7.5, 8 mM MgCl₂, 10 mM β -mercaptoethanol, 1 mM ATP and 0.2 μ l T4 DNA ligase (New England Biolabs Lot 17) for 16 hours at 16°C. The DNA was then used to transform frozen competent E.coli HB101 using standard methods (cf Methods in Enzymology Vol 68, pp 326-331) and transformants resistant to ampicillin selected on L-agar plates containing 100 μ g/ml ampicillin. Clones containing plasmids with the calcitonin sequence in the Bgl II site and in the correct orientation were identified and the constructions confirmed by DNA sequencing using the method of Maxam and Gilbert, W Proc. Natl. Acad. Sci. 74 560 (1977).



- 15 -

At this stage the calcitonin DNA translation reading frame is out of phase with the trpE gene reading frame and if this reading frame is being used by a ribosome, a stop codon (TAA) will be encountered after amino acid 322 of the trpE gene. Thus in contrast to the parent expression vector pCT37 which directs the overproduction of trpE gene product, these constructions would be expected to direct the overproduction of a novel truncated protein 323 amino acids long (i.e. about 35% Daltons M Wt).

That such a novel protein is indeed produced by E.coli harbouring these plasmids, and only under appropriate inducing conditions, has been established experimentally. Thus the final steps in the construction of plasmids directing the overproduction of the desired trpE-calcitonin (or calcitonin-gly) fusion proteins requires the calcitonin sequence to be brought into the same translation reading frame as the trpE gene. This can be achieved by adding $3n + 2$ nucleotides at the fusion junction between the trpE sequence and the calcitonin sequence (i.e. at the recreated unique Bgl II site). The synthetic octomer

5' G A T C C C G G

G G C C C T A G is suitable

since it is of the correct length ($8 = 3 \times 2 + 2$), has an appropriate 5' overhang for ligation into the Bgl II site and in addition creates a new unique restriction site for the enzyme Sau (5' CCCGGG 3') at the fusion junction. Thus the plasmids (10 μ g) were incubated in 6 mM Tris-HCl, 50 mM NaCl, 6 mM β -mercaptoethanol, 6 mM $MgCl_2$ with 10 units Bgl II and the phosphate groups removed from the linear molecules using CIAP as described above. Following phenol extraction and ethanol precipitation 0.1 μ g of the linear DNA was incubated with a 10-fold molar excess of the synthetic oligonucleotide



- 16 -

G A T C C C G G

G G C C T A G in 60 mM Tris-HCl pH 7.5, 8 mM MgCl₂, 10 mM β -mercaptoethanol, 1 mM ATP and 0.2 μ l T4 DNA ligase, for 16 hours at 16°C (Step ix Fig. 6). The ligation products were used to transform E.coli HB101 cells and ampicillin resistant transformed clones isolated on L-agar plates containing 100 μ g/ml ampicillin.

The majority of plasmids in these clones were found to contain more than one copy of the synthetic oligonucleotide at the fusion junction. 1 μ g of plasmids with 2 inserted linkers was linearised by incubation in 6 mM Tris-HCl, 6 mM MgCl₂, 6 mM β -mercaptoethanol and 20 mM KCl with 10 units of Sma I (supplied by the Boehringer Corporation (London) Ltd.), and following purification of the linear molecules from unlinearised plasmid, the DNA was religated with T4 DNA ligase and following a further round of transformation into HB101 plasmids with the structure shown at the bottom of Fig. 6, were isolated. As illustrated the DNA junction is now such that the calcitonin sequence is in phase with the trpE sequence and such plasmids therefore direct the production of a novel fusion protein of the expected size (about 38K Daltons M Wt; see Fig. 10). Verification that E.coli expressing this protein is expressing a human calcitonin sequence has been obtained by radioimmunoassay.



- 17 -

C. CALCITONIN FROM A TRP E-CALCITONIN FUSION PROTEIN

The ultimate aim of introducing the human calcitonin sequence into the trp E gene of E.coli in such a way that substantial quantities of a trp E- calcitonin fusion protein are produced is to liberate the calcitonin peptide in a commercially viable way. This may be done by several possible routes. As a first step the calcitonin peptide must be cleaved from the fusion protein. This can be achieved by the use of the enzyme trypsin which will cleave exactly at the fusion junction since the initial cysteine residue is preceded by an arginine. Calcitonin contains no arginines but one lysine residue at position 18 which would also be liable to trypsin cleavage. This lysine can however be protected by citraconic anhydride (Shine S, Fettes I, Nancy C.Y. Lan, Roberts, J.L. and Baxter, J.D. Nature 285, 456-461). The peptide liberated by this procedure differs from authentic calcitonin only in that in authentic calcitonin the C-terminal amino acid is a prolinamide rather than a proline (or proline-glycine) amino acid. Conversion of the liberated peptide into authentic calcitonin is possible through the use of the C-terminal modification activity of yeast carboxypeptidase Y (Breddam, K, Widmer, F and Johanson, J.T. Carlsberg Res. Commun. 45, 237-247 and 361-367 1980).

D. CONSTRUCTION OF VECTORS HAVING TRUNCATED TRP E GENES

PCT37 makes use of a naturally occurring BglII site in the trp E gene to form a fusion between the N terminus of the trp E gene product and calcitonin. This construction was used as the basis of constructing



- 18 -

a further family of trp E fusion vectors in which the 3' end of the trp E gene has been deleted. The trp E-calcitonin fusion plasmid (pCTD5) was linearised with the restriction enzyme SmaI (Fig.7). 20 μ g of

5 pCTD5 was incubated with 20 units of SmaI for 2 hours at 37°C in 20mMKCl, 6mM Tris-HCl (pH8.0), 6mM MgCl₂, 6mM β -mercaptoethanol, 100 μ g/ml bovine serum albumin. The reaction was stopped by phenol extraction followed by ethanol precipitation. The linear DNA (200 μ g/ μ l)

10 was then digested with Bal31. 10 μ g of DNA was digested with 1.25 units of Bal31 for varying time periods at 30°C in 600mM NaCl, 12mM Ca Cl₂ 12mM MgCl₂, 20mM Tris HCl (pH8.0), 1.0mM EDTA in a total volume of 50 μ l. 10 μ l aliquots were removed after 10 minute intervals

15 and the reaction stopped by phenol extraction. The 5' terminal phosphatase groups were removed by the addition of 3 units of calf intestinal alkaline phosphatase (CIAP) and incubation at 37°C for 30 mins, phosphatase activity was destroyed by two phenol and

20 two ether extractions. 100 μ g of the Bal31 digested DNA was incubated with 25 μ g of the kinased synthetic oligonucleotide linker R53 (CAAAAGATCTTTTG) in 60mM Tris HCl (pH7.5), 8mM MgCl₂, 10mM β -mercaptoethanol, 1mM ATP and 1.5 μ l of T4 DNA ligase. (New England

25 Biolabs Lot 20) for 16 hours at 16°C. The synthetic linker introduces a unique Bgl II site into the plasmid. Ampicillin resistant colonies were selected on L. agar plates containing 100 μ g/ml ampicillin following transformation into frozen competent E.coli HB101 (Methods

30 in Enzymology Vol.68 pp 326-331). The presence of BglIII restriction enzyme site in these plasmids was confirmed by restriction analysis.

The ability of these plasmids to express a novel

35 protein band of truncated trp E was investigated



- 19 -

in the following way. E.coli cells bearing a plasmid with a BglIII site were grown at 37°C in either L-broth in which the trp promoter is not induced or in minimal medium, (based on M9 salts) in which the promoter is induced (R.A. Hallewell and J.S. Emtage, Gene 9 27-47, 1980). The cells were harvested by centrifugation and dissolved in SDS sample buffer (Laemmli U.K., Nature 227, 580, 1970). Protein equivalent to 50µl of original culture was electrophoresed on a 12.5% (W/V) polyacrylamide gel and the separated proteins visualized by staining with Coomassie blue. In this way three plasmids were identified which gave truncated trpE proteins on induction.

The three plasmids 10g, 10i and 20a produced an inducible trpE protein band on Coomassie blue stained polyacrylamide gels of Mr 30K, 40K and 20K respectively. Restriction enzyme analysis indicated that the remaining N. Terminal portion of the trpE gene in these plasmids is 470 base pairs, 750 base pairs, and 380 base pairs respectively. In all cases the 5' portion of the gene is too small to encode the protein band seen on the Coomassie blue stained gel, and indicates that the trpE translation phase continues across the BglIII site. DNA sequence analysis of these constructions confirms this and indicates in the case of 10g and 10i that the trpE translation phase is maintained across the BglIII site and read through continues to the natural translation stop signal of the trpE gene.

The plasmid 20a, possesses two BglIII sites which generate a fragment of 690 base pairs. This fragment originated from the 5' region of the trpE gene (coordinates 140 to 160 inclusive; numbered according to Yanovsky et al., 1981, N.A.R. 9 6647-6668). The orientation of this fragment in the plasmid was not determined.



- 20 -

The nucleotide sequence analysis indicates that the phasing around the BglIII sites in these plasmids is different in each case. (Fig.8). The phasing around the BglIII site in 10i is the same as that in the first BglIII site of the authentic trpE gene. The three novel plasmids provide a series of vectors which will allow the cloning of any gene in phase with the truncated N.terminus of trpE in order to produce a fusion protein.

10

The formation of the trpE calcitonin fusions in the three vectors follows three paths depending on the phasing around the BglIII sites and these are outlined in (Fig.9).

15

The DNA encoding the human calcitonin gene is provided by cleavage of a precursor of the original trpE calcitonin fusion construction (in which the trpE and calcitonin are out of phase) with BglIII and PstI.

20 10µg of plasmids D13, (trpE-calcitonin) and E2, (trpE-calcitonin-gly) were individually incubated with 10 units of BglIII and PstII in 60mM NaCl, 10mM Tris-HCl (pH7.4), 10mM MgCl₂, 10mM β -mercaptoethanol, 100µg/ml bovine serum albumin at 37°C for 3 hours. Following
25 electrophoresis in low melting point agarose a 1.6Kb fragment containing the calcitonin gene sequence, was purified by three phenol extractions after the appropriate gel slice had been melted at 70°C for 10 minutes. The fragment was then concentrated by ethanol precipitation.
30 The three vectors were digested with BglIII and Pst as described above. The 5' phosphate groups from the linear fragments were removed by CIAP treatment. 250µg of cut vector were incubated with a 3M excess of the BglIII-PstI fragment from D or E2 in
35 the presence of 60mM Tris HCl (pH7.5), 8mM MgCl₂,



- 21 -

10mM β -mercaptoethanol, 1mM ATP and 1.5 μ l of T4 DNA ligase (New England biolabs, LOT 19) for 16 hours at 16°C. The DNA was then used to transform competent HB101 and transformants were selected for on L. agar plates containing 100 μ g/ml ampicillin. Clones containing the calcitonin sequence were identified by restriction enzyme analysis. In only one case (i.e. the 20a construction) was the phasing correct to produce a trpE-calcitonin fusion.

10

In order to get the phasing correct, in the other two cases, the following steps were performed. The 10g construct was linearised by cleavage with BglII and the 5' overhangs were then filled in. 800 μ g of plasmid was incubated with .2mM dGTP, .2mM d ATP, .2mM dTTP and 0.2,M dCTP, 10mM Tris-HCl (pH8.5) 10mM MgCl₂, and 2.5 units of E.coli DNA polymerase (large fragment Acc Klenow) at 20°C for 30 min. The reaction was stopped by heating to 70°C for 10 min. The blunt ended linear molecule was then religated using T4 DNA ligase. Ampicillin resistant transformants were screened for plasmids possessing a unique ClaI site indicating successful filling in of the GAT 5' overhang. The calcitonin should now be in phase with the trpE gene.

25

After linearising the 10i construct with BglII the plasmid was incubated with a 100 fold molar excess of synthetic oligonucleotide R70 (GATCCCGG) in 60mM Tris HCl, 8mM MgCl₂, 10mM β -mercaptoethanol, 1mM ATP and 2 μ l of T4 DNA ligase for 16 hours at 16°C. The ligation products were used to transform E.coli HB101 cells and ampicillin resistant clones were isolated on L. agar plates containing 100 μ g/ml ampicillin. The plasmids in these clones were screened for the presence of a unique SmaI site, indicating the presence of the

35



- 22 -

linker and the corresponding correct phasing between trpE and calcitonin. The production of a trpE calcitonin fusion was confirmed by Western blotting using an antibody raised to authentic human calcitonin (Towbin
5 et al, 1979 PNAS 76 4350). Each construction produced a band of the predicted molecular weight which was immunogenic.



- 23 -

CLAIMS:

1. A vector including a nucleotide sequence comprising, in phase, from the 5' end;
a trp promoter-operator,
5 a trpE ribosome binding site, and
a structural gene comprising a trpE gene, or a 5' portion thereof, and a gene coding for a heterologous polypeptide,
wherein the nucleotide sequence does not include a trp
10 attenuator.
2. A vector including a nucleotide sequence comprising in phase from the 5' end;
a trp promoter-operator,
a trpE ribosome binding site, and
15 a structural gene comprising a trpE gene, or a 5' portion thereof, a gene coding for a heterologous polypeptide, and a 3' portion of a trpE gene, without an intervening stop codon,
wherein the nucleotide sequence does not include a trp attenuator.
- 20 3. A process for the production of a polypeptide comprising the steps of;
culturing a host organism transformed with a vector according to claim 1 or 2, to obtain a fusion protein, and
25 cleaving the fusion protein to obtain the polypeptide.
4. A polypeptide produced by a process according to claim 3.
5. A fusion protein produced as an intermediate in a
30 process according to claim 3.
6. A host organism transformed with a vector according to claim 1 or 2.
7. A vector including a nucleotide sequence comprising in phase from the 5' end;
35 a trp promoter-operator,

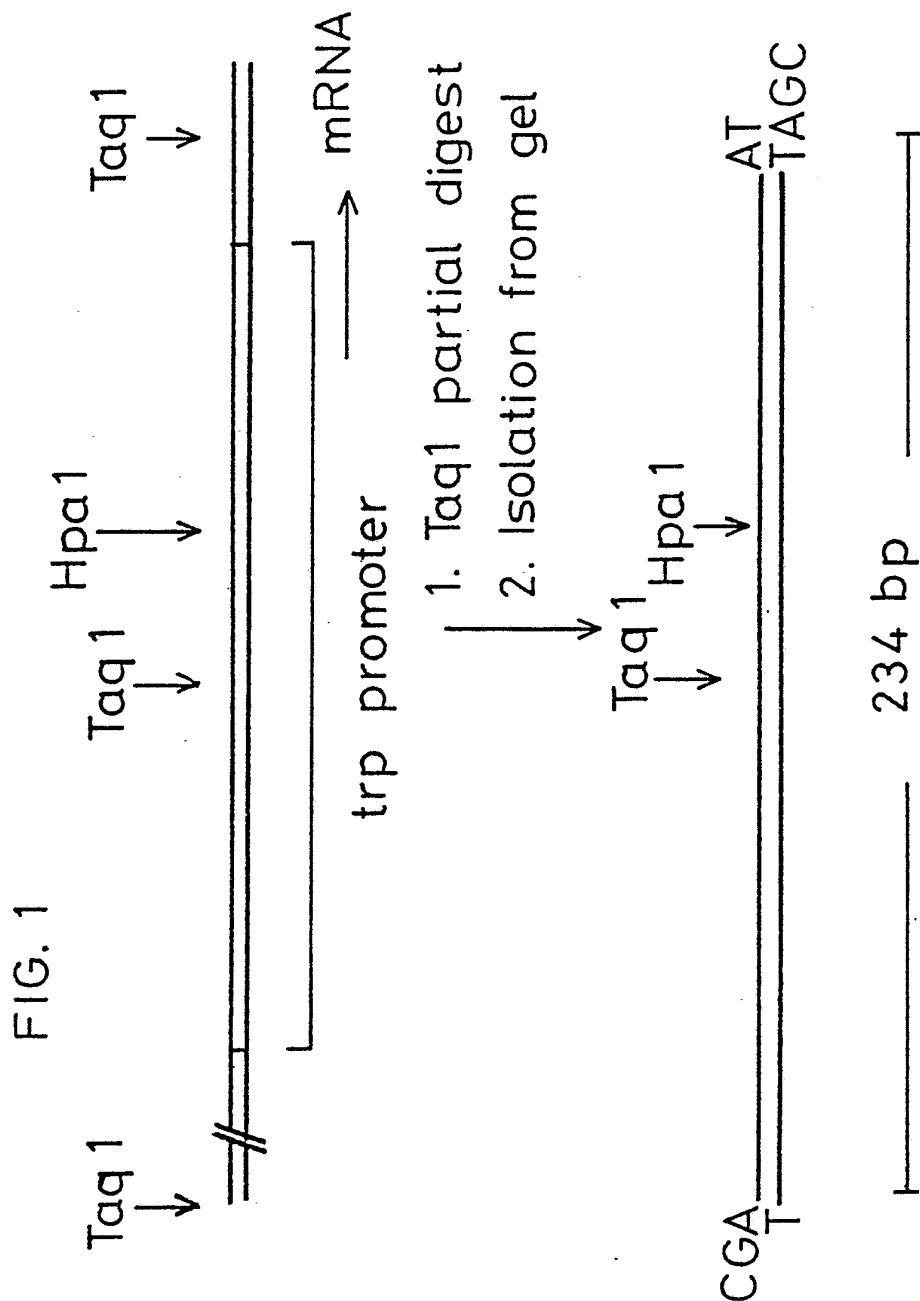


- 24 -

- a trpE ribosome binding site,
a trpE gene, or at least a 5' portion thereof, and
a restriction site suitable for the insertion of a
gene coding for a heterologous polypeptide,
5 wherein the nucleotide sequence does not include a trp
attenuator.
8. A vector including a nucleotide sequence compris-
ing in phase from the 5' end;
a trp promoter-operator,
10 a trpE ribosome binding site,
a trpE gene or at least a 5' portion thereof,
a restriction site suitable for the insertion of a
gene coding for a heterologous polypeptide, and
a 3' portion of a trpE gene,
15 wherein the nucleotide sequence does not include a trp
attenuator.



1/9



2/9

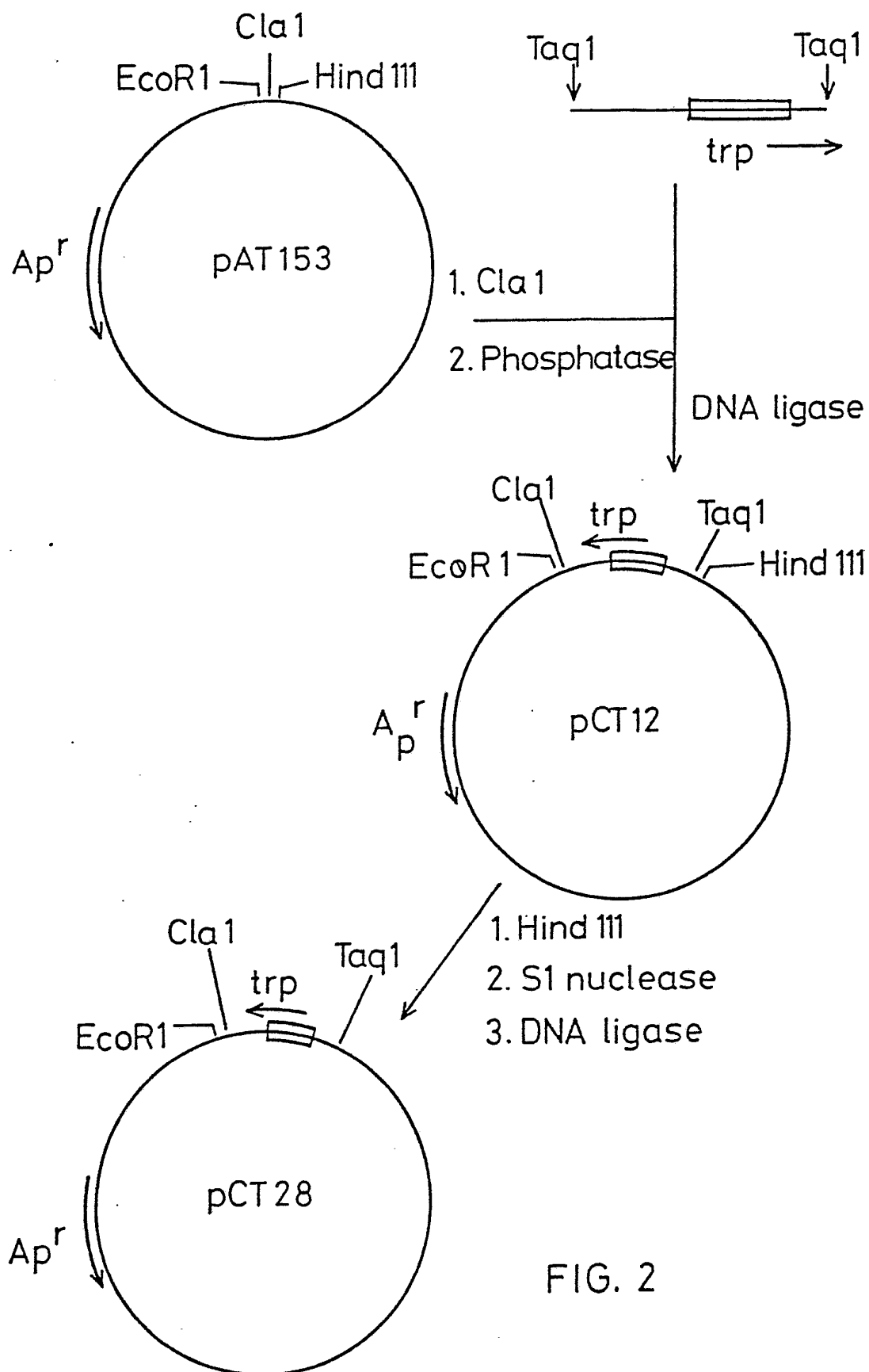
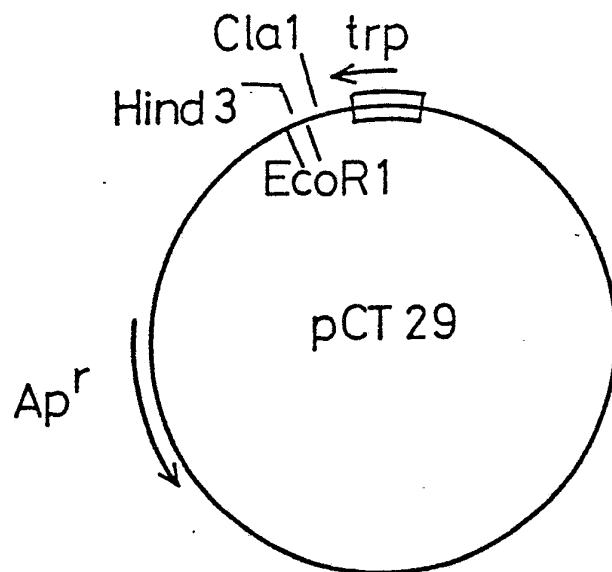


FIG. 2

3/9

pCT 28

- | | |
|-----------------|-------------------|
| 1. EcoR1 | 2. DNA polymerase |
| 3. DNA ligase + | pCCAAGCTTGG |
| 4. Hind III | 5. Agarose gel |
| | 6. DNA ligase |



- | |
|------------------|
| 1 Cla1 |
| 2 DNA polymerase |
| 3 Hind3 |

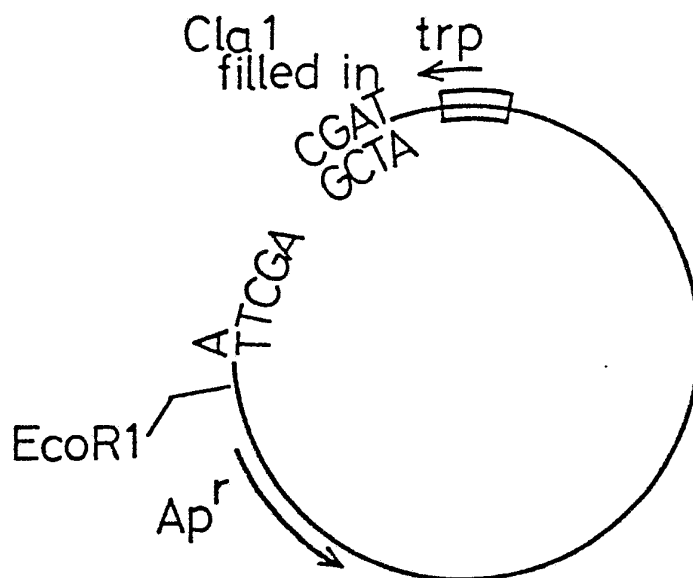


FIG. 3

4/9

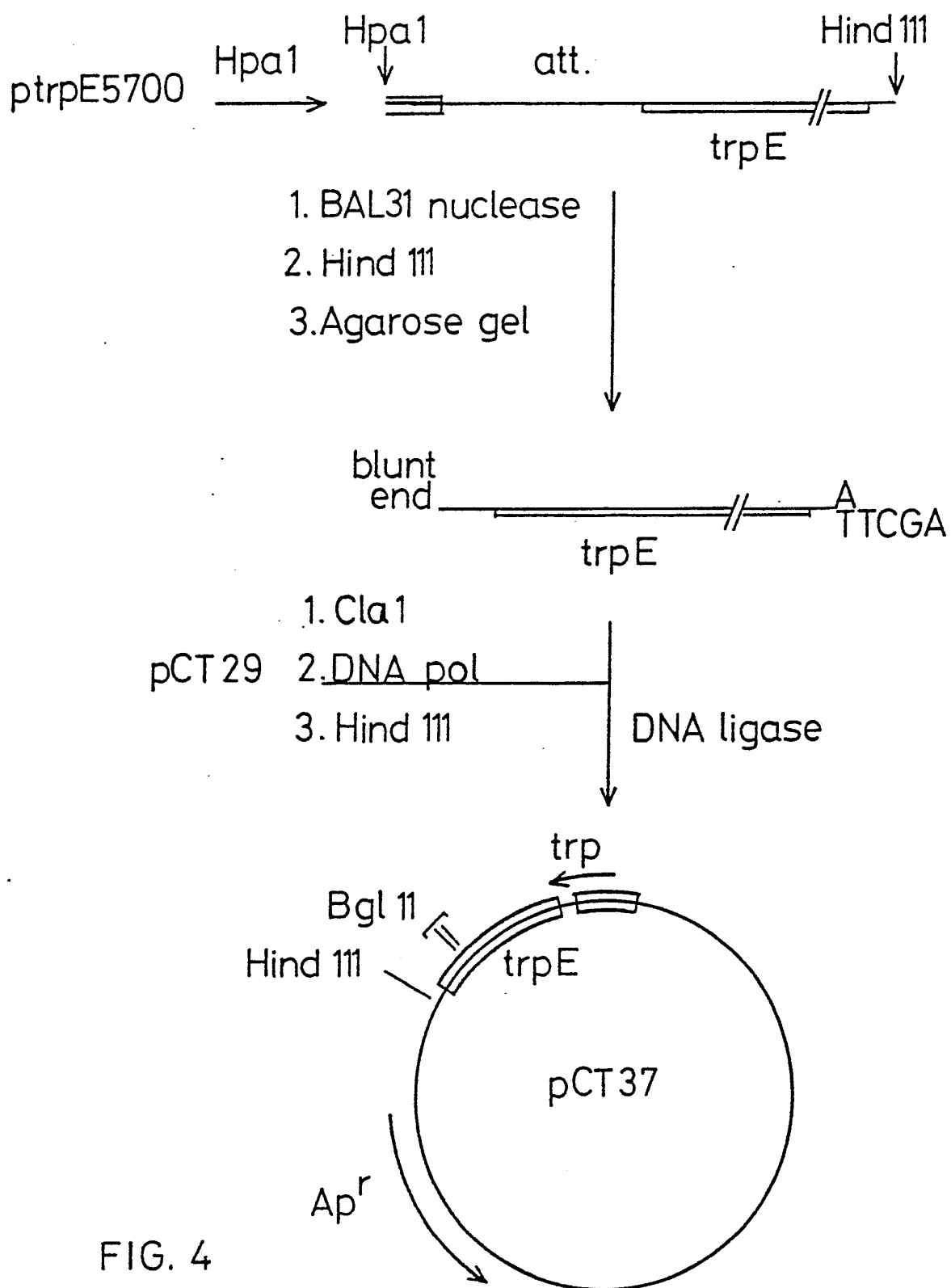


FIG. 4

5/9

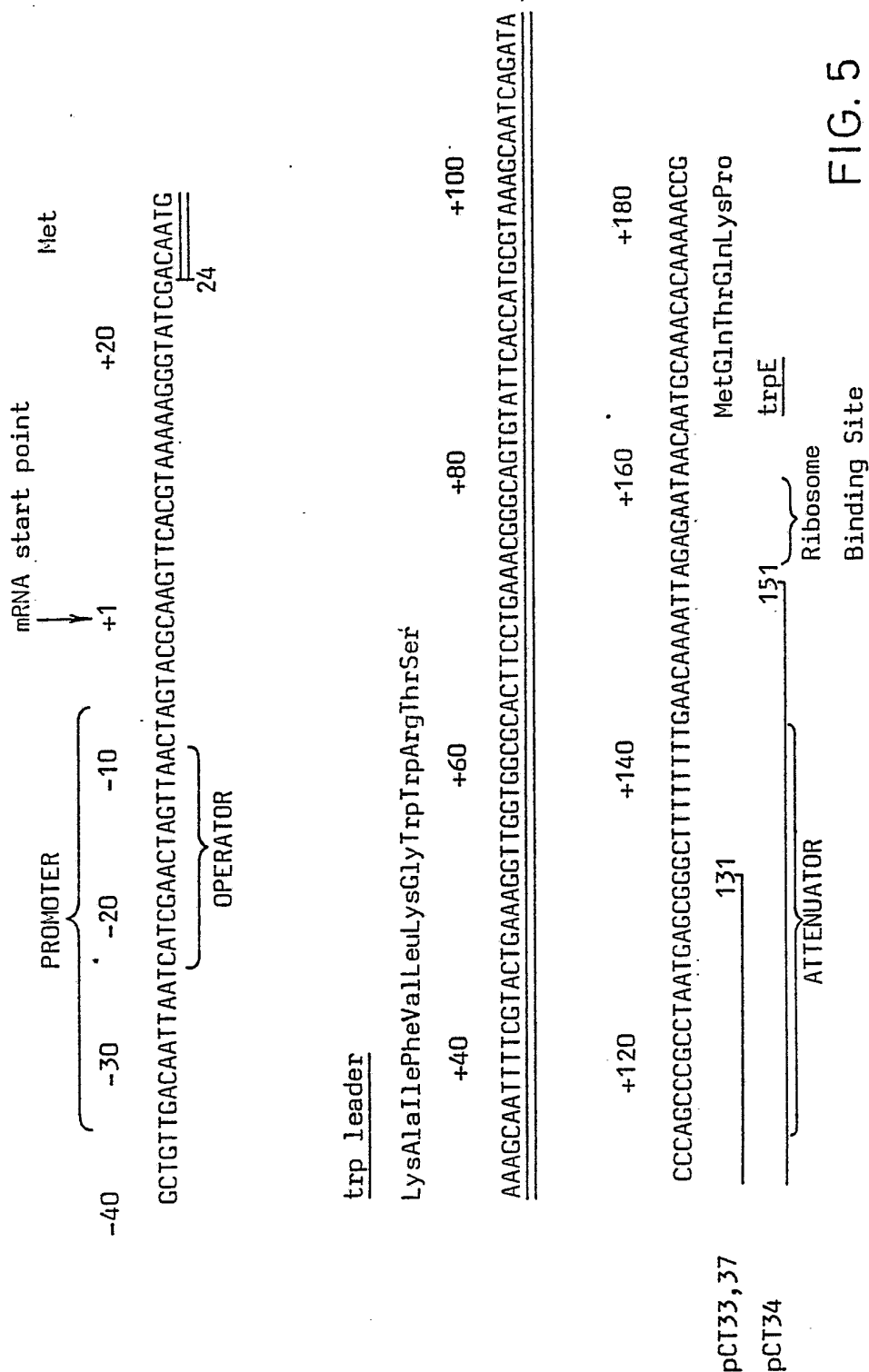
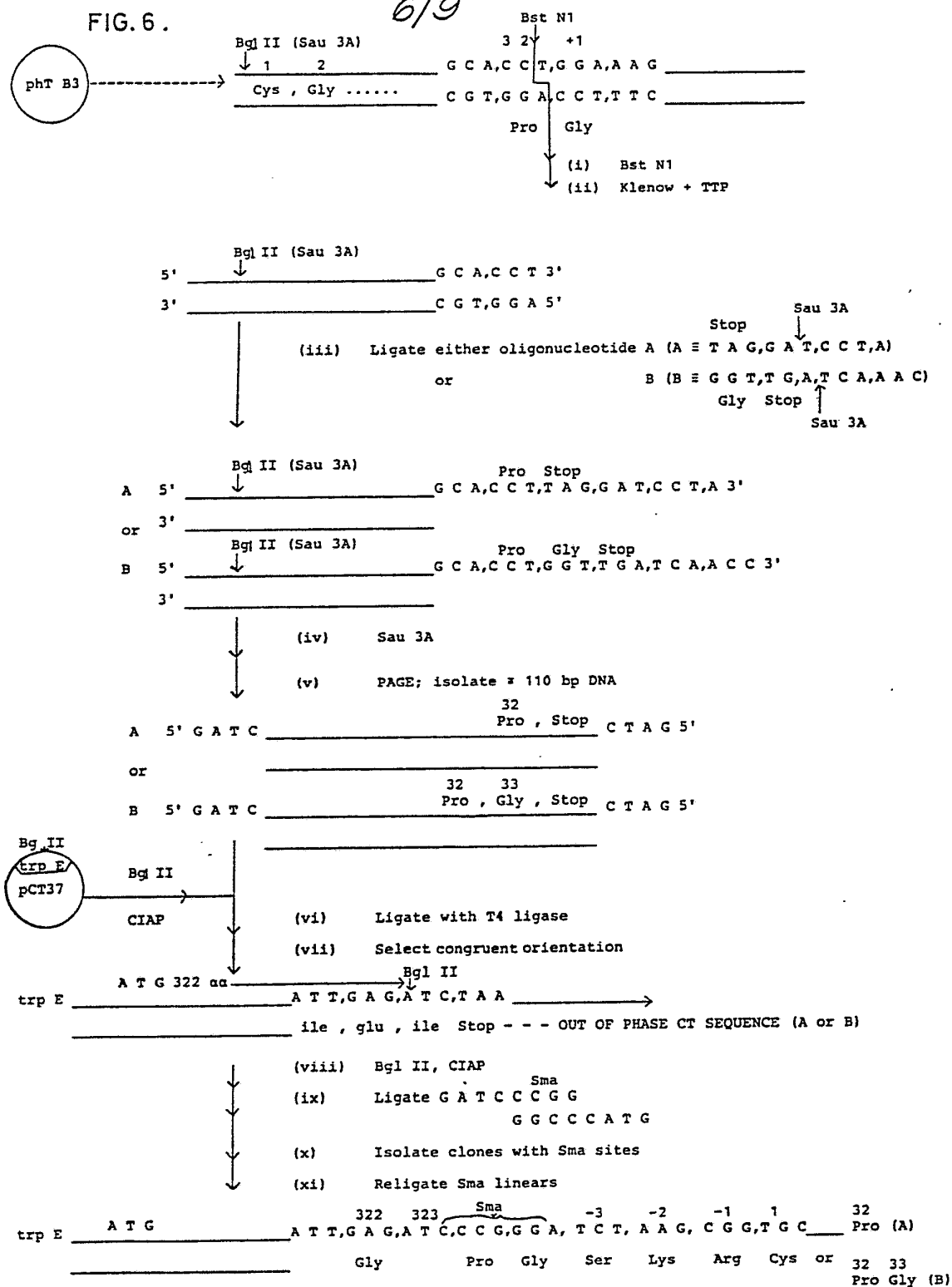


FIG. 5

6/9



7/9

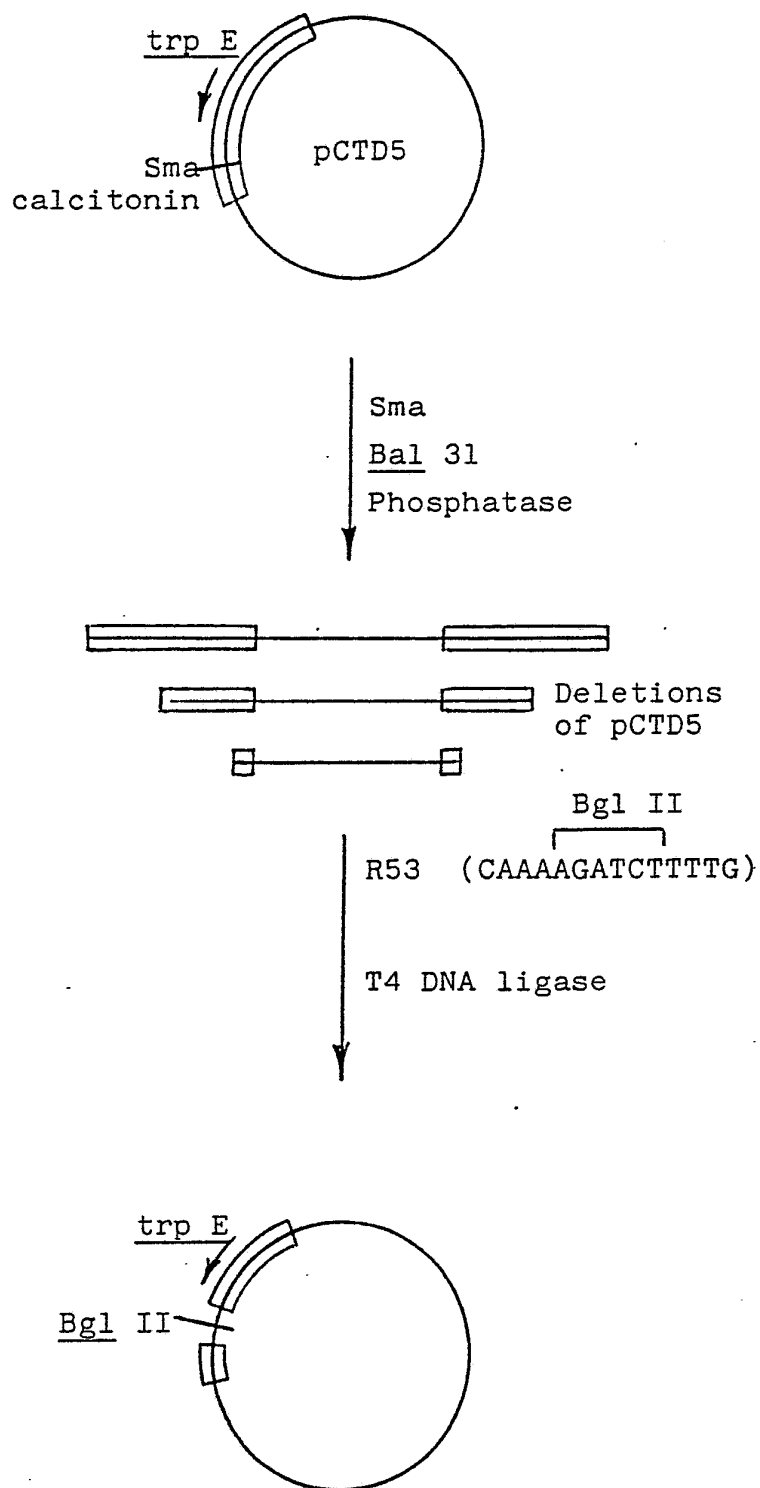


FIG. 7

8/9

FIG. 8

10g

Bgl II
 ATG-----TTA CCG CCA AAA GAT CTT TTG GTC-----TTC TGA
 Met/←160aa→/Leu Pro Pro Lys Asp Leu Leu Val/←137aa→/Phe xxx

10i

Bgl II
 ATG-----CTG CCC CAA AAG ATC TTT TGC GAA-----TTC TGA
 Met/←271aa→/Leu Pro Gln Lys Ile Phe Cys Glu/←114aa→/Phe xxx

20a

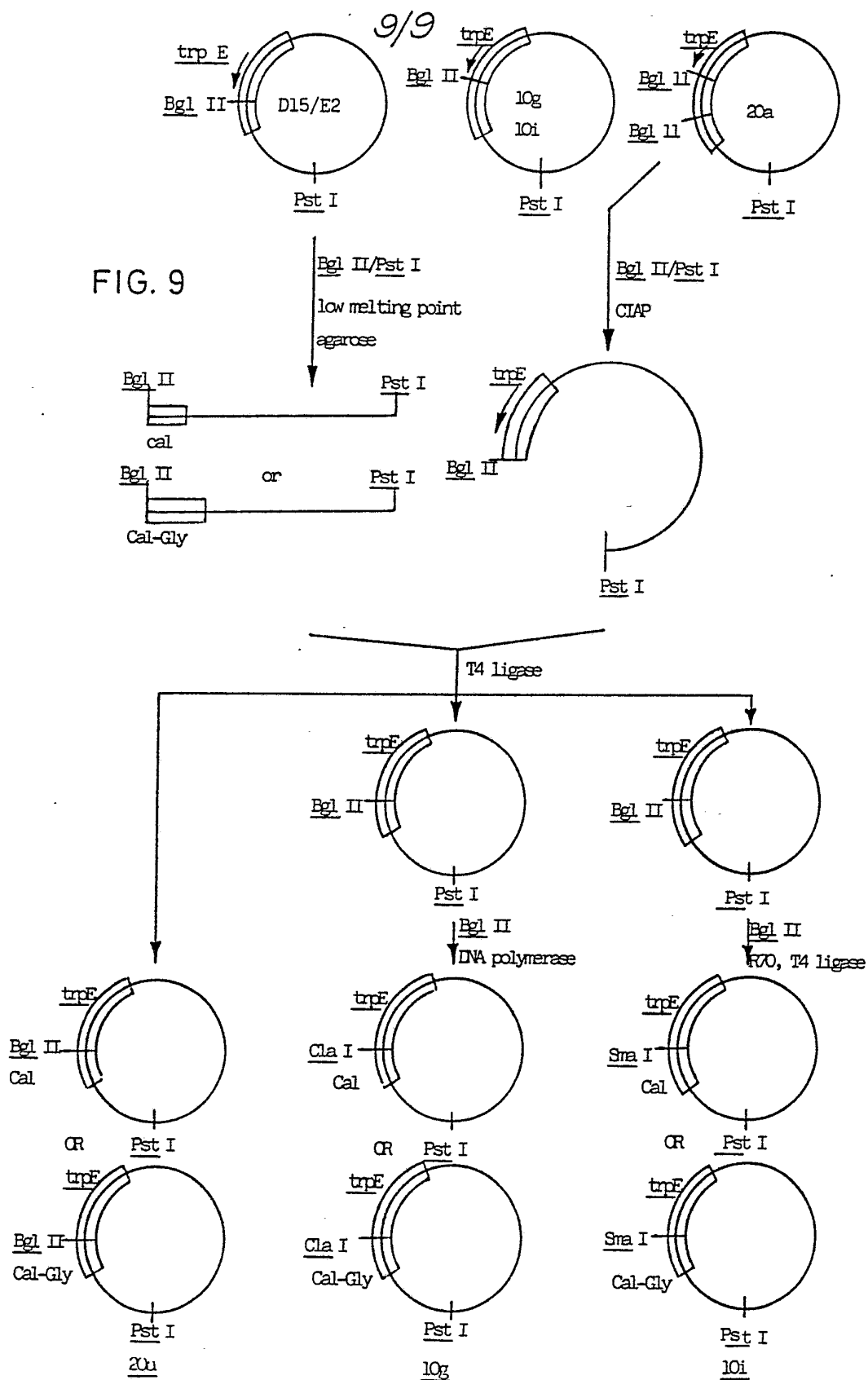
Bgl II
 ATG-----GAC GCC AAA AGA TCT TTT
 Met/←123aa→/Asp Ala Lys Arg Ser Phe

trpE

ATG/-----/CGC CAG ATT GAG ATC TAC CCG-----TTC TGA
 Met/←318→/Arg Gln Ile Glu Ile Tyr Pro/←194aa→/Phe xxx

SUBSTITUTE SHEET





INTERNATIONAL SEARCH REPORT

International Application No PCT/GB 83/00172

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all) ³		
According to International Patent Classification (IPC) or to both National Classification and IPC IPC ³ : C 12 N 15/00; C 12 P 21/02		
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁴		
Classification System	Classification Symbols	
IPC ³	C 12 N; C 12 P	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched ⁵		
III. DOCUMENTS CONSIDERED TO BE RELEVANT ¹⁴		
Category *	Citation of Document, ¹⁶ with Indication, where appropriate, of the relevant passages ¹⁷	Relevant to Claim No. ¹⁸
P, X	EP, A, 0070675 (R.K. CRAIG et al.) 26 January 1983 see the entire document (cited in the application) --	1, 3-7
A	EP, A, 0052002 (G.D. SEARLE) 19 mei 1982 -----	
<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>* Special categories of cited documents: ¹⁶</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="width: 45%;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"&" document member of the same patent family</p> </div> </div>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search ²	Date of Mailing of this International Search Report ¹	
11th October 1983	26 OCT. 1983	
International Searching Authority ¹	Signature of Authorized Officer ²⁰	
EUROPEAN PATENT OFFICE	G.L.M. Kruydenberg	

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON

INTERNATIONAL APPLICATION NO. PCT/GB 83/00172 (SA 5474)

This Annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 19/10/83

The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A- 0070675	26/01/83	EP-A- 0070186	19/01/83
		WO-A- 8300327	03/02/83
		WO-A- 8300346	03/02/83
		AU-A- 8589282	17/03/83
		AU-A- 8589382	17/03/83
EP-A- 0052002	19/05/82	AU-A- 7731281	20/05/82
		JP-A- 57163396	07/10/82