United States Patent [19]

Inouye et al.

[11] **Patent Number: 4,863,855**

[45] **Date of Patent: * Sep. 5, 1989**

[54] **NOVEL CLONING VEHICLES FOR POLYPEPTIDE EXPRESSION IN MICROBIAL HOSTS**

[75] Inventors: **Masayori Inouye**, Setauket, N.Y.; **Kenzo Nakamura**, Machida, Japan; **Yoshihiro Masui**, Port Jefferson Station, N.Y.

[73] Assignee: **The Research Foundation of State University of New York**, Albany, N.Y.

[*] Notice: The portion of the term of this patent subsequent to Feb. 17, 2004 has been disclaimed.

[21] Appl. No.: **378,481**

[22] Filed: **May 14, 1982**

[51] **Int. Cl.**[4] **C12P 21/00**; C12N 1/20; C12N 15/00

[52] **U.S. Cl.** **435/68**; 435/252.3; 435/252.33; 435/320; 435/172.3; 935/38; 935/39; 935/43; 935/44; 935/66; 935/72; 935/73

[58] **Field of Search** 435/68, 70, 91, 172, 435/253, 317, 172.3, 172.1, 320, 252.3, 252.31–252.35, 66, 72, 73; 935/38, 39, 41, 43

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,237,224	12/1980	Cohen et al.	435/68
4,643,969	2/1987	Inouye et al.	435/68
4,666,836	5/1987	Inouye et al.	435/172.3
4,757,013	7/1988	Inouye et al.	435/172.3

FOREIGN PATENT DOCUMENTS

0001929	5/1979	European Pat. Off. .

OTHER PUBLICATIONS

DeBoer et al: in *Promoters, Structure and Function,* Rodriguez et al (eds.), Praeger Special Studies, New York, 1982, pp. 462–481.
Hamming et al: Nucleic Acids Res. 7, 1019 (1979).
Nakamura et al: Chem. Abstr. 92:105971v (1980) of Cell 18, 1109 (1979).
Nakamura et al: J. Bacteriol. 138, 715 (1979).
Lee et al: J. Bacteriol. 146, 861 (1981).
Davis et al: Nature 283, 433 (1980).
Hall et al: Ann. Rev. Genet. 15, 91 (1981).
Goeddel et al., *Nature,* vol. 281, pp. 544–548, Oct. 1979.
Seeburg et al., *Nature,* vol. 276, pp. 795–798, Dec. 1978.
Villa-Komaroff et al., *Proc. Nat. Acad. Sci. U.S.A.,* vol. 75, pp. 3727–3731, Aug. 1978.
Tacon et al., *Molec. gen. Genet.,* vol. 177, pp. 427–438, 1980.

(List continued on next page.)

Primary Examiner—James Martinell
Attorney, Agent, or Firm—Gottlieb, Rackman & Reisman

[57] **ABSTRACT**

Methods and compositions are provided for regulated expression of polypeptides in transformed bacterial hosts. A novel class of plasmid cloning vehicles includes a DNA sequence coding for the desired polypeptide (or an insertion site therefor) linked for transcriptional expression in reading phase with one or more functional fragments derived from an outer membrane protein gene of a Gram-negative bacterium. The plasmids also include an inducible promoter sequence positioned in the proper orientation for transcriptional expression of the desired polypeptide, as well as a separate DNA sequence coding for a repressor molecule which can interact with the inducible promoter to prevent transcription therefrom. Expression of the desired polypeptide is under the control of both the constitutive promoter and the inducible promoter, although transcription from either promoter is normally blocked by the repressor molecule. However, the repressor can be selectively inactivated by means of an inducer molecule to permit transcriptional expression of the desired polypeptide from both promoters. The methods utilize such plasmids to introduce genetic capability into microorganisms for the production of proteins, such as medically or commercially useful hormones, enzymes, immunogenic proteins, or intermediates therefor, but only in the presence of an appropriate inducer.

40 Claims, 30 Drawing Sheets

OTHER PUBLICATIONS

Charney et al., *Nucleic Acids Research,* vol. 5, pp. 4479–4494, Nov. 1978.
Hallewell et al., *Gene,* vol. 9, pp. 27–47, 1980.
Henning et al., *Proc. Nat. Acad. Sci. U.S.A.,* vol. 76, No. 9, pp. 4360–4364, Sep. 1979.
Sato et al., *J. Bacteriol.,* vol. 139, No. 2, pp. 468–477, Aug. 1979.
Movva et al., *Proc. Nat. Acad. Sci. U.S.A.,* vol. 77, No. 7, pp. 3845–3849, Jul. 1980.
Talmadge et al., *Gene,* vol. 12, pp. 235–241, 1980.
Shimatake et al., *Nature,* vol. 292, pp. 128–132, Jul. 1981.
Gray, et al., *Nature,* vol. 295, pp. 503–508, Feb. 1982.

FIG. 1A

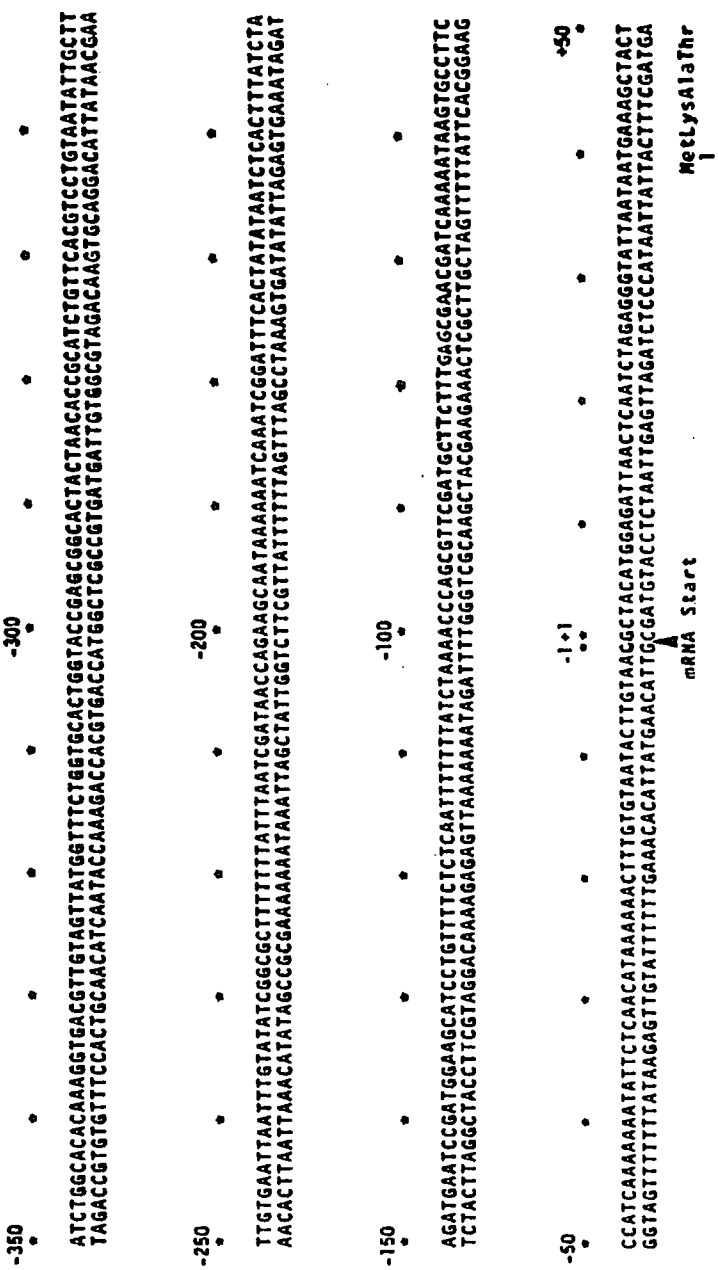
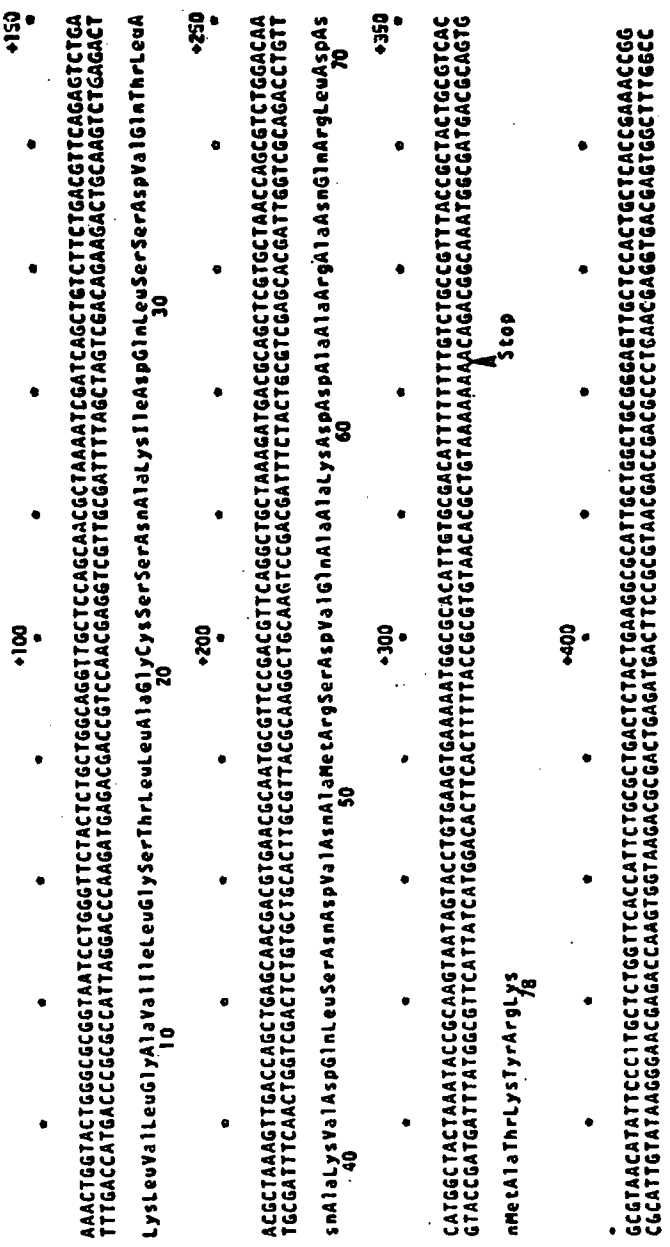

TGGCTGTGCAGAGCA
ACCGAGACGGTCTCGT
-350
-300
ATCTGGCACACAAAGGTGACGTTGTAGTTATGGTTCTGGTGCACTGGTACCGAGCGGCACTACTAACACCGCATCTGTTCAGGTCCTGTAATATTGCTT
TAGACCGTGTGTTTCCACTGCAACATCAATACCAAAGACCACGTGACCATGGCTCGCCGTGATGAATTGTGGCGTAGACAAGTGCAGGACATTATAACGAA
-250
-200
TTGTGAATTAATTTGTATATCGGCGCTTTTTTATTTAATCGATAACCAGAAGCAATAAAAAATCAAATCGGATTTCACTATATAATCTCACTTTATCTA
AACACTTAATTAAACATATAGCCGCGAAAAAAATAAATTAGCTATTGGTCTTCGTTATTTTTTAGTTTAGCCTAAAGTGATATATTAGAGTGAAATAGAT
-150
-100
AGATGAATCCGATGGAAGCATCCTGTTTTCTCTCAATTTTTTTATCTAAAACCCAGCGTTCGATGCTTCTTTGAGCGAACGATCAAAAATAAGTGCCTTC
TCTACTTAGGCTACCTTCGTAGGACAAAAGAGAGTTAAAAAAATAGATTTTGGGTCGCAAGCTACGAAGAAACTCGCTTAGTTTTTATTTCACGGAAG
-50
-1
+1
+50
CCATCAAAAAATATTCTCAACATAAAAAACTTTGTGTAATACTTGTAAGGCTACATGGAGATTAACTCAATCTAGAGGGTATTAATAATGAAAGCTACT
GGTAGTTTTTTTATAAGAGTTGTATATTTTTTGAAACACATTATGAACATTGCGATGTACCTCTAATTGAGTTAGATCTCCCATAATTATTACTTTCGATGA
mRNA Start
MetLysAlaThr
1

FIG. 1B


+100
+150
+200
+250
+300
+350
+400
LysLeuValLeuGlyAlaValIleLeuGlySerThrLeuLeuAlaGlyCysSerSerAsnAlaLysIleAspGlnLeuSerAspValGlnThrLeuA
10 20 30
snAlaLysValAspGlnLeuSerAsnAspValAsnAlaMetArgSerAspValGlnAlaAlaLysAspAspAlaAlaArgAlaAspAlaAsnGlnArgLeuAspAs
40 50 60 70
nMetAlaThrLysTyrArgLys
Stop

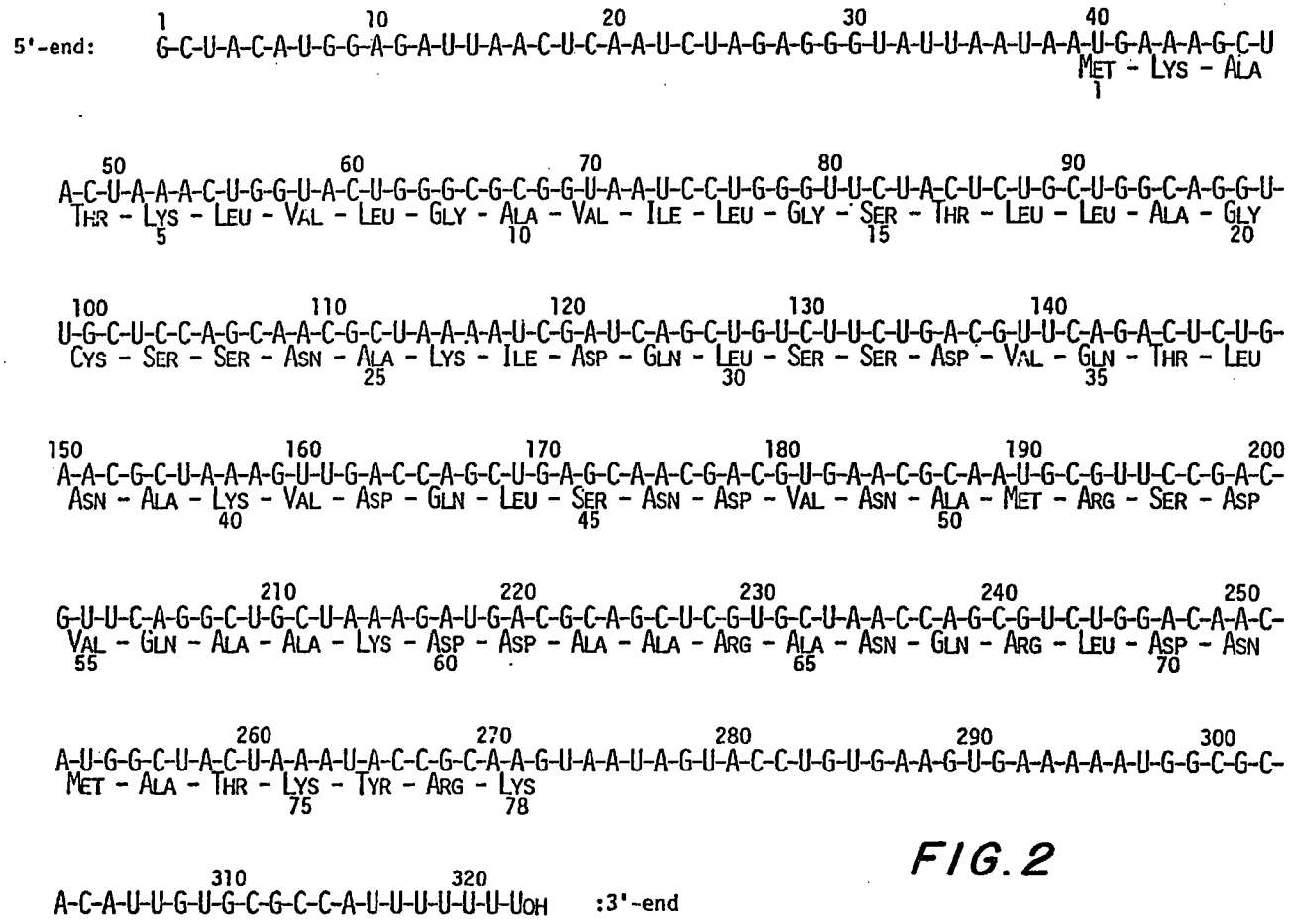

5'-end:
G-C-U-A-C-A-U-G-G-A-G-A-U-U-A-A-C-U-C-A-A-U-C-U-A-G-A-G-G-G-U-A-U-U-A-A-U-A-A-U-G-A-A-A-G-C-U
MET - LYS - ALA
A-C-U-A-A-A-C-U-G-G-U-A-C-U-G-G-G-C-G-C-G-G-U-A-A-U-C-C-U-G-G-G-U-U-C-U-A-C-U-C-U-G-C-U-G-G-C-A-G-G-U-
THR - LYS - LEU - VAL - LEU - GLY - ALA - VAL - ILE - LEU - GLY - SER - THR - LEU - LEU - ALA - GLY
U-G-C-U-C-C-A-G-C-A-A-C-G-C-U-A-A-A-A-U-C-G-A-U-C-A-G-C-U-G-U-C-U-U-C-U-G-A-C-G-U-U-C-A-G-A-C-U-C-U-G-
CYS - SER - SER - ASN - ALA - LYS - ILE - ASP - GLN - LEU - SER - SER - ASP - VAL - GLN - THR - LEU
A-A-C-G-C-U-A-A-A-G-U-U-G-A-C-C-A-G-C-U-G-A-G-C-A-A-C-G-A-C-G-U-G-A-A-C-G-C-A-A-U-G-C-G-U-U-C-C-G-A-C-
ASN - ALA - LYS - VAL - ASP - GLN - LEU - SER - ASN - ASP - VAL - ASN - ALA - MET - ARG - SER - ASP
G-U-U-C-A-G-G-C-U-G-C-U-A-A-A-G-A-U-G-A-C-G-C-A-G-C-U-C-G-U-G-C-U-A-A-C-C-A-G-C-G-U-C-U-G-G-A-C-A-A-C-
VAL - GLN - ALA - ALA - LYS - ASP - ASP - ALA - ALA - ARG - ALA - ASN - GLN - ARG - LEU - ASP - ASN
A-U-G-G-C-U-A-C-U-A-A-A-U-A-C-C-G-C-A-A-G-U-A-A-U-A-G-U-A-C-C-U-G-U-G-A-A-G-U-G-A-A-A-A-A-U-G-G-C-G-C-
MET - ALA - THR - LYS - TYR - ARG - LYS
A-C-A-U-U-G-U-G-C-G-C-C-A-U-U-U-U-U-U-UOH :3'-end

FIG. 2

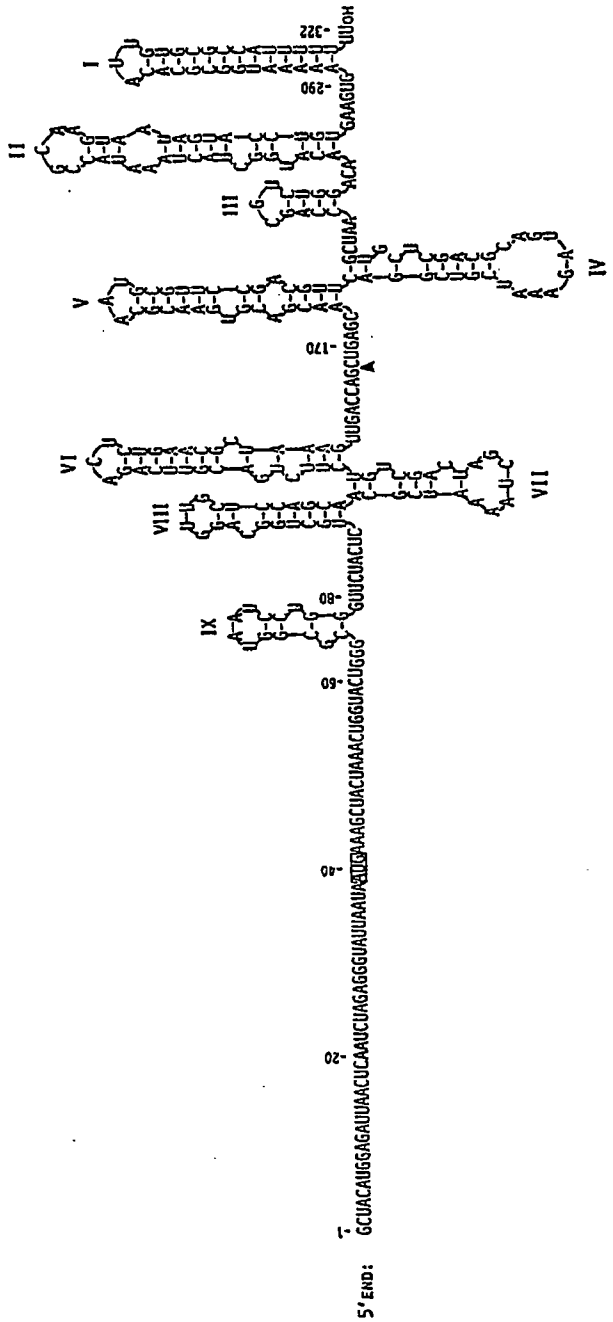

I
II
III
IV
V
VI
VII
VIII
IX
XI
-322
-290
-170
-80
-60
-40
-20
1
5' END:
GCUACAUGGAGAUUAACUCAAUCUAGAGGGUAUUAAUAGUUAAGCUACUAAACUGGUACUGGGGUUCUACUC
UUGACCAGCUGAGC
GCUAA
ACA
GAAGUG
UUOH

FIG. 3

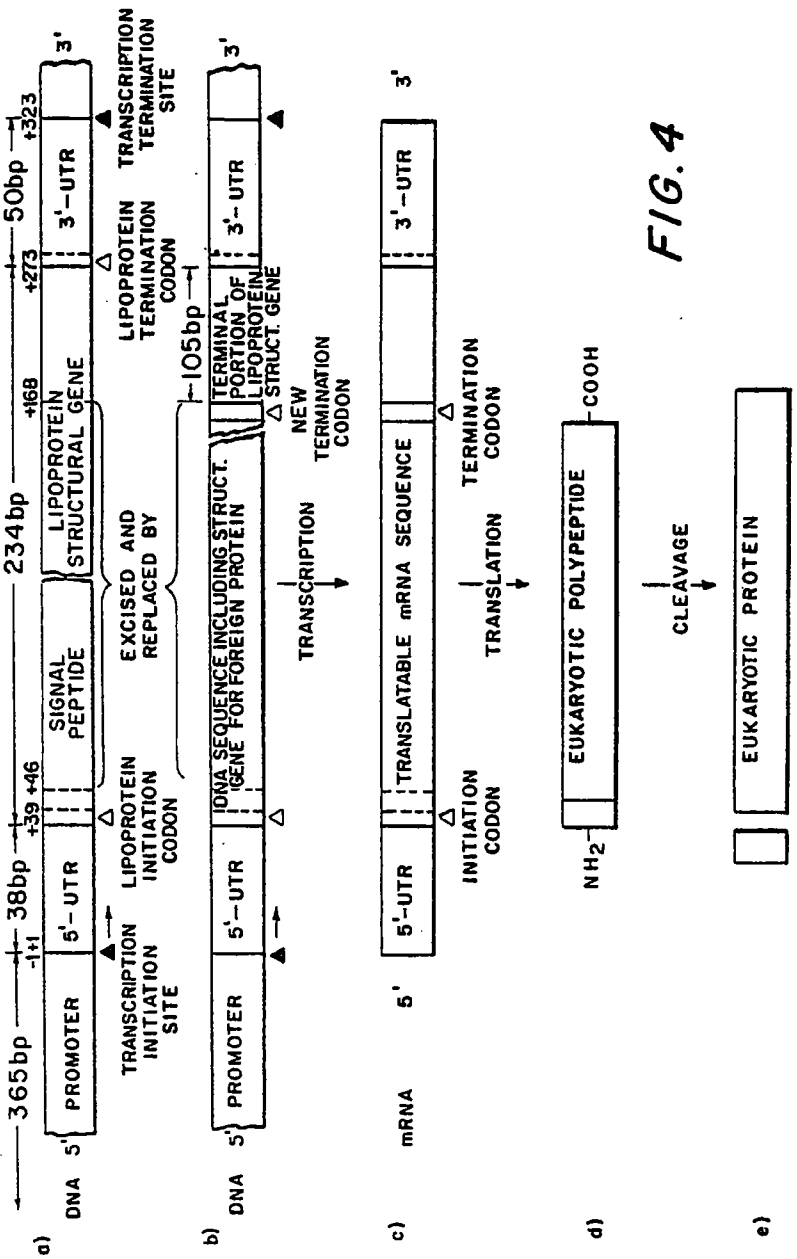

a)
DNA
5'
PROMOTER
5'-UTR
SIGNAL PEPTIDE
LIPOPROTEIN STRUCTURAL GENE
3'-UTR
3'
365bp
38bp
234bp
50bp
-1 +1
+39 +46
+168
+273
+323
TRANSCRIPTION INITIATION SITE
LIPOPROTEIN INITIATION CODON
EXCISED AND REPLACED BY
LIPOPROTEIN TERMINATION CODON
TRANSCRIPTION TERMINATION SITE
b)
DNA
5'
PROMOTER
5'-UTR
DNA SEQUENCE INCLUDING STRUCT. GENE FOR FOREIGN PROTEIN
105bp
TERMINAL PORTION OF LIPOPROTEIN STRUCT. GENE
3'-UTR
3'
NEW TERMINATION CODON
TRANSCRIPTION
c)
mRNA
5'
5'-UTR
TRANSLATABLE mRNA SEQUENCE
3'-UTR
3'
INITIATION CODON
TRANSLATION
TERMINATION CODON
d)
NH2
EUKARYOTIC POLYPEPTIDE
COOH
CLEAVAGE
e)
EUKARYOTIC PROTEIN

FIG. 4

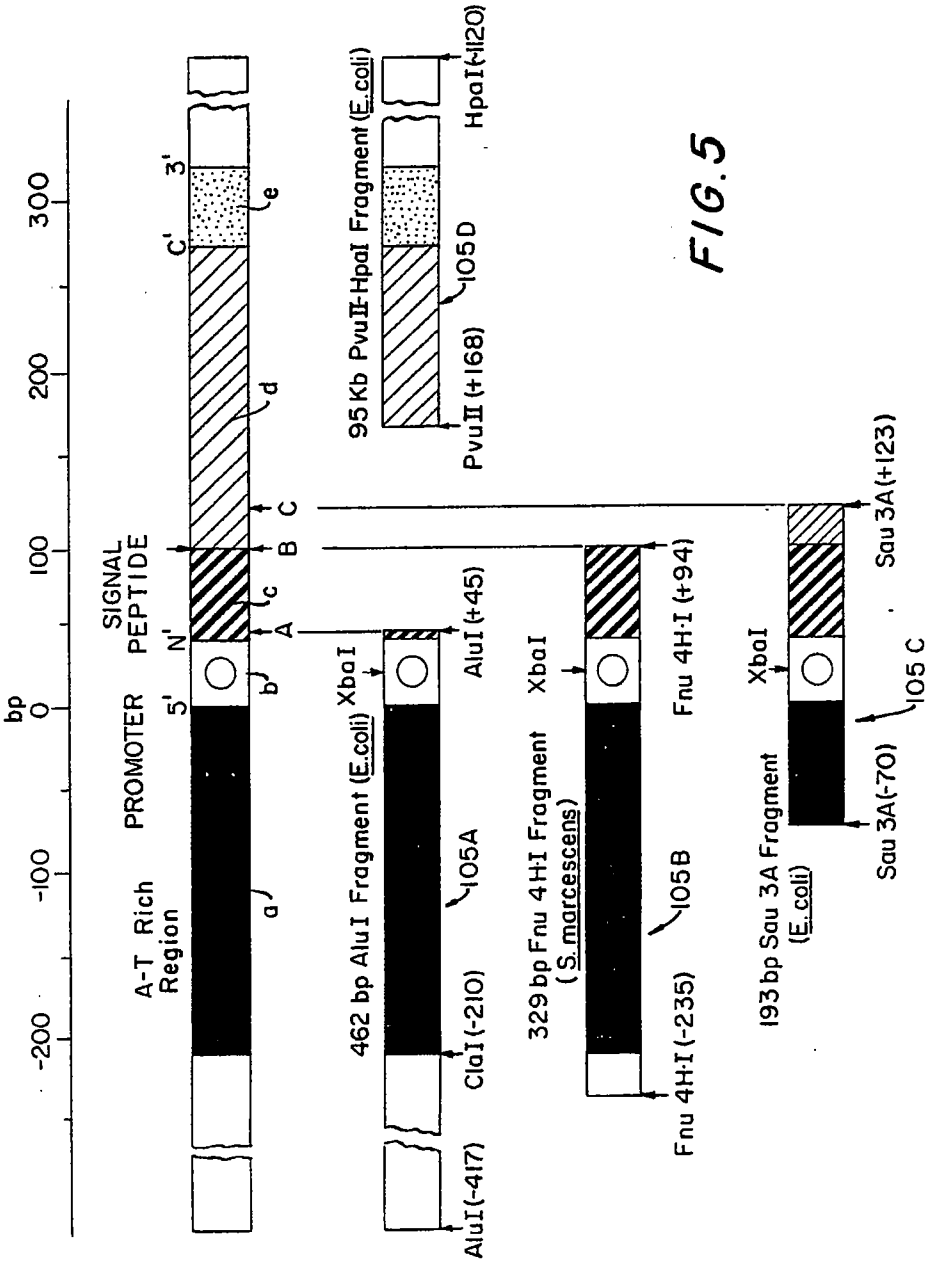

bp
300
200
100
0
-100
-200
A-T Rich Region
PROMOTER
SIGNAL PEPTIDE
5'
N'
C'
3'
a
b
c
d
e
A
B
C
105 C
Sau 3A(+123)
Sau 3A(-70)
193 bp Sau 3A Fragment (E. coli)
XbaI
105B
329 bp Fnu 4H·I Fragment (S. marcescens)
Fnu 4H·I (+94)
Fnu 4H·I (-235)
105A
462 bp AluI Fragment (E.coli)
AluI (+45)
ClaI (-210)
AluI (-417)
105D
95 Kb PvuII-HpaI Fragment (E.coli)
HpaI (+120)
PvuII (+168)

FIG. 5

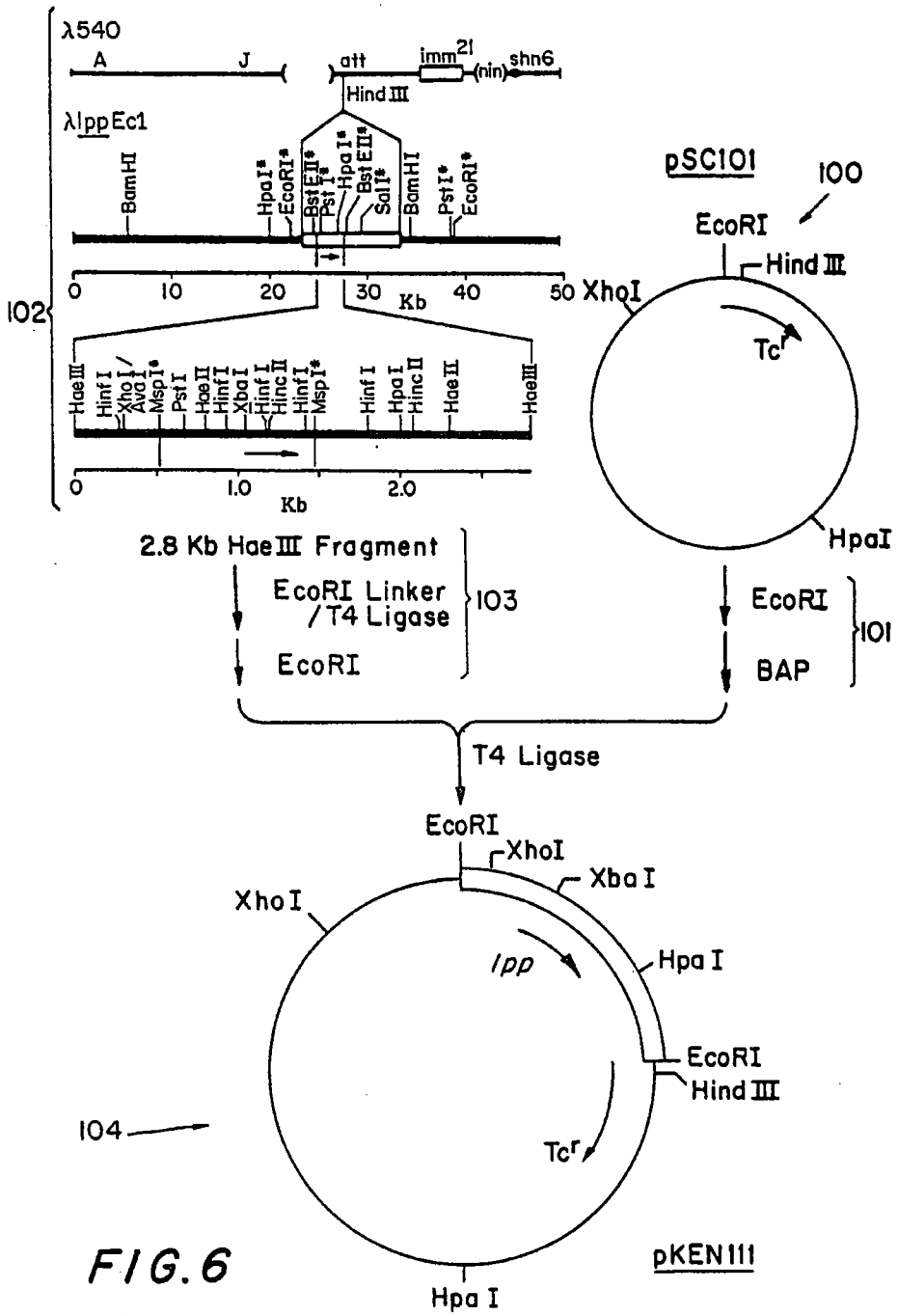
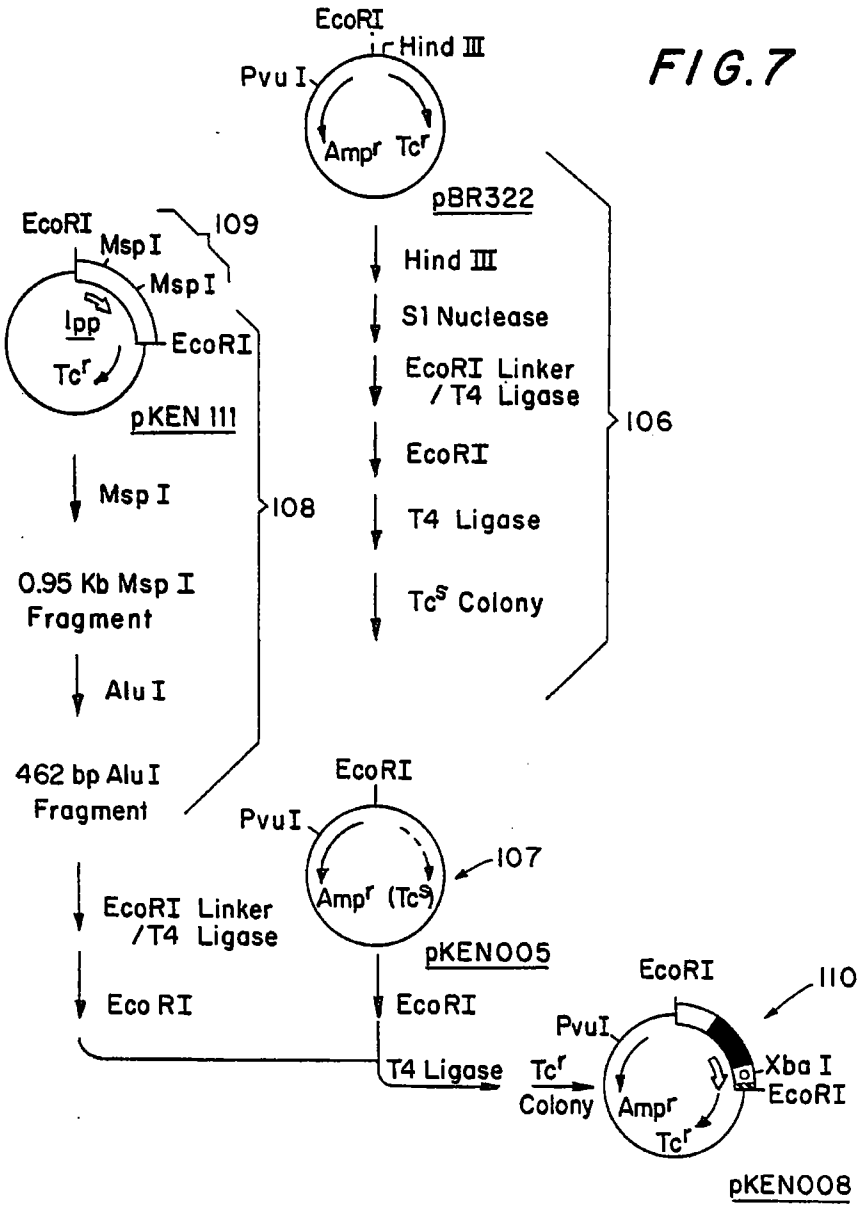

λ540
A
J
att
imm21
(nin)
shn6
Hind III
λlppEc1
BamHI
HpaI*
EcoRI*
BstEII*
PstI*
HpaI*
BstEII*
SalI*
BamHI
PstI*
EcoRI*
0
10
20
30
Kb
40
50
102
HaeIII
HinfI
XhoI/
AvaI*
MspI*
PstI
HaeII
HinfI
XbaI
HinfI
HincII
HinfI
MspI*
HinfI
HpaI
HincII
HaeII
HaeIII
0
1.0
Kb
2.0
pSC101
100
EcoRI
Hind III
XhoI
Tc^r
HpaI
2.8 Kb HaeIII Fragment
EcoRI Linker
/T4 Ligase
103
EcoRI
EcoRI
BAP
101
T4 Ligase
EcoRI
XhoI
Xba I
Xho I
lpp
Hpa I
EcoRI
Hind III
104
Tc^r
Hpa I
pKEN111

FIG. 6

FIG. 7


EcoRI
Hind III
Pvu I
Amp^r Tc^r
pBR322
Hind III
S1 Nuclease
EcoRI Linker
/ T4 Ligase
EcoRI
T4 Ligase
Tc^s Colony
106
EcoRI
Msp I
Msp I
109
lpp
Tc^r
EcoRI
pKEN 111
Msp I
108
0.95 Kb Msp I
Fragment
Alu I
462 bp Alu I
Fragment
EcoRI
PvuI
Amp^r (Tc^s)
107
pKEN005
EcoRI Linker
/T4 Ligase
Eco RI
EcoRI
T4 Ligase
EcoRI
110
PvuI
Tc^r
Colony
Amp^r
Tc^r
Xba I
EcoRI
pKEN008

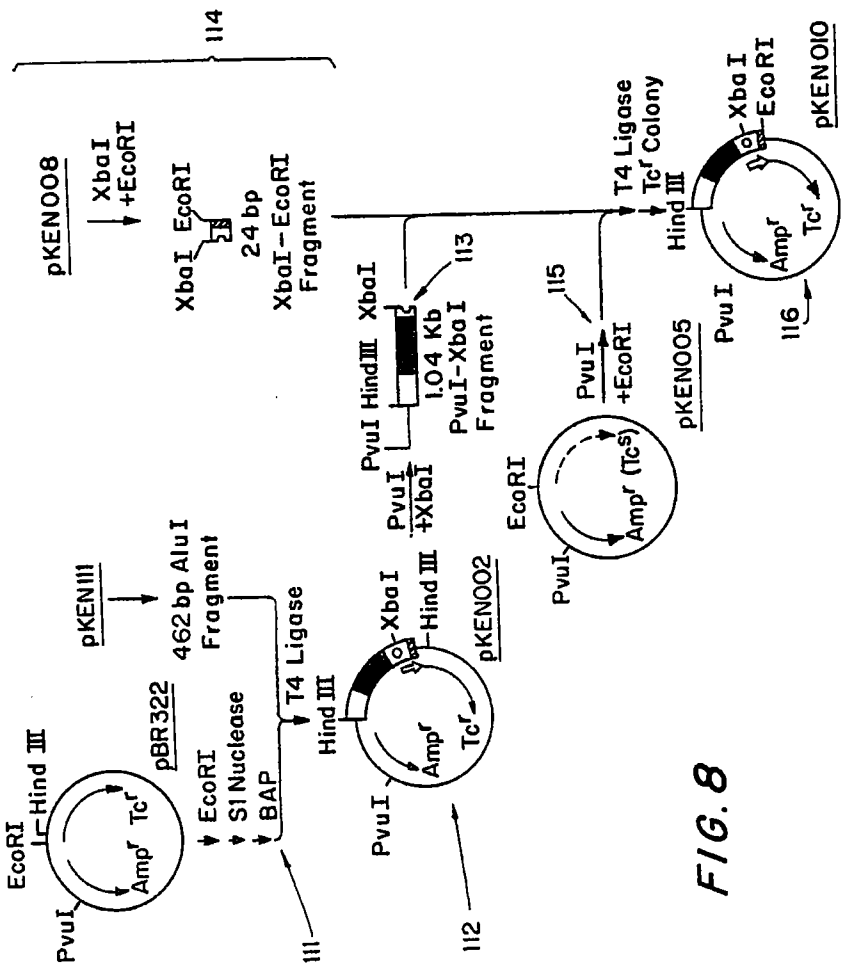
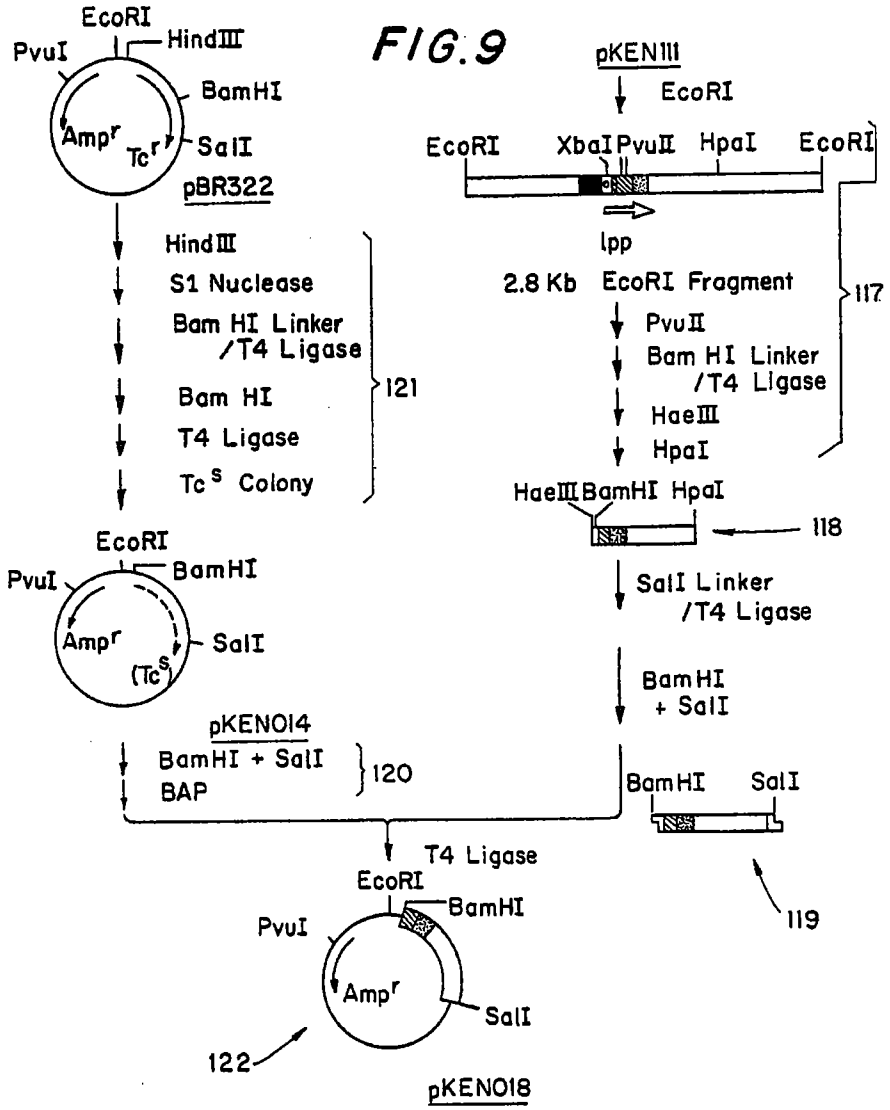

pKEN008
Xba I
+EcoRI
Xba I EcoRI
24 bp
Xba I-EcoRI
Fragment
114
pKEN111
462 bp Alu I
Fragment
EcoRI
Hind III
pBR322
Amp^r Tc^r
PvuI
EcoRI
S1 Nuclease
BAP
T4 Ligase
Hind III
111
Xba I
Hind III
pKEN002
Amp^r
Tc^r
PvuI
112
PvuI Hind III Xba I
1.04 Kb
PvuI-Xba I
Fragment
113
PvuI
+Xba I
EcoRI
Amp^r (Tc^s)
PvuI
115
PvuI
+EcoRI
pKEN005
T4 Ligase
Tc^r Colony
Hind III
Xba I
EcoRI
pKEN010
Amp^r
Tc^r
PvuI
116

FIG. 8

FIG. 9


EcoRI
PvuI
HindIII
BamHI
Amp^r
Tc^r
SalI
pBR322
HindIII
S1 Nuclease
Bam HI Linker /T4 Ligase
Bam HI
T4 Ligase
Tc^s Colony
121
EcoRI
PvuI
BamHI
Amp^r
(Tc^s)
SalI
pKEN014
BamHI + SalI
BAP
120
pKEN111
EcoRI
EcoRI
XbaI
PvuII
HpaI
EcoRI
lpp
2.8 Kb EcoRI Fragment
117
PvuII
Bam HI Linker /T4 Ligase
HaeIII
HpaI
HaeIII BamHI HpaI
118
SalI Linker /T4 Ligase
BamHI + SalI
BamHI
SalI
119
T4 Ligase
EcoRI
BamHI
PvuI
Amp^r
SalI
122
pKEN018

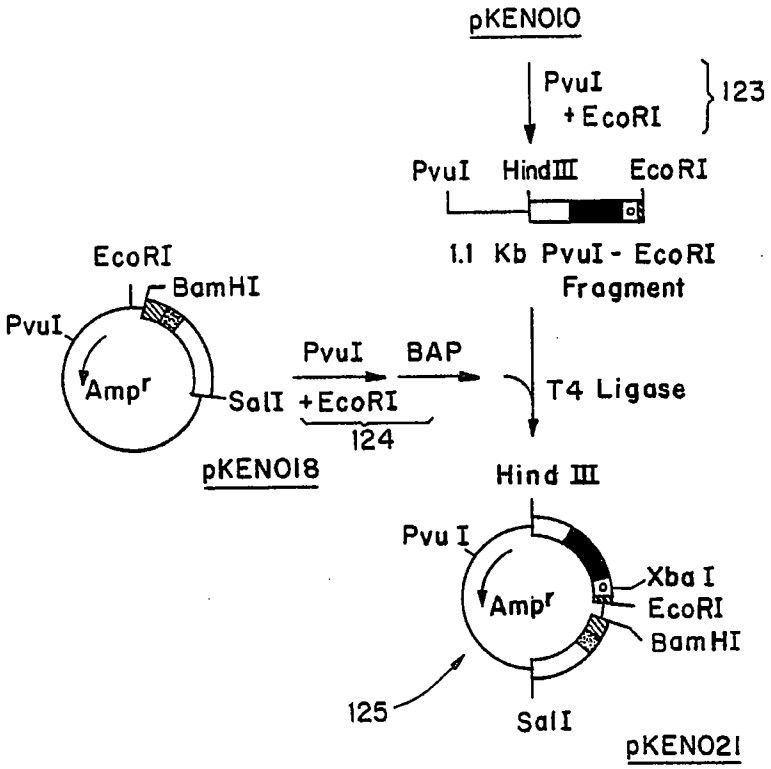
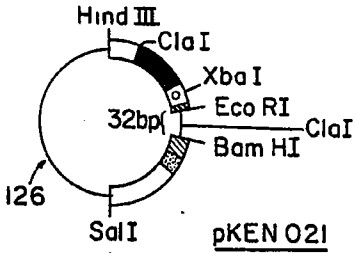
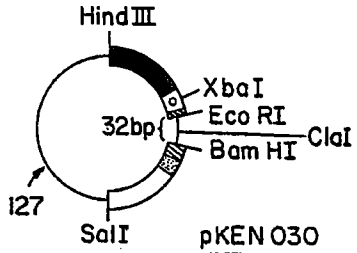

pKENO10
PvuI
+EcoRI
123
PvuI
HindIII
EcoRI
1.1 Kb PvuI - EcoRI
Fragment
EcoRI
BamHI
PvuI
Ampr
SalI
PvuI
+EcoRI
124
BAP
T4 Ligase
pKENO18
Hind III
Pvu I
Ampr
Xba I
EcoRI
BamHI
125
SalI
pKENO21

FIG. 10



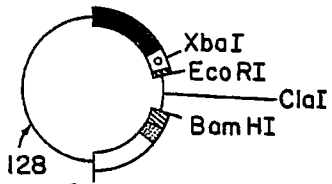
pKEN 021

- Cla I (partial)
- S1 Nuclease
- Hind III Linker / T4 Ligase
- Hind III
- T4 Ligase



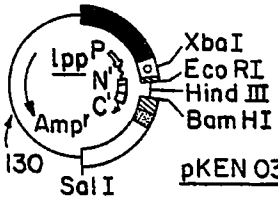
pKEN 030

- Hind III
- S1 Nuclease
- T4 Ligase



pKEN 033

- Eco RI + Bam HI
- T4 DNA Polymerase (4 dNTPs)
- Hind III Linker (p[CCAAGCTTGG]) / T4 Ligase } 129
- Hind III
- T4 Ligase



pKEN 037 (A-1)

FIG. 11

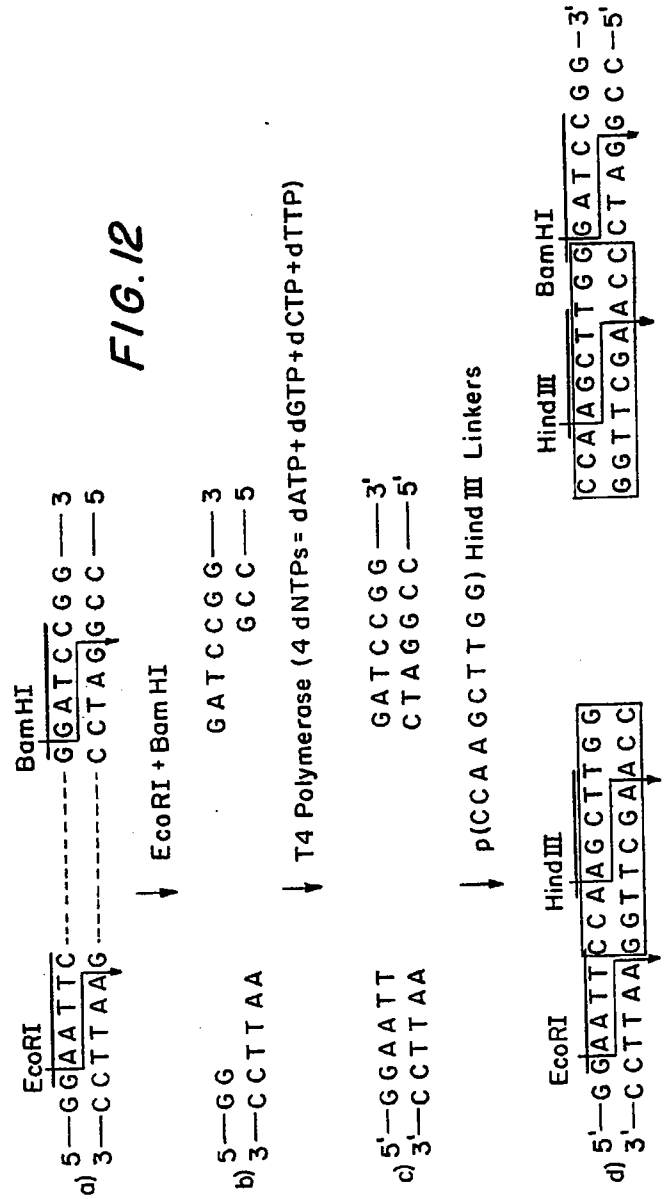

EcoRI
BamHI
a) 5—G G A A T T C ---------- G G A T C C — 3
3—C C T T A A G ---------- C C T A G G — 5
EcoRI + BamHI
b) 5—G G
3—C C T T A A
G A T C C G G — 3
G C C — 5
T4 Polymerase (4 dNTPs = dATP + dGTP + dCTP + dTTP)
c) 5'—G G A A T T
3'—C C T T A A
G A T C C G G —3'
C T A G G C C —5'
p(C C A A G C T T G G) HindIII Linkers
HindIII
BamHI
C C A A G C T T G G G A T C C G G —3'
G G T T C G A A C C C T A G G C C —5'
EcoRI
HindIII
d) 5'—G G A A T T C C A A G C T T G G
3'—C C T T A A G G T T C G A A C C
HindIII
T4 Ligase
EcoRI
HindIII
BamHI
e) 5'—G G A A T T C C A A G C T T G G A T C C G G —3'
3'—C C T T A A G G T T C G A A C C T A G G C C —5'

FIG. 12

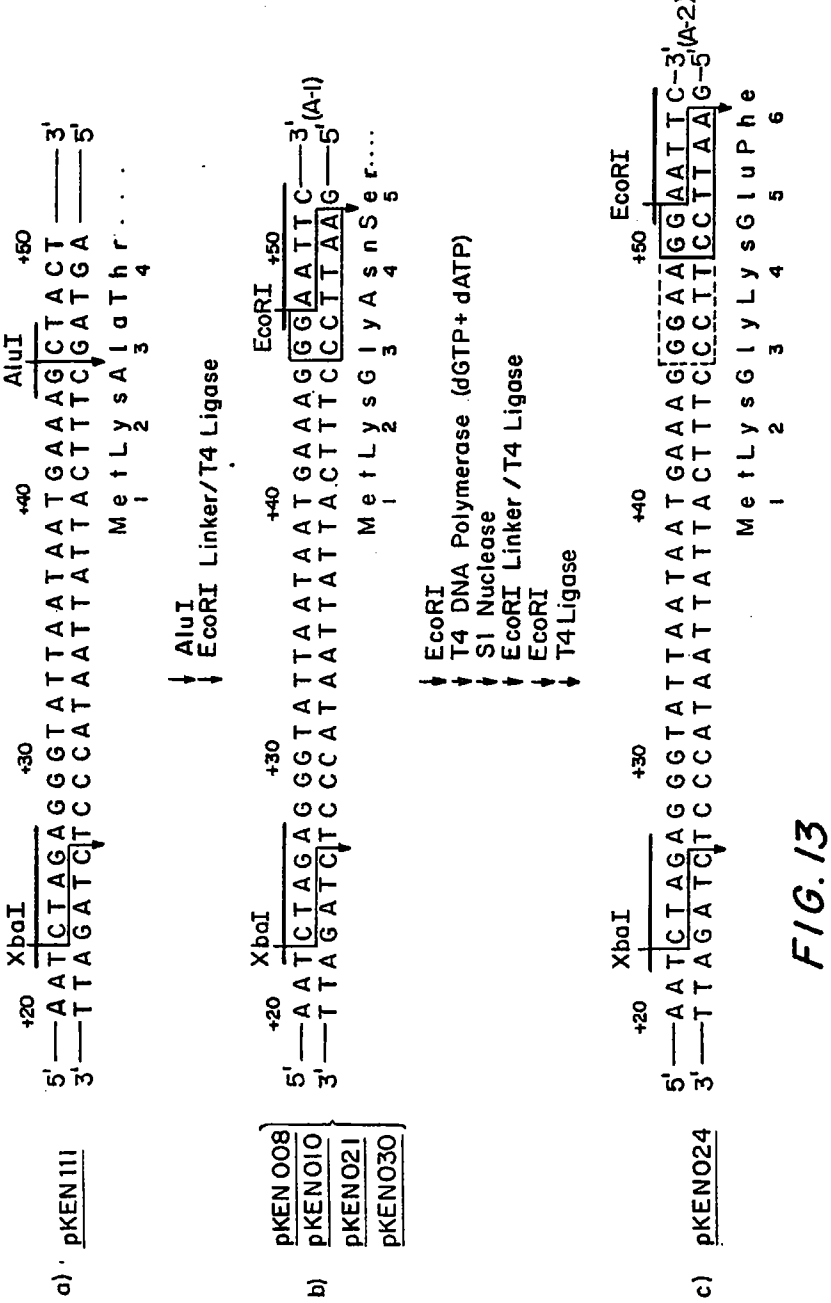

a) pKEN111
b) pKEN008
pKEN010
pKEN021
pKEN030
c) pKEN024
XbaI
AluI
EcoRI
+20
+30
+40
+50
Met Lys Ala Thr....
Met Lys Gly Asn Ser....
Met Lys Gly Lys Gly Glu Phe
AluI
EcoRI Linker/T4 Ligase
EcoRI
T4 DNA Polymerase (dGTP+dATP)
S1 Nuclease
EcoRI Linker/T4 Ligase
EcoRI
T4 Ligase
(A-1)
(A-2)

FIG. 13

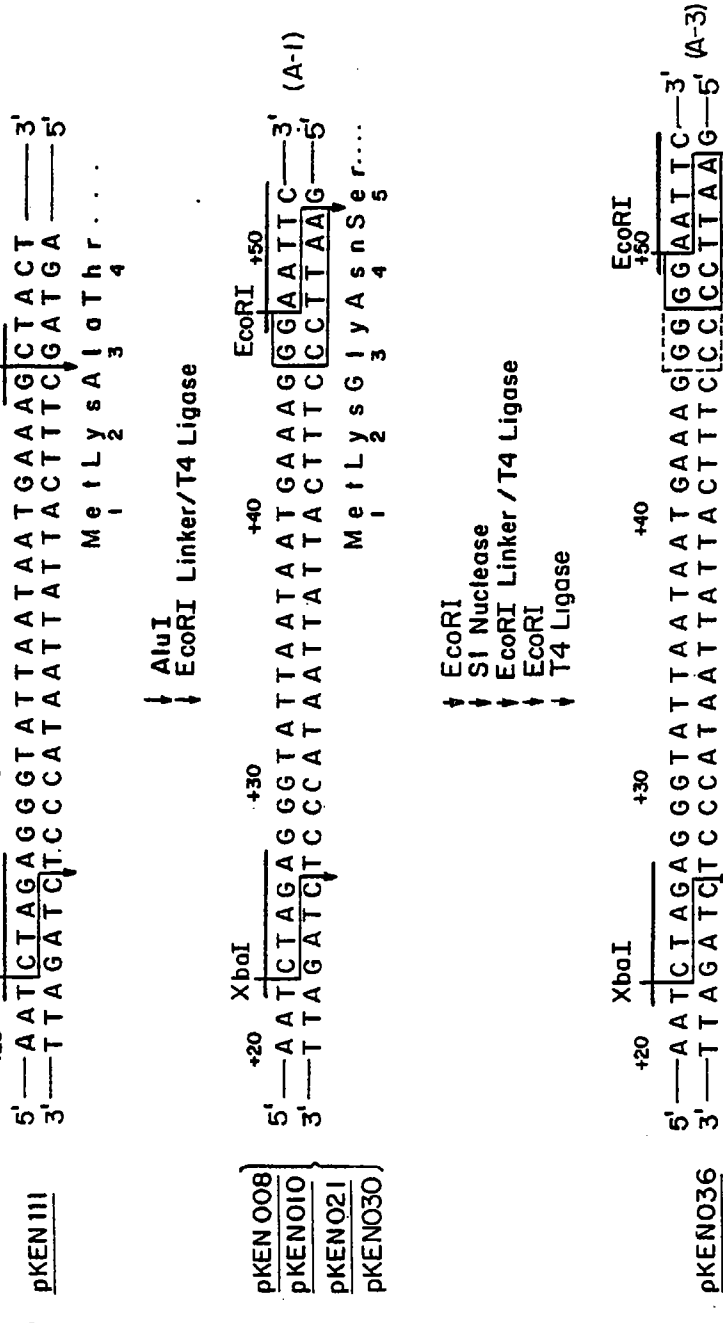

a) pKEN111
XbaI
AluI
+20
+30
+40
+50
5'—A A T C T A G A G G G T A T T A A T A A T G A A A G C T A C T —3'
3'—T T A G A T C T C C C A T A A T T A C T T T C G A T G A —5'
M e t L y s A l a T h r . . . .
1 2 3 4
AluI
EcoRI Linker/T4 Ligase
b)
pKEN008
pKEN010
pKEN021
pKEN030
XbaI
EcoRI
(A-1)
5'—A A T C T A G A G G G T A T T A A T A A T G A A A G G A A T T C—3'
3'—T T A G A T C T C C C A T A A T T A C T T T C C C T T A A G—5'
M e t L y s G l y A s n S e r . . . .
1 2 3 4 5
EcoRI
S1 Nuclease
EcoRI Linker / T4 Ligase
EcoRI
T4 Ligase
c) pKEN036
XbaI
EcoRI
(A-3)
5'—A A T C T A G A G G G T A T T A A T A A T G A A A G G G G A A T T C—3'
3'—T T A G A T C T C C C A T A A T T A C T T T C C C C T T A A G—5'
M e t L y s G l y G l y I l e . . . .
1 2 3 4 5

FIG. 14

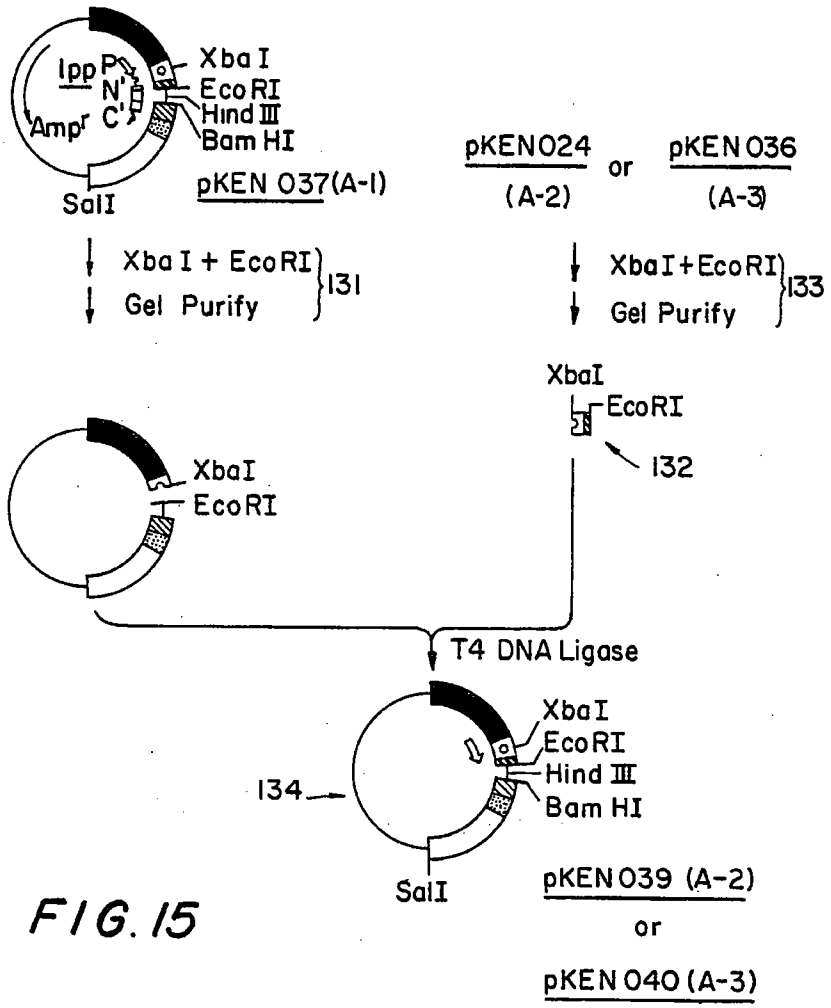

lpp P
N'
Amp^r C'
Xba I
EcoRI
Hind III
Bam HI
SalI
pKEN 037(A-1)
pKEN024 (A-2) or pKEN036 (A-3)
Xba I + EcoRI
Gel Purify
131
XbaI+EcoRI
Gel Purify
133
XbaI
EcoRI
132
XbaI
EcoRI
T4 DNA Ligase
XbaI
EcoRI
Hind III
Bam HI
134
SalI
pKEN039 (A-2)
or
pKEN040 (A-3)

FIG. 15

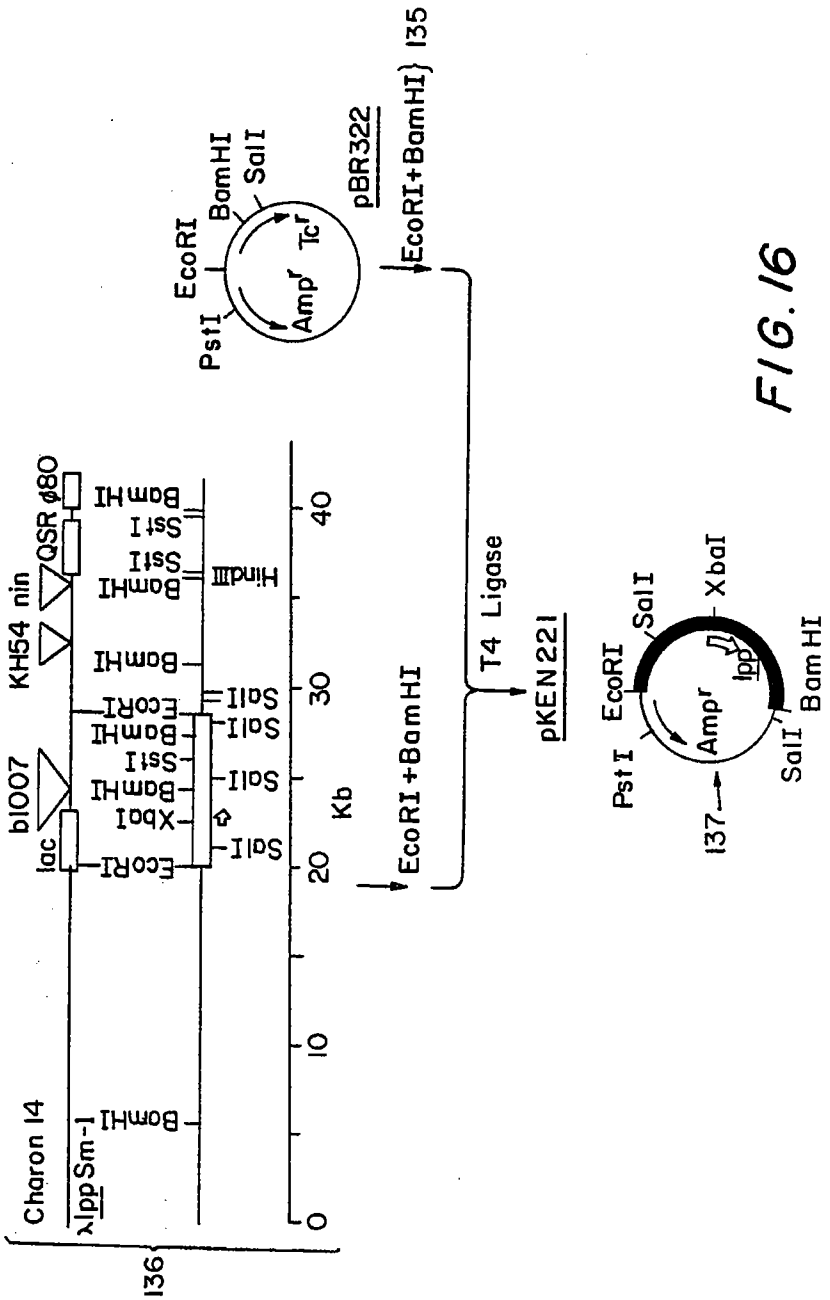

pBR322
EcoRI
BamHI
SalI
PstI
Amp^r
Tc^r
EcoRI+BamHI
135
T4 Ligase
pKEN221
EcoRI+BamHI
Charon 14
λlppSm-1
136
lac
b1007
KH54 nin
QSR ø80
BamHI
SstI
HindIII
EcoRI
SalI
XbaI
0
10
20
30
40
Kb
PstI
EcoRI
SalI
XbaI
lpp
Amp^r
BamHI
137

FIG. 16

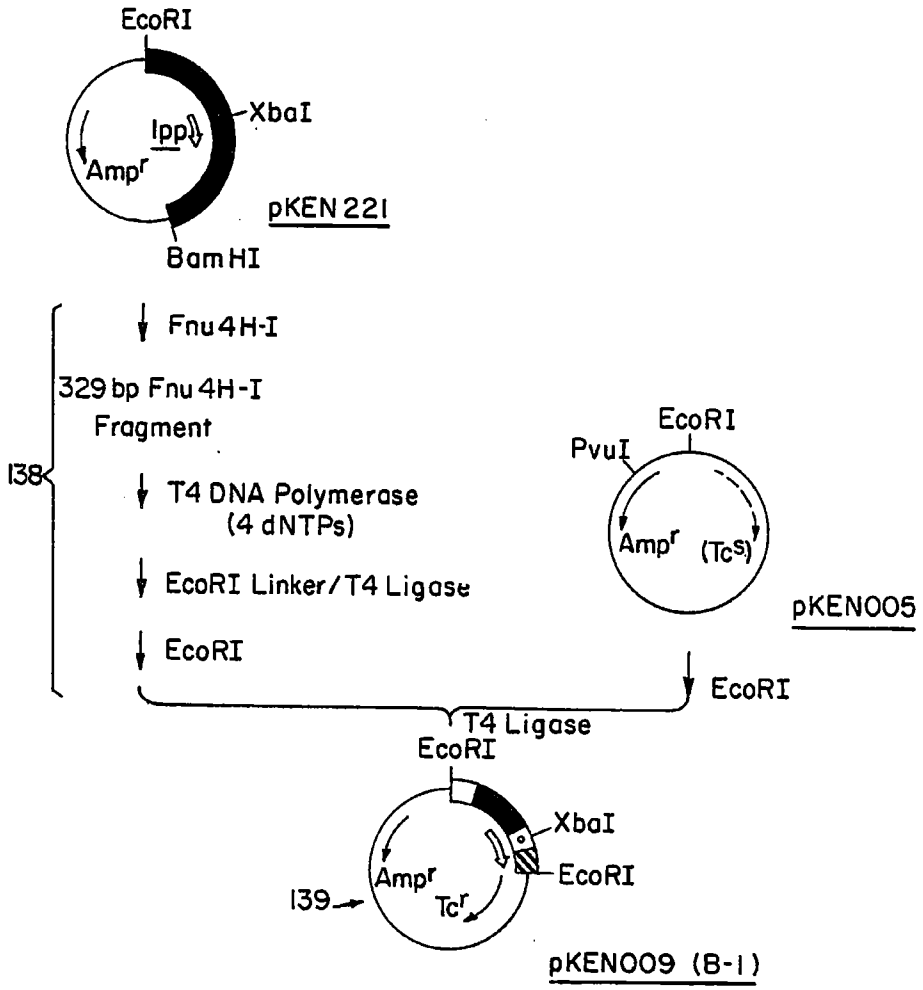

EcoRI
XbaI
lpp
Amp^r
pKEN 221
BamHI
Fnu4H-I
329 bp Fnu 4H-I
Fragment
138
T4 DNA Polymerase
(4 dNTPs)
EcoRI Linker/T4 Ligase
EcoRI
EcoRI
PvuI
Amp^r
(Tc^S)
pKENOO5
EcoRI
T4 Ligase
EcoRI
XbaI
EcoRI
Amp^r
Tc^r
139
pKENOO9 (B-1)

FIG. 17

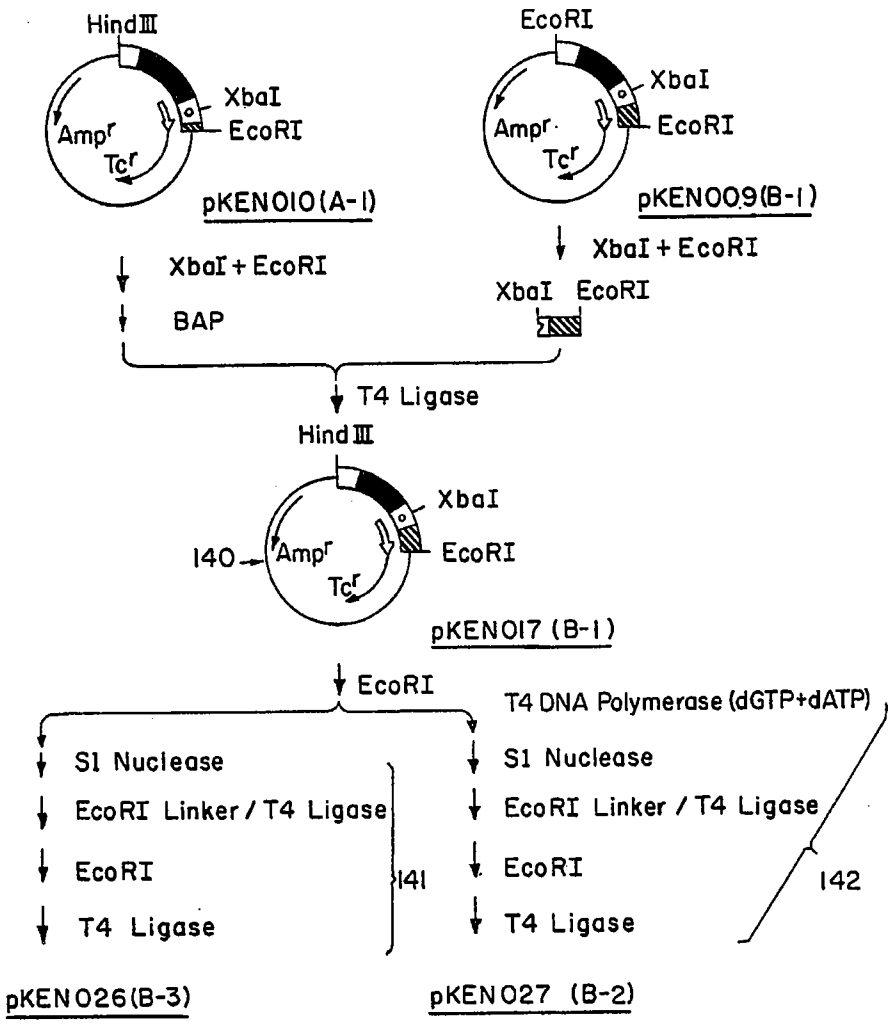

HindIII
XbaI
EcoRI
Ampr
Tcr
pKEN010(A-1)
XbaI + EcoRI
BAP
EcoRI
XbaI
EcoRI
Ampr
Tcr
pKEN009(B-1)
XbaI + EcoRI
XbaI EcoRI
T4 Ligase
HindIII
XbaI
EcoRI
140
Ampr
Tcr
pKEN017 (B-1)
EcoRI
T4 DNA Polymerase (dGTP+dATP)
S1 Nuclease
EcoRI Linker / T4 Ligase
EcoRI
T4 Ligase
141
S1 Nuclease
EcoRI Linker / T4 Ligase
EcoRI
T4 Ligase
142
pKEN026(B-3)
pKEN027 (B-2)

FIG. 18

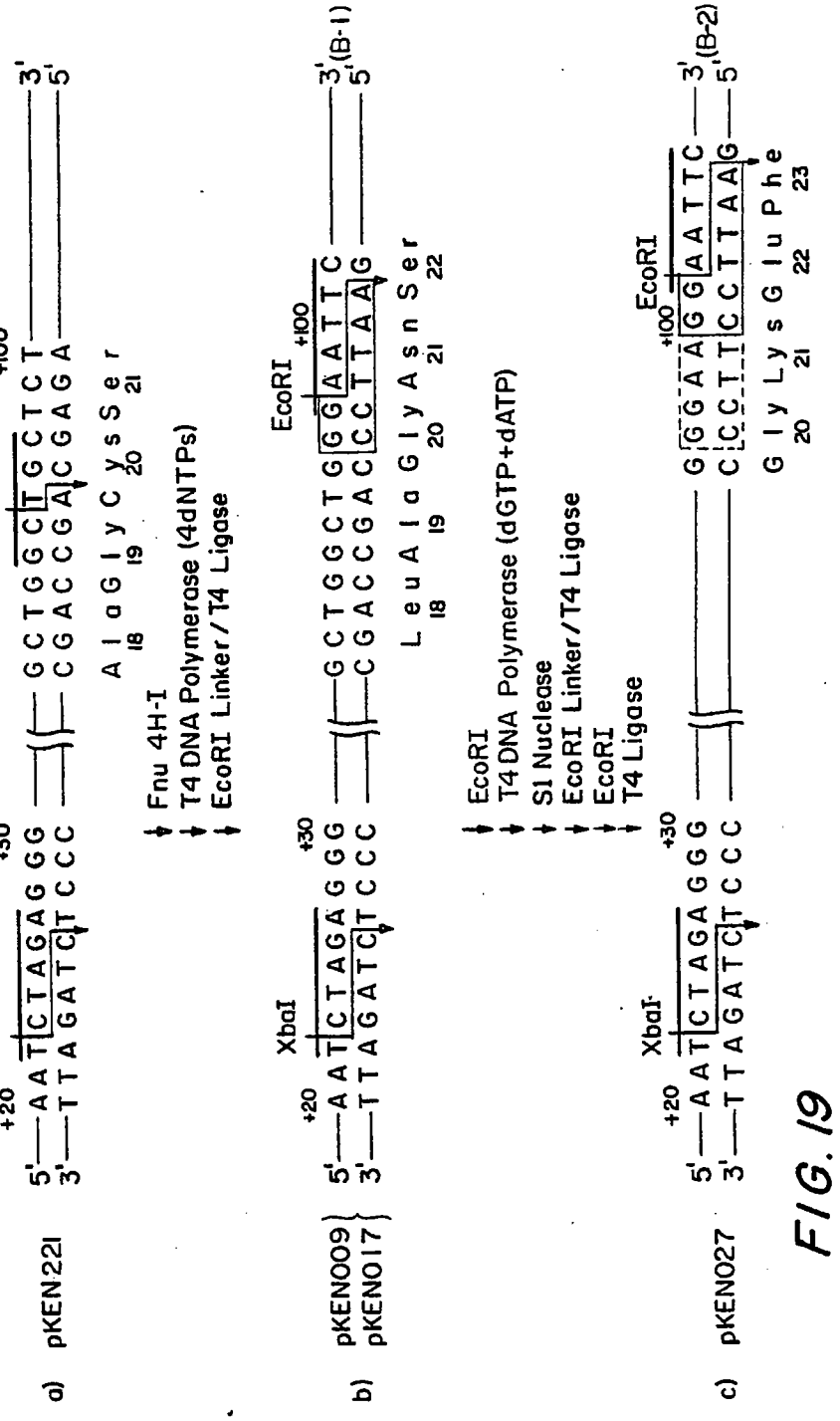

a) pKEN221
XbaI
Fnu4H-I
Fnu 4H-I
T4 DNA Polymerase (4dNTPs)
EcoRI Linker / T4 Ligase
b) pKEN009
pKEN017
EcoRI
T4 DNA Polymerase (dGTP+dATP)
S1 Nuclease
EcoRI Linker / T4 Ligase
EcoRI
T4 Ligase
c) pKEN027
(B-1)
(B-2)

FIG. 19

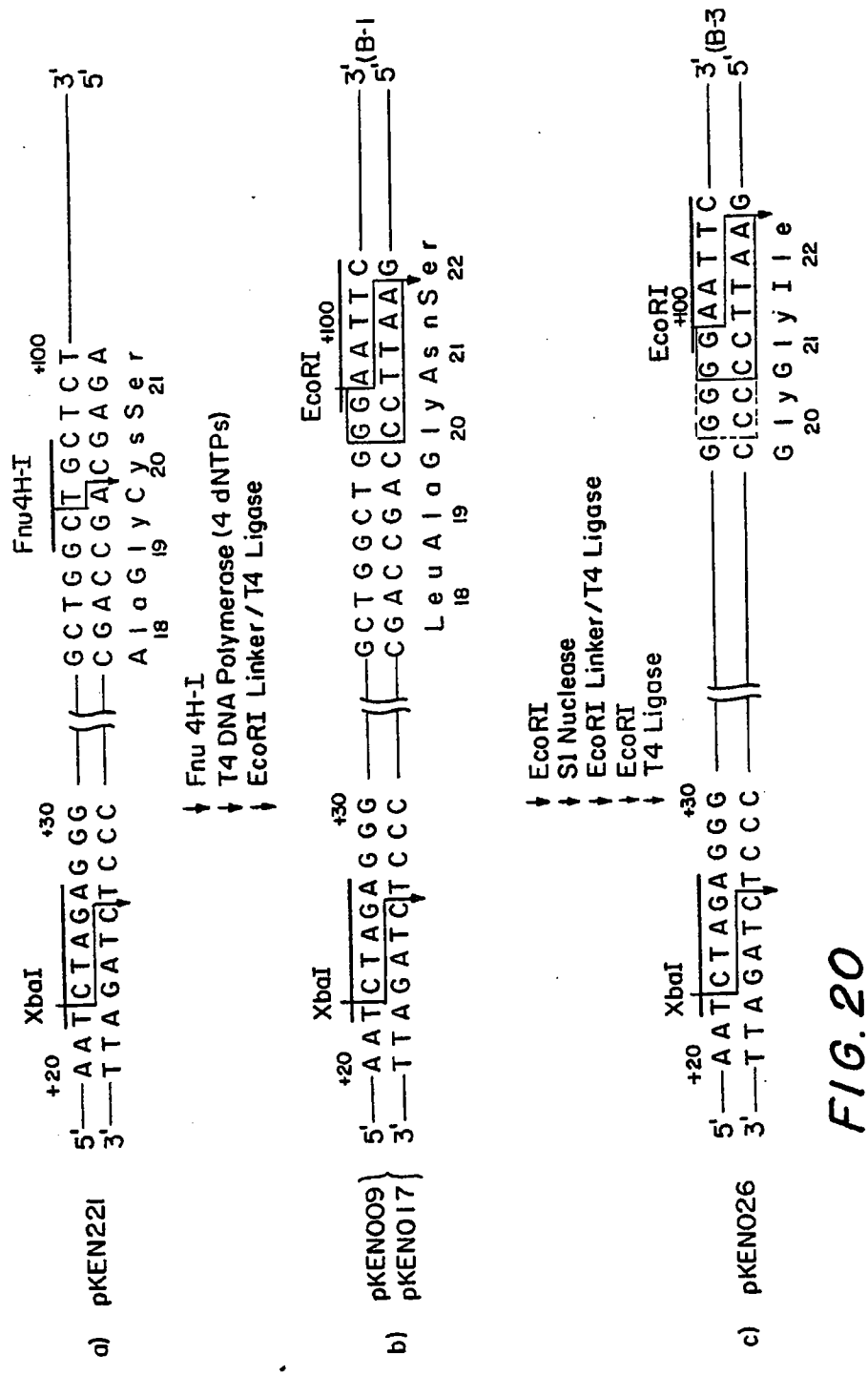

a) pKEN221
XbaI
Fnu4H-I
Fnu 4H-I
T4 DNA Polymerase (4 dNTPs)
EcoRI Linker / T4 Ligase
b) pKEN009
pKEN017
EcoRI
(B-1
EcoRI
S1 Nuclease
EcoRI Linker / T4 Ligase
EcoRI
T4 Ligase
c) pKEN026
(B-3

FIG. 20

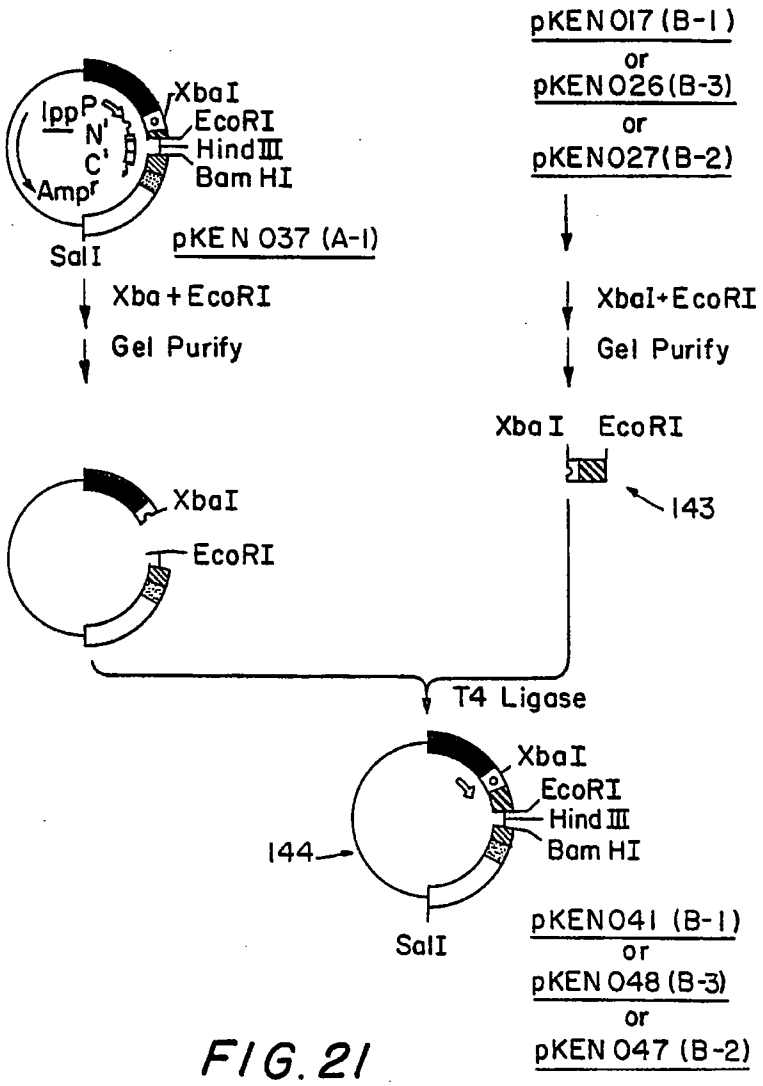

pKEN017 (B-1)
or
pKEN026 (B-3)
or
pKEN027 (B-2)
lpp
P
N'
C'
Amp^r
XbaI
EcoRI
Hind III
Bam HI
SalI
pKEN 037 (A-1)
Xba + EcoRI
Gel Purify
XbaI+EcoRI
Gel Purify
Xba I
EcoRI
143
XbaI
EcoRI
T4 Ligase
144
XbaI
EcoRI
Hind III
Bam HI
SalI
pKEN041 (B-1)
or
pKEN 048 (B-3)
or
pKEN 047 (B-2)

FIG. 21

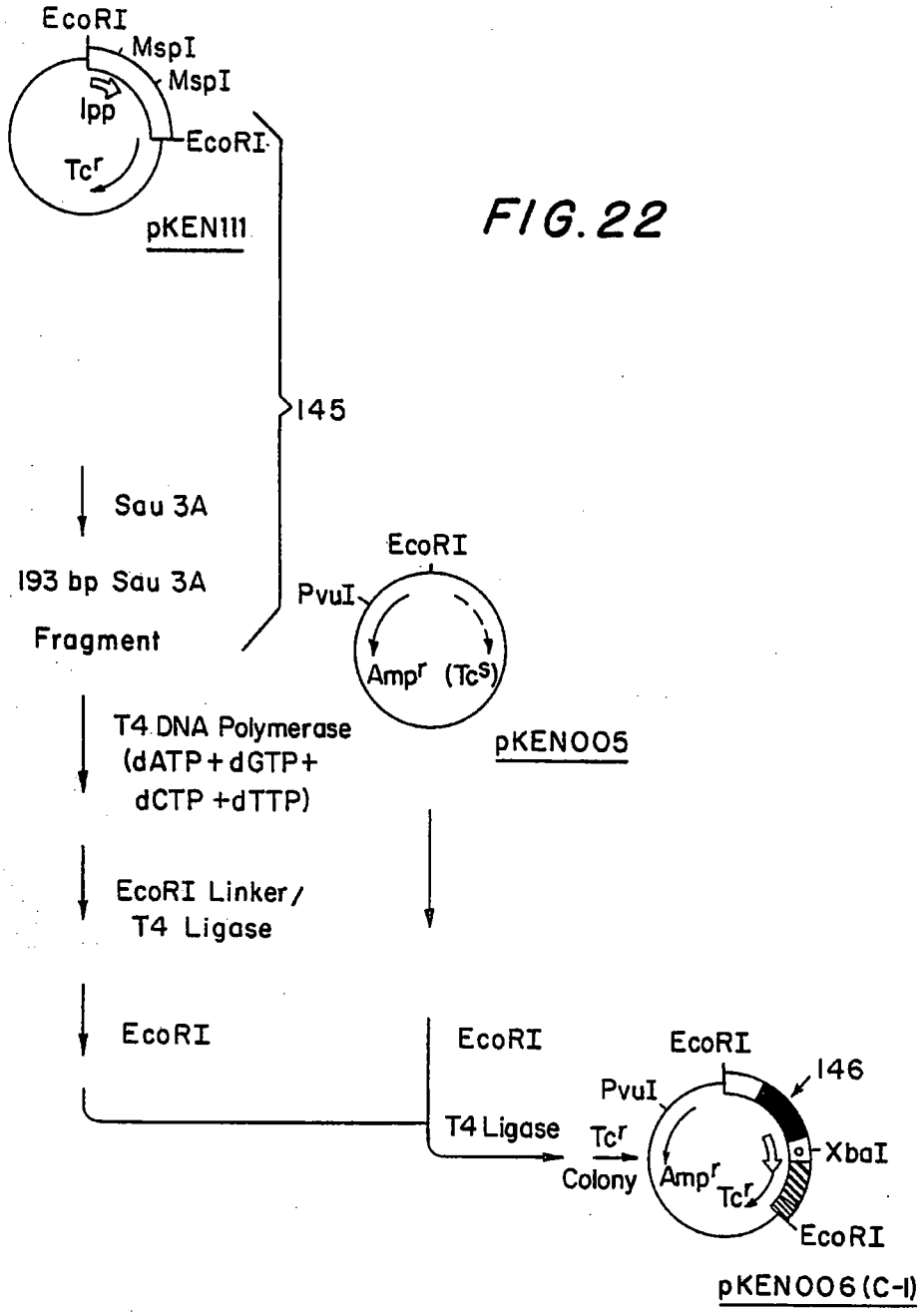

EcoRI
MspI
MspI
lpp
EcoRI
Tc^r
pKEN111
145
Sau 3A
193 bp Sau 3A
Fragment
EcoRI
PvuI
Amp^r (Tc^S)
pKENOO5
T4 DNA Polymerase
(dATP+dGTP+
dCTP +dTTP)
EcoRI Linker/
T4 Ligase
EcoRI
EcoRI
EcoRI
146
PvuI
T4 Ligase
Tc^r
Colony
Amp^r
Tc^r
XbaI
EcoRI
pKENOO6(C-I)

FIG. 22

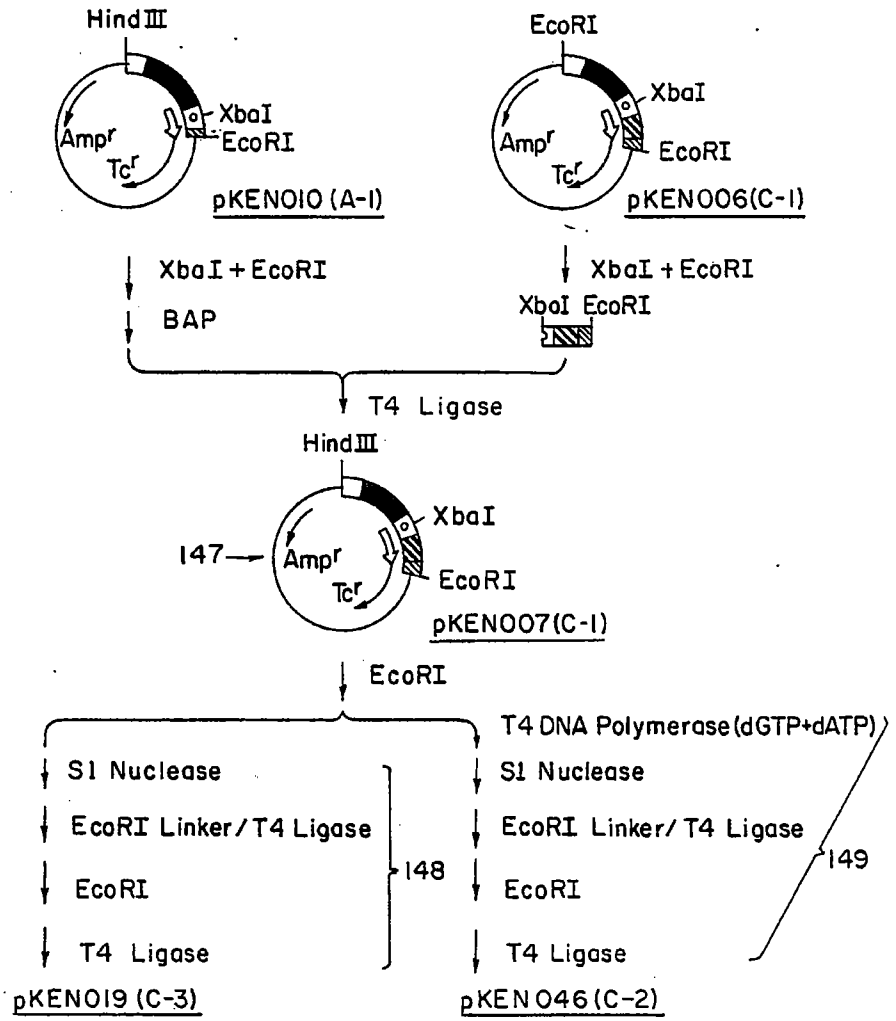

HindIII
XbaI
EcoRI
Ampr
Tcr
pKEN010 (A-1)
EcoRI
XbaI
EcoRI
Ampr
Tcr
pKEN006(C-1)
XbaI + EcoRI
BAP
XbaI + EcoRI
XbaI EcoRI
T4 Ligase
HindIII
147
XbaI
Ampr
Tcr
EcoRI
pKEN007(C-1)
EcoRI
T4 DNA Polymerase(dGTP+dATP)
S1 Nuclease
EcoRI Linker/T4 Ligase
EcoRI
T4 Ligase
148
S1 Nuclease
EcoRI Linker/T4 Ligase
EcoRI
T4 Ligase
149
pKEN019 (C-3)
pKEN046 (C-2)

FIG. 23

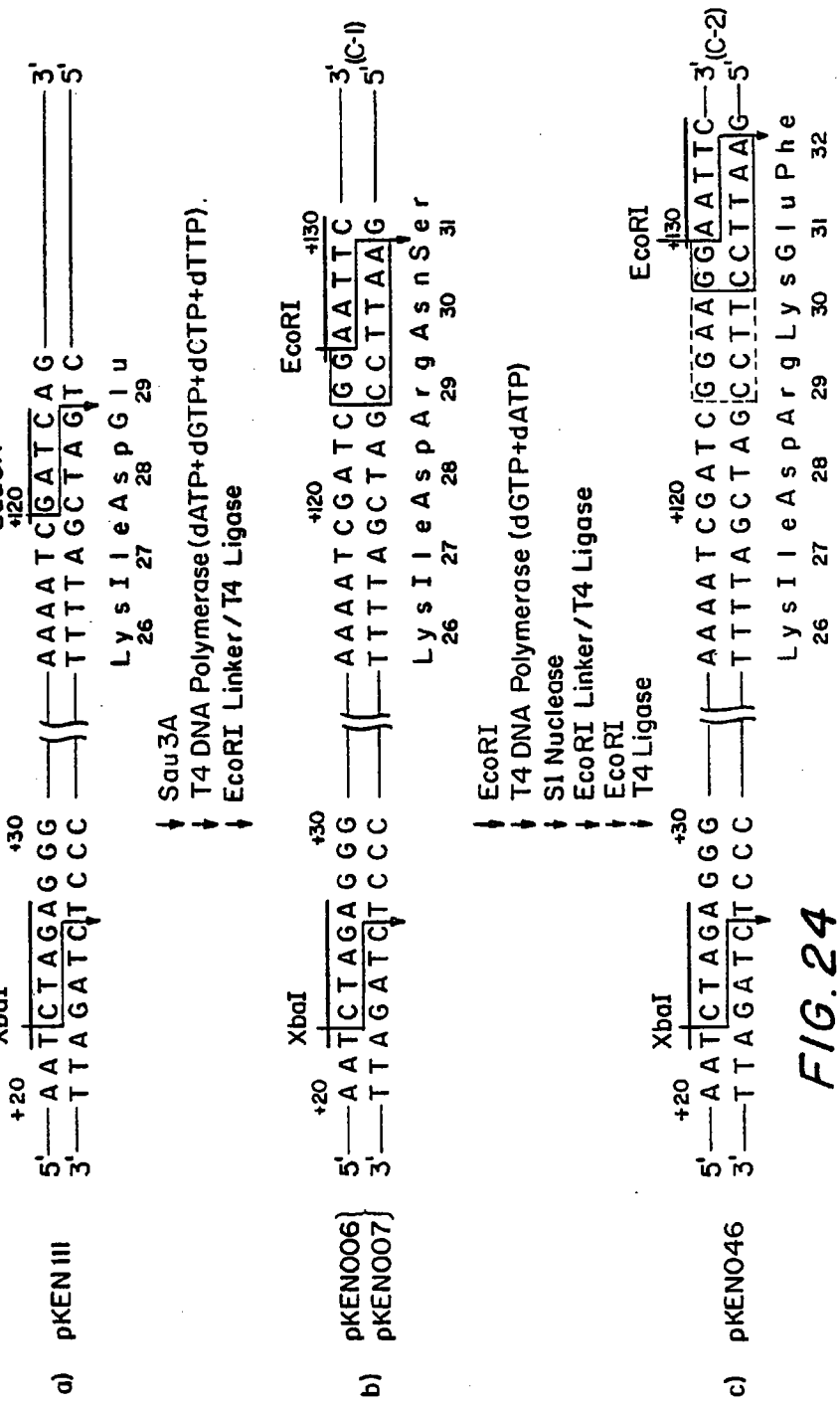

a) pKEN III
XbaI
Sau3A
Lys Ile Asp Glu
Sau3A
T4 DNA Polymerase (dATP+dGTP+dCTP+dTTP).
EcoRI Linker / T4 Ligase
b) pKEN006
pKEN007
XbaI
EcoRI
Lys Ile Asp Arg Asn Ser
EcoRI
T4 DNA Polymerase (dGTP+dATP)
S1 Nuclease
EcoRI Linker / T4 Ligase
EcoRI
T4 Ligase
c) pKEN046
XbaI
EcoRI
Lys Ile Asp Arg Lys Glu Phe

FIG. 24

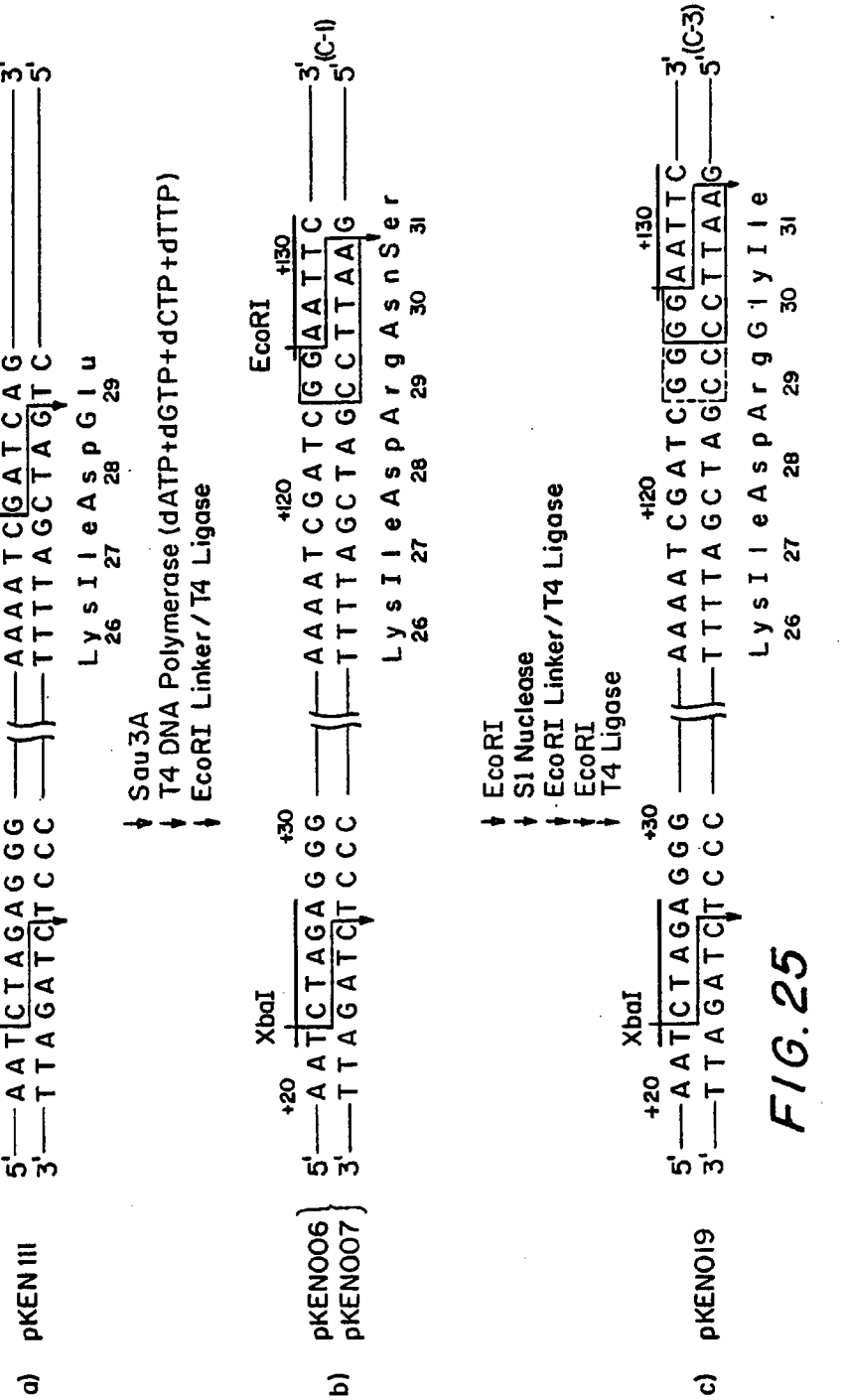

a) pKEN III
XbaI
Sau3A
Lys Ile Asp Glu
Sau3A
T4 DNA Polymerase (dATP+dGTP+dCTP+dTTP)
EcoRI Linker / T4 Ligase
b) pKEN006
pKEN007
XbaI
EcoRI
Lys Ile Asp Arg Asn Ser
EcoRI
S1 Nuclease
EcoRI Linker / T4 Ligase
EcoRI
T4 Ligase
c) pKEN019
XbaI
Lys Ile Asp Arg Gly Ile

FIG. 25

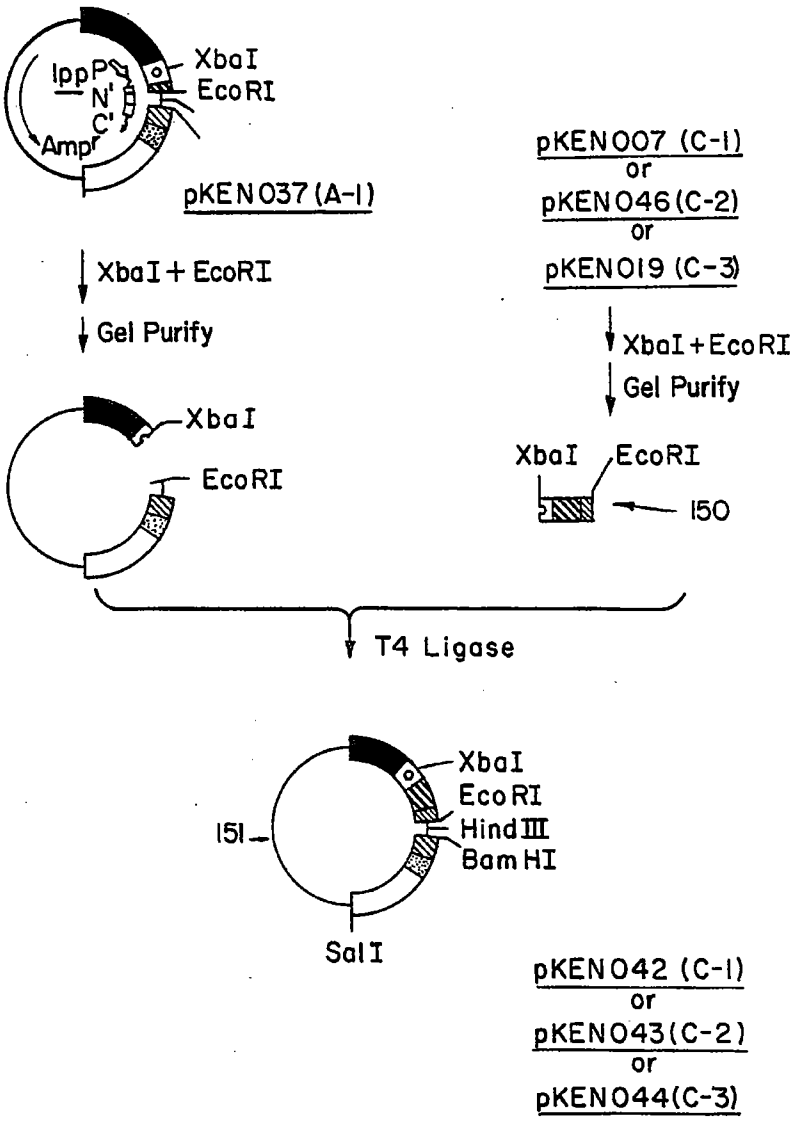

lpp
P
N'
C'
Amp^r
XbaI
EcoRI
pKEN037 (A-1)
XbaI + EcoRI
Gel Purify
XbaI
EcoRI
pKEN007 (C-1)
or
pKEN046 (C-2)
or
pKEN019 (C-3)
XbaI + EcoRI
Gel Purify
XbaI
EcoRI
150
T4 Ligase
151
XbaI
EcoRI
HindIII
BamHI
SalI
pKEN042 (C-1)
or
pKEN043 (C-2)
or
pKEN044 (C-3)

FIG. 26

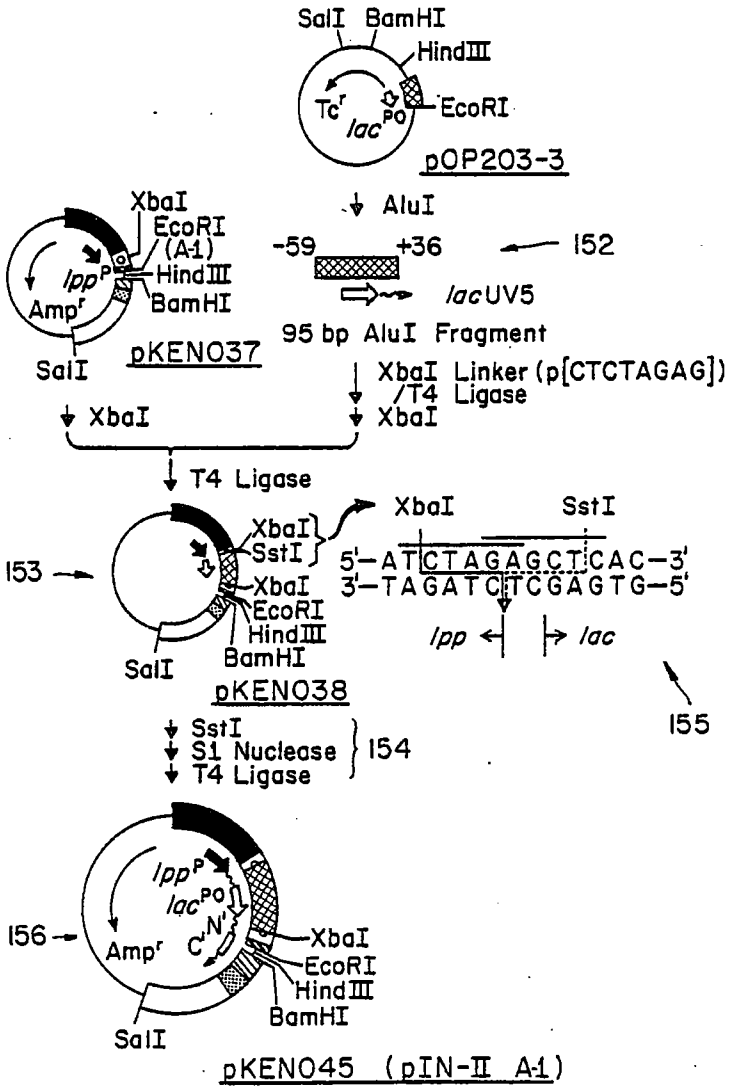

SalI BamHI
HindIII
EcoRI
Tc^r
lac^PO
pOP203-3
AluI
-59
+36
152
lacUV5
95 bp AluI Fragment
XbaI
EcoRI
(A-1)
HindIII
BamHI
lpp^P
Amp^r
SalI
pKEN037
XbaI Linker (p[CTCTAGAG])
/T4 Ligase
XbaI
XbaI
T4 Ligase
XbaI
SstI
XbaI
EcoRI
HindIII
BamHI
SalI
153
XbaI
SstI
5'—ATCTAGAGCTCAC—3'
3'—TAGATCTCGAGTG—5'
lpp
lac
pKEN038
155
SstI
S1 Nuclease
T4 Ligase
154
lpp^P
lac^PO
N'
C'
Amp^r
156
XbaI
EcoRI
HindIII
BamHI
SalI
pKEN045 (pIN-II A-1)

FIG. 27

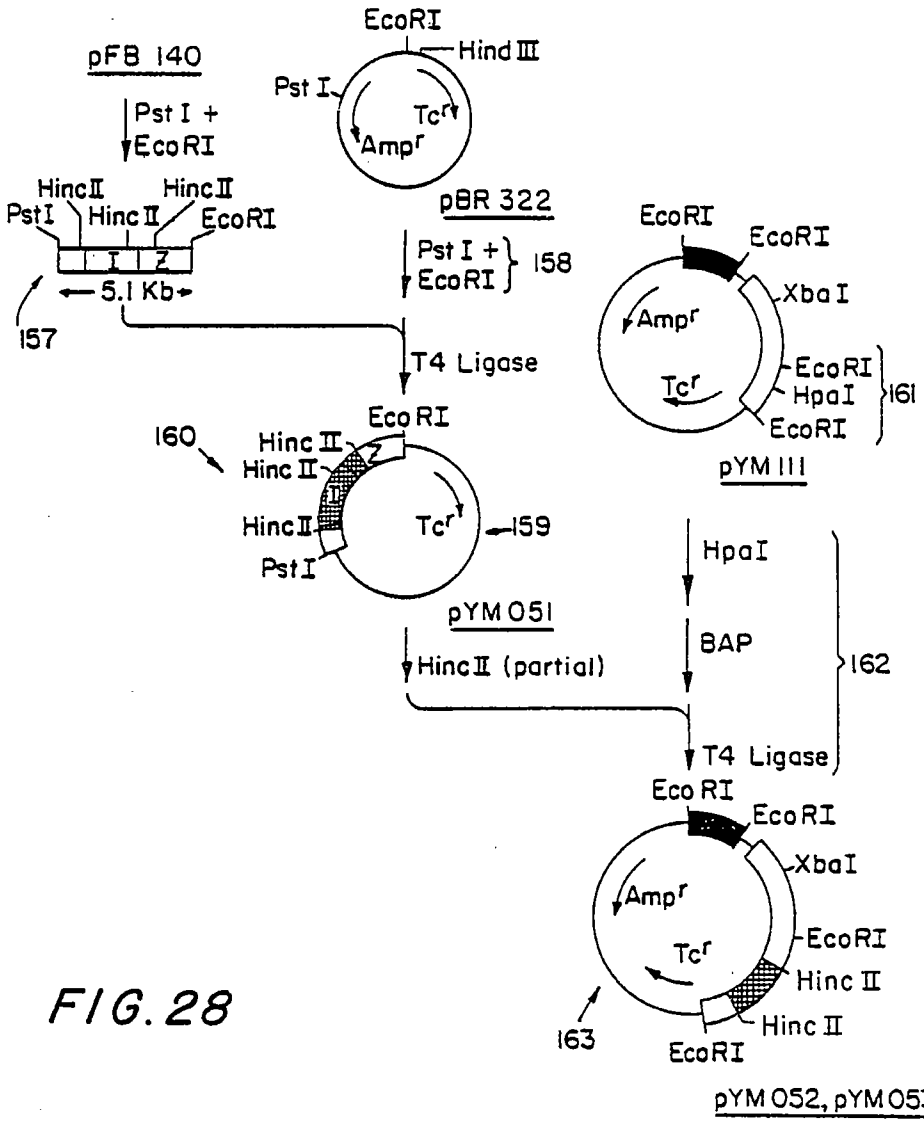

pFB 140
Pst I +
EcoRI
HincII
HincII
PstI
HincII
EcoRI
I
Z
5.1 Kb
157
EcoRI
Hind III
Pst I
Tc^r
Amp^r
pBR 322
Pst I +
EcoRI
158
T4 Ligase
160
EcoRI
Hinc II
Hinc II
HincII
PstI
Tc^r
159
pYM 051
HincII (partial)
EcoRI
EcoRI
XbaI
Amp^r
Tc^r
EcoRI
HpaI
EcoRI
161
pYM 111
HpaI
BAP
162
T4 Ligase
Eco RI
EcoRI
XbaI
Amp^r
Tc^r
EcoRI
Hinc II
Hinc II
EcoRI
163
pYM 052, pYM 053

FIG. 28

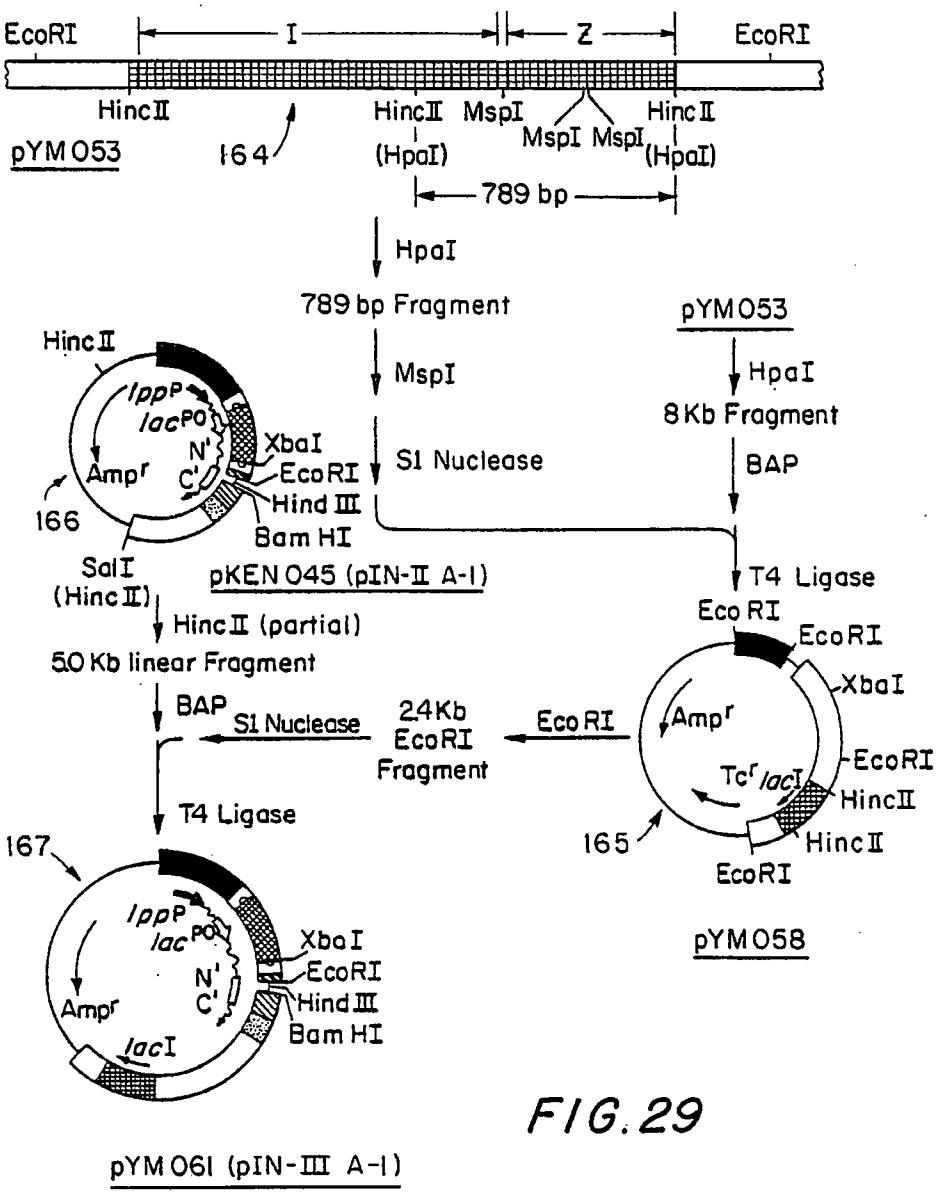

EcoRI
I
Z
EcoRI
HincII
HincII MspI
MspI MspI
HincII
pYM053
164
(HpaI)
(HpaI)
789 bp
HpaI
789 bp Fragment
pYM053
HincII
MspI
HpaI
8Kb Fragment
lppP
lacPO
N'
C'
Ampr
XbaI
EcoRI
Hind III
Bam HI
S1 Nuclease
BAP
166
SalI
(HincII)
pKEN045 (pIN-II A-1)
T4 Ligase
HincII (partial)
50Kb linear Fragment
BAP
S1 Nuclease
24Kb
EcoRI
Fragment
EcoRI
EcoRI
EcoRI
XbaI
Ampr
Tcr
lacI
EcoRI
HincII
HincII
165
EcoRI
pYM058
T4 Ligase
167
lppP
lacPO
XbaI
EcoRI
N'
C'
Hind III
Ampr
Bam HI
lacI
pYM061 (pIN-III A-1)

FIG. 29

NOVEL CLONING VEHICLES FOR POLYPEPTIDE EXPRESSION IN MICROBIAL HOSTS

This invention was made with Government support under Contract No. 5-R01-GM1904308 awarded by the National Institutes of Health. The Government has certain rights in this invention.

TECHNICAL FIELD

This invention relates generally to the field of recombinant genetics, and specifically to a novel class of plasmid cloning vehicles with which exogenous genes may be expressed in transformed bacterial hosts.

As is well-known in the art, genetic information is encoded on double-stranded deoxyribonucleic acid ("DNA") molecules ("genes") according to the sequence in which the DNA coding strand presents the characteristic bases of its repeating nucleotide components. The four nitrogenous bases that characterize the two strands of DNA nucleotides are linked in complementary pairs by hydrogen bonds to form the double helix of DNA: adenine (A) is linked to thymine (T) and guanine (G) to cytosine (C). "Expression" of the encoded information involves a two-part process. According to the dictates of certain control regions in the gene, an enzyme ("RNA polymerase") may be caused to move along the DNA coding strand, synthesizing messenger ribonucleic acid ("mRNA") in a process called "transcription." The DNA coding strand typically includes signals, which can be recognized by RNA polymerase, for both initiation and termination of transcription. In a subsequent "translation" step, the cell's ribosomes, in conjunction with transfer-RNA, convert the RNA "message" into proteins or "polypeptides," which determine cell form and function. Included in the information trnnscribed by mRNA from DNA are signals for the initiation and termination of ribosomal translation, as well as signals specifying 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" in which the characteristic bases of the nucleotides in each triplet or codon encode specific bits of information. For example, three nucleotides read as ATG (adenine-thymine-guanine) result in an mRNA signal which is interpreted as "start translation," while termination codons TAG, TAA and TGA are interpreted as "stop translation." Between the initiation codon and the termination codon lies the so-called "structural gene," the codons of which define the amino acid sequence ultimately translated. That definition proceeds according to the well-established "genetic code" (e.g., Watson, J. D., *Molecular Biology Of The Gene,* 3rd ed. [New York: W. A. Benjamin, Inc., 1976]), which specifies the codons for the various amino acids. Since there are 64 possible codon sequences but only twenty known amino acids, the genetic code is degenerate in the sense that different codons may yield the same amino acid. However, the code is precise in that for each amino acid there is at least one codon, and in that each codon yields a single amino acid and no other. Thus, for example, all of the codons, TTT, TTC, TTA and TTG, when read as such, encode for serine and no other amino acid. It will be apparent that during translation the proper reading phase or reading frame must be maintained in order to obtain the proper amino acid sequence in the polypeptide ultimately produced.

The DNA sequence within the control region of a gene which mediates the initiation of transcription is termed the "promoter" of the gene, while the specific signal encoded in the DNA following the structural gene at which transcription ends is defined as the "transcription termination site." Although the mechanisms which underlie the initiation and termination of transcriptionare not completely understood, it is believed that the promoter provides the site at which RNA polymerase must bind in order to initiate transcription, and that the effectiveness or "strength" of a particular promoter or terminator signal is determined by the efficiency with which RNA polymerase can recognize and interact with these signals. This in turn depends in large part upon the particular base sequence of the DNA at or near these sites (see, e.g., Rosenberg, M., et al., *Ann. Rev. Genet.,* 1979 13, 319–353).

The control regions of some genes may also include DNA sequences which can be recognized by certain effector molecules, the action of which can positively or negatively influence the interaction between RNA polymerase and DNA and thereby further regulate gene expression at the level of transcription. The expression of genetic information by such genes may, for example, be inhibited in the absence of a given substance, and is therefore termed "inducible." On the other hand, there also exist many genes (such as the lipoprotein gene of the Gram-negative bacterium *Escherichia coli* ["*E. coli*"]) whose control regions are not affected by effector molecules. The expression of genetic information by such genes is continuous during the lifetime of the cell, and is termed "constitutive." The control regions of such genes are generally comprised solely of a promoter signal and a terminator signal which immediately precede and follow, respectively, the DNA sequence to be transcribed.

The control regions cause mRNA synthesis to begin at a "transcription initiation site" located at or near the promoter, and to proceed until the transcription termination site is reached, producing an mRNA molecule of predetermined length with a base sequence complementary to the base sequence of the transcribed DNA. The DNA sequence between these two points defines not only the structural gene, the codons of which are ultimately translated for polypeptide expression, but also an "untranslated" region on either side of the structural gene.

Transcription therefore typically results in an mRNA molecule which carries a translatable RNA sequence, located between two untranslated regions. The untranslated region which precedes the structural sequence is known as the "5'-untranslated region," while the region which follows the structural signals is known as the "3'-untranslated region." As disclosed in detail hereinbelow, the DNA coding sequences for both of these untranslated regions, as well as the DNA coding sequences embodying the promoter signal and the terminator signal of certain genes, all of which may be referred to individually or collectively herein as "functional fragments" of those genes, may be effectively used in the creation of the novel cloning vehicles of the present invention.

As used herein, the term "cloning vehicle" defines a non-chromosomal double-stranded DNA molecule in "plasmid" form which can be replicated after being placed within a unicellular organism yy a process called

"transformation." An organism so transformed is called a "transformant." For present purposes, a "plasmid" is a circular non-chromosomal double-stranded DNA molecule derived from viruses or bacteria, the latter being termed "bacterial plasmids."

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 a recombinant plasmid may include DNA that codes for polypeptides not ordinarily produced by the organism susceptible to transformation by the recombinant plasmid, and the exogenous DNA may in some cases comprise human genetic material. Typically, 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 a desired phenotypical property. The recombinant moiety is inserted into a micro-organism 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, and generalized discussions of the subject appear in Cohen, S., *Scientific American* 233, 24–33 (July 1975), and in Gilbert, W., et al., *Scientific American* 242, 74–94 (April 1980). 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 inone way or another to facilitate ligation. The latter term refers to the formation of phosphodiester bonds between adjoining nucleotides, through the agency of a catalytic enzyme such as T4 DNA ligase. Thus, DNA fragments with "blunt" ends may be directly ligated. Alternatively, fragments containing complementary single strands at their adjoining ends are advantaged 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, or sometimes simply by "chewing back" one strand of a blunt end with an enzyme such as λ-exonuclease. Most commonly, however, such single strands may be formed by restriction endonucleases (also called restriction enzymes), which cleave the phosphodiester bonds in and around unique sequences of nucleotides of about 4–6 base pairs in length. Many restriction endonucleases and their recognition sequences are known, the so-called Eco RI endonuclease being one of the most widely employed.

Restriction endonucleases which cleave double-stranded DNA at unique sequences (e.g., at rotationally symmetric "palindromes") may leave cohesive termini. Thus, a plasmid or other cloning vehicle may be cleaved, leaving termini each comprising half of 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, synthetic DNA comprising cohesive termini may be provided for insertion into the cleaved vehicle. To discourage rejoinder of the vehicle's cohesive termini pending insertion of exogenous DNA, the termini can be digested with alkaline phosphatase, providing molecular selection for closure incorporating the exogenous fragment. Incorporation of a fragment in 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 when the fragment itself comprises termini respectively constituting half the recognition sequence of the same two different endonucleases.

As a result of wide-ranging work in recent years in recombinant DNA research, the prior art includes a number of successful and commercially viable schemes to express functional polypeptide products such as insulin, somatostatin and human and animal growth hormone. The present invention relates to an improvement of one of those schemes.

BACKGROUND ART

In U.S. Pat. No. 4,666,836, issued May 19, 1987 to M. Inouye and K. Nakamura, entitled "Novel Cloning Vehicles For Polypeptide Expression In Microbial Hosts," a class of recombinant bacterial plasmid cloning vehicles for expression of exogenous genes in transformed bacterial hosts is disclosed, comprising a DNA insert fragment coding for the desired polypeptide, linked in reading phase with one or more functional fragments derived from an outer membrane protein gene of any Gram-negative bacterium. In a preferred embodiment, the exogenous DNA codes for mammalian hormones, enzymes or immunogenic proteins (or intermediates therefor), the functional fragments are derived from the lipoprotein gene of *E. coli,* and the desired polypeptide is expressed in *E. coli* transformants preferred embodiment, the DNA sequence coding ired protein is linked with and is expressed in with four specific functional fragments associated with the *E. coli* lipoprotein gene, namely, the promoter, the 5'-untranslated region, the 3'-untranslated region and the transcription termination site of that gene.

These expression plasmids may also include a second promoter, preferably an inducible promoter and most preferably the *E. coli* β-galactosidase or "lac" promoter, which is inserted immediately downstream of the lipoprotein promoter so that the exogenous DNA is expressed only in the presence of a "lactose inducer." When induced, the DNA coding for the desired polypeptide is transcribed from both promoters, thereby increasing the yield of the desired product. Accordingly, both constitutive and inducible gene expression may be achieved using the cloning vehicles of the invention of U.S. Pat. No. 4,666,836.

However, it is disclosed in U.S. Pat. No. 4,666,836, that with the inducible cloning vehicles, special *E. coli* strains are preferred for use as transformants, specifically, those which can overproduce the lactose repressor molecule. In the wild-type *E. coli* cell, only about 10 copies of the lactose repressor molecule are maintained in the cell at any one time, which is just enough to repress (i.e., inhibit the expression of) the one lacZ gene normally contained in the cell. This is insufficient, however, to block the expression of the exogenous DNA cloned in an inducible expression plasmid of the invention of U.S. Pat. No. 4,666,836, since 10 to 20 copies of the cloning vehicle, each containing an active lac promoter, may exist in each cell at a given time. Therefore, much larger amounts of the lactose repressor are required, and for this purpose, the strain used for transformation is preferably a special *E. coli* strain JA221/F' lac^{Iq} lac+ pro+, which carries the mutant lac^{Iq} gene. The lac^{Iq} gene is a mutant of lacI, the "normal" gene

coding for the lactose repressor. The mutant gene overproduces the lactose repressor, providing about 100–150 molecules/cell at any given time. The lac^{Iq} gene is carried on the plasmid F-prime in this *E. coli* strain.

The fact that this scheme necessitates expression of the desired polypeptide in transformants carrying the plasmid F-prime gives rise to certain disadvantages. First of all, the class of recipients for the inducible expression plasmids of U.S. Pat. No. 4,666,836 is inherently limited to those *E. coli* strains which carry the lac^{I} q gene, since strains which lack this gene would not produce enough of the lactose repressor and would therefore continuously generate the desired expression product.

Secondly, the F-prime plasmid is a sex factor which causes *E. coli* cells to conjugate, resulting in transfer of the F-prime plasmid from one cell to another. The use of *E. coli* strains carrying this factor for eukaryotic gene cloning is complicated, thereby reducing still further the applicability of the scheme on which U.S. Pat. No. 4,666,836 is based.

Finally, since there are usually 2 or 3 copies of the F-prime plasmid in a cell (each of which maintains about 100–150 lactose repressor molecules), and since each cell also contains 10–20 copies of one of the inducible expression plasmids of U.S. Pat. No. 4,666,836 (each carrying a functional lac promoter), the ratio of repressor molecules to lac promoters will vary widely from cell to cell, and in some instances will not achieve complete repression of the desired expression product.

It is therefore the principal object of the present invention to provide a new class of plasmid cloning vehicles with which these disadvantages may be overcome.

DISCLOSURE OF INVENTION

In accordance with the primary objective of this invention, a class of recombinant bacterial plasmid cloning vehicles for expression of exogenous genes in transformed bacterial hosts is provided, each plasmid comprising a DNA insert fragment coding for the desired polypeptide, linked with one or more functional fragments derived from an outer membrane protein gene of a Gram-negative bacterium and also linked in reading phase with an inducible promoter fragment. Each plasmid also includes a DNA sequence coding for a protein capable of binding with and thereby repressing transcription from the inducible promoter fragment. In a preferred embodiment, the functional fragments are derived from the lipoprotein gene of *E. coli,* the inducible promoter fragment is the *E. coli* lac promoter, the DNA sequence for the repressor comprises an intact, functional *E. coli* lacI gene, and the desired polypeptide is expressed in *E. coli* transformants.

Most preferably, the present invention includes three sub-classes of plasmids, the members of each subclass containing one of three alternative insertion sites. In this manner, the selection of a particular plasmid or a particular sub-class of plasmids for gene expression can influence the ultimate location at which the expression product can be found and collected. Using one of these insertion sites, for example, the desired polypeptide can be expressed with a leader sequence located at the amino terminal which comprises the signal peptide of the *E. coli* lipoprotein, such that the desired product may be secreted through the cytoplasmic membrane and the signal peptide removed in vivo by processes native to the transformant, to yield the exogenous gene product. Using one or the other of the two remaining insertion sites, the expression product can be expected to be found either in the cytoplasm of the cell, or in the cell wall.

While the plasmids of each sub-class share a common insertion site, they differ from one another in their individual reading frames. Thus, each sub-class comprises three plasmids, whose reading frames in effect differ by one base pair, enabling the selection of any desired reading frame for each insertion site and thereby facilitating the use of the present invention with a wide variety of DNA insert fragments without the necessity of any direct modification of the reading frames of those fragments.

The exogenous DNA coding for the desired polypeptide is expressed in the plasmids of the present invention only in the presence of a lactose inducer. However, in the absence of a lactose inducer, the transcription of the cloned gene is completely repressed, due to the presence of a lacI gene on each copy of the expression plasmid existing in the host cell. Accordingly, inducible gene expression may be achieved using the cloning vehicles of the present invention, without the necessity of utilizing transformants carrying the F-prime factor in the manner disclosed in U.S. Pat. No. 4,666,836. Since the expression of genetic information in the cloning vehicles of the present invention is regulated from within each plasmid, the gene expression is termed "auto-regulated."

BRIEF DESCRIPTION OF DRAWINGS

The structure and function of the recombinant bacterial plasmids of the present invention, with which gene products such as human insulin may be expressed in bacterial transformants, is illustrated in the following specification, when taken in conjunction with the accompanying drawings wherein:

FIGS. 1A and 1B together are a schematic illustration of the 814-base pair DNA sequence encompassing the *E. coli* lipoprotein gene, in which the transcription initiation and termination sites are indicated by arrows (▲), and in which the 78 amino acid sequence of the prolipoprotein deduced from the DNA sequence is also shown, written below the corresponding codons of the DNA coding strand;

FIG. 2 shows the complete 322-nucleotide sequence of the lipoprotein mRNA of *E. coli,* in which the amino acid sequence of the prolipoprotein deduced from the mRNA sequence is also indicated, written below the corresponding codons of the nucleotide sequence;

FIG. 3 illustrates the proposed secondary structure of *E. coli* lipoprotein mRNA, in which the translation initiation codon is boxed;

FIG. 4 is a schematic outline of the process by which a eukaryotic protein or other desired polypeptide may be expressed using the cloning vehicles of the present invention, in which the transcription initiation and termination sites are indicated by arrows (▲) and the translation initiation and termination sites are indicated by arrows (Δ);

FIGS. 5–27 together comprise a schematic illustration of the preferred method for construction of the recombinant plasmid cloning vehicles of the invention, of U.S. Pat. No. 4,666,836, as disclosed therein, in which the relative positions of various restriction endonuclease cleavage sites are shown, and in which Amp^{r} and Tc^{r}, respectively, denote genes for ampicillin and tetracycline resistance; and

FIGS. 28 and 29 together comprise a schematic illustration of the preferred method of modification of one of the plasmids of U.S. Pat. No. 4,666,836 to obtain the corresponding plasmid of the present invention.

BEST MODE OF CARRYING OUT THE INVENTION

1. Summary Of Preliminary Research

As set forth in U.S. Pat. No. 4,666,836, recent investigation has shown that as a class, the major outer membrane proteins of Gram-negative bacteria are present in rather large quantities in each bacterial cell. For example, it has been found that the *E. coli* lipoprotein, which is one of the most extensively investigated membrane proteins, is also the most abundant protein in the cell in terms of numbers of molecules, there being approximately 700,000–750,000 lipoprotein molecules per cell. Since it has also been shown that there is only one structural gene for the lipoprotein of *E. coli,* extremely efficient machinery for lipoprotein gene expression, at the levels of both transcription and translation, is indicated. It is believed that the lipoprotein gene may be expressed at least ten times more efficiently than genes for ribosomal proteins. The presence of comparable quantities of other major outer membrane proteins in *E. coli,* such as the ompA protein, and the presence of comparable quantities of the maoor outer membrane proteins in other Gram-negative bacteria, such as the lipoprotein of *Serratia marcescens,* indicate that these systems may also have very efficient machinery for gene expression. Thus, while the discussion herein may refer in large part to the lipoprotein system in *E. coli,* it is to be understood that this invention extends to recombinant cloning vehicles which utilize the machinery for gene expression associated with any of the outer membrane protein genes of any Gram-negative bacterium.

Although the mechanisms which are responsible for the highly efficient expression of the *E. coli* lipoprotein gene are not yet completely understood, it is believed that several factors must contribute to the abundance of lipoprotein molecules in a bacterial cell. As shown in FIGS. 1A and 1B, the DNA nucleotide sequence of the lipoprotein gene of *E. coli* has recently been determined, an analysis of which has revealed many unique properties associated with the expression of this gene.

In particular, it has been found that in comparison with other known promoter sequences of *E. coli* genes, the lipoprotein promoter region shows a most striking feature, namely, an extremely high A-T content, which is believed likely to be essential for highly efficient transcription of the lipoprotein gene. The segment of 261 base pairs ("bp") preceding the transcription initiation site (from position −261 through position −1 as shown in FIG. 1A) has a very high A-T content of 70%, in contrast with 53% for the transcribed region (or mRNA region) of 322 base pairs (positions +1 to +322), 44% for a segment of 126 bp after the transcription termination site (positions +323 to +449), and 49% for the average A-T content of the *E. coli* chromosome. The A-T content of the segment from position −45 to position −1, within which the nucleotide sequence of the lipoprotein ("lpp") promoter appears to reside, is especially high (80%), and appears to be the highest among the *E. coli* promoter regions thus far sequenced. The A-T richness of the promoter sequence is considered to destabilize the helix structure of the DNA and thereby facilitate the RNA polymerase-mediated strand unwinding necessary for the initiation of transcription.

Apart from its A-T content, the lpp promoter also appears to contain a heptanucleotide sequence at positions −15 through −9 (only eight base pairs distal to the transcription initiation site) which is homologous to the generalized "Pribnow box," as well as a dodecanucleotide sequence at positions −38 through −27 which is homologous to the generalized "RNA polymerase recognition site." The homology of these sequences is striking, in that the Pribnow box sequence of the lpp promoter has only one base mismatching with the generalized sequence, while the recognition site sequence shows a mismatch of only 5 out of 12 bases of the generalized sequence. The importance of the specific base sequences at these sites for efficient transcription is well-documented, in that mutants with enhanced promoter efficiency show increased homology of these regions with the generalized sequences.

Further analysis of the DNA sequence of FIGS. 1A and 1B has revealed that besides having an extremely "strong" promoter, the lipoprotein gene also has an oligo-T transcription termination signal, located between positions +316 and +322, which is at least as efficient as all other *E. coli* transcription termination sites that have been studied. It is believed that this factor contributes to the overall efficiency of transcription by hastening the rate of mRNA production, and by limiting the size of the mRNA molecule which is transcribed from the DNA.

As shown in FIG. 2, the complete nucleotide sequence of the *E. coli* lipoprotein mRNA has also been determined, revealing that the mRNA has several unique features in its structure which appear to be important for efficient translation of the mRNA transcript. The mRNA consists of 322 nucleotides, 38 of which are in the 5'-untranslated region and 50 of which are in the 3'-untranslated region, leaving 234 nucleotides in the translated region which code for the lipoprotein precursor, or prolipoprotein. The mRNA sequence of FIG. 2 is complementary to the DNA sequence of FIGS. 1A and 1B, with the exception of the nucleotide at position 313 which is shown as C in FIG. 2 as determined by RNA sequencing, rather than A as determined by the DNA sequencing shown in FIG. 1B. The reason for this difference is not known at present.

The lipoprotein mRNA has been shown to be unusually stable, and it has been proposed that this stability is probably attributable to the formation of extensive secondary structures within the molecule. As shown in FIG. 3, the mRNA can form nine stable "hairpin" stem-and-loop structures (designated by Roman numerals I–IX), the most stable of which (I) is in the 3'-untranslated region. These secondary structures may be responsible for the longer functional half-life which has been observed for the lipoprotein mRNA in comparison with other *E. coli* mRNAs, and may thereby increase the availability of this molecule for ribosomal translation.

Furthermore, although 68% of the total nucleotides in the mRNA molecule are involved in the formation of the hairpin structures shown in FIG. 3, it should be noted that in the first 64 nucleotides from the 5' end there are no stable hairpin structures, whereas between the 65th nucleotide and the 3' end, 85% of the nucleotides are involved in the formation of hairpin structures. This is significant because in the 5'-untranslated region (positions +1 to +38) there appear to be two extensive