

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4139561号

(P4139561)

(45) 発行日 平成20年8月27日 (2008. 8. 27)

(24) 登録日 平成20年6月13日 (2008. 6. 13)

(51) Int. Cl.

F 1

C 1 2 N 15/09 (2006. 01)

C 1 2 N 15/00 Z N A A

C 1 2 N 1/21 (2006. 01)

C 1 2 N 1/21

請求項の数 12 (全 67 頁)

(21) 出願番号	特願2000-524410 (P2000-524410)	(73) 特許権者	500262197
(86) (22) 出願日	平成10年12月7日 (1998. 12. 7)		ユーロペイシェ ラボラトリウム フュ
(65) 公表番号	特表2002-503448 (P2002-503448A)		ール モレキュラーバイオロジー (イーエ
(43) 公表日	平成14年2月5日 (2002. 2. 5)		ムビーエル)
(86) 国際出願番号	PCT/EP1998/007945		ドイツ連邦共和国 ディー - 6 9 1 1 7 ハ
(87) 国際公開番号	W01999/029837		イデルベルク、メイヤーホフシュトラッセ
(87) 国際公開日	平成11年6月17日 (1999. 6. 17)		1
審査請求日	平成16年1月8日 (2004. 1. 8)	(74) 代理人	100088904
(31) 優先権主張番号	97121462. 2		弁理士 庄司 隆
(32) 優先日	平成9年12月5日 (1997. 12. 5)	(72) 発明者	スチュワート、フランシス
(33) 優先権主張国	欧州特許庁 (EP)		ドイツ連邦共和国 ディー - 6 9 1 8 1 ラ
(31) 優先権主張番号	98118756. 0		イメン、レーベンヴェク 3
(32) 優先日	平成10年10月5日 (1998. 10. 5)		
(33) 優先権主張国	欧州特許庁 (EP)		

最終頁に続く

(54) 【発明の名称】 新規DNAクローニング方法

(57) 【特許請求の範囲】

【請求項 1】

a) recEおよびrecT遺伝子を発現する能力を有し、かつ、recETに依存する機構を介して相同組換えを実行する能力を有する原核宿主細胞を準備する工程、

b) 該宿主細胞中で複製され得る環状である第1 DNA分子と、第1 DNA分子上の領域と配列相同な領域を少なくとも2個含む第2 DNA分子とを、recETに依存する機構を介する第1および第2 DNA分子間の相同組換えに好適な条件下で、該宿主細胞中で接触させる工程、ならびに

c) 該第1および第2 DNA分子間の相同組換えが生じた宿主細胞を選択する工程、

を含み、

第2 DNA分子はさらなる改変を行うことなく組換えを可能にする形態で宿主細胞中に導入される、原核細胞中でDNA分子をクローニングする方法。

【請求項 2】

recEおよび／またはrecT遺伝子の発現が調節性プロモーターの制御下にある、請求項1記載の方法。

【請求項 3】

recE遺伝子が、

(a) 図7Bに記載の位置1320(ATG)から2159(GAC)までの核酸配列、

(b) 図13Bに記載の位置1320(ATG)から1998(CGA)までの核酸配列、

(c) 遺伝コードの縮重範囲内にある、(a)もしくは(b)の核酸配列がコードするポ

リペプチドをコードする核酸配列、ならびに／または

(d) (a)、(b) および／もしくは (c) の核酸配列とストリンジェント条件下でハイブリダイズする核酸配列を含む核酸分子から選択され、
かつ／あるいは、

recT遺伝子が、

(e) 図7Bに記載の位置2155(ATG)から2961(GAA)までの核酸配列、

(f) 図13Bに記載の位置2086(ATG)から2868(GCA)までの核酸配列、

(g) 遺伝コードの縮重範囲内にある、(e) もしくは (f) の核酸配列がコードするポリペプチドをコードする核酸配列、ならびに／または

(h) (e)、(f) および／もしくは (g) の核酸配列とストリンジェント条件下でハイブリダイズする核酸配列を含む核酸分子から選択される、

請求項2記載の方法。

【請求項4】

大腸菌株がJC8679およびJC9604から選択される、請求項1記載の方法。

【請求項5】

recBCインヒビター遺伝子が、

(a) 図13Bに記載の位置3588(ATG)から4002(GTA)までの核酸配列、

(b) 遺伝コードの縮重範囲内にある、(a) の核酸配列がコードするポリペプチドをコードする核酸配列、ならびに／または

(c) (a) および／もしくは (b) の核酸配列とストリンジェント条件下でハイブリダイズする核酸配列

を含む核酸分子から選択され、宿主細胞がさらにrecBCインヒビター遺伝子を発現する能力を有する請求項4記載の方法。

【請求項6】

第2DNA分子が、2つの配列相同性領域の間に配置されたマーカー遺伝子を少なくとも1つ含有し、かつ、該マーカー遺伝子の発現によって相同組換えを検出する、請求項1～5のいずれか1項に記載の方法。

【請求項7】

宿主細胞中で複製され得る環状第1DNA分子を該原核宿主細胞中に導入する工程、ならびに

該第1DNA分子上の第3および第4領域にそれぞれ配列相同な第1および第2領域を含む第2DNA分子を該原核宿主細胞中に導入する工程、

を含んでなる、組換えDNA分子を作製するための請求項1記載の方法であって、

該宿主細胞が、組換えDNA分子を形成するようにrecETに依存する機構を介した相同組換えを実行する能力を有し、かつ

該組換えDNA分子は、第3領域と第4領域との間の配列が第2DNA分子の第1領域と第2領域との間の配列と置換されている第1DNA分子を含み、第2DNA分子はさらなる改変を行うことなく組換えを可能とする形態で該宿主細胞中に導入されることを特徴とする、
上記方法。

【請求項8】

染色体上に第1DNA分子を含む原核宿主細胞中に、該宿主染色体上の第1DNA分子の第3および第4領域とそれぞれ配列相同な第1および第2領域を含む第2DNA分子を導入する工程を含んでなる、組換えDNA分子を作製するための請求項1記載の方法であって、

該宿主細胞が、組換えDNA分子を形成するようにrecETに依存する機構を介した相同組換えを実行する能力を有し、かつ

該組換えDNA分子は、第3領域と第4領域との間の配列が第2DNA分子の第1領域と第2領域との間の配列と置換されている染色体上の第1DNA分子を含み、第2DNA分子はさらなる改変を行うことなく組換えを可能とする形態で該宿主細胞中に導入されることを特徴とする、
上記方法。

10

20

30

40

50

【請求項 9】

環状である第 1 DNA 分子と、第 1 DNA 分子上の領域と配列相同な領域を少なくとも 2 個含む第 2 DNA 分子とを組換えること、および、該第 2 DNA 分子は、さらなる改変を行うことなく組換えを可能とする形態で宿主細胞中に導入されることを特徴とする、recET に依存する機構を介する相同組換えを含むクローニング方法のための宿主細胞としての、recE および recT 遺伝子を発現する能力を有する原核細胞の使用。

【請求項 10】

環状である第 1 DNA 分子と、第 1 DNA 分子上の領域と配列相同な領域を少なくとも 2 個含む第 2 DNA 分子とを組換えること、および、該第 2 DNA 分子は、さらなる改変を行うことなく組換えを可能とする形態で宿主細胞中に導入されることを特徴とする、recET に依存する機構を介する相同組換えを含むクローニング方法のための、原核宿主細胞中での、recE および recT 遺伝子を発現する能力を有するベクター系の使用。

10

【請求項 11】

(a) recET に依存する機構を介して相同組換えを実行する能力を有する原核宿主細胞

、

(b) recE および recT 遺伝子を発現する能力を有するベクター系を含む該宿主細胞中での recE および recT 遺伝子の発現のための手段、ならびに

(c) 第 1 DNA 分子上の領域と配列相同な領域を少なくとも 2 個含む第 2 DNA 分子と組換えることができる環状第 1 DNA 分子である、該細胞中で複製され得るレシピエントクローニングビヒクル、

20

を含む、クローニングのための試薬キット。

【請求項 12】

第 1 および第 2 DNA 増幅プライマーをさらに含んでおり、第 1 DNA 増幅プライマーが環状レシピエントクローニングビヒクル上の第 3 領域に配列相同な第 1 領域を含み、かつ、第 2 DNA 増幅プライマーが環状レシピエントクローニングビヒクル上の第 4 領域に配列相同な第 2 領域を含む、請求項 11 記載のクローニングのための試薬キット。

【発明の詳細な説明】

【0001】

本発明は、少なくとも 2 つの DNA 分子間での相同組換え機構を使用する、DNA 分子をクローニングするための新規な方法に関する。さらに、DNA クローニングに好適な新規な試薬キットが提供される。

30

【0002】

細菌細胞中での外来 DNA のクローニングのための現行の方法は、通常、好適な細菌ベクターを準備し、このベクターを制限酵素で切断し、そのベクター中に外来 DNA 断片を *in vitro* で挿入するステップを含む。次に、生成した組換えベクターを使用して細菌を形質転換する。こうしたクローニング方法は約 20 年間使用されて成功しているが、これにはいくつかの欠点がある。これらの欠点として、特に、ベクター中に外来 DNA を挿入するために必要とされる *in vitro* ステップが、その外来 DNA 上またはベクター上で好適な制限部位が利用できない場合、非常に複雑で時間がかかることが多い点である。

【0003】

40

さらに、現行の方法は通常、外来 DNA 断片を配置するベクター中の好適な制限酵素切断部位の存在に依拠している。このことは最終クローニング産物について 2 つの制限を与えることになる。第 1 に、外来 DNA 断片はベクター中で通常 1 個または複数のこうした制限部位の位置にしか挿入することができない。したがって、クローニング産物は好適な制限部位の配置によって限定され、また好適な制限部位がないベクターの領域中へのクローニングは困難であり、不正確なことが多い。第 2 に、制限部位は典型的には長さが 4 から 8 塩基対であるので、使用する DNA 分子のサイズが増加するにつれて、多大な倍数で出現する。これは、大部分の現行のクローニング技術によって操作することができる DNA 分子のサイズの実用上の制限を意味している。特に、コスミド、BAC、PAC および P1 などのベクター中にクローン化された比較的大きなサイズの DNA を、制限酵素に基づく技術によって直接

50

操作するのは通常実現不能である。したがって、先行技術の欠点が少なくとも部分的に除去された新しいクローニング方法を提供する必要性がある。

【0004】

本発明において、recEおよびrecT遺伝子またはred α およびred β 遺伝子あるいはファージP22組換え系などの機能的に関連する遺伝子の産物を発現する能力がある細菌宿主細胞中で、2つのDNA分子間での効果的な相同組換え機構が使用可能な頻度で生じることが発見された (Kolodnerら、Mol.Microbiol.11(1994)23-30; Fenton,A.C.and Poteete,A.R.,Virology 134(1984) 148-160; Poteete,A.R.and Fenton,A.C.,Virology 134(1984) 161-167)。DNA断片のクローニングのこの新規な方法を「ETクローニング」と命名する。

【0005】

大腸菌 RecEおよびRecTタンパク質の同定および特性決定についてGillenら (J.Bacteriol.145(1981),521-532) およびHallら (J.Bacteriol.175(1993),277-287) が記載している。HallおよびKolodner (Proc.Natl.Acad.Sci.USA 91(1994),3205-3209) は、RecTタンパク質によって推進される、線状2本鎖DNAおよび相同性環状1本鎖DNAのin vitro相同的対形成および鎖交換を開示している。細胞中でのDNA分子のクローニングのためのこの方法の使用に対する引用はそこに見出だすことができない。

【0006】

大腸菌中の遺伝子組換えのrecET経路は知られている (Hall and Kolodner(1994),上記; Gillenら(1981),上記)。この経路は2つの遺伝子、recEおよびrecTの発現を必要とする。これらの遺伝子のDNA配列は公開されている (Hallら、上記)。RecEタンパク質は λ エクソまたは λ Red α などのバクテリオファージタンパク質に類似している (Gillenら、J.Mol.Biol.113(1977),27-41; Little,J.Biol.Chem.242(1967),679-686; Radding and Carter, J.Biol.Chem.246(1971),2513-2518; Joseph and Kolodner,J.Biol.Chem.258(1983),10418-10424)。RecTタンパク質は λ β タンパク質または λ Red β などのバクテリオファージタンパク質に類似している (Hallら、(1993),上記; Muniyappa and Radding,J.Biol.Chem.261(1986),7472-7478; Kmiec and Hollomon,J.Biol.Chem.256(1981),12636-12639)。上記の報文の内容を参照により、本明細書に組み入れる。

【0007】

Olinerら (Nucl.Acids Res.21(1993),5192-5197) は線状PCR産物および線状化プラスミドベクター間の分子間相同組換えによる、大腸菌中のPCR産物のin vivoクローニングについて記載している。原核生物中での相同組換えに基づく新規なクローニング方法を開発するためのその他の従来の手法もまた、ベクターを線状化するための制限酵素の使用 (Bubeckら、Nucleic Acids Res.21(1993),3601-3602; Olinerら、Nucleic Acids Res.21(1993),5192-5197; Degryse,Gene 170(1996),45-50) または宿主特異的recA依存組換え系 (Hamiltonら、J.Bacteriol.171(1989),4617-4622; Yangら、Nature Biotech.15(1997),859-865; Dabert and Smith,Genetics 145(1997),877-889) に依拠していた。これらの方法は適用性が非常に限定されており、実際上はほとんど使用されていない。

【0008】

本発明に従うDNAクローニングの新規な方法は制限酵素またはDNAリガーゼによるin vitro処理を必要とせず、したがってDNAクローニングの標準的な方法論とは基本的に差異がある。この方法はrecEおよびrecT遺伝子産物、またはred α およびred β 遺伝子産物、あるいは機能的に等価な遺伝子産物が関与する、大腸菌中の相同組換えの経路に依拠している。この方法はクローン化しようとするDNA断片である、好ましくは線状でかつ好ましくは染色体外DNA断片1個をエピソードまたは内在性宿主染色体若しくは染色体群のいずれかである第2の好ましくは環状DNAベクター分子と共有結合させるものである。したがって、これは2つの線状DNA断片または異なる組換え経路のいずれかの使用に依拠する相同組換えによる大腸菌中のクローニングについての従来の記載とは差異がある。

【0009】

本発明は、完全な>75 kbプラスミドおよび大腸菌染色体を含む大きなDNA分子を工学操作するための相同組換えを使用する、適応力がある方法を提供する。すなわち、ターゲット

10

20

30

40

50

の選択について、サイズまたは部位のいずれについても現実的に制限がない。したがって、高コピープラスミドからゲノムまで、宿主細胞中のどんなレシピエントDNAでも正確な変更を受けやすい。大きなDNA分子の工学操作の他に、本発明はDNA設計のための新規な、制限酵素に依存しない手法を大要とする。例えば、オリゴヌクレオチドの相同なアームを選択することによって、ターゲットエピソーム中の任意の2つの選択された塩基対間の欠失させることができる。同様に、選択されたDNA配列を選択された塩基対の位置に挿入して、例えば変更したタンパク質リーディングフレームを作製することができる。挿入および欠失の協奏的組合せならびにポイント突然変異もまた可能である。これらの戦略の適用は複雑で困難なDNAの構築、例えば真核細胞、例えばマウス胚性幹細胞中の相同組換えを意図する場合に特に適切である。さらに、本発明は所望のターゲット部位に正確に部位特異的組換え位置を定める簡単な方法を提供する。これは植物およびマウスなどのその他の生体系中の部位特異的組換えへの適用を単純化することとなる。

10

【0010】

本発明の主題の1つは、a) 相同組換えを実行する能力がある宿主細胞を準備し、b) 該宿主細胞中で複製され得る第1 DNA分子と、第1 DNA分子上の領域と配列相同な領域を少なくとも2個含む第2 DNA分子とを、該第1 および第2 DNA分子間の相同組換えに好適な条件下で接触させ、そしてc) 該第1 および第2 DNA分子間の相同組換えが生じた宿主細胞を選択する、ステップを含む、細胞中でのDNA分子のクローニング方法である。

【0011】

本発明の方法において、相同組換えは好ましくはrecET機構を介して生じる。すなわち、相同組換えが好ましくは大腸菌遺伝子recEおよびrecTから選択されるrecEおよびrecT遺伝子、またはファージ λ red α およびred β 遺伝子などの機能的に関連する遺伝子の遺伝子産物によって仲介される。

20

【0012】

本発明の方法に好適な宿主細胞は好ましくは細菌細胞、例えばグラム陰性細菌細胞である。さらに好ましくは、宿主細胞はSalmonella、KlebsiellaまたはEscherichiaなどの腸内細菌細胞である。最も好ましくは、宿主細胞は大腸菌細胞である。しかし、本発明のクローニング方法は菌類、植物または動物細胞などの真核細胞についても好適であることに留意すべきである。

【0013】

好ましくは、相同組換えおよびクローン化したDNAの増殖のために使用する宿主細胞は、例えばその中でrecEおよびrecT、またはred α およびred β 遺伝子の産物が発現される再菌株などのあらゆる細胞とすることができる。宿主細胞はrecEおよびrecT遺伝子をその宿主細胞の染色体上、または非染色体DNA、好ましくはベクター、例えばプラスミド上に位置させて含むことができる。好ましい事例において、RecEおよびRecT、またはRed α およびRed β 遺伝子産物は2つの別の調節性プロモーター、アラビノース誘導性BADプロモーター若しくはlacプロモーターなど、または非調節性プロモーターから発現される。あるいは、recEおよびrecT、またはred α およびred β 遺伝子は1個の調節性または非調節性プロモーターからのポリシストロン性mRNA上で発現される。好ましくは、その発現は調節性プロモーターによって制御される。

30

【0014】

recEまたはred α 遺伝子が調節性プロモーターの1つによって発現される実施形態もまた特に好ましい。ここでは、システムの組換え能力は必要な場合にのみ誘発され、その他の時点では可能な所望しない組換え反応が制限される。他方において、recTまたはred β 遺伝子はrecEまたはred α に対して過剰発現されるのが好ましい。これは強力な構成性プロモーター、例えばEM7プロモーターの使用、および／またはrecE若しくはred α 遺伝子よりも高コピー数のrecT若しくはred β の使用によって達成することができる。

40

【0015】

本発明の目的のためには、第1 および第2 DNA分子の相同組換えが、DNAでトランスフェクトした細胞 10^9 中で1個よりも多い細胞中に組換え産物をもたらすのに十分な効力を持つ

50

限り、どんなrecEおよびrecT遺伝子でも適している。recEおよびrecT遺伝子はあらゆる細菌株またはバクテリオファージから誘導することができ、あるいはそれらの突然変異体または変異型でもよい。好ましいのは大腸菌または大腸菌バクテリオファージから誘導されるrecEおよびrecT遺伝子であり、ラムダ型ファージ、例えばバクテリオファージλからのred α およびred β 遺伝子などである。

【0016】

さらに好ましくは、recEまたはred α 遺伝子が以下のものを含む核酸分子から選択される：

- (a) 図7Bに記載の位置1320(ATG)から2159(GAC)までの核酸配列、
- (b) 図14Bに記載の位置1320(ATG)から1998(CGA)までの核酸配列、
- (c) 遺伝コードの縮重範囲内にある、同一のポリペプチドをコードする核酸配列、ならびに／または
- (d) (a)、(b)および／若しくは(c)からの核酸配列とストリンジェント条件下でハイブリダイズする核酸配列。

10

【0017】

さらに好ましくは、recTまたはred β 遺伝子が以下のものを含む核酸分子から選択される：

- (a) 図7Bに記載の位置2155(ATG)から2961(GAA)までの核酸配列、
- (b) 図14Bに記載の位置2086(ATG)から2868(GCA)までの核酸配列、
- (c) 遺伝コードの縮重範囲内にある、同一のポリペプチドをコードする核酸配列、ならびに／または
- (d) (a)、(b)および／若しくは(c)からの核酸配列とストリンジェント条件下でハイブリダイズする核酸配列。

20

【0018】

本発明は既定の配列の突然変異体および変異型、例えば天然に存在する突然変異体および変異型、または遺伝子工学操作によって取得される突然変異体および変異型をも包含することに留意すべきである。さらに、図7Bに示すrecE遺伝子は天然タンパク質のアミノ酸位置588-866をコードする、すでに切断された遺伝子であることに留意すべきである。突然変異体および変異型は、好ましくは図7Bおよび13Bに示すrecEおよびrecT配列、ならびに図14Bに示すred α およびred β 配列に少なくとも60%、好ましくは少なくとも70%、さらに好ましくは少なくとも80%の核酸配列相同性を有する。

30

【0019】

本発明において、ストリンジェント条件下でのハイブリダイゼーションは好ましくは下記のSambrookら(1989)にしたがって定義されるもので、0.1x SSC、0.5% SDS中、55℃、好ましくは62℃、そしてさらに好ましくは68℃で30分の洗浄後、検出し得るハイブリダイゼーションシグナルを含む。

【0020】

好ましい事例において、recEおよびrecT遺伝子は大腸菌K12株中に存在する相当する内在性遺伝子およびその誘導体から、またはバクテリオファージから誘導される。特に、sbca突然変異を保有する株が好適である。こうした菌株の例はJC8679およびJC9604である(Gillenら(1981)、上記)。あるいは、相当する遺伝子をラムダ型ファージまたはファージP22などのその他のコリファージから取得してもよい。

40

【0021】

JC8679およびJC9604の遺伝子型はSex (Hfr, F+, F-, またはF') : F-である。JC8679は以下の突然変異を含む：recBC21, recC22, sbca23, thr-1, ara-14, leuB6, DE(gpt-proA)62, lacY1, tsx-33, gluV44(AS), galK2(0c), LAM-, his-60, relA1, rpsL31(strR), xylA5, mtl-1, argE3(0c) およびthi-1。JC9604は同一の突然変異およびさらに突然変異recA56を含む。

【0022】

さらに、recEおよびrecT、またはred α およびred β 遺伝子を第1のドナー源、例えばドナー細菌細胞から単離して、第2の受容体源、例えば受容体細菌若しくは真核細胞中に形質

50

転換し、そこで組換えDNA手法によって発現させることができることも留意すべきである。

【0023】

本発明の1実施形態において、使用する宿主細胞はsbcA突然変異を有する細菌株、例えば上記の大腸菌株JC8679およびJC9604の1つである。しかし、本発明の方法はsbcA突然変異を有する宿主細胞または類似の細胞に限定されるものではない。驚くべきことに、本発明のクローニング方法がsbcA突然変異を持たないrecBC+またはrecBC-細胞中、例えば大腸菌recBC+細胞などの原核recBC+宿主細胞中でも機能することが発見された。この事例において、好ましくはその中でrecBC型エキソヌクレアーゼインヒビター遺伝子の産物が発現する宿主細胞を使用する。好ましくは、そのエキソヌクレアーゼインヒビターは宿主のrecB Cシステムまたはその等価物を阻害する能力があるものである。こうしたエキソヌクレアーゼインヒビター遺伝子の好適な例は λ red γ 遺伝子 (Murphy, J. Bacteriol. 173(1991), 5808-5821) およびそれぞれその機能的等価物、例えばファージP22 (Murphy, J. Biol. Chem. 269(1994), 22507-22516) などの別のコリファージから取得することができるものである。

【0024】

さらに好ましくは、エキソヌクレアーゼインヒビター遺伝子が以下のものを含む核酸分子から選択される：

(a) 図14Aに記載の位置3588(ATG)から4002(GTA)までの核酸配列、

(b) 遺伝コードの縮重範囲内にある、同一のポリペプチドをコードする核酸配列、ならびに／または

(d) (a) および／若しくは(b)からの核酸配列と(上記の定義の)ストリンジェント条件下でハイブリダイズする核酸配列。

【0025】

驚くべきことに、recBC+およびrecBC-両株中でのエキソヌクレアーゼインヒビター遺伝子の発現の結果、クローニング効率が顕著に改善することが発見された。

【0026】

本発明に従うクローニング方法は第1 DNA分子および第2 DNA分子間の相同組換えを利用する。第1 DNA分子は宿主細胞中で機能し得る複製起点、例えば大腸菌複製起点を保有するあらゆるDNA分子が可能である。さらに、第1 DNA分子は宿主細胞中で複製される得る形態で存在する。この第1 DNA分子、すなわちベクターはその宿主細胞中で機能し得る複製起点を含有するあらゆる染色体外DNA分子が可能であり、例えば単一の、低、中若しくは高コピープラスミドを含むプラスミド、またはコスミド、P1、BAC若しくはPACベクター技術を基礎とするその他の染色体外環状DNA分子である。こうしたベクターの例は例えばSambrook ら (Molecular Cloning, Laboratory Manual, 2nd Edition(1989), Cold Spring Harbor Laboratory Press) およびLoannouら (Nature Genet. 6(1994), 84-89) またはそれらの中に引用された参照中に記載されている。第1 DNA分子は宿主細胞染色体、特に大腸菌染色体でも可能である。好ましくは、第1 DNA分子は2本鎖DNA分子である。

【0027】

第2 DNA分子は好ましくは線状DNA分子で、第1 DNA分子上の領域と相同な配列、好ましくは同一の配列領域を少なくとも2つ含む。これらの相同または同一領域はそれぞれ好ましくは少なくとも15ヌクレオチド、さらに好ましくは少なくとも20ヌクレオチド、最も好ましくはそれぞれ少なくとも30ヌクレオチドである。長さが約40またはそれ以上、例えば60またはそれ以上のヌクレオチドの配列相同領域を使用した場合、特に良好な結果が得られた。2つの配列相同領域は線状DNA分子上に1つが一方の末端、そしてもう1つが他方の末端になるように配置することができるが、内部に配置することもできる。好ましくは、第2 DNA分子も2本鎖DNA分子である。

【0028】

2つの配列相同領域は実験による設計にしたがって選択される。第1 DNA分子のどの領域を第2 DNA分子上に配置する2つの配列相同領域として選択するかについては、相同組換え現象が第1 DNA分子の複製起点を欠失させてはならないこと以外には、制限はない。相

10

20

30

40

50

同組換え反応のために十分な配列相同性が維持されている限り、配列相同領域が非同一配列領域によって遮断されていてもよい。ターゲット部位に対して非相同配列領域を有する配列相同アームを使用することによって、置換、例えばポイント突然変異、挿入および／または欠失などの突然変異をETクローニングによってターゲット部位に導入することができる。

【0029】

細菌細胞中にクローン化しようとする第2の外来DNA分子はあらゆる起源から誘導することができる。例えば、第2 DNA分子をPCRなどの核酸増幅反応によって合成してもよい。この場合、増幅のためのプライマーとして作用する3'末端の配列の他に、増幅を開始させるために使用するDNAオリゴヌクレオチドの両方が2つの相同領域の一方または他方を含有する。この設計のオリゴヌクレオチドを使用する場合、DNA増幅産物は増幅に好適な任意のDNA配列が可能であり、この他に各末端に配列相同領域の1つを有することになる。

10

【0030】

第2 DNA分子の作製の特定の例は、ある表現型の差異、特に抗生物質耐性を細菌宿主細胞に与えるように作用する遺伝子の増幅である。この操作法の単純な変形として、PCRプライマー配列および配列相同領域の他に別の配列を含むオリゴヌクレオチドの使用を含む。また別の単純な変法は増幅産物を作製するための増幅プライマーの3以上の使用である。また別の単純な変法は増幅産物を作製するための増幅反応の2以上の使用である。さらに別の変法はPCR以外の方法によって、例えばエンドヌクレアーゼまたは制限酵素切断によって任意のDNA起源からの断片を線状化して取得されるDNA断片の使用である。

20

【0031】

第2 DNA分子は必ずしも単一の種類のDNA分子である必要はないことに注意すべきである。異種集団の第2 DNA分子の使用、例えばゲノムなどのDNAライブラリーまたはcDNAライブラリーの作製もちろん可能である。

【0032】

本発明の方法に第1および第2 DNA分子のin vivoでの接触を含ませてもよい。本発明の1実施形態において、あらかじめ第1ベクターDNA分子を保有させてある細菌株中に第2 DNA断片を形質転換させる。別の実施形態において、第2 DNA分子および第1 DNA分子を細菌宿主中に同時形質転換する前に in vitroで混合する。本発明のこの2つの実施形態を図1に図式的に示す。その形質転換の方法は当技術分野で知られている任意の方法とすることができる（例えばSambrookら、上記）。しかし、好ましい形質転換または同時形質転換の方法はエレクトロポレーション法である。

30

【0033】

ETクローニング機構を介する第1および第2 DNA分子間の相同組換えに好適な条件下で第1および第2 DNA分子を接触させた後、該第1および第2 DNA分子間の相同組換えが生じた宿主細胞を選択する。この選択操作はいくつかの異なる方法で実施することができる。次の3つの好ましい選択方法を図2に示し、以下に詳細に記載する。

【0034】

第1の選択方法において、配列相同な2つの領域の間に配置するマーカーのための遺伝子を保有する第2 DNA断片を利用し、そのマーカー遺伝子の発現によって相同組換えを検出可能にする。マーカー遺伝子は宿主中または第1 DNA分子から発現されない表現型マーカーのための遺伝子とすることができる。ETクローニングによる組換えに際し、第2 DNA断片の確実な獲得によって与えられる宿主株の表現型における変化からETクローニング産物を同定する。

40

【0035】

好ましい事例において、表現型マーカーは抗生物質に対して耐性を与える遺伝子、特にカナマイシン、アンピリシン、クロラムフェニコール、テトラサイクリン、または使用する細菌株に対して殺菌または静菌効果を示すあらゆるその他の物質に耐性を与える遺伝子である。

【0036】

50

単純な変法は使用する細菌宿主株内に存在するある欠失を補足する遺伝子の使用である。例えば、宿主株がある代謝補充物がないと増殖できないように突然変異させる。この補充物が存在しないとき、第2 DNA断片上の1 遺伝子が突然変異による欠損を補足して増殖を可能にする。ETクローニングステップの結果、この目的のDNA再編成を有するエピソームを含有する細胞のみが増殖することになる。

【0037】

別の例において、宿主株は、コドンの1つがオープンリーディングフレームを切断する停止コドンとなるように突然変異させた表現型マーカー遺伝子を保有する。この表現型マーカー遺伝子からの全長タンパク質の発現には、発現されるとその停止コドンを認識して全オープンリーディングフレームの翻訳を可能にする、サプレッサーtRNA遺伝子の導入を必要とする。ETクローニングステップによってこのサプレッサーtRNA遺伝子が導入され、その表現型マーカー遺伝子の発現についての選択または同定によって、成功した組換え体を同定する。これらの場合、ETクローニングステップの結果、この目的のDNA再編成を有する細胞のみが増殖することになる。

10

【0038】

さらに別の単純な変法は、コロニーの色または形態に容易に検出し得る変化を与えるリポーター遺伝子の使用である。好ましい1 事例において、緑色蛍光タンパク質（GFP）を使用して、GFPの蛍光放出によってETクローニング産物を保有するコロニーを同定することができる。別の好ましい事例において、lacZ遺伝子を使用して、培地にX-galを添加したときのコロニーの青色の発色によってETクローニング産物を保有するコロニーを同定することができる。

20

【0039】

第2の選択方法において、ETクローニングによる第2 DNA断片の第1 DNA分子中への挿入が第1 DNA分子上に存在するマーカーの発現を変更する。この実施形態において、第1 DNA分子は配列相同な2つの領域の間に少なくとも1個のマーカー遺伝子を含有し、変更された発現、例えばそのマーカー遺伝子の発現の欠如によって相同組換えを検出することができる。

【0040】

好ましい適用例において、第1 DNA分子上に存在するマーカーはsacB、ccdBまたはテトラサイクリン耐性遺伝子などの対抗選択性遺伝子産物である。これらの場合、第2 DNA断片による形質転換または第2 DNA断片および第1 DNA分子の同時形質転換後、ETクローニングによって改変されなかった第1 DNA分子を保有する細菌細胞を培地上に接種したとき、対抗選択性マーカーの発現が宿主に対して毒性または静菌性効果をもたらす。ETクローニングステップの結果、この目的のDNA再編成を有する第1 DNA分子を含有する細菌細胞のみが増殖することになる。

30

【0041】

別の好ましい適用例において、第1 DNA分子がコロニーの色または形態を容易に検出し得る変化を与えるリポーター遺伝子を保有する。好ましい1 事例において、第1 DNA分子上に緑色蛍光タンパク質（GFP）を存在させて、GFPの蛍光放出の差異によって、ETクローニング産物を持つかまたは持たない第1 DNA分子を保有するコロニーを識別することができる。別の好ましい事例において、lacZ遺伝子を第1 DNA分子上に存在させて、培地にX-galを添加したときの青色または白色といったコロニーの色によってETクローニング産物を持つかまたは持たない第1 DNA分子を保有するコロニーを識別することができる。

40

【0042】

第3の選択法において、ETクローニングによる第2 DNA断片の第1 DNA分子への組み込みにより、部位特異的なリコンビナーゼに対する標的部位が取り除かれる。本明細書中では該標的部位をRT（リコンビナーゼの標的の意）と言い、該部位は配列が相同的な二つの領域間の第1 DNA分子上に存在する。相同組換え事象は、標的部位の除去によって検出され得る。

【0043】

50

ETクローニング産物非存在下では、RTは対応する部位特異的リコンビナーゼによる使用のために利用できる。このRTが存在するかしないかの差はETクローニング産物の選択における基盤となる。このRTおよび対応する部位特異的リコンビナーゼの存在下では、部位特異的リコンビナーゼはこのRTにおける組換えを媒介して宿主の表現型を変化させ、容易に観察できる表現型が増殖または存在できないようにする。このRTの非存在下では、対応する部位特異的リコンビナーゼは組換えを媒介できない。

【0044】

好ましい事例においては、第2 DNA断片が標的とする第1 DNA分子は2つのRTを含む。そのうちの1つは抗生物質耐性遺伝子に隣接するがその一部ではない。第2 DNA断片は、このRTを取り除くように設計される。対応する部位特異的リコンビナーゼにさらされると、ETクローニング産物を担持しない第1 DNA分子は、抗生物質耐性遺伝子を取り除くRT間での部位特異的組換え反応を受ける。そのため第1 DNA分子は、対応する抗生物質に対する耐性を伝達できない。ETクローニング産物を含む第1 DNA分子、または他の原因で部位特異的組換えを受けなかった第1 DNA分子のみが、抗生物質耐性を伝達する。

【0045】

別の好ましい事例においては、第2 DNA断片のETクローニングによって取り除かれるRTは、使用する宿主株に存在する欠損を補う遺伝子に隣接する。別の好ましい事例では、第2 DNA断片のETクローニングによって取り除かれるRTは、コロニーの色または形態における容易に検出可能な程度の変化を伝達するレポーター遺伝子に隣接する。

【0046】

別の好ましい事例においては、第2 DNA断片のETクローニングによって取り除かれるRTは、第1 エピソームDNA分子上のどこにでも存在し、該エピソームは宿主ゲノムに組み込まれると細菌宿主細胞の生存とは両立しない複製起点を持つ。この場合、宿主ゲノムは第2のRTを担持しており、そのRTは対応する部位特異的リコンビナーゼが宿主ゲノム中に位置するRTにそのRTを介してエピソームを組み込み得るような突然変異を受けたRTであってもそうでなくてもよい。その他の好ましいRTとして、リゾルベース/トランスポザーゼ・クラスの部位特異的リコンビナーゼに対するRTが挙げられる。RTには、部位特異的組換えの従来の実施例に記載されたもの、それらの天然のまたは突然変異を受けた変種も含む。

【0047】

好ましい部位特異的リコンビナーゼは、Cre、FLP、Kwまたはインテグラーゼ・クラスのいかなる部位特異的リコンビナーゼをも含む。その他の好ましい部位特異的リコンビナーゼには、リゾルベース/トランスポザーゼ・クラスの部位特異的リコンビナーゼがある。

【0048】

宿主細胞中での部位特異的リコンビナーゼの発現法には制限がない。好ましい方法では、部位特異的リコンビナーゼの発現を、ETクローニング効率の最適化に従って発現が誘導および停止されるように調節する。この場合、部位特異的リコンビナーゼ遺伝子を、宿主ゲノムに組み込むかまたはエピソーム上に担持させてもよい。別の好ましい事例では、宿主細胞から除去されるように部位特異的リコンビナーゼを条件的複製起点を持つエピソームから発現させる。

【0049】

別の好ましい事例では、上記の3つ選択方法の内の少なくとも2つを組合せる。特に好ましい事例は、上記の第1の選択方法の2工程に続いて第2の選択方法を使用するものである。この組合せの使用においては最も簡単なことだが、クローニングされるDNA断片が、表現型変化の獲得によって正確なETクローニング産物を第1工程において同定できるような遺伝子または遺伝子群を含む必要がある。第2工程では、第2周期のETクローニング産物が同定可能なように、第1工程で導入された遺伝子また遺伝子群の発現を変化させる。好ましい実施例では、使用する遺伝子はテトラサイクリン耐性遺伝子であり、第1工程のETクローニング産物はテトラサイクリン耐性の獲得によって同定される。第2工程では、テトラサイクリン遺伝子発現の減失は、塩化ニッケル、フザリン酸またはテトラサイクリン遺伝子が発現したときに宿主細胞に対して毒性を有する他のあらゆる薬剤に対する感応

性の減失によって同定される。この2工程の工程は、第1に宿主の表現型変化をもたらす遺伝子の組み込みにより、第2に関連する表現型の変化の減失により、もっとも単純には第1工程で組み込まれたいくつかのDNA配列の除去により、ETクローニング産物の同定を可能にする。この方法によって、ETクローニング産物の同定のために使用する遺伝子は挿入され、続いて取り除かれ、これらの遺伝子を含まないETクローニング産物が残る。

【0050】

本発明の他の実施形態では、ETクローニングを以下の工程を含む遺伝子組み換え法に用いてもよい。すなわち a) RecEおよびRecTまたはRed α およびRed β タンパク質の供給源の提供、b) 適当な宿主細胞中で複製され得る第1 DNA分子と該第1 DNA分子上の領域と配列相同性を持つ領域を少なくとも二つ含む第2 DNA分子との、該第1 および第2 DNA分子間の相

10

【0051】

RecEおよびRecT、またはRed α およびRed β タンパク質の供給源は、精製されたまたは部分的に精製されたRecEおよびRecTもしくはRed α およびRed β タンパク質、またはRecEおよびRecTもしくはRed α およびRed β タンパク質を含む細胞抽出物のいずれかでよい。

【0052】

本実施形態における相同組換え事象はin vitroで起こり得る。例えば、相同組換え事象に必要な他の成分を含む細胞抽出物を提供するときである。しかしながら相同組換えはin vivoでも起こり得る。例えば、RecEおよびRecTもしくはRed α およびRed β タンパク質、または抽出物を宿主細胞 (recET陽性であってもなくてもよく、またred α β 陽性であってもなくてもよい) 中に導入し、宿主細胞中のDNA分子に接触させることによる。組換えがin vitroで生じるとき、DNA分子の選択は、組換え混合物を適した宿主細胞中に形質転換し、陽性クローンを前記のように選択することにより達成される。組換えがin vivoで生じるとき、上記選択法を直接利用できる。

20

【0053】

本発明の更なる主題は、相同組換えを含むクローニング法のための宿主細胞として、細胞、好ましくは細菌細胞、最も好ましくはrecEおよびrecTまたはred α およびred β 遺伝子を発現し得る大腸菌細胞の使用である。

【0054】

依然として、本発明の更なる主題は宿主細胞中のrecEおよびrecTまたはred α およびred β 遺伝子を発現し得るベクター系および該ベクターの相同組換えを含むクローニング法への使用である。好ましくは、ベクター系は上述のように、例えば λ red γ 遺伝子などへのエキソヌクレアーゼ阻害遺伝子をも発現し得る。該ベクター系は少なくとも一つのベクターを含みうる。recEおよびrecTまたはred α およびred β 遺伝子は、好ましくは単一ベクター上に位置し、より好ましくは両方の遺伝子と同一であり得る調節可能なプロモーター、またはそれぞれの遺伝子に対する単一のプロモーターの制御下にある。特に好ましいのは、recEまたはred α 遺伝子に対してrecTまたはred β 遺伝子を過剰に発現し得るベクター系である。

30

【0055】

依然として本発明の更なる主題は、相同組換えを含むクローニング法におけるRecEおよびRecTまたはRed α およびRed β タンパク質の供給源の使用である。

40

【0056】

本発明の更に別の主題は、

(a) 宿主細胞、好ましくは細菌宿主細胞、

(b) 宿主細胞中のrecEおよびrecTまたはred α およびred β 遺伝子の発現手段、例えばベクター系を含むもの、ならびに

(c) レシピエントクローニングビヒクル、例えば、上記細胞中で複製され得るベクター、からなる試薬キットである。

【0057】

50

一方、本発明の方法における第1 DNA分子に対応するレシピエントクローニングビヒクルは、細菌細胞中に既に存在し得る。他方、それは細菌細胞と離れた状態でも存在し得る。

【0058】

別の実施形態では試薬キットは、

(a) RecEおよびRecT、またはRed α およびRed β タンパク質の供給源、ならびに

(b) 宿主細胞中で増殖し得るレシピエントクローニングビヒクル、ならびに

(c) 任意に、前記レシピエントクローニングビヒクルの増殖に適した宿主細胞、からなる。

【0059】

試薬キットはさらに、特にレシピエントETクローニング産物が少なくとも一つの部位特異的リコンビナーゼ標的部位を有するときに、宿主細胞中で部位特異的リコンビナーゼを発現する手段を含むことが好ましい。さらに、試薬キットはETクローニングに使用する線状DNA断片の供給源としての使用に適したDNA分子を含んでいても良い。好ましくは、該DNA分子は、線状断片のPCR増殖のための鋳型として、また制限酵素分解により線状DNA断片を放出する特別に設計されたDNAベクターとして、また陽性対照またはその他の役割で使用するためにキット中に含まれる調製された線状断片として提供される。さらに試薬キットは、前記ベクターに相同的な領域を含む核酸増幅プライマーを含んでいてもよい。好ましくは、この相同的な領域は核酸増幅プライマーの5'末端に位置する。

【0060】

本発明をさらに次の配列表、図、実施例によって例示する。

【0061】

図および実施例

配列番号1は、プラスミドpBAD24-recET（図7）の核酸配列を示す。

【0062】

配列番号2／3は、pBAD24-recETの位置1320～2162に存在する末端切断型(truncated) recE遺伝子(t-recE)の核酸配列およびアミノ酸配列を示す。

【0063】

配列番号4／5は、pBAD24-recETの位置2155～2972に存在するrecT遺伝子の核酸配列およびアミノ酸配列を示す。

【0064】

配列番号6／7は、pBAD24-recETの位置974～996に示される配列に対する相補鎖上に存在するaraC遺伝子の核酸配列およびアミノ酸配列を示す。

【0065】

配列番号8／9は、pBAD24-recETの位置3493～4353に存在するbla遺伝子の核酸配列およびアミノ酸配列を示す。

【0066】

配列番号10は、プラスミドpBAD-ET γ （図13）の核酸配列を示す。

【0067】

配列番号11は、遺伝子red α (1320～2000)、red β (2086～2871)およびred γ (3403～3819)をコードする領域と同様に、プラスミドpBAD- α β γ （図14）の核酸配列を示す。

【0068】

配列番号12-14は、Red α 、Red β およびRed γ タンパク質のアミノ酸配列をそれぞれ示す。red γ 配列はpBAD-ET γ （図13）およびpBAD- α β γ （図14）のそれぞれに存在する。

【0069】

1. 方法

1.1. 直鎖断片の調製

直鎖DNA断片を増幅するために、標準的なPCR反応の条件を用いた。用いたプライマーの塩基配列は表1に示した。

【0070】

表1

10

20

30

40

50

pJP5603(Penfold and Pemberton, Gene 118(1992), 145-146)由来のTn5-neo遺伝子が、オリゴペアであるa/bとc/dを用いて増幅された。pMAK705(Hashimoto-Gotoh and Sekiguchi, J.Bacteriol.131(1977), 405-412)由来のクロラムフェニコール(cm)耐性遺伝子が、プライマーペアであるe/fとn/pを用いて増幅された。FRTまたはloxP部位には含まれたTn5-新遺伝子はpKaZまたはpKaX(<http://www.embl-heidelberg.de/ExternalInfo/stewart>)からオリゴペアi/h, g/h,およびj/kを用いて増幅されたものである。pIB279(Blomfield et al., Mol.Microbiol.5(1991),1447-1457)由来のsacB-neoカセットはオリゴペアl/mを用いて増幅された。pSVpaZ11(Buchholz et al., Nucleic Acids Res.24(1996), 4234-4262)由来のlacZ遺伝子断片はオリゴペアl*/m*を用いて増幅された。PCR産物は、QIAGEN PCR Purification Kitを用いて精製し、H₂O₂にて溶出した後に、Dpn Iにより残りの鋳型DNAをすべて消化した。消化後、PCR産物をフェノール：CHCl₃により一度抽出し、そしてエタノール沈殿後、H₂O中に約0.5 µg/µlの濃度で再懸濁した。

10

【0071】

1.2 コンピテント細胞の調製とエレクトロポレーション

一晩培養した飽和状態の培養物をLB培地にて50倍に希釈し、OD600が0.5となるまで増殖させ、氷中で15分間冷却した。細菌細胞を0℃、7000rpmで10分間遠心した。ペレットを氷冷10%グリセロール中に再懸濁し、再度遠心した(7000rpm、-5℃で10分)。この操作をさらに2回繰り返し、細胞ペレットを等量の氷冷10%グリセロールに懸濁した。50 µlのアリコート管を液体窒素中で凍結し、-80℃で保存した。細胞を氷上で解凍し、1 µl DNA溶液(同時形質転換のために、0.3 µgのプラスミドと0.2 µgのPCR産物; または形質転換のために、0.2 µgのPCR産物を含む)を添加した。エレクトロポレーションは、氷冷したキューベットとBio-Rad Gene Pulserを200オームに設定したPulse Controllerとともに25 µFD、2.3kVの設定で用いて行った。エレクトロポレーションののち、LB培地(1ml)を添加した。細胞を37℃にて1時間振とうしながらインキュベートし、抗生物質プレート上に塗布した。

20

【0072】

1.3 RecEおよびRecT発現の誘導

pBAD24-recETを保有する大腸菌JC5547を、0.2%グルコース、100 µg/mlアンピシリンを含むLB培地中で一晩培養した。1/100接種により、1つ(0)は0.2%グルコースを含むようにして、計5本のLBを平行して始めた。培養物を37℃で振とうしながら4時間インキュベートし、回収の3, 2, 1, または1/2時間前に0.1%L-アラビノースを添加し、続いて前述のように処理した。回収の直前に、100 µlを採取し、10%SDS-ポリアクリルアミドゲルにて分析した。回収の90分前に、Hoxa-P1とpBAD-ETγを保有する大腸菌NS3145を0.1%L-アラビノースにより誘導した。

30

【0073】

1.4 FLPおよびCre発現プラスミドの一時的形質転換

FLPおよびCre発現プラスミドである、pSC101温度感受性オリジンをベースとする705-Creおよび705-FLP(Buchholz et al, Nucleic Acids Res. 24(1996), 3118-3119)を、塩化ルビジウムコンピテント細菌細胞中に形質転換した。細胞を、25 µg/mlクロラムフェニコールプレート上に塗布し、30℃で2日間増殖させた。そこから、コロニーを採り、抗生物質を含まないL-アガープレートに再度塗布し、40℃で一晩インキュベートした。単一コロニーについて、さまざまな抗生物質プレート上で分析したところ、そのすべてが期待どおりクロラムフェニコールおよびカナマイシン耐性を欠失していた。

40

【0074】

1.5 sacB発現のスクロース対抗選択

図11に記載されているように調製された大腸菌JC9604lacZ株を、sacB-neoPCR断片およびpSVpaX1(Buchholz et al, Nucleic Acids Res, 24(1996), 4256-4262)により同時形質転換した。100 µg/mlアンピシリン、50 µg/mlカナマイシンプレート上で選択した後、pSVpaX-sacB-neoプラスミドを単離し、pSVpaX1からプライマーl*/m*を用いて増幅されたPCR断片とともに、新鮮なJC9604lacZ細胞に同時形質転換した。オリゴm*はBamH1部位を生じる

50

サイレントな点突然変異を保持していた。細胞を、7%スクロース、100 μ g/mlアンピシリン、40 μ g/ml X-galを含むプレート上に塗布し、28°Cで2日間インキュベートした。スクロースプレート上で増殖させた青色コロニーと白色コロニーを計数し、さらに制限酵素による分析を行った。

【0075】

1.6 他の方法

DNA調製およびサザン分析は、標準的な手順に従って行った。ハイブリダイゼーションプローブは、Tn5 neo遺伝子(PvuII)、Hoxa3遺伝子(両方のHindIII断片)、lacZ遺伝子(pSVpa X1由来のEcoRIおよびBamHI断片)、cm遺伝子(pMAK705由来のBstBI断片) およびP1ベクター断片 (P1ベクター由来の2.2kb EcoRI断片) より単離された断片のランダムなプライミ

10

【0076】

2. 結果

2.1 大腸菌における組換え現象の確認

大腸菌におけるフレキシブルな相同組換え反応を確認するために、直鎖DNAと環状DNA間の組換えに基づくアッセイを設計した (図1, 図3)。Tn5カナマイシン耐性遺伝子 (neo) をもつ直鎖DNAはPCRにより調製した (図3a)。はじめに、neoのPCR増幅に用いられたオリゴヌクレオチドは、5'末端にプラスミド中の選択した領域と相同の42ヌクレオチドを、3'末端にPCRプライマーとして提供するための18ヌクレオチドを有する60量体であった。直鎖DNA および環状DNAは等モルの比で混合し、さまざまな大腸菌宿主に同時形質転換した。相同組換えはsbcA大腸菌宿主のなかでのみ検出された。アンピシリン/カナマイシン両耐性のコロニー (図3b)の95%以上が期待された相同組換えプラスミドを含んでいることが、制限酵素消化と塩基配列決定によって示された。neo遺伝子のゲノムへの組み込みによる、カナマイシン耐性の低いバックグラウンドのみがみられた (結果示さず)。

20

【0077】

直鎖および環状の組換え反応は2つの方法で特色づけられた。相同的アームの長さで組換え効率の関係は単純であり、長さが長いほど効率が高かった (図3c)。効率は、試した範囲内で、60bpまで増加した。レシピエント側のプラスミド中の二つの選択された相同部位間の距離の効率を調べた (図3d)。同一の左相同的アームをさまざまな右相同的アームと組み合わせることで8つのPCR断片のセットを作製した。右相同的アームは、プラスミド塩基配列から、左から0-3100bpとなるように選択した。挿入産物(0)はほとんど有効でないが、あとはすべてのケースから、4倍未満の差異はあるものの、正しい産物が容易に得られた。相同的アームが片方しかないPCR断片は反応しないので、正しい産物をえるには相同的アームの両方が存在することも必要であった。

30

【0078】

2.2 RecEとRecTの含有

宿主の遺伝子型およびこの相同組換え反応の間の関係は、種々の組換え要素の欠損した大腸菌株の一覧を利用してさらに系統的に調べた(表2)。

【0079】

表2

40

2つのsbcA株、JC8679およびJC9604のみが意図した組換え産物を有し、RecAは必要ではなかった。sbcA株では、RecEとRecTの発現は活性化されている。JC8679とJC8691の比較から、recEへの依存性が推論される。特筆すべきことに、JC9387中では組換え産物が見出されず、これは、sbcBCバックグラウンドでは、50ヌクレオチドの相同的アームに基づく相同組換えを起こすには至らないことを示唆する。

【0080】

RecEおよびRecTが含まれることを示すために、recETオペロンの一部を誘導性発現ベクター中にクローン化し、pBAD24-recETを作製した (図6a)。このときRecEの最初の588アミノ酸は無くてもよいので、recE遺伝子のN末端を切断した。recBC株、JC5547をpBAD24-recETで形質転換し、コンピテント細胞の回収の前に、培地に、種々の時点でアラビノースを加

50

えてRecE/RecTの誘導を経時的に行った。回収されたコンピテント細胞のバッチは、ゲル電気泳動によりタンパク質発現について評価し（図6c）、直鎖DNA断片と内因性pBAD24-recETプラスミドとの間の組換えについて評価した。RecE/RecT発現の増加に従って組換えも増加したが、RecE/RecTの誘導がなければ、組換え産物は見出されなかった。この実験はまた、直鎖DNAおよび環状DNAの同時形質転換は必ずしも必要でなく、環状のレシピエントは宿主中に内在するものでもよいことを示した。図3、6および表2に示された結果から我々はRecEおよびRecTがrecBC大腸菌において、とても有用な相同組換え反応を、有用な頻度で仲介するとの結論を導いた。RecEとRecTが含まれるので、我々は、この直鎖DNAおよび環状DNAの組換え法を“ETクローニング”と呼ぶ。

【0081】

10

2.3 ETクローニングの大型標的DNAへの応用

大型DNAエピソームが大腸菌中で操作できることを示すために、インタクトなマウスHoxa複合体の少なくとも59kbを含む（DNA配列決定およびサザンブロッティングにより確認された）、 $a > 76\text{kb}$ のP1クローンを、sbcAバックグラウンドをもつ大腸菌株(JC9604)に導入し、2周期の“ETクローニング”に供した。第一周期では、P1ベクター上のTn903カナマイシン耐性遺伝子をアンピシリン耐性遺伝子に置き換えた（図4）。第2周期では、Hoxa3とa4遺伝子間を標的とし、これは、Hoxa3の近位配列プロモーターの上流の2塩基対間にneo遺伝子を挿入するか、Hoxa3とa4遺伝子の間の6203bpを削除することによって行った（図8a）。挿入および削除を行ったETクローニング産物は容易に得られ（図8b, レーン2, 3および5）、2周期のETクローニングが、この大型大腸菌エピソームにおいて、明瞭な予想外の組換えを起こすことなく、正確に行われたことを示した。

20

【0082】

大腸菌染色体中の遺伝子を標的とすることで、ETクローニングの一般的な応用の可能性をさらに調べた（図9a）。正しい組換え産物と正しくない組換え産物の比が β -ガラクトシダーゼ発現を評価することにより決定できるように、JC9604の β -ガラクトシダーゼ(lacZ)遺伝子を選択した。標準的条件（ $0.2\mu\text{g}$ PCR断片； $50\mu\text{l}$ コンピテント細胞）で、24個の1次コロニーができ、 β -ガラクトシダーゼ発現（図9b）とDNA分析（図9c, レーン3-6）によりそのうちの20個が正しかった。

【0083】

2.4 操作配列を除去するための2次的な組換え反応

30

上記のETクローニングの産物は、選択マーカー遺伝子を含んでいなければならないという点で、制約がある。この制約を取り除くためのさらなる組換え工程を利用する2つの異なる方法を開発した。第一の方法では、FlpまたはCreリコンビナーゼにより仲介される、部位特異的組換えを採用した。図8および9の実験で、Flp組換え標的部位(FRT)またはCre組換え標的部位(loxP)が、直鎖基質中のneo遺伝子をはさむように包含された。FRTまたはloxP間の組換えは、部位特異的組換えののち、温度を変えることでプラスミドの単純な消去ができるように、pSC101温度感受性複製起点(Hashimoto-Gotoh and Sekiguchi, J. Bacteriol. 131(1977), 405-412)をもつプラスミドにより発現される、FlpまたはCreを用いてなされた。ETおよびFlp組換えののち、正しく組み換えられたHoxa P1ベクターを、他の組換え産物の出現をみずに回収した（図8, レーン4および6）。同様に、Creリコンビナーゼも標的となったlacZ対立遺伝子（図9, レーン7～10）を正しく組換えた。このように部位特異的組換えは、容易にETクローニングと組み合わせることができ、これにより、操作上の一連の過程を省略することができ、またDNA操作の点において、34bpの部位特異的組換え標的部位を残すことが可能である。選択マーカー遺伝子を除去する第2の方法では、いかなる操作上の過程をも経ていないDNA産物を残すために、ポジティブおよび対抗選択の過程を組み合わせた2周期のETクローニングを用いた。

40

【0084】

さらにこの実験は、ETクローニングが操作された領域内でフレームシフトや点突然変異などの小規模の突然変異を促進するかどうかを、 β -ガラクトシダーゼ活性にもとづく機能テストによって評価できるように設計した。第1の周期で、pSVpaX1のlacZ遺伝子は、neo

50

および*B. subtilis* *sacB* (Blomfieldら, *Mol. Microbiol.* 5 (1991), 1447-1457) 遺伝子を含む3.3kbのPCR断片により分断されたことが、カナマイシン耐性に対する選択によって確かめられた(図10a)。他のポジティブに選択された組換え産物に対して上に記載したように、現実には、選択されたコロニーはすべて白色であって(図10b)、*LacZ*の分断に成功していることを示しており、DNA分析によって、17個中17個が正しい組換え体であることが確かめられた。第2周期では*lacZ*を修復するように設計された1.5kb PCR断片が*sacB* 遺伝子に対する対抗選択により導入された。*lacZ*の修復は*Bam*H1制限部位を生じるサイレントな点突然変異を含んでいた。スクロース耐性コロニーのうち約4分の1が β -ガラクトシダーゼを発現し、分析された(17個中17個; 図10c) すべてが*Bam*H1点突然変異とともに、修復された*lacZ*遺伝子を保有していた。スクロース耐性コロニーの残りの4分の3は β -ガラクトシダーゼ発現せず、分析されたすべて(17個中17個; 図10c) がさまざまな大規模な突然変異を起こしていて、そのうちには、ETクローニング産物に類似のものはみられなかった。このように2周期の、*lacZ*遺伝子を標的とするETクローニングにおいて、小規模の突然変異による β -ガラクトシダーゼ活性の障害はみられず、*RecE/RecT* 組換えが高度な正確さをもって進行することが示唆された。対抗選択の工程でみられた不正確な産物の有意な存在は、対抗選択遺伝子の発現を阻止するいかなる突然変異をも選択されてしまうことによる、対抗選択の利用に随伴する限界であるといえる。周知の通り、すべての不正確な産物は広範囲の突然変異によるものであるから、正確なET産物とはDNA解析により容易に区別できた。またべつの実験で(図5) 我々は、*ccdB*遺伝子に対する対抗選択による、*pZero2.1*(*In Vitro* Gen)へのETクローニングにより、バックグラウンドの少ない不正確な産物(8%)が得られることを観察しており、これはETクローニング効率に影響するものとは異なる要因により、対抗選択のバックグラウンドがばらつくことを示唆していた。

【0085】

2.5 大腸菌宿主間のETクローニングの転移

上記の実験は、*sbcA*突然変異が*recBC*のサプレッサーとして同定されていたことから(Barbouroら, *Proc. Natl. Acad. Sci. USA* 67 (1970), 128-135; Clark, *Genetics* 78(1974), 259-271)、*recBC*大腸菌宿主中で行った。しかしながら、多くの有用な大腸菌株はP1, BAC, またはPACエピソームの増殖に通常使用される株も含めた*recBC+*である。*recBC+*株へのETクローニングの転用を行うために、我々は、*pBAD-ET γ* および *pBAD- $\alpha\beta\gamma$* を開発した(図13, 14)。これらのプラスミドはETクローニングの可動性に重要な3つの特徴を有する。第一は、*RecBC*が大腸菌の主要なエクソヌクレアーゼであり、導入された直鎖状の断片を分解する。それゆえに*RecBC*阻害剤の*Red γ* (Murphy, *J. Bacteriol.* 173(1991), 5808-5821)を保有させた。第2に誘導可能なプロモーターのもとに*recE*または*red α* を配置することで、*RecE/RecT*または*Red α /Red β* の組換え能力を制御した。その結果、必要なときに、ETクローニングを誘導でき、他のときには、望まない組換え現象は制限される。第3に*recE*または*Red α* ではみられないが、*RecT*または*Red β* が過剰に発現されるときに、ETクローニング効率が上昇することを観察した。それゆえに我々は*recT*または*red β* を強力な構成要素であるEM7プロモーターの下流に配置した。

【0086】

pBAD-ET γ を初期の*Hoxa* P1エピソームを保持する、NS3145大腸菌中に、形質転換した(図11a)。P1ベクター主鎖中の領域は*n*および*p*相同的アームにはさまれたクロラムフェニコール耐性遺伝子(*cm*)のPCR増幅標的とされた。ポジティブに選択されたETクローニング反応について上述したように、ほとんど(>90%)のクロラムフェニコール耐性コロニーが正確であった。周知のように、形質転換された直鎖状のDNAに関しては、*sbcA*宿主JC9604中の同じエピソームを標的とする類似の単純な実験に比べた場合、*pBAD-ET γ* を利用したETクローニング全体の効率は、3倍近く上昇した。これは、*RecT*の過剰発現によりETクローニング効率が上昇するという我々の観察と一致する。

【0087】

recA-, *recBC+*大腸菌株DK1において、*pBAD-ET γ* から発現された*RecE/RecT*、および*pBAD-*

α β γ から発現されたRed α /Red β に仲介されたETクローニング効率を比較した (図 1 2) 。 pBAD-ET γ またはpBAD- α β γ による大腸菌DK1の形質転換ののち、図6a, c, に示したと同様、pBADベクターのbla遺伝子をクロラムフェニコール遺伝子と入れ替えるための実験を行った。pBAD-ET γ またはpBAD- α β γ はいずれも、RecEおよびRed α のアラビノースによる誘導に対する反応、および目的とされた多くの事象に関して、同様のETクローニング効率を示した。

【0088】

【表1】

表1 PCR用オリゴの配列

10

Figure 3ab
左(left) TGACCCCTCACAAGGAGACGACCTTCCATGACCGAGTACAAGAGGGATGTAACGCACTGA
右(right) TACAAATGTGGTATGGCTGATTATGATCCTCTAGAGTCGGTGCTCACTGCCCGCTTTCCA
鋳型 : pJP5603
標的ベクター : pSV-paz11

Figure 3c
a-left: CTTCATGACCGAGTACAAGAGGGATGTAACGCACTGA
a-right: ATGATCCTCTAGAGTCGGTGCTCACTGCCCGCTTTCCA
b-left: AGACGACCTTCCATGACCGAGTACAAGAGGGATGTAACGCACTGA
b-right: GCTGATTATGATCCTCTAGAGTCGGTGCTCACTGCCCGCTTTCCA
c-left: CACAAGGAGACGACCTTCCATGACCGAGTACAAGAGGGATGTAACGCACTGA
c-right: TGGTATGGCTGATTATGATCCTCTAGAGTCGGTGCTCACTGCCCGCTTTCCA
d-left: TGACCCCTCACAAGGAGACGACCTTCCATGACCGAGTACAAGAGGGATGTAACGCACTGA
d-right: TACAAATGTGGTATGGCTGATTATGATCCTCTAGAGTCGGTGCTCACTGCCCGCTTTCCA
e-left: CACGCCCCTGACCCCTCACAAGGAGACGACCTTCCATGACCGAGTACAAGAGGGATGTAACGCACTGA
e-right: TAAAACCTCTACAAATGTGGTATGGCTGATTATGATCCTCTAGAGTCGGTGCTCACTGCCCGCTTTCCA
f-left: TCCCCTGACCCACGCCCCTGACCCCTCACAAGGAGACGACCTTCCATGACCGAGTACAAGAGGGATGTAACGCACTGA
f-right: TAAAGCAAGTAAAACCTCTACAAATGTGGTATGGCTGATTATGATCCTCTAGAGTCGGTGCTCACTGCCCGCTTTCCA
template: pJP5603
targeting vector: pSV-paz11

20

Figure 3d
a-left: TCATCCTCTGCATGGTCAGGTCATGGATGAGCAGACGATGGTGCAAGGATCAAGGGCTGCTAAAGGAA
a-right: TAATGCGAACAGCGCACGGCGTTAAAGTTGTTCTGCTTCATCAGCAGGATGGCGAAGAACTCCAGCAT
b-left: CACGAGCATCATCCTCTGCATGGTCAGGTCATGGATGAGCAGACGATGGCAAGGGCTGCTAAAGGAA
b-right: TAATGCGAACAGCGCACGGCGTTAAAGTTGTTCTGCTTCATCAGCAGGATGGCGAAGAACTCCAGCAT
c-left: TTAACCGTCACGAGCATCATCCTCTGCATGGTCAGGTCATGGATGAGCACAAGGGCTGCTAAAGGAA
c-right: TAATGCGAACAGCGCACGGCGTTAAAGTTGTTCTGCTTCATCAGCAGGATGGCGAAGAACTCCAGCAT
d-left: TGCTGCTGAACGGCAAGCCGTTGCTGATTTCGAGGCGTTAACCGTCACGACAAGGGCTGCTAAAGGAA
d-right: TAATGCGAACAGCGCACGGCGTTAAAGTTGTTCTGCTTCATCAGCAGGATGGCGAAGAACTCCAGCAT
e-left: TCTCTATCGTGGTGGTTGAACTGCACACCGCCGACGGCACGCTGATTCAAGGGCTGCTAAAGGAA
e-right: TAATGCGAACAGCGCACGGCGTTAAAGTTGTTCTGCTTCATCAGCAGGATGGCGAAGAACTCCAGCAT
f-left: TGGAGTGACGGCAGTTATCTGGAAGATCAGGATATGTGGCGGATGAGCGCAAGGGCTGCTAAAGGAA

30

40

f-right:
TAATGCGAACAGCGCACGGCGTTAAAGTTGTTCTGCTTCATCAGCAGGATGGCGAAGAACTCCAGCAT
g-left:
TGCATTCTAGTTGTGGTTTGCCAAACTCATCAATGTATCTTATCATGTCAAGGGCTGCTAAAGGAA
g-right:
TAATGCGAACAGCGCACGGCGTTAAAGTTGTTCTGCTTCATCAGCAGGATGGCGAAGAACTCCAGCAT
h-left:
TGCATTCTAGTTGTGGTTTGCCAAACTCATCAATGTATCTTATCATGTCAAGGGCTGCTAAAGGAA
h-right:
TATTTTGGACACCAGACCAACTGGTAATGGTAGCGACCGGCTCAGCTGGCGAAGAACTCCAGCAT
template: pJP5603
targeting vector: pSV-paz11

Figure 4

left:
TCAAGTGAGAAATACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCCACCAGC
TGGTATGGCTGATTATGATC
right:
TCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTGGGCAATCAGGTGCGACA
ATCTACCACCAGCTCTTTTCTACGGGGTCTGACGC
template: pBR322
targeting vector: Hoxa-P1

10

Figure 5

left:
TGCAGCTGGCAGCAGAGTTTCCGACTGGAAAGCGGGCAGTTAATACGACTCACTATAGGGAGAACA
GGAAACAGCTATGCCCATAAACCCAGAGTA
right:
TGCGCCGCTACAGGGCGCGTCCATTCGCCATTCAGGCCTGACTCACTAGTGATGGTGATGGTGATGTGG
GGGTGCCGCTCAGT
template: pmtrx (a pBluescript vector carrying mouse trithorax cDNA)
targeting vector: pZero2.1

20

Figure 6

left:
TGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGGAGAAAAAATCACT
GGATATACCACCG
right:
TACAGGGCGCGTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACGCCCCGCCCTGC
CACTCATCGCA
template: pMAK705
targeting vector: pBAD-24 backbone Amp resistant gene

Figure 8

i:
TGCCAAGCTTGACCCACTGTGGAAGTTCCAAAAAGCGGAAGGCTCTTGAGCTACTTCACTAACAAC
CGG
g:
TCACCATCTTCGGGCCATTTGTAGACTGGAATATTTGAGCTATGAGTGTGCTACTTCACTAACAACCG
G
h:
TGGCCCCAGGGTGACGCGGACATGGAGTTGTCGCCAGGGCACTGGTCCATGAGAGTGCCAAGCTACTC
GCGAC
template: pKaZ
targeting vector: Hoxa-P1

30

Figure 9

j:

TAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCT
TTGCCGTGGTTTATAACTTCGTATAGCATACATTATACGAAGTTATGGGCTGCTAAAGGAAGCGGAACAC
G

k:

TGGCAGTTCAGGCCAATCCGCGCCGGATGCGGTGTATCGCTCGCCACTTCAACATCAACGGTAATCGCC
ATTTGACCATATAACTTCGTATAATGTATGCTATACGAAGTTATCCCCAGAGTCCCGCTCAGAAGAACT
template: pJP5603

targeting vector: JC9604 chromosome

Figure 10

l:

TAGCTTGGCACTGGCCGTGTTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCCATCAC
ATATACCTGCCGTTCACTAT

m:

TATCGGTGGCCGTGGTGTGCGCTCCGCCGCTTCATACTGCACCGGGCGGGAAGGCGATTCCGAAGCCC
AACCTTTCATAGAAGCC

template: pIB279

targeting vector: pSV-paX1

l*: GCTTGGCACTGGCCGTGTTTTACAACGTCGTGACTGGGAA

m*:

TCGGTGGCCGTGGTGTGCGCTCCGCCGCTTCATACTGCACCGGGCGGGAAGGATCCACAGATTGATC
CAGCGATACAGC

template: pSV-paz11

targeting vector: pSV-sacB-neo

Figure 11

n:

TACCGCATTAAGCTTATCGATGATAAGCTGTCAAACATGAGAATTGACCCGGAACCTTCTCGAGGAA
GTTCTATTCTCTAGAAAGTATAGGAACTCCGAATAAATACCTGTACGGAAGATCACTT

p:

TTCCCTCAAGAATTTTACTCTGTGAGAAACGGCCTTAACGACGTAGTCGAGGGACCTAGAAGTTCCTAT
ACTTTCTAGAGAATAGGAACTTCATTATCACTTATTCAGGCGTAGCACCAGGCG

template: pMAK705

targeting vector: Hoxa-P1

Figure 12

left:

TGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGGAGAAAAAATCACT
GGATATACCACCG

right:

TACAGGGCGCGTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACGCCCCGCCCTGC
CACTCATCGCA

template: pMAK705

targeting vector: pBAD-24 backbone Amp resistant gene

【表 2】

大腸菌株	遺伝子型	Amp+Kan	Amp x 10 ⁸ /μg
JC8679	<i>recBC sbcA</i>	318	2.30
JC9604	<i>recA recBC sbcA</i>	114	0.30
JC8691	<i>recBC sbcA recE</i>	0	0.37
JC5547	<i>recA recBC</i>	0	0.37
JC5519	<i>recBC</i>	0	1.80
JC15329	<i>recA recBC sbcBC</i>	0	0.03
JC9387	<i>recBC sbcBC</i>	0	2.20
JC8111	<i>recBC sbcBC recF</i>	0	2.40
JC9366	<i>recA</i>	0	0.37
JC13031	<i>recJ</i>	0	0.45

【図面の簡単な説明】

【図1】 ETクローニングに好ましい方法をダイアグラムによって示す。クローニングする線状DNA断片を、レシピエントエピソーム中の配列およびPCR反応のプライミングのための配列、にマッチする(match)ように選んだ左の相同的アーム、ならびにレシピエントエピソーム中の別の配列およびPCR反応のプライミングのための配列にマッチするように選択した右の相同的アームを含むオリゴヌクレオチドプライマーを用いるPCRで合成する。PCR反応の産物は、本明細書中では選択マーカー遺伝子(sm1)であるが、結果として左右の相同的アームに隣接し、同時形質転換の前にin vitroでエピソームと混合し得る。また、標的エピソームを持っている宿主細胞中に形質転換し得る。宿主細胞はrecEおよびrecT遺伝子の産物を含む。ETクローニング産物はレシピエントエピソーム上の2つの選択マーカー-sm1およびsm2の組み合わせによって同定する。

【図2】 ETクローニング産物の3つの同定法を示す。1つ目は(図の左)、ETクローニングによる、宿主に表現型の違いをもたらす遺伝子の獲得を示す。該遺伝子は、ここでは選択マーカー遺伝子(sm)である。2つ目は(図の中央)、ETクローニングによる宿主に表現型の違いをもたらす遺伝子の欠失を示す。該遺伝子はここでは、対抗選択マーカー遺伝子(counter-sm)である。3つ目はETクローニングによる部位特異的リコンビナーゼ(SSR)に対する標的部位(RT、環状エピソーム上に三角形で示す)の欠失を示す。この場合、正しいETクローニング産物は、SSRが選択マーカー遺伝子(sm)を欠失させるのに必要な標的部位の1つを欠失する。SSRが該sm遺伝子を欠失できなかったことにより、正確なETクローニング産物を同定する。

【図3】 ETクローニングの簡単な実施例を提示する。

(a)上のパネル—PCRによってカナマイシン耐性遺伝子を増幅するように、およびレシピエントベクターに存在する相同的アームに隣接するように図1に記載のとおり設計されたオリゴヌクレオチドから合成されたPCR産物(左レーン)をin vitroでレシピエントベクター(第2レーン)と混合し、recET + 大腸菌宿主中に同時形質転換した。レシピエントベクターはアンピシリン耐性遺伝子を担持する。(b)PCR産物単独(0.2 μg)またはベクター単独(0.3 μg)によるsbcA大腸菌株JC9604の形質転換はアンピシリンおよびカナマイシンによる2重選択(amp+kan)に対する耐性を伝達しない。しかしながら、PCR産物とベクターの両方を同時形質転換すると、2重耐性コロニーが産生される。これらのコロニーの95%より多くが、相同的アームの選択によりカナマイシン遺伝子がレシピエントベクターに正確に導入されている正確なETクローニング産物を含む。(a)の右側の2レーンは、ETクローニング前後のレシピエントベクターのPvu II制限酵素消化を示す。(c) bと同様、ただし6つのPCR産物(それぞれ0.2 μg)を、pSVpaZ11(それぞれ0.3 μg)でJC9604中に同時形質転換し、Amp+KanプレートまたはAmpプレート上にプレーティングしたことを除く。結果を、組換え産物を表すAmp+Kan耐性コロニー数をコンピテント細胞調製物のプラスミド形質転換率を表すAmp-耐性コロニー数で割って 10^6 をかけてプロットした。PCR産物は、相同的アームの長さが異なることを除いてa-b PCR産物と同じであった。結果は、コンピテント細胞およびDNAの同じ群を用いた5回の実験から得られたものである。エラーバーは標準偏差を示す。(d)50bpの相同的アームに隣接する8つの産物を、pSVpaZ11とともにJC9604中に同時形質転換した。8つのPCR産物のすべてが、同一の左相同的アームおよび増幅されたneo遺伝子を含む。右の相同的アームは、(0)に隣接するpSVpaZ11配列から、またはより離れた(7~3100bp)左の相同的アームから選択した。4回の実験から結果を得た。

【図4】 選択マーカーを交換するためのおよそ100kb P1ベクター中のETクローニングカナマイシン耐性遺伝子を選択マーカーとして用い、少なくとも70kbのマウスHox a遺伝子クラスターを含むP1クローンを使用した。ETクローニングの前に、このエピソームはカナマイシン耐性(上のパネル、左上)をアンピシリン感受性の(上のパネル、右上)宿主大腸菌に伝達する。カナマイシン耐性遺伝子をアンピシリン耐性遺伝子に置き換えるように設計した線状DNA断片を図1で概要を示したようにPCRで作製し、レシピエントHox a/P1

ベクターが常在する大腸菌宿主細胞中に形質転換した。ETクローニングの結果、カナマイシン耐性遺伝子は欠失し、カナマイシン感受性は回復した（上のパネル、左下）。そして、アンピシリン耐性を獲得した（上のパネル、右下）。DNA組換えが正確になされたことは、ETクローニング前後で単離されたDNAを制限酵素分解、およびサザンブロット分析することによって確認した（下のパネル）。

【図5】 対抗選択マーカー除去のためのETクローニング

図1および2に概要を示したようにカナマイシン耐性遺伝子を含むように作製したPCR断片（上のパネル、左、第3レーン）はその選択された相同的アームによって、ベクターpZero-2.1に存在する対抗選択ccdB遺伝子を欠失させるように指令される。recE/recT+大腸菌宿主中に形質転換する前に、PCR産物とpZeroベクターをin vitroで混合した（上のパネル、左、第1レーン）。pZero-2.1単独で形質転換しカナマイシン選択培地上にプレーティングすると、コロニーは殆ど増殖しなかった（下のパネル、左）。pZero-2.1とPCR産物を同時形質転換すると、PvuII消化（上のパネル、右）によって視覚化したとき、意図した分子事象が起こったことを示すETクローニング産物が現われた（下のパネル、右）。

【図6】 エピソームからのrecEおよびrecTの誘導発現によって媒介されるETクローニング

RecE/RecTは、線状および環状DNA分子間の相同組換えを媒介する。(a)プラスミドpBAD24-recETを大腸菌JC5547中に形質転換した。そして続いて回収の前にL-アラビノースを示した回数加えることによってRecE/RecT発現を誘導した後、コンピテント細胞群を調製した。PCR産物をオリゴヌクレオチドeおよびfを用いてpMAK705のクロラムフェニコール耐性遺伝子(cm)およびpBAD24-recETのアンピシリン耐性遺伝子(bla)と隣接するように選択された50bp相同的アームを含むように作製し、続いて形質転換し、組換え体をクロラムフェニコールプレートで同定した。(b)アラビノースを、pBAD24-recETで形質転換したJC5547の培養物に、コンピテント細胞調製のために回収する直前に、種々の回数で直接加えた。タンパク質発現の総量はSDS-PAGEおよびクマシーブルー染色で分析した。(c)形質転換およびアリコートしたKanプレートへのプレーティングの際に、5 pgのpZero2.1を含むこと、カナマイシン耐性を伝達することにより決定される形質転換率につて、PCR産物1 μgあたりのクロラムフェニコール耐性コロニーの数を対照に対して規準化した。

【図7】 Aには、プラスミドpBAD24-recETをダイアグラムで示す。該プラスミドは、誘導性のBADプロモーター (P_{BAD}) 制御下で、遺伝子recE(末端切断型形態)およびrecTを含む。このプラスミドはさらに、アンピシリン耐性遺伝子 (Amp^r) およびaraC遺伝子を含む。

Bには、pBAD24-recETの核酸配列およびタンパク質をコードする部分を示す。

【図8】 複数の組換え工程による大きな大腸菌エピソームの遺伝子操作

a. 組換え反応のスキーム。JC9604に常在するマウスHoxa複合体のP1クローンをneo遺伝子および2つのFlp組換え標的 (FRT) を含むPCR産物で組換えて改変した。2つのPCR産物は、1つがgおよびh相同的アームに隣接し（挿入）、もう1つがiおよびh相同的アームに隣接する（欠失）ことを除いて、同じであった。第2工程では、neo遺伝子がFRT間のFlp組み換えによって、温度感受性起点 (ts ori) pSC101をベースとするFlp発現プラスミドの一時的な形質転換によって、取り除かれた。b上のパネル；各工程における3つの独立したコロニーから得たP1 DNA調製物のEcoRIによる消化を示すエチジウムブロミドで染色したアガロースゲル。真ん中のパネル；neo遺伝子プローブにハイブリダイズした上のパネルのサザンブロット。下のパネル；組換え部位を視覚化するためにHoxa3プローブにハイブリダイズした上のパネルのサザンブロット。レーン1；大腸菌株NS3145中で増殖した起源(original)Hoxa3P1クローン。レーン2；該P1ベクターに常在するTn903カナマイシン耐性遺伝子のアンピシリン耐性遺伝子との置換により、8.1kbバンド（レーン1）が9.0kbに増加した。レーン3；Tn5-neo遺伝子をHoxa3の上流のg-h相同的アームと共に挿入すると、6.7kbのバンド（レーン1,2）が、9.0kbに増加した。レーン4、Flpリコンビナーゼはg-h neo遺伝子を欠失させ、9.0kbのバンド（レーン3）が6.7kbのバンドまで減少した。レーン5、Hoxa3-4遺伝子間の6kbのDNAをi-hneo遺伝子と置き換えることで欠失させると、6.7kb

バンド（レーン2）が4.5kbに減少した。レーン6、Flpリコンビナーゼがi-h neo遺伝子を欠失させ、4.5kbバンドが2.3kbバンドに減少した。

【図9】 大腸菌染色体の遺伝子操作。A. 組換え反応のスキーム。大腸菌染色体の7.8'にあるJC9604の内因性lacZ遺伝子は（関連するAva I部位および座標とともに展開した形で示される）、示すように、相同的アームjおよびk、およびloxP部位に隣接するneo遺伝子を含むPCR断片の標的となった。neo遺伝子を組み込むとAvaI部位を含むほとんどのlacZ遺伝子を取り除かれ、1443および3027bpバンドが3277bpバンドに変化した。第2工程では、pSC101温度感受性起点（ts ori）に基づくCre発現プラスミドの一時的な形質転換によるloxP間のCre組換えによって、neo遺伝子を取り除かれた。Creリコンビナーゼによるneo遺伝子の除去によって3277バンドが2111bpに減少した。b. βガラクトシダーゼ発現をX-Galプレート上の線条接種コロニーによって評価した。最上列の3つの条班は宿主JC9604株内のβ-ガラクトシダーゼ発現を示し（w.t.）、下の3つの条班は（Km）は24の独立した最初のコロニーを示す。そのなかの20は意図した組換え事象を表すβ-ガラクトシダーゼ発現の減失を表す。c. AvaIで消化した大腸菌染色体DNAの、全lacZコード領域から作製したランダムプライムプローブを用いたサザン分析；レーン1,2、w.t.；レーン3~6、j-k neo遺伝子挿入後の独立した4つの白色コロニー；レーン7~10、Cre発現プラスミドで一時的に形質転換した後の同一の4つのコロニー。

10

【図10】 点突然変異を導入するための2周期のETクローニング

a 組換え反応のスキーム。pSVpaX1のlacZ遺伝子を図9の実験で作製された株JC9604lacZ中で分断し、p1B279に対するPCRによって合成されかつlおよびm相同的アームに隣接するsacB-neo遺伝子カセットによって、内因性lacZ遺伝子の発現を除去し、競合する配列を取り除いた。pSV-sacB-neoと称する組換え体をAmp+Kanプレート上で選択した。そして、pSV-sacB-neoのlacZ遺伝子を、完全なlacZ遺伝子から作製したPCR断片によりl*およびm*相同的アームを用いて修復した。m*相同的アームは、BamH1部位を創るCからGへのサイレント変化を含有した。pSVpaX1*と称する組換え体を7%スクロースを用いてsacB遺伝子に対する対抗選択によって同定した。b. pSVpaX1からのβ-ガラクトシダーゼの発現はpSV-sacB-neo中では中断し、pSVpaX1*中では復活した。発現をX-galプレート上で分析した。pSV-sacB-neoおよびpSVpaX1*それぞれの3つの独立したコロニーを示す。c スクロースによる対抗選択後に採取した独立コロニーから調製したDNAをBamH1で消化したもののエチジウムブロミド染色アガロースゲル。すべてのβ-ガラクトシダーゼ発現コロニー（青色）は、導入されたBamH1制限部位を有した（上のパネル）。すべての白色コロニーは大規模な再配列（rearrangement）を表し、特徴的な1.5kbのBamH1制限断片を有する産物はなかった（下のパネル）。

20

30

【図11】 大きなエピソームを改変するためのETクローニングのrecBC+宿主への転換

a. 移動性ET系を有するプラスミドpBAD-ETγのスキームおよびHoxa P1エピソームを標的とするのに使用するストラテジー。pBAD-ETγはpBAD24をベースとしており、(i)アラビノース誘導性P_{BAD}プロモーター下の末端切断型(truncated)recE遺伝子(t-recE)、(ii)EM7プロモーター下のrecT遺伝子、および(iii)Tn5プロモーター下のredγ遺伝子、を含む。pBAD-ETγはHoxa P1エピソームを含むrecA大腸菌株、NS3145中に形質転換された。アラビノース誘導後、コンピテント細胞を調製し、nおよびp相同的アームに隣接するクロラムフェニコール耐性遺伝子(cm)を有するPCR産物で形質転換した。nおよびpは、P1ベクターのセグメントと組換えられるように選択した。b. 組換え標的部位を視覚化するためにP1ベクターから作製したプローブ（上のパネル）およびクロラムフェニコール耐性遺伝子から作製したプローブ（下のパネル）、とハイブリダイズしたDNAをPvu IIで消化したもののサザンブロット。レーン1、ETクローニングの前にHoxaP1エピソームを有する細胞から調製したDNA。レーン2~17、16個のクロラムフェニコール耐性コロニーから調製したDNA。

40

【図12】 pBAD-ETγ内のrecE/recT遺伝子を用いたときとpBAD-αβγ内のredα/redβ遺伝子を用いたときのETクローニングの比較

示したプラスミドpBAD-ETγまたはpBAD-αβγを、大腸菌recA-、recBC+株、DK1中に形質転換し、図6に記載のようにクロラムフェニコール遺伝子の標的として、ETクローニング

50

効率を評価した。アラビノースによるタンパク質発現誘導は1時間だった。

【図13】 Aは、プラスミドpBAD-ET γ をダイアグラムにより示す。

Bは、pBAD-ET γ の核酸配列およびタンパク質コード部分を示す。

【図14】 Aは、プラスミドpBAD- $\alpha\beta\gamma$ をダイアグラムにより示す。このプラスミドはrecEおよびrecT遺伝子がred α およびred β 遺伝子に置き換わっていることを除いて図13に示すプラスミドと実質的に一致する。

Bは、pBAD- $\alpha\beta\gamma$ の核酸配列およびタンパク質コード部分を示す。

【配列表】

SEQUENCE LISTING

10

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: European Molecular Biology Laboratory
(EMBL)
(B) STREET: Meyerhofstrasse 1
(C) CITY: Heidelberg
(E) COUNTRY: DE
(F) POSTAL CODE (ZIP): D-69117

(ii) TITLE OF INVENTION: Novel DNA Cloning Method

(iii) NUMBER OF SEQUENCES: 14

20

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30
(EPO)

(vi) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: EP 97121562.2
(B) FILING DATE: 05-DEC-1997

(vi) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: EP 98118756.0
(B) FILING DATE: 05-OCT-1998

30

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6150 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(vii) IMMEDIATE SOURCE:

(B) CLONE: pBAD24-recET

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION: complement (96..974)
(D) OTHER INFORMATION: /product= "araC"

40

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION: 1320..2162
(D) OTHER INFORMATION: /product= "t-recE"

(ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION:2155..2972
 (D) OTHER INFORMATION:/product= "recT"

(ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION:3493..4353
 (D) OTHER INFORMATION:/product= "bla"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ATCGATGCAT AATGTGCCTG TCAAATGGAC GAAGCAGGGA TTCTGCAAAC CCTATGCTAC	60
TCCGTCAAGC CGTCAATTGT CTGATTCGTT ACCAATTATG ACAACTTGAC GGCTACATCA	120
TTCACTTTTT CTTCACAACC GGCACGGAAC TCGCTCGGGC TGGCCCCGGT GCATTTTTTA	180
AATACCCGCG AGAAATAGAG TTGATCGTCA AAACCAACAT TGCACCGAC GGTGGCGATA	240
GGCATCCGGG TGGTGTCTAA AAGCAGCTTC GCCTGGCTGA TACGTTGGTC CTCGCGCCAG	300
CTTAAGACGC TAATCCCTAA CTGCTGGCGG AAAAGATGTG ACAGACGCGA CGGCGACAAG	360
CAAACATGCT GTGCGACGCT GCGGATATCA AAATTGCTGT CTGCCAGGTG ATCGCTGATG	420
TACTGACAAG CCTCGCGTAC CCGATTATCC ATCGGTGGAT GGAGCGACTC GTTAATCGCT	480
TCCATGCGCC GCAGTAACAA TTGCTCAAGC AGATTTATCG CCAGCAGCTC CGAATAGCGC	540
CCTFCCCCTT GCCCGGCGTT AATGATTGTC CCAAACAGGT CGCTGAAATG CGGCTGGTGC	600
GCTTCATCCG GCGGAAAGAA CCCCCTATTG GCAAATATTG ACGGCCAGTT AAGCCATTCA	660
TGCCAGTAGG CGCGCGGACG AAAGTAAACC CACTGGTGAT ACCATTGCGG AGCCTCCGGA	720
TGACGACCGT AGTGATGAAT CTCTCTGGC GGAACAGCA AAATATCACC CGGTCGGCAA	780
ACAAATTCTC GTCCCTGATT TTTCACCACC CCCTGACCGC GAATGGTGAG ATTGAGAATA	840
TAACCTTTCA TTCCAGCGG TCGGTCGATA AAAAAATCGA GATAACCGTT GGCCTCAATC	900
GGCGTTAAAC CCGCCACCAG ATGGGCATTA AACGAGTATC CCGGCAGCAG GGGATCATTT	960
TGCGCTTCAG CCATACTTTT CATACTCCCG CCATTCAGAG AAGAAACCAA TTGTCCATAT	1020
TGCATCAGAC ATTGCCGTCA CTGCGTCTTT TACTGGCTCT TCTCGCTAAC CAAACCGTA	1080
ACCCCGCTTA TTAAGAGCAT TCTGTAACAA AGCGGGACCA AAGCCATGAC AAAACGCGT	1140
AACAAAAGTG TCTATAATCA CGGCAGAAAA GTCCACATTG ATTATTTGCA CGGCGTCACA	1200
CTTTGCTATG CCATAGCATT TTTATCCATA AGATTAGCGG ATCCTACCTG ACGCTTTTTA	1260
TCGCAACTCT CTACTGTTTC TCCATACCCG TTTTTTTGGG CTAGCAGGAG GAATTCACCA	1320
TGGATCCCGT AATCGTAGAA GACATAGAGC CAGGTATTTA TTACGGAATT TCGAATGAGA	1380
ATTACCACGC GGGTCCCGGT ATCAGTAAGT CTCAGCTCGA TGACATTGCT GATACTCCGG	1440
CACTATATTT GTGGCGTAAA AATGCCCCCG TGGACACCAC AAAGACAAAA ACGCTCGATT	1500
TAGGAACTGC TTTCCACTGC CGGGTACTTG AACCGGAAGA ATTCAGTAAC CGCTTTATCG	1560

10

20

30

TAGCACCTGA ATTTAACCGC CGTACAAACG CCGGAAAAGA AGAAGAGAAA GCGTTTCTGA	1620
TGGAAATGCGC AAGCACAGGA AAAACGGTTA TCACTGCGGA AGAAGGCCGG AAAATTGAAC	1680
TCATGTATCA AAGCGTTATG GCTTTGCCGC TGGGGCAATG GCTTGTGAA AGCGCCGGAC	1740
ACGTGAATC ATCAATTTAC TGGGAAGATC CTGAAACAGG AATTTTGTGT CGGTGCCGTC	1800
CGGACAAAAT TATCCCTGAA TTTCACTGGA TCATGGACGT GAAACTACG GCGGATATTC	1860
AACGATTCAA AACCGCTTAT TACGACTACC GCTATCACGT TCAGGATGCA TTCTACAGTG	1920
ACGGTTATGA AGCACAGTTT GGAGTGCAGC CAACTTTCGT TTTTCTGGTT GCCAGCACAA	1980
CTATTGAATG CGGACGTTAT CCGGTTGAAA TTTTCATGAT GGGCGAAGAA GCAAACTGG	2040
CAGGTCAACA GGAATATCAC CGCAATCTGC GAACCTGTC TGAATGCTG AATACCGATG	2100
AATGGCCAGC TATTAAGACA TTATCACTGC CCCGCTGGGC TAAGGAATAT GCAAATGACT	2160
AAGCAACCAC CAATCGCAA AGCCGATCTG CAAAAAATC AGGGAAACCG TGCACCAGCA	2220
GCAGTTAAAA ATAGCGACGT GATTAGTTTT ATTAACCAGC CATCAATGAA AGAGCAACTG	2280
GCAGCAGCTC TTCCACGCCA TATGACGGCT GAACGTATGA TCCGTATCGC CACCACAGAA	2340
ATTCTGTAAG TTCCGGCGTT AGGAACTGT GACACTATGA GTTTTGTGAG TCGGATCGTA	2400
CAGTGTTCAC AGCTCGGACT TGAGCCAGGT AGCGCCCTCG GTCATGCATA TTTACTGCCT	2460
TTTGTAATA AAAACGAAAA GAGCGTAAA AAGAACGTC AGCTAATCAT TGGCTATCGC	2520
GGCATGATTG ATCTGGCTCG CCGTTCTGGT CAAATCGCCA GCCTGTCAGC CCGTGTGTC	2580
CGTGAAGGTG ACGAGTTTAG CTTTCAATTT GGCCTTGATG AAAAGTTAAT ACACCGCCCC	2640
GGAGAAAACG AAGATGCCCC GGTACCCAC GTCTATGCTG TCGCAAGACT GAAAGACGGA	2700
GGTACTCAGT TTGAAGTTAT GACGCGCAA CAGATTGAGC TGGTGCAGC CCTGAGTAAA	2760
GCTGGTAATA ACGGGCCGTG GGTAATCAC TGGGAAGAAA TGGCAAAGAA AACGGCTATT	2820
CGTCGCCTGT TCAATATTTT GCCCGTATCA ATTGAGATCC AGCGTGCAGT ATCAATGGAT	2880
GAAAAGGAAC CACTGACAA CGATCCTGCA GATTCCTCTG TATTAACCGG GGAATACAGT	2940
GTAATCGATA ATTCAGAGGA ATAGATCTAA GCTTGGCTGT TTTGGCGGAT GAGAGAAGAT	3000
TTTCAGCCTG ATACAGATTA AATCAGAACG CAGAAGCGGT CTGATAAAAC AGAATTTGCC	3060
TGGCGGCAGT AGCGCGGTGG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG	3120
TAGCGCCGAT GGTAGTGTGG GGTCTCCCA TGCGAGAGTA GGGAACTGCC AGGCATCAA	3180
TAAAACGAAA GGCTCAGTCG AAAGACTGGG CCTTTCGTTT TATCTGTTGT TTGTCGGTGA	3240
ACGCTCTCCT GAGTAGGACA AATCCGCCG GAGCGGATT GAACGTTGCG AAGCAACGGC	3300
CCGGAGGGTG GCGGGCAGGA CGCCCGCCAT AAAGTGCAG GCATCAAATT AAGCAGAAGG	3360
CCATCCTGAC GGATGGCCTT TTTGCGTTTC TACAACTCT TTTGTTTATT TTTCTAAATA	3420
CATTCAAATA TGTATCCGCT CATGAGACAA TAACCTGAT AAATGCTTCA ATAATATTGA	3480
AAAAGGAAGA GTATGAGTAT TCAACATTC CGTGTGCCC TTATTCCCTT TTTGCGGCA	3540
TTTGCCTTC CTGTTTTTGC TCACCCAGAA ACGCTGGTGA AAGTAAAAGA TGCTGAAGAT	3600

10

20

30

CAGTTGGGTG CACGAGTGGG TTACATCGAA CTGGATCTCA ACAGCGGTAA GATCCTTGAG 3660
 AGTTTTTCGCC CCGAAGAACG TTTTCCAATG ATGAGCACTT TTAAAGTTCT GCTATGTGGC 3720
 GCGGTATTAT CCCGTGTTGA CGCCGGGCAA GAGCAACTCG GTCGCCGCAT ACACTATTCT 3780
 CAGAACTGACT TGGTTGAGTA CTCACCAGTC ACAGAAAAGC ATCTTACGGA TGGCATGACA 3840
 GTAAGAGAAT TATGCAGTGC TGCCATAACC ATGAGTGATA ACACTGCGGC CAACTTACTT 3900
 CTGACAACGA TCGGAGGACC GAAGGAGCTA ACCGCTTTTT TGCACAACAT GGGGGATCAT 3960
 GTAACTCGCC TTGATCGTTG GGAACCGGAG CTGAATGAAG CCATACCAA CGACGAGCGT 4020
 GACACCACGA TGCCTGTAGC AATGGCAACA ACGTTGCGCA AACTATTAACT TGGCGAACTA 4080
 CTTACTCTAG CTTCCCGGCA ACAATTAATA GACTGGATGG AGGCGGATAA AGTTGCAGGA 4140
 CCACCTCTGC GCTCGGCCCT TCCGGCTGGC TGGTTTATTG CTGATAAATC TGGAGCCGGT 4200
 GAGCGTGGGT CTCGCGGTAT CATTGCAGCA CTGGGGCCAG ATGGTAAGCC CTCCCGTATC 4260
 GTAGTTATCT ACACGACGGG GAGTCAGGCA ACTATGGATG AACGAAATAG ACAGATCGCT 4320
 GAGATAGGTG CCTCACTGAT TAAGCATTGG TAACTGTCAG ACCAAGTTTA CTCATATATA 4380
 CTTTAGATTG ATTTACGCGC CCTGTAGCGG CGCATTAAAGC GCGGCGGGTG TGGTGGTTAC 4440
 GCGCAGCGTG ACCGCTACAC TTGCCAGCGC CCTAGCGCCC GCTCCTTTCTG CTTTCTTCCC 4500
 TTCCTTTCTC GCCACGTTCT CCGGCTTTCC CCGTCAAGCT CTAAATCGGG GGCTCCCTTT 4560
 AGGGTTCCGA TTTAGTGCTT TACGGCACCT CGACCCCAA AACTTGATT TGGGTGATGG 4620
 TTCACGTAGT GGGCCATCGC CCTGATAGAC GGTTTTTCGC CTTTGACGT TGGAGTCCAC 4680
 GTTCTTTAAT AGTGGACTCT TGTTCCAAAC TTGAACAACA CTCAACCCTA TCTCGGGCTA 4740
 TTCTTTTGAT TTATAAGGGA TTTTGCCGAT TTCGGCCTAT TGGTTAAAAA ATGAGCTGAT 4800
 TTAACAAAAA TTTAACGCGA ATTTTAACAA AATATTAAAG TTTACAATTT AAAAGGATCT 4860
 AGGTGAAGAT CCTTTTTGAT AATCTCATGA CCAAAATCCC TTAACGTGAG TTTTCGTTCC 4920
 ACTGAGCGTC AGACCCCGTA GAAAAGATCA AAGGATCTTC TTGAGATCCT TTTTTCTGTC 4980
 GCGTAATCTG CTGCTTGCAA ACAAAGAAC CACCGCTACC AGCGGTGGTT TGGTTGCCGG 5040
 ATCAAGAGCT ACCAACTCTT TTTCCGAAGG TAACTGGCTT CAGCAGAGCG CAGATACCAA 5100
 ATACTGTCCT TCTAGTGTAG CCGTAGTTAG GCCACCACTT CAAGAACTCT GTAGCACCGC 5160
 CTACATACCT CGCTCTGCTA ATCCTGTTAC CAGTGGCTGC TGCCAGTGGC GATAAGTCGT 5220
 GTCTTACCGG GTTGGACTCA AGACGATAGT TACCGGATAA GCGCAGCGG TCGGGCTGAA 5280
 CGGGGGGTTT GTGCACACAG CCCAGCTTGG AGCGAACGAC CTACACCGAA CTGAGATACC 5340
 TACAGCGTGA GCTATGAGAA AGCGCCACGC TTCCCGAAGG GAGAAAGGCG GACAGGTATC 5400
 CGGTAAGCGG CAGGGTCGGA ACAGGAGAGC GCACGAGGGA GCTTCCAGGG GGAAACGCCT 5460
 GGTATCTTTA TAGTCTGTC GGGTTTCGCC ACCTCTGACT TGAGCGTCGA TTTTGTGAT 5520
 GCTCGTCAGG GGGGCGGAGC CTATGGAAAA ACGCCAGCAA CGCGGCCTTT TTACGGTTCC 5580
 TGGCCTTTTG CTGGCCTTTT GCTCACATGT TCTTCTCTGC GTTATCCCTT GATTCTGTGG 5640

10

20

30

ATAACCGTAT TACCGCCTTT GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACGACCGAGC 5700
 GCAGCGAGTC AGTGAGCGAG GAAGCGGAAG AGCGCCTGAT GCGGTATTTT CTCCTTACGC 5760
 ATCTGTGCGG TATTTACACAC CGCATAGGGT CATGGCTGCG CCCCACACCC CGCCAACACC 5820
 CGCTGACGCG CCCTGACGGG CTGTCTGCT CCCGGCATCC GCTTACAGAC AAGCTGTGAC 5880
 CGTCTCCGGG AGCTGCATGT GTCAGAGGTT TTCACCGTCA TCACCGAAAC GCGCGAGGCA 5940
 GCAAGGAGAT GGCGCCCAAC AGTCCCCCGG CCACGGGGCC TGCCACCATA CCCACGCCGA 6000
 AACAAAGCGCT CATGAGCCCG AAGTGGCGAG CCCGATCTTC CCCATCGGTG ATGTGCGCGA 6060
 TATAGGCGCC AGCAACCGCA CCTGTGGCGC CGGTGATGCC GGCCACGATG CGTCCGCGCT 6120
 AGAGGATCTG CTCATGTTTG ACAGCTTATC 6150

10

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 843 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: both

(vii) IMMEDIATE SOURCE:

- (B) CLONE: t-recE

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..843
- (D) OTHER INFORMATION: /product= "t-recE"

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

ATG GAT CCC GTA ATC GTA GAA GAC ATA GAG CCA GGT ATT TAT TAC GGA 48
 Met Asp Pro Val Ile Val Glu Asp Ile Glu Pro Gly Ile Tyr Tyr Gly
 1 5 10 15

ATT TCG AAT GAG AAT TAC CAC GCG GGT CCC GGT ATC AGT AAG TCT CAG 96
 Ile Ser Asn Glu Asn Tyr His Ala Gly Pro Gly Ile Ser Lys Ser Gln
 20 25 30

CTC GAT GAC ATT GCT GAT ACT CCG GCA CTA TAT TTG TGG CGT AAA AAT 144
 Leu Asp Asp Ile Ala Asp Thr Pro Ala Leu Tyr Leu Trp Arg Lys Asn
 35 40 45

GCC CCC GTG GAC ACC ACA AAG ACA AAA ACG CTC GAT TTA GGA ACT GCT 192
 Ala Pro Val Asp Thr Thr Lys Thr Lys Thr Leu Asp Leu Gly Thr Ala
 50 55 60

TTC CAC TGC CGG GTA CTT GAA CCG GAA GAA TTC AGT AAC CGC TTT ATC 240
 Phe His Cys Arg Val Leu Glu Pro Glu Glu Phe Ser Asn Arg Phe Ile
 65 70 75 80

GTA GCA CCT GAA TTT AAC CGC CGT ACA AAC GCC GGA AAA GAA GAA GAG 288
 Val Ala Pro Glu Phe Asn Arg Arg Thr Asn Ala Gly Lys Glu Glu Glu
 85 90 95

AAA GCG TTT CAG ATG GAA TGC GCA AGC ACA GGA AAA ACG GTT ATC ACT 336
 Lys Ala Phe Leu Met Glu Cys Ala Ser Thr Gly Lys Thr Val Ile Thr
 100 105 110

30

GCG GAA GAA GGC CGG AAA ATT GAA CTC ATG TAT CAA AGC GTT ATG GCT 384
 Ala Glu Glu Gly Arg Lys Ile Glu Leu Met Tyr Gln Ser Val Met Ala
 115 120 125

TTG CCG CTG GGG CAA TGG CTT GTT GAA AGC GCC GGA CAC GCT GAA TCA 432
 Leu Pro Leu Gly Gln Trp Leu Val Glu Ser Ala Gly His Ala Glu Ser
 130 135 140

TCA ATT TAC TGG GAA GAT CCT GAA ACA GGA ATT TTG TGT CGG TGC CGT 480
 Ser Ile Tyr Trp Glu Asp Pro Glu Thr Gly Ile Leu Cys Arg Cys Arg
 145 150 155 160

CCG GAC AAA ATT ATC CCT GAA TTT CAC TGG ATC ATG GAC GTG AAA ACT 528
 Pro Asp Lys Ile Ile Pro Glu Phe His Trp Ile Met Asp Val Lys Thr
 165 170 175

ACG GCG GAT ATT CAA CGA TTC AAA ACC GCT TAT TAC GAC TAC CGC TAT 576
 Thr Ala Asp Ile Gln Arg Phe Lys Thr Ala Tyr Tyr Asp Tyr Arg Tyr
 180 185 190

CAC GTT CAG GAT GCA TTC TAC AGT GAC GGT TAT GAA GCA CAG TTT GGA 624
 His Val Gln Asp Ala Phe Tyr Ser Asp Gly Tyr Glu Ala Gln Phe Gly
 195 200 205

GTG CAG CCA ACT TTC GTT TTT CTG GTT GCC AGC ACA ACT ATT GAA TGC 672
 Val Gln Pro Thr Phe Val Phe Leu Val Ala Ser Thr Thr Ile Glu Cys
 210 215 220

GGA CGT TAT CCG GTT GAA ATT TTC ATG ATG GGC GAA GAA GCA AAA CTG 720
 Gly Arg Tyr Pro Val Glu Ile Phe Met Met Gly Glu Glu Ala Lys Leu
 225 230 235 240

GCA GGT CAA CAG GAA TAT CAC CGC AAT CTG CGA ACC CTG TCT GAC TGC 768
 Ala Gly Gln Gln Glu Tyr His Arg Asn Leu Arg Thr Leu Ser Asp Cys
 245 250 255

CTG AAT ACC GAT GAA TGG CCA GCT ATT AAG ACA TTA TCA CTG CCC CGC 816
 Leu Asn Thr Asp Glu Trp Pro Ala Ile Lys Thr Leu Ser Leu Pro Arg
 260 265 270

TGG GCT AAG GAA TAT GCA AAT GAC TAA 843
 Trp Ala Lys Glu Tyr Ala Asn Asp *
 275 280

10

20

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 281 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

30

- (ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Met Asp Pro Val Ile Val Glu Asp Ile Glu Pro Gly Ile Tyr Tyr Gly
 1 5 10 15

Ile Ser Asn Glu Asn Tyr His Ala Gly Pro Gly Ile Ser Lys Ser Gln
 20 25 30

Leu Asp Asp Ile Ala Asp Thr Pro Ala Leu Tyr Leu Trp Arg Lys Asn
 35 40 45

Ala Pro Val Asp Thr Thr Lys Thr Lys Thr Leu Asp Leu Gly Thr Ala
 50 55 60

Phe His Cys Arg Val Leu Glu Pro Glu Glu Phe Ser Asn Arg Phe Ile
 65 70 75 80
 Val Ala Pro Glu Phe Asn Arg Arg Thr Asn Ala Gly Lys Glu Glu Glu
 85 90 95
 Lys Ala Phe Leu Met Glu Cys Ala Ser Thr Gly Lys Thr Val Ile Thr
 100 105 110
 Ala Glu Glu Gly Arg Lys Ile Glu Leu Met Tyr Gln Ser Val Met Ala
 115 120 125
 Leu Pro Leu Gly Gln Trp Leu Val Glu Ser Ala Gly His Ala Glu Ser
 130 135 140
 Ser Ile Tyr Trp Glu Asp Pro Glu Thr Gly Ile Leu Cys Arg Cys Arg
 145 150 155 160
 Pro Asp Lys Ile Ile Pro Glu Phe His Trp Ile Met Asp Val Lys Thr
 165 170 175
 Thr Ala Asp Ile Gln Arg Phe Lys Thr Ala Tyr Tyr Asp Tyr Arg Tyr
 180 185 190
 His Val Gln Asp Ala Phe Tyr Ser Asp Gly Tyr Glu Ala Gln Phe Gly
 195 200 205
 Val Gln Pro Thr Phe Val Phe Leu Val Ala Ser Thr Thr Ile Glu Cys
 210 215 220
 Gly Arg Tyr Pro Val Glu Ile Phe Met Met Gly Glu Glu Ala Lys Leu
 225 230 235 240
 Ala Gly Gln Gln Glu Tyr His Arg Asn Leu Arg Thr Leu Ser Asp Cys
 245 250 255
 Leu Asn Thr Asp Glu Trp Pro Ala Ile Lys Thr Leu Ser Leu Pro Arg
 260 265 270
 Trp Ala Lys Glu Tyr Ala Asn Asp *
 275 280

10

20

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 810 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: both

30

- (vii) IMMEDIATE SOURCE:
- (B) CLONE: recT

- (ix) FEATURE:
- (A) NAME/KEY: CDS
 - (B) LOCATION: 1..810
 - (D) OTHER INFORMATION: /product= "recT"

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

ATG ACT AAG CAA CCA CCA ATC GCA AAA GCC GAT CTG CAA AAA ACT CAG
 Met Thr Lys Gln Pro Pro Ile Ala Lys Ala Asp Leu Gln Lys Thr Gln
 285 290 295

48

GGA AAC CGT GCA CCA GCA GCA GTT AAA AAT AGC GAC GTG ATT AGT TTT Gly Asn Arg Ala Pro Ala Ala Val Lys Asn Ser Asp Val Ile Ser Phe 300 305 310	96	
ATT AAC CAG CCA TCA ATG AAA GAG CAA CTG GCA GCA GCT CTT CCA CGC Ile Asn Gln Pro Ser Met Lys Glu Gln Leu Ala Ala Ala Leu Pro Arg 315 320 325	144	
CAT ATG ACG GCT GAA CGT ATG ATC CGT ATC GCC ACC ACA GAA ATT CGT His Met Thr Ala Glu Arg Met Ile Arg Ile Ala Thr Thr Glu Ile Arg 330 335 340 345	192	
AAA GTT CCG GCG TTA GGA AAC TGT GAC ACT ATG AGT TTT GTC AGT GCG Lys Val Pro Ala Leu Gly Asn Cys Asp Thr Met Ser Phe Val Ser Ala 350 355 360	240	10
ATC GTA CAG TGT TCA CAG CTC GGA CTT GAG CCA GGT AGC GCC CTC GGT Ile Val Gln Cys Ser Gln Leu Gly Leu Glu Pro Gly Ser Ala Leu Gly 365 370 375	288	
CAT GCA TAT TTA CTG CCT TTT GGT AAT AAA AAC GAA AAG AGC GGT AAA His Ala Tyr Leu Leu Pro Phe Gly Asn Lys Asn Glu Lys Ser Gly Lys 380 385 390	336	
AAG AAC GTT CAG CTA ATC ATT GGC TAT CGC GGC ATG ATT GAT CTG GCT Lys Asn Val Gln Leu Ile Ile Gly Tyr Arg Gly Met Ile Asp Leu Ala 395 400 405	384	
CGC CGT TCT GGT CAA ATC GCC AGC CTG TCA GCC CGT GTT GTC CGT GAA Arg Arg Ser Gly Gln Ile Ala Ser Leu Ser Ala Arg Val Val Arg Glu 410 415 420 425	432	20
GGT GAC GAG TTT AGC TTC GAA TTT GGC CTT GAT GAA AAG TTA ATA CAC Gly Asp Glu Phe Ser Phe Glu Phe Gly Leu Asp Glu Lys Leu Ile His 430 435 440	480	
CGC CCG GGA GAA AAC GAA GAT GCC CCG GTT ACC CAC GTC TAT GCT GTC Arg Pro Gly Glu Asn Glu Asp Ala Pro Val Thr His Val Tyr Ala Val 445 450 455	528	
GCA AGA CTG AAA GAC GGA GGT ACT CAG TTT GAA GTT ATG ACG CGC AAA Ala Arg Leu Lys Asp Gly Gly Thr Gln Phe Glu Val Met Thr Arg Lys 460 465 470	576	
CAG ATT GAG CTG GTG CGC AGC CTG AGT AAA GCT GGT AAT AAC GGG CCG Gln Ile Glu Leu Val Arg Ser Leu Ser Lys Ala Gly Asn Asn Gly Pro 475 480 485	624	
TGG GTA ACT CAC TGG GAA GAA ATG GCA AAG AAA ACG GCT ATT CGT CGC Trp Val Thr His Trp Glu Glu Met Ala Lys Lys Thr Ala Ile Arg Arg 490 495 500 505	672	30
CTG TTC AAA TAT TTG CCC GTA TCA ATT GAG ATC CAG CGT GCA GTA TCA Leu Phe Lys Tyr Leu Pro Val Ser Ile Glu Ile Gln Arg Ala Val Ser 510 515 520	720	
ATG GAT GAA AAG GAA CCA CTG ACA ATC GAT CCT GCA GAT TCC TCT GTA Met Asp Glu Lys Glu Pro Leu Thr Ile Asp Pro Ala Asp Ser Ser Val 525 530 535	768	
TTA ACC GGG GAA TAC AGT GTA ATC GAT AAT TCA GAG GAA TAG Leu Thr Gly Glu Tyr Ser Val Ile Asp Asn Ser Glu Glu * 540 545 550	810	

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 270 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```

Met Thr Lys Gln Pro Pro Ile Ala Lys Ala Asp Leu Gln Lys Thr Gln
 1          5          10          15
Gly Asn Arg Ala Pro Ala Ala Val Lys Asn Ser Asp Val Ile Ser Phe
          20          25          30
Ile Asn Gln Pro Ser Met Lys Glu Gln Leu Ala Ala Ala Leu Pro Arg
          35          40          45
His Met Thr Ala Glu Arg Met Ile Arg Ile Ala Thr Thr Glu Ile Arg
          50          55          60
Lys Val Pro Ala Leu Gly Asn Cys Asp Thr Met Ser Phe Val Ser Ala
          65          70          75          80
Ile Val Gln Cys Ser Gln Leu Gly Leu Glu Pro Gly Ser Ala Leu Gly
          85          90          95
His Ala Tyr Leu Leu Pro Phe Gly Asn Lys Asn Glu Lys Ser Gly Lys
          100          105          110
Lys Asn Val Gln Leu Ile Ile Gly Tyr Arg Gly Met Ile Asp Leu Ala
          115          120          125
Arg Arg Ser Gly Gln Ile Ala Ser Leu Ser Ala Arg Val Val Arg Glu
          130          135          140
Gly Asp Glu Phe Ser Phe Glu Phe Gly Leu Asp Glu Lys Leu Ile His
          145          150          155          160
Arg Pro Gly Glu Asn Glu Asp Ala Pro Val Thr His Val Tyr Ala Val
          165          170          175
Ala Arg Leu Lys Asp Gly Gly Thr Gln Phe Glu Val Met Thr Arg Lys
          180          185          190
Gln Ile Glu Leu Val Arg Ser Leu Ser Lys Ala Gly Asn Asn Gly Pro
          195          200          205
Trp Val Thr His Trp Glu Glu Met Ala Lys Lys Thr Ala Ile Arg Arg
          210          215          220
Leu Phe Lys Tyr Leu Pro Val Ser Ile Glu Ile Gln Arg Ala Val Ser
          225          230          235          240
Met Asp Glu Lys Glu Pro Leu Thr Ile Asp Pro Ala Asp Ser Ser Val
          245          250          255
Leu Thr Gly Glu Tyr Ser Val Ile Asp Asn Ser Glu Glu *
          260          265          270

```

(2) INFORMATION FOR SEQ ID NO: 6:

10

20

30

40

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 876 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: both
 (D) TOPOLOGY: both
- (vii) IMMEDIATE SOURCE:
 (B) CLONE: araC
- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: complement (1..876)
 (D) OTHER INFORMATION: /product= "araC"
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

10

```

TGACAACTTG ACGGCTACAT CATTCACTTT TTCTTCACAA CCGGCACGGA ACTCGCTCGG      60
GCTGGCCCCG GTGCATTTTT TAAATACCCG CGAGAAATAG AGTTGATCGT CAAAACCAAC      120
ATTGCGACCG ACGGTGGCGA TAGGCATCCG GGTGGTGCTC AAAAGCAGCT TCGCCTGGCT      180
GATACGTTGG TCCTCGCGCC AGCTTAAGAC GCTAATCCCT AACTGCTGGC GGAAAAGATG      240
TGACAGACGC GACGGCGACA AGCAAACATG CTGTGCGACG CTGGCGATAT CAAAATTGCT      300
GTCTGCCAGG TGATCGCTGA TGTACTGACA AGCCTCGCGT ACCCGATTAT CCATCGGTGG      360
ATGGAGCGAC TCGTTAATCG CTTCCATGCG CCGCAGTAAC AATTGCTCAA GCAGATTAT      420
CGCCAGCAGC TCCGAATAGC GCCCTTCCCC TTGCCCGGCG TTAATGATTT GCCCAAACAG      480
GTCGCTGAAA TCGCGCTGGT GCGCTTCATC CGGGCGAAAG AACCCCGTAT TGGCAAATAT      540
TGACGGCCAG TTAAGCCATT CATGCCAGTA GGCGCGCGGA CGAAAGTAAA CCCACTGGTG      600
ATACCATTGG CGAGCCTCCG GATGACGACC GTAGTGATGA ATCTCTCTCG GCGGGAACAG      660
CAAAATATCA CCCGGTCGGC AAACAAATTC TCGTCCCTGA TTTTTCACCA CCCCCTGACC      720
GCGAATGGTG AGATTGAGAA TATAACCTTT CATTCCCAGC GGTCGGTGA TAAAAAATC      780
GAGATAACCG TTGGCCTCAA TCGGCGTTAA ACCCGCCACC AGATGGGCAT TAAACGAGTA      840
TCCCGGCAGC AGGGGATCAT TTTGCGCTTC AGCCAT                                876

```

20

(2) INFORMATION FOR SEQ ID NO: 7:

30

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 292 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```

Met Ala Glu Ala Gln Asn Asp Pro Leu Leu Pro Gly Tyr Ser Phe Asn
 1           5           10          15
Ala His Leu Val Ala Gly Leu Thr Pro Ile Glu Ala Asn Gly Tyr Leu
          20          25          30

```

Asp Phe Phe Ile Asp Arg Pro Leu Gly Met Lys Gly Tyr Ile Leu Asn
 35 40 45
 Leu Thr Ile Arg Gly Gln Gly Val Val Lys Asn Gln Gly Arg Glu Phe
 50 55 60
 Val Cys Arg Pro Gly Asp Ile Leu Leu Phe Pro Pro Gly Glu Ile His
 65 70 75 80
 His Tyr Gly Arg His Pro Glu Ala Arg Glu Trp Tyr His Gln Trp Val
 85 90 95
 Tyr Phe Arg Pro Arg Ala Tyr Trp His Glu Trp Leu Asn Trp Pro Ser
 100 105 110
 Ile Phe Ala Asn Thr Gly Phe Phe Arg Pro Asp Glu Ala His Gln Pro
 115 120 125
 His Phe Ser Asp Leu Phe Gly Gln Ile Ile Asn Ala Gly Gln Gly Glu
 130 135 140
 Gly Arg Tyr Ser Glu Leu Leu Ala Ile Asn Leu Leu Glu Gln Leu Leu
 145 150 155 160
 Leu Arg Arg Met Glu Ala Ile Asn Glu Ser Leu His Pro Pro Met Asp
 165 170 175
 Asn Arg Val Arg Glu Ala Cys Gln Tyr Ile Ser Asp His Leu Ala Asp
 180 185 190
 Ser Asn Phe Asp Ile Ala Ser Val Ala Gln His Val Cys Leu Ser Pro
 195 200 205
 Ser Arg Leu Ser His Leu Phe Arg Gln Gln Leu Gly Ile Ser Val Leu
 210 215 220
 Ser Trp Arg Glu Asp Gln Arg Ile Ser Gln Ala Lys Leu Leu Leu Ser
 225 230 235 240
 Thr Thr Arg Met Pro Ile Ala Thr Val Gly Arg Asn Val Gly Phe Asp
 245 250 255
 Asp Gln Leu Tyr Phe Ser Arg Val Phe Lys Lys Cys Thr Gly Ala Ser
 260 265 270
 Pro Ser Glu Phe Arg Ala Gly Cys Glu Glu Lys Val Asn Asp Val Ala
 275 280 285
 Val Lys Leu Ser
 290

10

20

30

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 861 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: both

- (vii) IMMEDIATE SOURCE:
- (B) CLONE: bla

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION:1..861

(D) OTHER INFORMATION:/product= "bla"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

ATG AGT ATT CAA CAT TTC CGT GTC GCC CTT ATT CCC TTT TTT GCG GCA Met Ser Ile Gln His Phe Arg Val Ala Leu Ile Pro Phe Phe Ala Ala 295 300 305	48
TTT TGC CTT CCT GTT TTT GCT CAC CCA GAA ACG CTG GTG AAA GTA AAA Phe Cys Leu Pro Val Phe Ala His Pro Glu Thr Leu Val Lys Val Lys 310 315 320	96
GAT GCT GAA GAT CAG TTG GGT GCA CGA GTG GGT TAC ATC GAA CTG GAT Asp Ala Glu Asp Gln Leu Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp 325 330 335 340	144
CTC AAC AGC GGT AAG ATC CTT GAG AGT TTT CGC CCC GAA GAA CGT TTT Leu Asn Ser Gly Lys Ile Leu Glu Ser Phe Arg Pro Glu Glu Arg Phe 345 350 355	192
CCA ATG ATG AGC ACT TTT AAA GTT CTG CTA TGT GGC GCG GTA TTA TCC Pro Met Met Ser Thr Phe Lys Val Leu Cys Gly Ala Val Leu Ser 360 365 370	240
CGT GTT GAC GCC GGG CAA GAG CAA CTC GGT CGC CGC ATA CAC TAT TCT Arg Val Asp Ala Gly Gln Glu Leu Gly Arg Arg Ile His Tyr Ser 375 380 385	288
CAG AAT GAC TTG GTT GAG TAC TCA CCA GTC ACA GAA AAG CAT CTT ACG Gln Asn Asp Leu Val Glu Tyr Ser Pro Val Thr Glu Lys His Leu Thr 390 395 400	336
GAT GGC ATG ACA GTA AGA GAA TTA TGC AGT GCT GCC ATA ACC ATG AGT Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala Ala Ile Thr Met Ser 405 410 415 420	384
GAT AAC ACT GCG GCC AAC TTA CTT CTG ACA ACG ATC GGA GGA CCG AAG Asp Asn Thr Ala Ala Asn Leu Leu Leu Thr Thr Ile Gly Gly Pro Lys 425 430 435	432
GAG CTA ACC GCT TTT TTG CAC AAC ATG GGG GAT CAT GTA ACT CGC CTT Glu Leu Thr Ala Phe Leu His Asn Met Gly Asp His Val Thr Arg Leu 440 445 450	480
GAT CGT TGG GAA CCG GAG CTG AAT GAA GCC ATA CCA AAC GAC GAG CGT Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu Arg 455 460 465	528
GAC ACC ACG ATG CCT GTA GCA ATG GCA ACA ACG TTG CGC AAA CTA TTA Asp Thr Thr Met Pro Val Ala Met Ala Thr Thr Leu Arg Lys Leu Leu 470 475 480	576
ACT GGC GAA CTA CTT ACT CTA GCT TCC CGG CAA CAA TTA ATA GAC TGG Thr Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp 485 490 495 500	624
ATG GAG GCG GAT AAA GTT GCA GGA CCA CTT CTG CGC TCG GCC CTT CCG Met Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro 505 510 515	672
GCT GGC TGG TTT ATT GCT GAT AAA TCT GGA GCC GGT GAG CGT GGG TCT Ala Gly Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser 520 525 530	720

10

20

30

CGC GGT ATC ATT GCA GCA CTG GGG CCA GAT GGT AAG CCC TCC CGT ATC	768
Arg Gly Ile Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile	
535 540 545	
GTA GTT ATC TAC ACG ACG GGG AGT CAG GCA ACT ATG GAT GAA CGA AAT	816
Val Val Ile Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn	
550 555 560	
AGA CAG ATC GCT GAG ATA GGT GCC TCA CTG ATT AAG CAT TGG TAA	861
Arg Gln Ile Ala Glu Ile Gly Ala Ser Leu Ile Lys His Trp *	
565 570 575	

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 287 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

Met Ser Ile Gln His Phe Arg Val Ala Leu Ile Pro Phe Phe Ala Ala	
1 5 10 15	
Phe Cys Leu Pro Val Phe Ala His Pro Glu Thr Leu Val Lys Val Lys	
20 25 30	
Asp Ala Glu Asp Gln Leu Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp	
35 40 45	20
Leu Asn Ser Gly Lys Ile Leu Glu Ser Phe Arg Pro Glu Glu Arg Phe	
50 55 60	
Pro Met Met Ser Thr Phe Lys Val Leu Leu Cys Gly Ala Val Leu Ser	
65 70 75 80	
Arg Val Asp Ala Gly Gln Glu Gln Leu Gly Arg Arg Ile His Tyr Ser	
85 90 95	
Gln Asn Asp Leu Val Glu Tyr Ser Pro Val Thr Glu Lys His Leu Thr	
100 105 110	
Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala Ala Ile Thr Met Ser	
115 120 125	
Asp Asn Thr Ala Ala Asn Leu Leu Thr Thr Ile Gly Gly Pro Lys	
130 135 140	30
Glu Leu Thr Ala Phe Leu His Asn Met Gly Asp His Val Thr Arg Leu	
145 150 155 160	
Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu Arg	
165 170 175	
Asp Thr Thr Met Pro Val Ala Met Ala Thr Thr Leu Arg Lys Leu Leu	
180 185 190	
Thr Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp	
195 200 205	
Met Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro	
210 215 220	
Ala Gly Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser	

40

225	230	235	240
Arg Gly Ile Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile			
	245	250	255
Val Val Ile Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn			
	260	265	270
Arg Gln Ile Ala Glu Ile Gly Ala Ser Leu Ile Lys His Trp *			
	275	280	285

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7195 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: both

10

(vii) IMMEDIATE SOURCE:

- (B) CLONE: pBAD-ETgamma

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 3588..4004
- (D) OTHER INFORMATION: /product= "red gamma"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

ATCGATGCGAT AATGTGCCTG TCAAATGGAC GAAGCAGGGA TTCTGCAAAC CCTATGCTAC	60
TCCGTCAAGC CGTCAATTGT CTGATTCGTT ACCAATTATG ACAACTTGAC GGCTACATCA	120
TTCACTTTTT CTTACAACCC GGCACGGAAC TCGCTCGGGC TGGCCCCGGT GCATTTTTTA	180
AATACCCGCG AGAAATAGAG TTGATCGTCA AAACCAACAT TCGGACCGAC GGTGGCGATA	240
GGCATCCGGG TGGTGCTCAA AAGCAGCTTC GCCTGGCTGA TACGTTGGTC CTCGCGCCAG	300
CTTAAGACGC TAATCCCTAA CTGCTGGCGG AAAAGATGTG ACAGACGCGA CGGCGACAAG	360
CAACATGCT GTGCGACGCT GGCATATCA AAATTGCTGT CTGCCAGGTG ATCGCTGATG	420
TACTGACAAG CCTCGCGTAC CCGATTATCC ATCGGTGGAT GGAGCGACTC GTTAATCGCT	480
TCCATGCGCC GCAGTAACAA TTGCTCAAGC AGATTTATCG CCAGCAGCTC CGAATAGCGC	540
CCTTCCCCTT GCCCGCGGTT AATGATTGTC CCAAACAGGT CGCTGAAATG CGGCTGGTGC	600
GCTTCATCCG GCGGAAAGAA CCCCCTATTG GCAAATATTG ACGGCCAGTT AAGCCATTCA	660
TGCCAGTAGG CGCGCGGACG AAAGTAAACC CACTGGTGAT ACCATTGCGG AGCCTCCGGA	720
TGACGACCGT AGTGATGAAT CTCTCCTGGC GGAACAGCA AAATATCACC CGGTCGGCAA	780
ACAAATCTC GTCCCTGATT TTTCACCACC CCCTGACCGC GAATGGTGAG ATTGAGAATA	840
TAACCTTTCA TTCCCGCGG TCGGTCGATA AAAAAATCGA GATAACCGTT GGCCTCAATC	900
GGCGTTAAAC CCGCCACCAG ATGGGCATTA AACGAGTATC CCGGCAGCAG GGGATCATT	960
TGCGCTTCAG CCATACTTTT CATACTCCCG CCATTCAGAG AAGAAACCAA TTGTCCATAT	1020

20

30

TGCATCAGAC ATTGCCGTCA CTGCGTCTTT TACTGGCTCT TCTCGCTAAC CAAACCGGTA	1080
ACCCCGCTTA TTAAGCAT TCTGTAAACAA AGCGGGACCA AAGCCATGAC AAAACGCGT	1140
AACAAAAGTG TCTATAATCA CGGCAGAAAA GTCCACATTG ATTATTTGCA CGGCGTCACA	1200
CTTTGCTATG CCATAGCATT TTTATCCATA AGATTAGCGG ATCCTACCTG ACGCTTTTTA	1260
TCGCAACTCT CTACTGTTTC TCCATACCCG TTTTTTTGGG CTAGCAGGAG GAATTCACCA	1320
TGGATCCCGT AATCGTAGAA GACATAGAGC CAGGTATTTA TTACGGAATT TCGAATGAGA	1380
ATTACCACGC GGGTCCCGGT ATCAGTAAGT CTCAGCTCGA TGACATTGCT GATACTCCGG	1440
CACTATATTT GTGGCGTAAA AATGCCCCCG TGGACACCAC AAAGACAAAA ACGCTCGATT	1500
TAGGAACTGC TTTCCACTGC CGGTACTTG AACCAGGAAGA ATTCAGTAAC CGCTTTATCG	1560
TAGCACCTGA ATTTAACCGC CGTACAAACG CCGGAAAAGA AGAAGAGAAA GCGTTTCTGA	1620
TGGAATGCGC AAGCACAGGA AAAACGGTTA TCAGTCCGGA AGAAGGCCGG AAAATTGAAC	1680
TCATGTATCA AAGCGTTATG GCTTTGCCGC TGGGGCAATG GCTTGTTGAA AGCGCCGGAC	1740
ACGCTGAATC ATCAATTTAC TGGGAAGATC CTGAAACAGG AATTTTGTGT CGGTGCCGTC	1800
CGGACAAAAT TATCCCTGAA TTTCACTGGA TCATGGACGT GAAAACCTACG GCGGATATTC	1860
AACGATTCAA AACCCTTAT TACGACTACC GCTATCACGT TCAGGATGCA TTCTACAGTG	1920
ACGGTTATGA AGCAGAGTTT GGAGTGCAGC CAACTTTCGT TTTTCTGTTT GCCAGCACAA	1980
CTATTGAATG CGGACGTTAT CCGGTTGAAA TTTTCATGAT GGGCGAAGAA GCAAACTGG	2040
CAGGTCAACA GGAATATCAC CGCAATCTGC GAACCTGTC TGAATGCCTG AATACCGATG	2100
AATGGCCAGC TATTAAGACA TTATCACTGC CCCGCTGGGC TAAGGAATAT GCAAATGACT	2160
AGATCTCGAG GTACCCGAGC ACGTGTGAC AATTAATCAT CGGCATAGTA TATCGGCATA	2220
GTATAATACG ACAAGGTGAG GAACTAAACC ATGGCTAAGC AACCACCAAT CGCAAAAGCC	2280
GATCTGCAAA AAACCTAGGG AAACCGTGCA CCAGCAGCAG TTAATAATAG CGACGTGATT	2340
AGTTTTATTA ACCAGCCATC AATGAAAGAG CAACTGGCAG CAGCTCTTCC ACGCCATATG	2400
ACGGCTGAAC GTATGATCCG TATCGCCACC ACAGAAATTC GTAAAGTTCC GCGGTTAGGA	2460
AACGTGACA CTATGAGTTT TGTCAGTGCG ATCGTACAGT GTTACAGCT CGGACTTGAG	2520
CCAGGTAGCG CCCTCGGTCA TGCATATTTA CTGCCCTTTG GTAATAAAAA CGAAAAGAGC	2580
GGTAAAAAGA ACGTTCAGCT AATCATTGGC TATCGCGGCA TGATTGATCT GGCTCGCCGT	2640
TCTGGTCAAA TCGCCAGCCT GTCAGCCCGT GTTGTCCTGT AAGGTGACGA GTTTAGCTTC	2700
GAATTTGGCC TTGATGAAAA GTTAATACAC CGCCCGGGAG AAAACGAAGA TGCCCCGTT	2760
ACCCACGTCT ATGCTGTGCG AAGACTGAAA GACGGAGGTA CTCAGTTTGA AGTTATGACG	2820
CGCAACAGA TTGAGCTGGT GCGCAGCCTG AGTAAAGCTG GTAATAACGG GCCGTGGGTA	2880
ACTCACTGGG AAGAAATGGC AAAGAAAACG GCTATTCGTC GCCTGTTCAA ATATTTGCCC	2940
GTATCAATTG AGATCCAGCG TGCAGTATCA ATGGATGAAA AGGAACCACT GACAATCGAT	3000
CCTGCAGATT CCTCTGTATT AACCGGGGAA TACAGTGTA TCGATAATTC AGAGGAATAG	3060

10

20

30

ATCTAAGCTT CCTGCTGAAC ATCAAAGGCA AGAAAACATC TGTGTGCAAA GACAGCATCC	3120
TTGAACAAGG ACAATTAACA GTTAACAAAT AAAAAACGCA AAGAAAATGC CGATATCCTA	3180
TTGGCATTITTT CTTTTATTTC TTATCAACAT AAAGGTGAAT CCCATACCTC GAGCTTCACG	3240
CTGCCGCAAG CACTCAGGGC GCAAGGGCTG CTAAGAGGAA GCGGAACACG TAGAAAGCCA	3300
GTCCGCAGAA ACGGTGCTGA CCCCAGGATGA ATGTCAGCTA CTGGGCTATC TGGACAAGGG	3360
AAAACGCAAG CGCAAAGAGA AAGCAGGTAG CTTGCAGTGG GCTTACATGG CGATAGCTAG	3420
ACTGGGCGGT TTTATGGACA GCAAGCGAAC CGGAATTGCC AGCTGGGGCG CCTCTGGTA	3480
AGGTTGGGAA GCCCTGCAAA GTAAACTGGA TGGCTTTCTT GCCGCCAAGG ATCTGATGGC	3540
GCAGGGGATC AAGATCTGAT CAAGAGACAG GATGAGGATC GTTTCGCATG GATATTAATA	3600
CTGAAACTGA GATCAAGCAA AAGCATTCAC TAACCCCTT TCCTGTTTTT CTAATCAGCC	3660
CGGCATTTTCG CGGGCGATAT TTTCACAGCT ATTTTCAGGAG TTCAGCCATG AACGCTTATT	3720
ACATTCAGGA TCGTCTTGAG GCTCAGAGCT GGGCGCGTCA CTACCAGCAG CTCGCCCCGTG	3780
AAGAGAAAGA GGCAGAACTG GCAGACGACA TGGAAAAAGG CCTGCCCCAG CACCTGTTTG	3840
AATCGCTATG CATCGATCAT TTGCAACGCC ACGGGGCCAG CAAAAATCC ATTACCCGTG	3900
CGTTTGATGA CGATGTTGAG TTTCAGGAGC GCATGGCAGA ACACATCCGG TACATGGTTG	3960
AAACCATTGC TCACCACCAG GTTGATATTG ATTCAGAGGT ATAAAACGAG TAGAAGCTTG	4020
GCTGTTTTGG CGGATGAGAG AAGATTTTCA GCCTGATACA GATTAAATCA GAACGCAGAA	4080
GCGGTCTGAT AAAACAGAAT TTGCCTGGCG GCAGTAGCGC GGTGGTCCCA CCTGACCCCA	4140
TGCCGAACCTC AGAAGTGAAG CGCCGTAGCG CCGATGGTAG TGTGGGTCTT CCCCATGCGA	4200
GAGTAGGGAA CTGCCAGGCA TCAAATAAAA CGAAAGGCTC AGTCGAAAGA CTGGGCCTTT	4260
CGTTTTATCT GTTGTGTTGC GGTGAACGCT CTCCTGAGTA GGACAAATCC GCCGGGAGCG	4320
GATTTGAACG TTGCGAAGCA ACGGCCCGGA GGGTGGCGGG CAGGACGCCC GCCATAAACT	4380
GCCAGGCATC AAATTAAGCA GAAGGCCATC CTGACGGATG GCCTTTTTGC GTTCTACAA	4440
ACTCTTTTGT TTATTTTCTT AAATACATTC AAATATGTAT CCGCTCATGA GACAATAACC	4500
CTGATAAATG CTTCAATAAT ATTGAAAAAG GAAGAGTATG AGTATTCAAC ATTTCCGTGT	4560
CGCCCTTATT CCCTTTTTTG CGGCATTTTG CTTTCTGTT TTTGCTCACC CAGAAACGCT	4620
GGTGAAAGTA AAAGATGCTG AAGATCAGTT GGGTGCACGA GTGGGTTACA TCGAACTGGA	4680
TCTCAACAGC GGTAAGATCC TTGAGAGTTT TCGCCCCGAA GAACGTTTTT CAATGATGAG	4740
CACTTTTAAA GTTCTGCTAT GTGGCGCGGT ATTATCCCGT GTTGACGCCG GGCAAGAGCA	4800
ACTCGGTCGC CGCATACACT ATTCTCAGAA TGACTTGTTT GAGTACTCAC CAGTCACAGA	4860
AAAGCATCTT ACGGATGGCA TGACAGTAAG AGAATTATGC AGTGCTGCCA TAACCATGAG	4920
TGATAACACT GCGGCCAACT TACTTCTGAC AACGATCGGA GGACCGAAGG AGCTAACCGC	4980
TTTTTTGCAC AACATGGGGG ATCATGTAAC TCGCCTTGAT CGTTGGGAAC CGGAGCTGAA	5040
TGAAGCCATA CCAAACGACG AGCGTGACAC CACGATGCCT GTAGCAATGG CAACAACGTT	5100

10

20

30

GCGCAAACTA TTAAGTGGCG AACTACTTAC TCTAGCTTCC CGGCAACAAT TAATAGACTG 5160
 GATGGAGGCG GATAAAGTTG CAGGACCACT TCTGCGCTCG GCCCTTCCGG CTGGCTGGTT 5220
 TATTGCTGAT AAATCTGGAG CCGGTGAGCG TGGGTCTCGC GGTATCATTG CAGCACTGGG 5280
 GCCAGATGGT AAGCCCTCCC GTATCGTAGT TATCTACACG ACGGGGAGTC AGGCAACTAT 5340
 GGATGAACGA AATAGACAGA TCGCTGAGAT AGGTGCCTCA CTGATTAAGC ATTGGTAACT 5400
 GTCAGACCAA GTTTACTCAT ATATACTTTA GATTGATTTA CGCGCCCTGT AGCGGCGCAT 5460
 TAAGCGCGGC GGGTGTGGTG GTTACGCGCA GCGTGACCGC TACACTTGCC AGCGCCCTAG 5520
 CGCCCGCTCC TTTCGCTTTC TTCCCTTCCT TTCTCGCCAC GTTCGCGGC TTTCCCGTC 5580
 AAGCTCTAA TCGGGGCTC CCTTTAGGT TCCGATTAG TGCTTTACGG CACCTCGACC 5640
 CCAAAAACT TGATTGGGT GATGGTTCAC GTAGTGGGCC ATCGCCCTGA TAGACGTTT 5700
 TTCGCCCTT GACGTTGGAG TCCACGTTCT TTAATAGTGG ACTCTTGTT CAAACTTGAA 5760
 CAACACTCAA CCCTATCTCG GGCTATTCTT TTGATTATA AGGGATTTG CCGATTTCCG 5820
 CCTATTGGTT AAAAAATGAG CTGATTAAAC AAAAATTAA CGCGAATTT AACAAAATAT 5880
 TAACGTTTAC AATTAAAAG GATCTAGGTG AAGATCCTT TTGATAATCT CATGACCAA 5940
 ATCCCTTAA GTGAGTTTT GTTCCACTGA GCGTCAGACC CCGTAGAAA GATCAAAGGA 6000
 TCTTCTTGAG ATCCTTTTT TCTGCGCGTA ATCTGCTGCT TGCAAACAAA AAAACCACCG 6060
 CTACCAGCGG TGGTTGTTT GCCGGATCAA GAGCTACCAA CTCTTTTTCC GAAGGTAAC 6120
 GGCTTCAGCA GAGCGCAGAT ACCAAATACT GTCCTTCTAG TGTAGCCGTA GTTAGGCCAC 6180
 CACTTCAAGA ACTCTGTAGC ACCGCTACA TACCTCGCTC TGCTAATCCT GTTACCAGTG 6240
 GCTGCTGCCA GTGGCGATAA GTCGTGTCTT ACCGGGTTGG ACTCAAGACG ATAGTTACCG 6300
 GATAAGGCGC AGCGGTCGGG CTGAACGGGG GGTTCGTGCA CACAGCCAG CTTGGAGCGA 6360
 ACGACCTACA CCGAAGTGA ATACCTACAG CGTGAGCTAT GAGAAAGCGC CACGCTTCCC 6420
 GAAGGGAGAA AGGCGGACAG GTATCCGTA AGCGGCAGGG TCGGAACAGG AGAGCGCACG 6480
 AGGGAGCTTC CAGGGGAAA CGCCTGGTAT CTTTATAGTC CTGTCGGGTT TCGCCACCTC 6540
 TGACTTGAGC GTCGATTTTT GTGATGCTCG TCAGGGGGG GAGCCTATG GAAAAACGCC 6600
 AGCAACGCGG CTTTTTACG GTTCTGGCC TTTTGCTGGC CTTTGCTCA CATGTTCTTT 6660
 CCTGCGTTAT CCCCTGATTC TGTGGATAAC CGTATTACCG CCTTTGAGTG AGCTGATACC 6720
 GCTCGCCGCA GCCGAACGAC CGAGCGCAGC GAGTCAGTGA GCGAGGAAGC GGAAGAGCGC 6780
 CTGATGCGGT ATTTTCTCCT TACGCATCTG TCGGTATTT CACACCGCAT AGGTCATGG 6840
 CTGCGCCCCG ACACCCGCCA ACACCCGCTG ACGCGCCCTG ACGGGCTTGT CTGCTCCCGG 6900
 CATCCGCTTA CAGACAAGCT GTGACCGTCT CCGGGAGCTG CATGTGTCAG AGGTTTTTAC 6960
 CGTCATCACC GAAACGCGCG AGGCAGCAAG GAGATGGCG CCAACAGTCC CCCGGCCACG 7020
 GGGCCTGCCA CCATACCCAC GCCGAAACAA GCGCTCATGA GCGGAAAGTG GCGAGCCCGA 7080
 TCTTCCCAT CCGTGATGTC GCGGATATAG GCGCCAGCAA CCGCACCTGT GCGCCCGGTG 7140

10

20

30

ATGCCGGCCA CGATGCGTCC GCGTAGAGG ATCTGCTCAT GTTGGACAGC TTATC

7195

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7010 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: both

(vii) IMMEDIATE SOURCE:

- (B) CLONE: pBAD-alpha-beta-gamma

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:1320..2000
- (D) OTHER INFORMATION:/product= "red alpha"

10

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:2086..2871
- (D) OTHER INFORMATION:/product= "red beta"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:3403..3819
- (D) OTHER INFORMATION:/product= "red gamma"

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

ATCGATGCAT AATGTGCCTG TCAAATGGAC GAAGCAGGGA TTCTGCAAAC CCTATGCTAC 60
TCCGTCAAGC CGTCAATTGT CTGATTCGTT ACCAATTATG ACAACTTGAC GGCTACATCA 120
TTCACCTTTTT TTTCACAACC GGCACGGAAC TCGCTCGGGC TGGCCCCGGT GCATTTTTTA 180
AATACCCGCG AGAAATAGAG TTGATCGTCA AAACCAACAT TGCGACCGAC GGTGGCGATA 240
GGCATCCGGG TGGTGTCTAA AAGCAGCTTC GCCTGGCTGA TACGTTGGTC CTCGCGCCAG 300
CTTAAGACGC TAATCCCTAA CTGCTGGCGG AAAAGATGTG ACAGACGCGA CGGCGACAAG 360
CAAACATGCT GTGCGACGCT GCGGATATCA AAATTGCTGT CTGCCAGGTG ATCGCTGATG 420
TACTGACAAG CCTCGCGTAC CCGATTATCC ATCGGTGGAT GGAGCGACTC GTTAATCGCT 480
TCCATGCGCC GCAGTAACAA TTGCTCAAGC AGATTTATCG CCAGCAGCTC CGAATAGCGC 540
CCTTCCCCTT GCGCGCGGTT AATGATTGTC CCAAACAGGT CGCTGAAATG CGGCTGGTGC 600
GCTTCATCCG GCGGAAAGAA CCGCGTATTG GCAAATATTG ACGGCCAGTT AAGCCATTCA 660
TGCCAGTAGG CGCGCGGACG AAAGTAAACC CACTGGTGAT ACCATTGCGG AGCCTCCGGA 720
TGACGACCGT AGTGATGAAT CTCTCCTGGC GGGAACAGCA AAATATCACC CGGTGCGCAA 780
ACAAATTCTC GTCCCTGATT TTTCACCACC CCCTGACCGC GAATGGTGAG ATTGAGAATA 840
TAACCTTTCA TTCCCAGCGG TCGGTCGATA AAAAAATCGA GATAACCGTT GGCCTCAATC 900
GGCGTTAAAC CCGCCACCAG ATGGGCATTA AACGAGTATC CCGGCAGCAG GGGATCATTT 960

30

TGCCTTCAG CCATACTTTT CATACTCCCG CCATTCAGAG AAGAAACCAA TTGTCCATAT 1020
 TGCATCAGAC ATTGCCGTCA CTGCGTCTTT TACTGGCTCT TCTCGCTAAC CAAACCGGTA 1080
 ACCCCGCTTA TTAAAAGCAT TCTGTACAA AGCGGGACCA AAGCCATGAC AAAAACGCGT 1140
 AACAAAAGTG TCTATAATCA CGGCAGAAAA GTCCACATTG ATTATTTGCA CGGCGTCACA 1200
 CTTTGCTATG CCATAGCATT TTTATCCATA AGATTAGCGG ATCCTACCTG ACGCTTTTTA 1260
 TCGCAACTCT CTACTGTTTC TCCATACCCG TTTTTTTGGG CTAGCAGGAG GAATTCACC 1319
 ATG ACA CCG GAC ATT ATC CTG CAG CGT ACC GGG ATC GAT GTG AGA GCT 1367
 Met Thr Pro Asp Ile Ile Leu Gln Arg Thr Gly Ile Asp Val Arg Ala
 290 295 300
 GTC GAA CAG GGG GAT GAT GCG TGG CAC AAA TTA CGG CTC GGC GTC ATC 1415
 Val Glu Gln Gly Asp Asp Ala Trp His Lys Leu Arg Leu Gly Val Ile
 305 310 315
 ACC GCT TCA GAA GTT CAC AAC GTG ATA GCA AAA CCC CGC TCC GGA AAG 1463
 Thr Ala Ser Glu Val His Asn Val Ile Ala Lys Pro Arg Ser Gly Lys
 320 325 330 335
 AAG TGG CCT GAC ATG AAA ATG TCC TAC TTC CAC ACC CTG CTT GCT GAG 1511
 Lys Trp Pro Asp Met Lys Met Ser Tyr Phe His Thr Leu Leu Ala Glu
 340 345 350
 GTT TGC ACC GGT GTG GCT CCG GAA GTT AAC GCT AAA GCA CTG GCC TGG 1559
 Val Cys Thr Gly Val Ala Pro Glu Val Asn Ala Lys Ala Leu Ala Trp
 355 360 365
 GGA AAA CAG TAC GAG AAC GAC GCC AGA ACC CTG TTT GAA TTC ACT TCC 1607
 Gly Lys Gln Tyr Glu Asn Asp Ala Arg Thr Leu Phe Glu Phe Thr Ser
 370 375 380
 GGC GTG AAT GTT ACT GAA TCC CCG ATC ATC TAT CGC GAC GAA AGT ATG 1655
 Gly Val Asn Val Thr Glu Ser Pro Ile Ile Tyr Arg Asp Glu Ser Met
 385 390 395
 CGT ACC GCC TGC TCT CCC GAT GGT TTA TGC AGT GAC GGC AAC GGC CTT 1703
 Arg Thr Ala Cys Ser Pro Asp Gly Leu Cys Ser Asp Gly Asn Gly Leu
 400 405 410 415
 GAA CTG AAA TGC CCG TTT ACC TCC CGG GAT TTC ATG AAG TTC CGG CTC 1751
 Glu Leu Lys Cys Pro Phe Thr Ser Arg Asp Phe Met Lys Phe Arg Leu
 420 425 430
 GGT GGT TTC GAG GCC ATA AAG TCA GCT TAC ATG GCC CAG GTG CAG TAC 1799
 Gly Gly Phe Glu Ala Ile Lys Ser Ala Tyr Met Ala Gln Val Gln Tyr
 435 440 445
 AGC ATG TGG GTG ACG CGA AAA AAT GCC TGG TAC TTT GCC AAC TAT GAC 1847
 Ser Met Trp Val Thr Arg Lys Asn Ala Trp Tyr Phe Ala Asn Tyr Asp
 450 455 460
 CCG CGT ATG AAG CGT GAA GGC CTG CAT TAT GTC GTG ATT GAG CGG GAT 1895
 Pro Arg Met Lys Arg Glu Gly Leu His Tyr Val Val Ile Glu Arg Asp
 465 470 475
 GAA AAG TAC ATG GCG AGT TTT GAC GAG ATC GTG CCG GAG TTC ATC GAA 1943
 Glu Lys Tyr Met Ala Ser Phe Asp Glu Ile Val Pro Glu Phe Ile Glu
 480 485 490 495
 AAA ATG GAC GAG GCA CTG GCT GAA ATT GGT TTT GTA TTT GGG GAG CAA 1991
 Lys Met Asp Glu Ala Leu Ala Glu Ile Gly Phe Val Phe Gly Glu Gln
 500 505 510

10

20

30

TGG CGA TAG ATCCGGTACC CGAGCACGTG TTGACAATTA ATCATCGGCA 2040
Trp Arg *

TAGTATATCG GCATAGTATA ATACGACAAG GTGAGGAACT AAACC ATG AGT ACT 2094
Met Ser Thr
1

GCA CTC GCA ACG CTG GCT GGG AAG CTG GCT GAA CGT GTC GGC ATG GAT 2142
Ala Leu Ala Thr Leu Ala Gly Lys Leu Ala Glu Arg Val Gly Met Asp
5 10 15

TCT GTC GAC CCA CAG GAA CTG ATC ACC ACT CTT CGC CAG ACG GCA TTT 2190
Ser Val Asp Pro Gln Glu Leu Ile Thr Thr Leu Arg Gln Thr Ala Phe
20 25 30 35

AAA GGT GAT GCC AGC GAT GCG CAG TTC ATC GCA TTA CTG ATC GTT GCC 2238 10
Lys Gly Asp Ala Ser Asp Ala Gln Phe Ile Ala Leu Leu Ile Val Ala
40 45 50

AAC CAG TAC GGC CTT AAT CCG TGG ACG AAA GAA ATT TAC GCC TTT CCT 2286
Asn Gln Tyr Gly Leu Asn Pro Trp Thr Lys Glu Ile Tyr Ala Phe Pro
55 60 65

GAT AAG CAG AAT GGC ATC GTT CCG GTG GTG GGC GTT GAT GGC TGG TCC 2334
Asp Lys Gln Asn Gly Ile Val Pro Val Val Gly Val Asp Gly Trp Ser
70 75 80

CGC ATC ATC AAT GAA AAC CAG CAG TTT GAT GGC ATG GAC TTT GAG CAG 2382
Arg Ile Ile Asn Glu Asn Gln Gln Phe Asp Gly Met Asp Phe Glu Gln
85 90 95

GAC AAT GAA TCC TGT ACA TGC CGG ATT TAC CGC AAG GAC CGT AAT CAT 2430 20
Asp Asn Glu Ser Cys Thr Cys Arg Ile Tyr Arg Lys Asp Arg Asn His
100 105 110 115

CCG ATC TGC GTT ACC GAA TGG ATG GAT GAA TGC CGC CGC GAA CCA TTC 2478
Pro Ile Cys Val Thr Glu Trp Met Asp Glu Cys Arg Arg Glu Pro Phe
120 125 130

AAA ACT CGC GAA GGC AGA GAA ATC ACG GGG CCG TGG CAG TCG CAT CCC 2526
Lys Thr Arg Gly Arg Glu Ile Thr Gly Pro Trp Gln Ser His Pro
135 140 145

AAA CGG ATG TTA CGT CAT AAA GCC ATG ATT CAG TGT GCC CGT CTG GCC 2574
Lys Arg Met Leu Arg His Lys Ala Met Ile Gln Cys Ala Arg Leu Ala
150 155 160

TTC GGA TTT GCT GGT ATC TAT GAC AAG GAT GAA GCC GAG CGC ATT GTC 2622
Phe Gly Phe Ala Gly Ile Tyr Asp Lys Asp Glu Ala Glu Arg Ile Val
165 170 175

GAA AAT ACT GCA TAC ACT GCA GAA CGT CAG CCG GAA CGC GAC ATC ACT 2670 30
Glu Asn Thr Ala Tyr Thr Ala Glu Arg Gln Pro Glu Arg Asp Ile Thr
180 185 190 195

CCG GTT AAC GAT GAA ACC ATG CAG GAG ATT AAC ACT CTG CTG ATC GCC 2718
Pro Val Asn Asp Glu Thr Met Gln Glu Ile Asn Thr Leu Leu Ile Ala
200 205 210

CTG GAT AAA ACA TGG GAT GAC GAC TTA TTG CCG CTC TGT TCC CAG ATA 2766
Leu Asp Lys Thr Trp Asp Asp Asp Leu Leu Pro Leu Cys Ser Gln Ile
215 220 225

TTT CGC CGC GAC ATT CGT GCA TCG TCA GAA CTG ACA CAG GCC GAA GCA 2814
Phe Arg Arg Asp Ile Arg Ala Ser Ser Glu Leu Thr Gln Ala Glu Ala
230 235 240

GTA AAA GCT CTT GGA TTC CTG AAA CAG AAA GCC GCA GAG CAG AAG GTG Val Lys Ala Leu Gly Phe Leu Lys Gln Lys Ala Ala Glu Gln Lys Val 245 250 255	2862
GCA GCA TAG ATCTCGAGAA GCTTCCTGCT GAACATCAAA GGCAAGAAAA Ala Ala * 260	2911
CATCTGTTGT CAAAGACAGC ATCCTTGAAC AAGGACAATT AACAGTTAAC AAATAAAAAC	2971
GCAAAAGAAA ATGCCGATAT CCTATGGCA TTTTCTTTTA TTTCTTATCA ACATAAAGGT	3031
GAATCCCATTA CCTCGAGCTT CACGCTGCCG CAAGCACTCA GGGCGCAAGG GCTGCTAAAA	3091
GGAAGCGGAA CACGTAGAAA GCCAGTCCGC AGAAACGGTG CTGACCCCGG ATGAATGTCA	3151
GCTACTGGGC TATCTGGACA AGGGAAAACG CAAGCGCAAA GAGAAAGCAG GTAGCTTGCA	3211
GTGGGCTTAC ATGGCGATAG CTAGACTGGG CGGTTTTATG GACAGCAAGC GAACCGGAAT	3271
TGCCAGCTGG GGCGCCCTCT GGTAAGGTTG GGAAGCCCTG CAAAGTAAAC TGGATGGCTT	3331
TCTTGCCGCC AAGGATCTGA TGGCGCAGGG GATCAAGATC TGATCAAGAG ACAGGATGAG	3391
GATCGTTTCG C ATG GAT ATT AAT ACT GAA ACT GAG ATC AAG CAA AAG CAT Met Asp Ile Asn Thr Glu Thr Glu Ile Lys Gln Lys His 1 5 10	3441
TCA CTA ACC CCC TTT CCT GTT TTC CTA ATC AGC CCG GCA TTT CGC GGG Ser Leu Thr Pro Phe Pro Val Phe Leu Ile Ser Pro Ala Phe Arg Gly 15 20 25	3489
CGA TAT TTT CAC AGC TAT TTC AGG AGT TCA GCC ATG AAC GCT TAT TAC Arg Tyr Phe His Ser Tyr Phe Arg Ser Ser Ala Met Asn Ala Tyr Tyr 30 35 40 45	3537
ATT CAG GAT CGT CTT GAG GCT CAG AGC TGG GCG CGT CAC TAC CAG CAG Ile Gln Asp Arg Leu Glu Ala Gln Ser Trp Ala Arg His Tyr Gln Gln 50 55 60	3585
CTC GCC CGT GAA GAG AAA GAG GCA GAA CTG GCA GAC GAC ATG GAA AAA Leu Ala Arg Glu Lys Glu Ala Glu Leu Ala Asp Asp Met Glu Lys 65 70 75	3633
GGC CTG CCC CAG CAC CTG TTT GAA TCG CTA TGC ATC GAT CAT TTG CAA Gly Leu Pro Gln His Leu Phe Glu Ser Leu Cys Ile Asp His Leu Gln 80 85 90	3681
CGC CAC GGG GCC AGC AAA AAA TCC ATT ACC CGT GCG TTT GAT GAC GAT Arg His Gly Ala Ser Lys Lys Ser Ile Thr Arg Ala Phe Asp Asp Asp 95 100 105	3729
GTT GAG TTT CAG GAG CGC ATG GCA GAA CAC ATC CGG TAC ATG GTT GAA Val Glu Phe Gln Glu Arg Met Ala Glu His Ile Arg Tyr Met Val Glu 110 115 120 125	3777
ACC ATT GCT CAC CAC CAG GTT GAT ATT GAT TCA GAG GTA TAA Thr Ile Ala His His Gln Val Asp Ile Asp Ser Glu Val * 130 135	3819
AACGAGTAGA AGCTTGGCTG TTTTGGCGGA TGAGAGAAGA TTTTCAGCCT GATACAGATT	3879
AAATCAGAAC GCAGAAGCGG TCTGATAAAA CAGAATTTGC CTGGCGGCAG TAGCGCGGTG	3939
GTCCACCTG ACCCCATGCC GAACTCAGAA GTGAAACGCC GTAGCGCCGA TGGTAGTGTG	3999
GGGTCTCCCC ATGCGAGAGT AGGGAAGTGC CAGGCATCAA ATAAACGAA AGGCTCAGTC	4059

10

20

30

GAAAGACTGG GCCTTTCGTT TTATCTGTTG TTTGTCGGTG AACGCTCTCC TGAGTAGGAC 4119
 AAATCCGCCG GGAGCGGATT TGAACGTTGC GAAGCAACGG CCCGGAGGGT GGCGGGCAGG 4179
 ACGCCCGCCA TAAACTGCCA GGCATCAAAT TAAGCAGAAG GCCATCCTGA CGGATGGCCT 4239
 TTTTGCCTTT CTACAACTC TTTTGTATTAT TTTTCTAAAT ACATTCAAAT ATGTATCCGC 4299
 TCATGAGACA ATAACCCTGA TAAATGCTTC AATAATATTG AAAAAGGAAG AGTATGAGTA 4359
 TTCAACATTT CCGTGTGCGC CTTATTCCCT TTTTGCCTT ATTTTGCCTT CCTGTTTTTG 4419
 CTCACCCAGA AACGCTGGTG AAAGTAAAAG ATGCTGAAGA TCAGTTGGGT GCACGAGTGG 4479
 GTTACATCGA ACTGGATCTC AACAGCGGTA AGATCCTTGA GAGTTTTTCGC CCCGAAGAAC 4539
 GTTTTCCAAT GATGAGCACT TTAAAGTTC TGCTATGTGG CGCGGTATTA TCCCGTGTTG 4599
 ACGCCGGGCA AGAGCAACTC GGTGCGCGCA TACACTATTC TCAGAATGAC TTGGTTGAGT 4659
 ACTCACCAGT CACAGAAAAG CATCTTACGG ATGGCATGAC AGTAAGAGAA TTATGCAGTG 4719
 CTGCCATAAC CATGAGTGAT AACACTGCGG CCAACTTACT TCTGACAACG ATCGGAGGAC 4779
 CGAAGGAGCT AACCCTTTTT TTGCACAACA TGGGGGATCA TGTAACCTCGC CTTGATCGTT 4839
 GGGAAACCGGA GCTGAATGAA GCCATACCAA ACGACGAGCG TGACACCACG ATGCTGTAG 4899
 CAATGGCAAC AACGTTGCGC AAACATATTA CTGGCGAACT ACTTACTCTA GCTTCCCGGC 4959
 AACAAATTAAT AGACTGGATG GAGGCGGATA AAGTTGCAGG ACCACTTCTG CGCTCGGCCC 5019
 TTCCGGCTGG CTGGTTTATT GCTGATAAAT CTGGAGCCGG TGAGCGTGGG TCTCGCGGTA 5079
 TCATTGCAGC ACTGGGGCCA GATGGTAAGC CCTCCCGTAT CGTAGTTATC TACACGACGG 5139
 GGAGTCAGGC AACTATGGAT GAACGAAATA GACAGATCGC TGAGATAGGT GCCTCACTGA 5199
 TTAAGCATTG GTAACGTGCA GACCAAGTTT ACTCATATAT ACTTTAGATT GATTACGCG 5259
 CCCTGTAGCG GCGCATTAAAG CGCGGCGGGT GTGGTGGTTA CGCGCAGCGT GACCGCTACA 5319
 CTTGCCAGCG CCCTAGCGCC CGCTCCTTTC GCTTCTTCC CTTCCTTCT CTGCCAGTTC 5379
 GCCGGCTTTC CCCGTCAAGC TCTAAATCGG GGGCTCCCTT TAGGGTTCCG ATTTAGTGCT 5439
 TTACGGCACC TCGACCCCAA AAAACTTGAT TTGGGTGATG GTTCACGTAG TGGGCCATCG 5499
 CCCTGATAGA CGGTTTTTCG CCCTTTGACG TTGGAGTCCA CGTCTTTTAA TAGTGGACTC 5559
 TTGTTCCAAA CTTGAACAAC ACTCAACCCT ATCTCGGGCT ATTCTTTTGA TTTATAAGGG 5619
 ATTTTGCCGA TTTGCGCCTA TTGGTTAAAA AATGAGCTGA TTTAACAAAA ATTTAACGCG 5679
 AATTTTAACA AAATATTAAC GTTTACAATT TAAAAGGATC TAGGTGAAGA TCCTTTTGA 5739
 TAATCTCATG ACCAAAATCC CTTAACGTGA GTTTTCGTTC CACTGAGCGT CAGACCCCGT 5799
 AGAAAAGATC AAAGGATCTT CTTGAGATCC TTTTTTCTG CGCGTAATCT GCTGCTTGCA 5859
 AACAAAAAAA CCACCGCTAC CAGCGGTGGT TTGTTTGCCG GATCAAGAGC TACCAACTCT 5919
 TTTTCCGAAG GTAACGCTC TCAGCAGAGC GCAGATACCA AATACTGTCC TTCTAGTGTA 5979
 GCCGTAGTTA GGCCACCACT TCAAGAACTC TGTAGCACCG CCTACATACC TCGCTCTGCT 6039
 AATCCTGTGA CAGTGGCTG CTGCCAGTGG CGATAAGTCG TGTCTTACCG GGTGGAATC 6099

10

20

30

AAGACGATAG TTACCGGATA AGGCGCAGCG GTCGGGCTGA ACGGGGGGTT CGTGCACACA 6159
 GCCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG AGCTATGAGA 6219
 AAGCGCCACG CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGGTCGG 6279
 AACAGGAGAG CGCACGAGGG AGCTTCCAGG GGGAAACGCC TGGTATCTTT ATAGTCCTGT 6339
 CGGGTTTCGC CACCTCTGAC TTGAGCGTCG ATTTTGTGA TGCTCGTCAG GGGGGCGGAG 6399
 CCTATGGAAA AACGCCAGCA ACGCGGCCTT TTTACGGTTC CTGGCCTTTT GCTGGCCTTT 6459
 TGCTCACATG TTCTTTCCTG CGTTATCCCC TGATTCTGTG GATAACCGTA TTACCGCCTT 6519
 TGAGTGAGCT GATACCGCTC GCCGCAGCCG AACGACCGAG CGCAGCGAGT CAGTGAGCGA 6579
 GGAAGCGGAA GAGCGCCTGA TCGGGTATTT TCTCCTTACG CATCTGTGCG GTATTTCACA 6639
 CCGCATAGGG TCATGGCTGC GCCCCGACAC CCGCCAACAC CCGCTGACGC GCCCTGACGG 6699
 GCTTGTCTGC TCCCGGCATC CGCTTACAGA CAAGCTGTGA CCGTCTCCGG GAGCTGCATG 6759
 TGTGAGAGGT TTTACCGTTC ATCACCAGAA CGCGCGAGGC AGCAAGGAGA TGGCGCCCAA 6819
 CAGTCCCCCG GCCACGGGGC CTGCCACCAT ACCCAGCGCG AAACAAGCGC TCATGAGCCC 6879
 GAAGTGGCGA GCCCGATCTT CCCCATCGGT GATGTCGGCG ATATAGGCGC CAGCAACCGC 6939
 ACCTGTGGCG CCGGTGATGC CGGCCACGAT GCGTCCGGCG TAGAGGATCT GCTCATGTTT 6999
 GACAGCTTAT C 7010

10

(2) INFORMATION FOR SEQ ID NO: 12:

20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 227 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Thr Pro Asp Ile Ile Leu Gln Arg Thr Gly Ile Asp Val Arg Ala
 1 5 10 15
 Val Glu Gln Gly Asp Asp Ala Trp His Lys Leu Arg Leu Gly Val Ile
 20 25 30
 Thr Ala Ser Glu Val His Asn Val Ile Ala Lys Pro Arg Ser Gly Lys
 35 40 45
 Lys Trp Pro Asp Met Lys Met Ser Tyr Phe His Thr Leu Leu Ala Glu
 50 55 60
 Val Cys Thr Gly Val Ala Pro Glu Val Asn Ala Lys Ala Leu Ala Trp
 65 70 75 80
 Gly Lys Gln Tyr Glu Asn Asp Ala Arg Thr Leu Phe Glu Phe Thr Ser
 85 90 95
 Gly Val Asn Val Thr Glu Ser Pro Ile Ile Tyr Arg Asp Glu Ser Met
 100 105 110
 Arg Thr Ala Cys Ser Pro Asp Gly Leu Cys Ser Asp Gly Asn Gly Leu
 115 120 125

30

Glu Leu Lys Cys Pro Phe Thr Ser Arg Asp Phe Met Lys Phe Arg Leu
 130 135 140
 Gly Gly Phe Glu Ala Ile Lys Ser Ala Tyr Met Ala Gln Val Gln Tyr
 145 150 155 160
 Ser Met Trp Val Thr Arg Lys Asn Ala Trp Tyr Phe Ala Asn Tyr Asp
 165 170 175
 Pro Arg Met Lys Arg Glu Gly Leu His Tyr Val Val Ile Glu Arg Asp
 180 185 190
 Glu Lys Tyr Met Ala Ser Phe Asp Glu Ile Val Pro Glu Phe Ile Glu
 195 200 205
 Lys Met Asp Glu Ala Leu Ala Glu Ile Gly Phe Val Phe Gly Glu Gln
 210 215 220
 Trp Arg *
 225

10

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 262 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Met Ser Thr Ala Leu Ala Thr Leu Ala Gly Lys Leu Ala Glu Arg Val
 1 5 10 15
 Gly Met Asp Ser Val Asp Pro Gln Glu Leu Ile Thr Thr Leu Arg Gln
 20 25 30
 Thr Ala Phe Lys Gly Asp Ala Ser Asp Ala Gln Phe Ile Ala Leu Leu
 35 40 45
 Ile Val Ala Asn Gln Tyr Gly Leu Asn Pro Trp Thr Lys Glu Ile Tyr
 50 55 60
 Ala Phe Pro Asp Lys Gln Asn Gly Ile Val Pro Val Val Gly Val Asp
 65 70 75 80
 Gly Trp Ser Arg Ile Ile Asn Glu Asn Gln Gln Phe Asp Gly Met Asp
 85 90 95
 Phe Glu Gln Asp Asn Glu Ser Cys Thr Cys Arg Ile Tyr Arg Lys Asp
 100 105 110
 Arg Asn His Pro Ile Cys Val Thr Glu Trp Met Asp Glu Cys Arg Arg
 115 120 125
 Glu Pro Phe Lys Thr Arg Glu Gly Arg Glu Ile Thr Gly Pro Trp Gln
 130 135 140
 Ser His Pro Lys Arg Met Leu Arg His Lys Ala Met Ile Gln Cys Ala
 145 150 155 160
 Arg Leu Ala Phe Gly Phe Ala Gly Ile Tyr Asp Lys Asp Glu Ala Glu
 165 170 175

20

30

Arg Ile Val Glu Asn Thr Ala Tyr Thr Ala Glu Arg Gln Pro Glu Arg
 180 185 190
 Asp Ile Thr Pro Val Asn Asp Glu Thr Met Gln Glu Ile Asn Thr Leu
 195 200 205
 Leu Ile Ala Leu Asp Lys Thr Trp Asp Asp Asp Leu Leu Pro Leu Cys
 210 215 220
 Ser Gln Ile Phe Arg Arg Asp Ile Arg Ala Ser Ser Glu Leu Thr Gln
 225 230 235 240
 Ala Glu Ala Val Lys Ala Leu Gly Phe Leu Lys Gln Lys Ala Ala Glu
 245 250 255
 Gln Lys Val Ala Ala *
 260

10

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 139 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Met Asp Ile Asn Thr Glu Thr Glu Ile Lys Gln Lys His Ser Leu Thr
 1 5 10 15
 Pro Phe Pro Val Phe Leu Ile Ser Pro Ala Phe Arg Gly Arg Tyr Phe
 20 25 30
 His Ser Tyr Phe Arg Ser Ser Ala Met Asn Ala Tyr Tyr Ile Gln Asp
 35 40 45
 Arg Leu Glu Ala Gln Ser Trp Ala Arg His Tyr Gln Gln Leu Ala Arg
 50 55 60
 Glu Glu Lys Glu Ala Glu Leu Ala Asp Asp Met Glu Lys Gly Leu Pro
 65 70 75 80
 Gln His Leu Phe Glu Ser Leu Cys Ile Asp His Leu Gln Arg His Gly
 85 90 95
 Ala Ser Lys Lys Ser Ile Thr Arg Ala Phe Asp Asp Asp Val Glu Phe
 100 105 110
 Gln Glu Arg Met Ala Glu His Ile Arg Tyr Met Val Glu Thr Ile Ala
 115 120 125
 His His Gln Val Asp Ile Asp Ser Glu Val *
 130 135

20

30

Table 1: Sequences of Oligos for PCR

Figure 3ab

left: TGACCCCTCACAAGGAGACGACCTTCCATGACCGAGTACAAGAGGGATGTAACGCACTGA

right: TACAAATGTGGTATGGCTGATTATGATCCTCTAGAGTCGGTGCTCACTGCCCCGCTTTCCA

template: pJP5603

targeting vector: pSV-paz11

Figure 3c

a-left: CTTCCATGACCGAGTACAAGAGGGATGTAACGCACTGA

a-right: ATGATCCTCTAGAGTCGGTGCTCACTGCCCCGCTTTCCA

b-left: AGACGACCTTCCATGACCGAGTACAAGAGGGATGTAACGCACTGA

b-right: GCTGATTATGATCCTCTAGAGTCGGTGCTCACTGCCCCGCTTTCCA

c-left: CACAAGGAGACGACCTTCCATGACCGAGTACAAGAGGGATGTAACGCACTGA

c-right: TGGTATGGCTGATTATGATCCTCTAGAGTCGGTGCTCACTGCCCCGCTTTCCA

d-left: TGACCCCTCACAAGGAGACGACCTTCCATGACCGAGTACAAGAGGGATGTAACGCACTGA

d-right: TACAAATGTGGTATGGCTGATTATGATCCTCTAGAGTCGGTGCTCACTGCCCCGCTTTCCA

e-left:

CACGCCCCTGACCCCTCACAAGGAGACGACCTTCCATGACCGAGTACAAGAGGGATGTAACGCACTGA

e-right:

TAAAACCTCTACAAATGTGGTATGGCTGATTATGATCCTCTAGAGTCGGTGCTCACTGCCCCGCTTTCCA

f-left:

TCCCCTGACCCACGCCCCTGACCCCTCACAAGGAGACGACCTTCCATGACCGAGTACAAGAGGGATGT
AACGCACTGA

f-right:

TAAAGCAAGTAAAACCTCTACAAATGTGGTATGGCTGATTATGATCCTCTAGAGTCGGTGCTCACTGCC
CGCTTTCCA

template: pJP5603

targeting vector: pSV-paz11

Figure 3d

a-left:

TCATCCTCTGCATGGTCAGGTCATGGATGAGCAGACGATGGTGCAAGGATCAAGGGCTGCTAAAGGAA

a-right:

TAATGCGAACAGCGCACGGCGTTAAAGTTGTTCTGCTTCATCAGCAGGATGGCGAAGAACTCCAGCAT

b-left:

CACGAGCATCATCCTCTGCATGGTCAGGTCATGGATGAGCAGACGATGGCAAGGGCTGCTAAAGGAA

b-right:

TAATGCGAACAGCGCACGGCGTTAAAGTTGTTCTGCTTCATCAGCAGGATGGCGAAGAACTCCAGCAT

c-left:

TTAACGTCACGAGCATCATCCTCTGCATGGTCAGGTCATGGATGAGCACAAGGGCTGCTAAAGGAA

c-right:

TAATGCGAACAGCGCACGGCGTTAAAGTTGTTCTGCTTCATCAGCAGGATGGCGAAGAACTCCAGCAT

d-left:

TGCTGCTGAACGGCAAGCCGTTGCTGATTCGAGGCGTTAACCGTCACGACAAGGGCTGCTAAAGGAA

d-right:

TAATGCGAACAGCGCACGGCGTTAAAGTTGTTCTGCTTCATCAGCAGGATGGCGAAGAACTCCAGCAT

e-left:

TCTCTATCGTGCGGTGGTTGAACTGCACACCGCCGACGGCACGCTGATTCAAGGGCTGCTAAAGGAA

e-right:

TAATGCGAACAGCGCACGGCGTTAAAGTTGTTCTGCTTCATCAGCAGGATGGCGAAGAACTCCAGCAT

f-left:

TGGAGTGACGGCAGTTATCTGGAAGATCAGGATATGTGGCGGATGAGCGCAAGGGCTGCTAAAGGAA

10

20

30

f-right:
 TAATGCGAACAGCGCACGGCGTTAAAGTTGTTCTGCTTCATCAGCAGGATGGCGAAGAACTCCAGCAT
 g-left:
 TGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCAAGGGCTGCTAAAGGAA
 g-right:
 TAATGCGAACAGCGCACGGCGTTAAAGTTGTTCTGCTTCATCAGCAGGATGGCGAAGAACTCCAGCAT
 h-left:
 TGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCAAGGGCTGCTAAAGGAA
 h-right:
 TATTTTGGACACCAGACCAACTGGTAATGGTAGCGACCGGCGCTCAGCTGGCGAAGAACTCCAGCAT
 template: pJP5603
 targeting vector: pSV-paz11

Figure 4

left:
 TCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAAATGGCAAAAGCTTATGCCCACCAGC
 TGGTATGGCTGATTATGATC
 right:
 TCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACA
 ATCTACCACCAGCTCTTTTCTACGGGGTCTGACGC
 template: pBR322
 targeting vector: Hoxa-PI

10

Figure 5

left:
 TGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTTAATACGACTCACTATAGGGAGAACA
 GGAAACAGCTATGCCCATAAACACCCAGAGTA
 right:
 TGCGCCGCTACAGGGCGCGTCCATTCGCCATTCAGGCCTGACTCACTAGTGATGGTGATGGTGATGTGG
 GGGGTGCCGCTCAGT
 template: pmtrx (a pBluescript vector carrying mouse trithorax cDNA)
 targeting vector: pZero2.1

20

Figure 6

left:
 TGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGGAGAAAAAATCACT
 GGATATACCACCG
 right:
 TACAGGGCGCGTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACGCCCCGCCCTGC
 CACTCATCGCA
 template: pMAK705
 targeting vector: pBAD-24 backbone Amp resistant gene

30

Figure 8

i:
 TGCCAAGCTTGACCCACTGTGGAAGTGTCCAAAAAGCGGGAAGGCTCTTGAGCTACTTCACTAACAAC
 CGG
 g:
 TCACCATCTTCGGGCCATTTGTAGACTGGAATATTTGAGCTATGAGTGTGCTACTTCACTAACAACCG
 G
 h:
 TGGCCCCAGGGTGACGCGGACATGGAGTTGTGCCAGGGCACTGGTCCATGAGAGTGCCAAGCTACTC
 GCGAC
 template: pKaZ
 targeting vector: Hoxa-PI

Figure 9

j:

TAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCT
TTGCCTGGTTTATAAAGTTCGTATAGCATACATTATACGAAGTTATGGGCTGCTAAAGGAAGCGGAACAC
G

k:

TGGCAGTTCAGGCCAATCCGCGCCGGATGCGGTGTATCGCTCGCCACTTCAACATCAACGGTAATCGCC
ATTTGACCATATAACTTCGTATAATGTATGCTATACGAAGTTATCCCCAGAGTCCCGCTCAGAAGAACT

template: pJP5603

targeting vector: JC9604 chromosome

Figure 10

l:

TAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCCATCAC
ATATACCTGCCGTTCACTAT

m:

TATCGGTGGCCGTGGTGTGCGGCTCCGCCGCCCTTCATACTGCACCGGGCGGGAAGGCGATTCCGAAGCCC
AACCTTTCATAGAAGCC

template: pIB279

targeting vector: pSV-paX1

l*: GCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAA

m*:

TCGGTGGCCGTGGTGTGCGGCTCCGCCGCCCTTCATACTGCACCGGGCGGGAAGGATCCACAGATTTGATC
CAGCGATACAGC

template: pSV-paz11

targeting vector: pSV-sacB-neo

Figure 11

n:

TACCGCATTAAGCTTATCGATGATAAGCTGTCAAACATGAGAATTGACCCGGAACCCTTCTCGAGGAA
GTTCTATTCTCTAGAAAAGTATAGGAAGTCCGAATAAATACCTGTGACGGAAGATCACTT

p:

TTCCCTCAAGAATTTTACTCTGTGAGAAACGGCCCTTAACGACGTAGTCGAGGGACCTAGAAGTTCCTAT
ACTTTCTAGAGAATAGGAAGTTCATTATCACTTATTCAGGCGTAGCACCAGGCG

template: pMAK705

targeting vector: Hoxa-P1

Figure 12

left:

TGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGGAGAAAAAATCACT
GGATATACCACCG

right:

TACAGGGCGCGTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACGCCCCGCCCTGC
CACTCATCGCA

template: pMAK705

targeting vector: pBAD-24 backbone Amp resistant gene

10

20

30

【図 1】

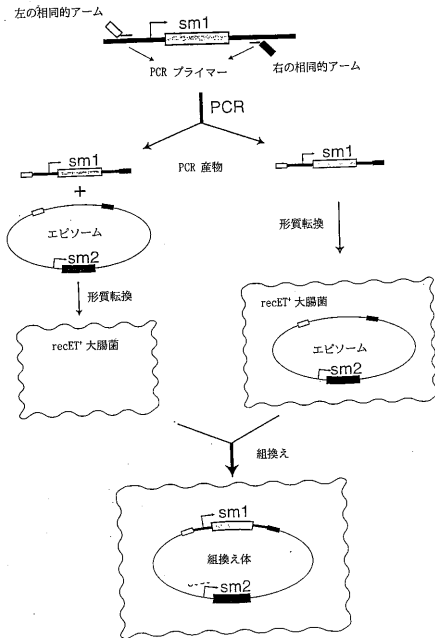


Figure 1

【図 2】

組換え体を選択する3つの方法

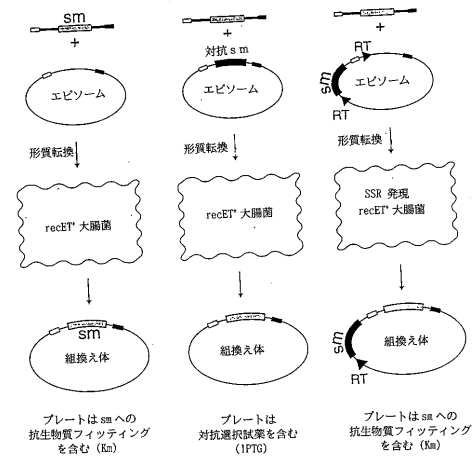
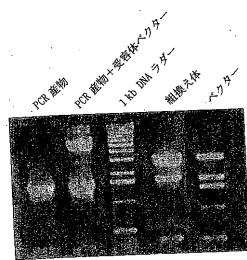


Figure 2

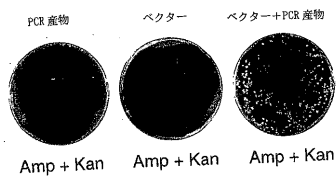
【図 3】

Figure 3

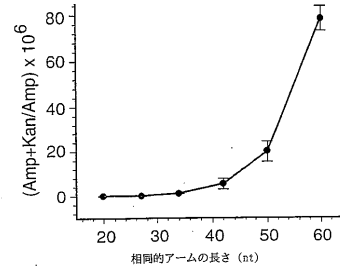
a



b



c



d

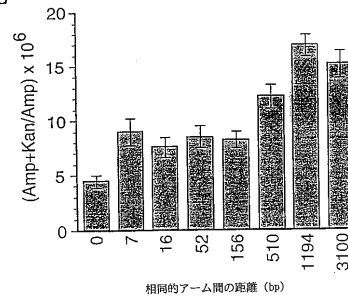


Figure 3

【図 4】

Figure 4a

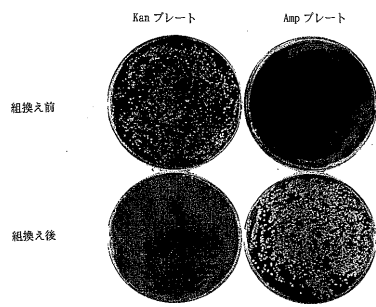
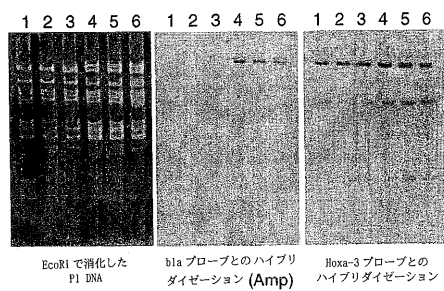


Figure 4b



レーン 1 : NS3145 起源(original)細菌株中の P1-Hox クローンのうちの 1 つ (Kan 耐性)
 レーン 2-3 : 相同組換えの前の JC9604 中の P1-Hox クローンのうちの 2 つ (Kan 耐性)
 レーン 4-6 : 相同組換えの後の JC9604 中の P1-Hox クローンのうちの 3 つ (Amp 耐性)

【図 6】

Figure 6

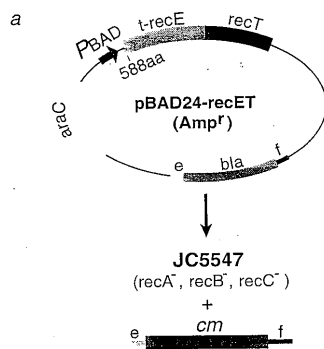
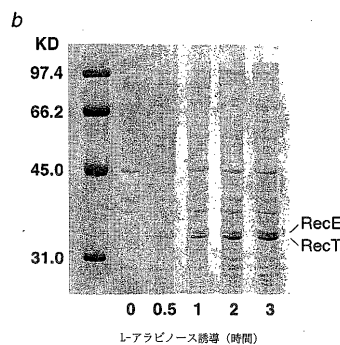
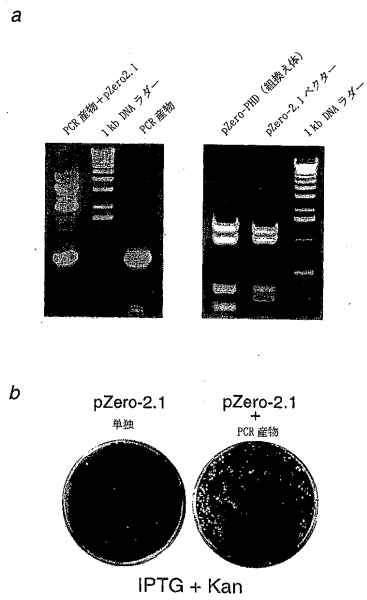


Figure 6



【図 5】

Figure 5



c

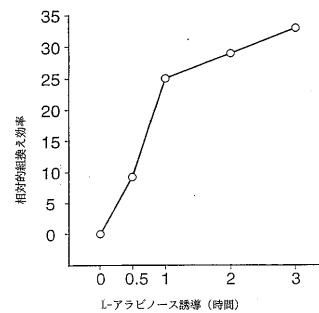


Figure 6

【図 7】

Figure 7a

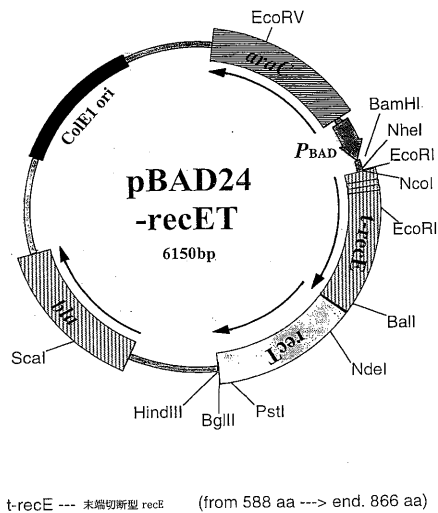


Figure 7b (続き)

450 CAT CGG TGG ATG GAG CGA CTC GTT AAT CGC TTC
 175 Met Pro Pro His Leu Ser Glu Asn Ile Ala Glu
 483 CAT GCG CCG CAG TAA CAA TTG CTC AAG CAG ATT
 164 Met Arg Arg Leu Leu Leu Gln Glu Leu Leu Asn
 516 TAT GCG CAG CAG CTC CGA ATA GCG CCC TTC CCC
 153 Ile Ala Leu Leu Glu Ser Tyr Arg Gly Glu Gly
 549 TTG CCC GGC GTT AAT GAT TTG CCC AAA CAG GTC
 142 Gln Gly Ala Asn Ile Ile Gln Gly Phe Leu Asp
 582 GCT GAA ATG CCG CTG GTG GCG TTC ATC CCG GCG
 131 Ser Phe His Pro Gln His Ala Glu Asp Pro Arg
 615 AAA GAA CCC CGT ATT GGC AAA TAT TGA CGG CCA
 120 Phe Phe Gly Thr Asn Ala Phe Ile Ser Pro Trp
 648 GTT AAG CCA TTC ATG CCA GTA GGC GCG CCG ACG
 109 Asn Leu Trp Glu His Trp Tyr Ala Arg Pro Arg
 681 AAA GTA AAC CCA CTG GTG ATA CCA TTC GCG AGC
 98 Phe Tyr Val Trp Gln His Tyr Trp Glu Arg Ala
 714 CTC CGG ATG ACG ACC GTA GTG ATG AAT CTC TCC
 87 Glu Pro His Arg Gly Tyr His His Ile Glu Gly
 747 TGG CCG GAA CAG CAA AAT ATC ACC CCG TCG GCA
 76 Pro Pro Phe Leu Leu Ile Asp Gly Pro Arg Cys
 780 AAC AAA TTC TCG TCC CTG ATT TTT CAC CAC CCC
 65 Val Phe Glu Arg Gly Gln Asn Lys Val Val Gly
 813 CTG ACC GCG AAT GGT GAG ATT GAG AAT ATA ACC
 54 Gln Gly Arg Ile Thr Leu Asn Leu Ile Tyr Gly
 846 TTT CAT TCC CAG CCG TCG GTC GAT AAA AAA ATC
 43 Lys Met Gly Leu Pro Arg Asp Ile Phe Phe Asp

Figure 7b

1 ATCGATGCATAATGTGCGCTGCAAAATGGAAGCAGGGATTC
 44 TGCAAAACCCATGCTACTCCGTCAAGCCGTCAATTGTCTGATT
 87 CGTTACCAA TTA TGA CAA CIT GAC GGC TAC ATC
 293 Ser Leu Lys Val Ala Val Asp
 120 ATT CAC TTT TTC TTC ACA ACC GGC ACG GAA CTC
 285 Asn Val Lys Glu Glu Cys Gly Ala Arg Phe Glu
 153 GCT CCG GCT GGC CCC GGT GCA TTT TTT AAA TAC
 274 Ser Pro Ser Ala Gly Thr Cys Lys Lys Phe Val
 186 CCG CGA GAA ATA GAG TTG ATC GTC AAA ACC AAC
 263 Arg Ser Phe Tyr Leu Gln Asp Asp Phe Gly Val
 219 ATT GCG ACC GAC GGT GGC GAT AGG CAT CCG GGT
 252 Asn Arg Gly Val Thr Ala Ile Pro Met Arg Thr
 252 GGT GCT CAA AAG CAG CTT CCG CTG GCT GAT ACG
 241 Thr Ser Leu Leu Leu Lys Ala Gln Ser Ile Arg
 285 TTG GTC CTC GCG CCA GCT TAA GAC GCT AAT CCC
 230 Gln Asp Glu Arg Trp Ser Leu Val Ser Ile Gly
 318 TAA CTG CTG GCG GAA AAG ATG TGA CAG ACG CGA
 219 Leu Gln Gln Arg Phe Leu His Ser Leu Arg Ser
 351 CCG CGA CAA GCA AAC ATG CTG TGC GAC GCT GGC
 208 Pro Ser Leu Cys Val His Gln Ala Val Ser Ala
 EcoRV
 384 GAT ATC AAA ATT GCT GTC TCC CAG GTG ATC GCT
 197 Ile Asp Phe Asn Ser Asp Ala Leu His Asp Ser
 417 GAT GTA CTG ACA AGC CTC GCG TAC CCG ATT ATC
 186 Ile Tyr Gln Cys Ala Glu Arg Val Arg Asn Asp

Figure 7b (続き)

879 GAG ATA ACC GIT GGC CTC AAT CCG CGT TAA ACC
 32 Leu Tyr Gly Asn Ala Glu Ile Pro Thr Leu Gly
 912 GCG CAC CAG ATG GGC ATT AAA CGA GTA TCC CCG
 21 Ala Val Leu His Ala Asn Phe Ser Tyr Gly Pro
 945 CAG CAG GGG ATC ATT TTG CCG TTC AGC CAT
 10 Leu Leu Pro Asp Asn Gln Ala Glu Ala Met
 975 ACTTTTCATA CTCCGCCAT TCAGAGAAGA AACCAATTGT
 1015 CCATATTCGA TCAGACATTC CCGTCACTGC GTCCTTTACT
 1055 GGCTCTTCTC GCTAACCAAA CCGGTAAACC CGCTTATTTAA
 1095 AAGCATTTCTG TAACAAAGCG GGACCAAGC CATGACAAAA
 1135 ACGCGTAACA AAAGTGTCCTA TAATCACGGC AGAAAAGTCC
 1175 ACATTGATTA TTTCACCGGC GTCACACITTT GCTATGCCAT
 1215 AGCATTTTITA TCCATAAGAT TAGCGGATCC TACCTGACGC
 1255 TTTTATTCGC AACTCTCTAC TGTTCTCTCCA TACCCGTTT
 1295 TTTGGGCTAG CAGGAGGAAT TCACC ATG GAT CCC GTA
 1 Met Asp Pro Val
 1332 ATC GTA GAA GAC ATA GAG CCA GGT ATT TAT TAC
 5 Ile Val Glu Asp Ile Glu Pro Gly Ile Tyr Tyr
 1365 GGA ATT TCG AAT GAG AAT TAC CAC GCG GGT CCC
 16 Gly Ile Ser Asn Glu Asn Tyr His Ala Gly Pro
 1398 GGT ATC AGT AAG TCT CAG CTC GAT GAC ATT GCT

Figure 7b (続き)

27▶Gly Ile Ser Lys Ser Gln Leu Asp Asp Ile Ala
 1431 GAT ACT CCG GCA CTA TAT TTG TGG CGT AAA AAT
 38▶Asp Thr Pro Ala Leu Tyr Leu Trp Arg Lys Asn
 1464 GCC CCC GTG GAC ACC ACA AAG ACA AAA ACG CTC
 49▶Ala Pro Val Asp Thr Thr Lys Thr Lys Thr Leu
 1497 GAT TTA GGA ACT GCT TTC CAC TGC CGG GTA CTT
 60▶Asp Leu Gly Thr Ala Phe His Cys Arg Val Leu
 EcoRI
 1530 GAA CCG GAA GAA TTC AGT AAC CGC TTT ATC GTA
 71▶Glu Pro Glu Glu Phe Ser Asn Arg Phe Ile Val
 1563 GCA CCT GAA TTT AAC CGC CGT ACA AAC GCC GGA
 82▶Ala Pro Glu Phe Asn Arg Arg Thr Asn Ala Gly
 1596 AAA GAA GAA GAG AAA GCG TTT CTG ATG GAA TGC
 93▶Lys Glu Glu Glu Lys Ala Phe Leu Met Glu Cys
 1629 GCA AGC ACA GGA AAA ACG GTT ATC ACT GCG GAA
 104▶Ala Ser Thr Gly Lys Thr Val Ile Thr Ala Glu
 1662 GAA GGC CGG AAA ATT GAA CTC ATG TAT CAA AGC
 115▶Glu Gly Arg Lys Ile Glu Leu Met Tyr Gln Ser

Figure 7b (続き)

225▶Gly Arg Tyr Pro Val Glu Ile Phe Met Met Gly
 2025 GAA GAA GCA AAA CTG GCA GGT CAA CAG GAA TAT
 236▶Glu Glu Ala Lys Leu Ala Gly Gln Gln Glu Tyr
 2058 CAC CGC AAT CTG CGA ACC CTG TCT GAC TGC CTG
 247▶His Arg Asn Leu Arg Thr Leu Ser Asp Cys Leu
 Ball
 2091 AAT ACC GAT GAA TGG CCA GCT ATT AAG ACA TTA
 258▶Asn Thr Asp Glu Trp Pro Ala Ile Lys Thr Leu
 2124 TCA CTG CCC CGC TGG GCT AAG GAA TAT GCAA
 269▶Ser Leu Pro Arg Trp Ala Lys Glu Tyr AlaA
 2155 ATG ACT AAG CAA CCA CCA ATC GCA AAA GCC GAT
 1▶Met Thr Lys Gln Pro Pro Ile Ala Lys Ala Asp
 279▶s nAs p...
 2188 CTG CAA AAA ACT CAG GGA AAC CGT GCA CCA GCA
 12▶Leu Gln Lys Thr Gln Gly Asn Arg Ala Pro Ala
 2221 GCA GTT AAA AAT AGC GAC GTG ATT AGT TTT ATT
 23▶Ala Val Lys Asn Ser Asp Val Ile Ser Phe Ile
 2254 AAC CAG CCA TCA ATG AAA GAG CAA CTG GCA GCA
 34▶Asn Gln Pro Ser Met Lys Glu Gln Leu Ala Ala
 Ndel
 2287 GCT CTT CCA CGC CAT ATG ACG GCT GAA CGT ATG
 45▶Ala Leu Pro Arg His Met Thr Ala Glu Arg Met

Figure 7b (続き)

1695 GTT ATG GCT TTG CCG CTG GGG CAA TGG CTT GTT
 126▶Val Met Ala Leu Pro Leu Gly Gln Trp Leu Val
 1728 GAA AGC GCC GGA CAC GCT GAA TCA TCA ATT TAC
 137▶Glu Ser Ala Gly His Ala Glu Ser Ser Ile Tyr
 1761 TGG GAA GAT CCT GAA ACA GGA ATT TTG TGT CGG
 148▶Trp Glu Asp Pro Glu Thr Gly Ile Leu Cys Arg
 1794 TGC CGT CCG GAC AAA ATT ATC CCT GAA TTT CAC
 159▶Cys Arg Pro Asp Lys Ile Ile Pro Glu Phe His
 1827 TGG ATC ATG GAC GTG AAA ACT ACG GCG GAT ATT
 170▶Trp Ile Met Asp Val Lys Thr Thr Ala Asp Ile
 1860 CAA CGA TTC AAA ACC GCT TAT TAC GAC TAC CGC
 181▶Gln Arg Phe Lys Thr Ala Tyr Tyr Asp Tyr Arg
 1893 TAT CAC GTT CAG GAT GCA TTC TAC AGT GAC GGT
 192▶Tyr His Val Gln Asp Ala Phe Tyr Ser Asp Gly
 1926 TAT GAA GCA CAG TTT GGA GTG CAG CCA ACT TTC
 203▶Tyr Glu Ala Gln Phe Gly Val Gln Pro Thr Phe
 1959 GTT TTT CTG GTT GCC AGC ACA ACT ATT GAA TGC
 214▶Val Phe Leu Val Ala Ser Thr Thr Ile Glu Cys
 1992 GGA CGT TAT CCG GTT GAA ATT TTC ATG ATG GGC

Figure 7b (続き)

2320 ATC CGT ATC GCC ACC ACA GAA ATT CGT AAA GTT
 56▶Ile Arg Ile Ala Thr Thr Glu Ile Arg Lys Val
 2353 CCG GCG TTA GGA AAC TGT GAC ACT ATG AGT TTT
 67▶Pro Ala Leu Gly Asn Cys Asp Thr Met Ser Phe
 2386 GTC AGT GCG ATC GTA CAG TGT TCA CAG CTC GGA
 78▶Val Ser Ala Ile Val Gln Cys Ser Gln Leu Gly
 2419 CTT GAG CCA GGT AGC GCC CTC GGT CAT GCA TAT
 89▶Leu Glu Pro Gly Ser Ala Leu Gly His Ala Tyr
 2452 TTA CTG CCT TTT GGT AAT AAA AAC GAA AAG AGC
 100▶Leu Leu Pro Phe Gly Asn Lys Asn Glu Lys Ser
 2485 GGT AAA AAG AAC GTT CAG CTA ATC ATT GGC TAT
 111▶Gly Lys Lys Asn Val Gln Leu Ile Ile Gly Tyr
 2518 CGC GGC ATG ATT GAT CTG GCT CGC CGT TCT GGT
 122▶Arg Gly Met Ile Asp Leu Ala Arg Arg Ser Gly
 2551 CAA ATC GCC AGC CTG TCA GCC CGT GAT GTC CGT
 133▶Gln Ile Ala Ser Leu Ser Ala Arg Val Val Arg
 2584 GAA GGT GAC GAG TTT AGC TTC GAA TTT GGC CTT
 144▶Glu Gly Asp Glu Phe Ser Phe Glu Phe Gly Leu
 2617 GAT GAA AAG TTA ATA CAC CGC CCG GGA GAA AAC
 155▶Asp Glu Lys Leu Ile His Arg Pro Gly Glu Asn
 2650 GAA GAT GCC CCG GTT ACC CAC GTC TAT GCT GTC
 166▶Glu Asp Ala Pro Val Thr His Val Tyr Ala Val
 2683 GCA AGA CTG AAA GAC GGA GGT ACT CAG TTT GAA
 177▶Ala Arg Leu Lys Asp Gly Gly Thr Gln Phe Glu
 2716 GTT ATG ACG CGC AAA CAG ATT GAG CTG GTG CGC
 188▶Val Met Thr Arg Lys Gln Ile Glu Leu Val Arg

Figure 7b (続き)

2749 AGC CTG AGT AAA GCT GGT AAT AAC GGG CCG TGG
 199▶Ser Leu Ser Lys Ala Gly Asn Asn Gly Pro Trp
 2782 GTA ACT CAC TGG GAA GAA ATG GCA AAG AAA ACG
 210▶Val Thr His Trp Glu Glu Met Ala Lys Lys Thr
 2815 GCT ATT CGT CGC CTG TTC AAA TAT TTG CCC GTA
 221▶Ala Ile Arg Arg Leu Phe Lys Tyr Leu Pro Val
 2848 TCA ATT GAG ATC CAG CGT GCA GTA TCA ATG GAT
 232▶Ser Ile Glu Ile Gln Arg Ala Val Ser Met Asp
 PstI
 2881 GAA AAG GAA CCA CTG ACA ATC GAT CCT GCA GAT
 243▶Glu Lys Glu Pro Leu Thr Ile Asp Pro Ala Asp
 2914 TCC TCT GTA TTA ACC GGG GAA TAC AGT GTA ATC
 254▶Ser Ser Val Leu Thr Gly Glu Tyr Ser Val Ile
 BglII HindIII
 2947 GAT AAT TCA GAG GAA TAG ATCTAAGCTT
 265▶Asp Asn Ser Glu Glu ...
 2975 GGCTGTTTTG GCGGATGAGA GAAGATTTTC AGCCTGATAC
 3015 AGATTAAATC AGAACGCAGA AGCGGTCTGA TAAACACAGAA
 3055 TTTGCCTGGC GGCAGTAGCG CGGTGGTCCC ACCTGACCCC
 3095 ATGCCGAAC TACAAGTGAA ACGCCGTAGC GCCGATGGTA
 3135 GTGTGGGGTC TCCCCATGCG AGAGTAGGGA ACTGCCAGGC
 3175 ATCAATAAAA ACGAAAGGCT CAGTCGAAAG ACTGGGCCIT
 3215 TCGTTTATAT TGTGTGTTGT CGGTGAACGC TCTCTGAGT
 3255 AGGACAAATC CGCCGGGAGC GGATTTGAAC GTTGCGAAGC
 3295 AACGCCCCCG AGGGTGGCGC GCAGGACGCC CGCATAAAC
 3335 TGCCAGGCAT CAAATTAAAG AGAAGGCCAT CCTGACGGAT

Figure 7b (続き)

3856 AGT GCT GCC ATA ACC ATG AGT GAT AAC ACT GCG
 122▶Ser Ala Ala Ile Thr Met Ser Asp Asn Thr Ala
 3889 GCC AAC TTA CTT CTG ACA ACG ATC GGA GGA CCG
 133▶Ala Asn Leu Leu Leu Thr Thr Ile Gly Gly Pro
 3922 AAG GAG CTA ACC GCT TTT TTG CAC AAC ATG GGG
 144▶Lys Glu Leu Thr Ala Phe Leu His Asn Met Gly
 3955 GAT CAT GTA ACT CGC CTT GAT CGT TGG GAA CCG
 155▶Asp His Val Thr Arg Leu Asp Arg Trp Glu Pro
 3988 GAG CTG AAT GAA GCC ATA CCA AAC GAC GAG CGT
 166▶Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu Arg
 4021 GAC ACC ACG ATG CCT GTA GCA ATG GCA ACA ACG
 177▶Asp Thr Thr Met Pro Val Ala Met Ala Thr Thr
 4054 TTG CGC AAA CTA TTA ACT GGC GAA CTA CTT ACT
 188▶Leu Arg Lys Leu Leu Thr Gly Glu Leu Leu Thr
 4087 CTA GCT TCC CGG CAA CAA TTA ATA GAC TGG ATG
 199▶Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp Met
 4120 GAG GCG GAT AAA GTT GCA GGA CCA CTT CTG CGC
 210▶Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg
 4153 TCG GCC CTT CCG GCT GGC TGG TTT AIT GCT GAT
 221▶Ser Ala Leu Pro Ala Gly Trp Phe Ile Ala Asp
 4186 AAA TCT GGA GCC GGT GAG CGT GGG TCT CGC GGT
 232▶Lys Ser Gly Ala Gly Glu Arg Gly Ser Arg Gly
 4219 ATC ATT GCA GCA CTG GGG CAA GAT GGT AAG CCC
 243▶Ile Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro
 4252 TCC CGT ATC GTA GTT ATC TAC ACG ACG GGG AGT
 254▶Ser Arg Ile Val Val Ile Tyr Thr Thr Gly Ser

Figure 7b (続き)

3375 GGCCTTTTTG CGTTCTTACA AACTCTTTTG TTTATTTTTT
 3415 TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC
 3455 CCTGATAAAT GCTTCAATAA TATTGAAAAA GGAAGAGT AT
 1▶Me
 3495 G AGT ATT CAA CAT TTC CGT GTC GCC CTT ATT
 1▶t Ser Ile Gln His Phe Arg Val Ala Leu Ile
 3526 CCC TTT TTT GCG GCA TTT TGC CTT CCT GTT TTT
 12▶Pro Phe Phe Ala Ala Phe Cys Leu Pro Val Phe
 3559 GCT CAC CCA GAA ACG CTG GTG AAA GTA AAA GAT
 23▶Ala His Pro Glu Thr Leu Val Lys Val Lys Asp
 3592 GCT GAA GAT CAG TTG GGT GCA CGA GTG GGT TAC
 34▶Ala Glu Asp Gln Leu Gly Ala Arg Val Gly Tyr
 3625 ATC GAA CTG GAT CTC AAC AGC GGT AAG ATC CTT
 45▶Ile Glu Leu Asp Leu Asn Ser Gly Lys Ile Leu
 3658 GAG AGT TTT CGC CCC GAA GAA CGT TTT CCA ATG
 56▶Glu Ser Phe Arg Pro Glu Glu Arg Phe Pro Met
 3691 ATG AGC ACT TTT AAA GTT CTG CTA TGT GGC GCG
 67▶Met Ser Thr Phe Lys Val Leu Leu Cys Gly Ala
 3724 GTA TTA TCC CGT GTT GAC GCC GGG CAA GAG CAA
 78▶Val Leu Ser Arg Val Asp Ala Gly Gln Glu Gln
 3757 CTC GGT CGC CGC ATA CAC TAT TCT CAG AAT GAC
 89▶Leu Gly Arg Arg Ile His Tyr Ser Gln Asn Asp
 Scal
 3790 TTG GTT GAG TAC TCA CCA GTC ACA GAA AAG CAT
 100▶Leu Val Glu Tyr Ser Pro Val Thr Glu Lys His
 3823 CTT ACG GAT GGC ATG ACA GTA AGA GAA TTA TGC
 111▶Leu Thr Asp Gly Met Thr Val Arg Glu Leu Cys

Figure 7b (続き)

4285 CAG GCA ACT ATG GAT GAA CGA AAT AGA CAG ATC
 265▶Gln Ala Thr Met Asp Glu Arg Asn Arg Gln Ile
 4318 GCT GAG ATA GGT GCC TCA CTG ATT AAG CAT TGG
 276▶Ala Glu Ile Gly Ala Ser Leu Ile Lys His Trp
 4351 TAA CTGTCAGACC AAGTTTACTC ATATATACIT
 287▶...
 4384 TAGATTGATT TACGCGCCCT GTAGCGGCGC ATTAAGCGCG
 4424 GCGGTGTGG TGGTTACGCG CAGCGTGACC GCTACACTTG
 4464 CCAGCGCCCT AGCGCCCGCT CCTTTGCGTT TCTCCCTTC
 4504 CTTCTCGCC ACGTTCGCGC GCTTTCCCGC TCAGCTCTTA
 4544 AATCGGGGGC TCCCTTTAGG GTTCOGATT AGTGCTTTAC
 4584 GGCACCTCGA CCCCCAAAAA CTTGATTGG GTGATGGTTC
 4624 ACGTAGTGGG CCATCGCCCT GATAGACGGT TTTTCGCCCT
 4664 TTGACGTGG AGTCCACGTT CTTTAATAGT GGACTCTTGT
 4704 TCCAAACTTG AACAACTAC AACCTATCT CCGGCTATTC
 4744 TTTTGATTTA TAAGGGATT TGCCGATTTC GGCTATTGG
 4784 TTAATAAATG AGCTGATTTA ACAAATTTT AACCGGAATT
 4824 TTAACAAAAA ATTAACGTTT ACAATTTAAA AGGATCTAGG
 4864 TGAAGATCCT TTTTGATAAT CTCATGACCA AATCCCTTA
 4904 ACGTAGATT TCGTTCCACT GAGCGTCAGA CCCCCTAGAA
 4944 AAGATCAAG GATCTTCTTG AGATCCCTTT TTTCTGCGCG
 4984 TAATCTGCTG CTTCGAAACA AAAAAACCAC CGTACCAGC
 5024 GGTGGTTTGT TTGCCGGATC AAGAGCTACC AACTCTTTTT

Figure 7b (続き)

5064 CCGAAGGTAA CTGGCTTCAG CAGAGCGCAG ATACCAAATA
 5104 CTGTCTTCT AGTGTAGCCG TAGTTAGGCC ACCACTTCAA
 5144 GAACCTCTGA GCACCGCCTA CATACTCGC TCTGCTAATC
 5184 CTGTTACCAG TGGCTGCTGC CAGTGGCGAT AAGTCGTGTC
 5224 TTACCGGGTT GGACTCAAGA CGATAGTTAC CGGATAAGGC
 5264 GCAGCGGTCG GGCTGAACGG GGGTTCGTG CACACAGCCC
 5304 AGCTTGGAGC GAACGACCTA CACCGAACTG AGATACCTAC
 5344 AGCGTGAGCT ATGAGAAAGC GCCACGCTTC CCGAAGGGAG
 5384 AAAGGGGAC AGGTATCCGG TAAGCGGCAG GGTGGAACA
 5424 GGAGAGCGCA CGAGGAGCTT TCCAGGGGGA AACGCTGGT
 5464 ATCTTTATAG TCCTGTCCGG TTTCGCCACC TCTGACTTGA
 5504 GCGTCGATTT TTGTGATGCT CGTCAGGGGG GCGGAGCCTA
 5544 TGGAAAAACG CCAGCAACGC GGCCTTTTITA CGGTCTCTGG
 5584 CCTTTTGCTG GCCTTTTGCT CACATGTTCT TTCCTGCGTT
 5624 ATCCCCTGAT TCTGTGGATA ACCGTATTAC CGCCTTGTAG
 5664 TGAGCTGATA CCGCTCCGCG CAGCCGAACG ACCGAGCGCA
 5704 GCGAGTCAGT GAGCGAGGAA GCGGAAGAGC GCCTGATGCG
 5744 GTATTTTCTC CTTACGCATC TGTGCGGTAT TTCACACCGC
 5784 ATAGGGTCAT GGCTGCGCCC CGACACCCGC CAACACCCGC
 5824 TGACGCGCCC TGACGGGCTT GTCTGCTCCC GGCATCCGCT
 5864 TACAGACAAG CTGTGACCGT CTCGGGAGC TGCGATGTGC
 5904 AGAGGTTTTC ACCGTATCA CCGAAACGCG CGAGGCAGCA
 5944 AGGAGATGGC GCCCAACAGT CCCCCGGCCA CCGGGCCTGC

Figure 7b (続き)

5984 CACCATACCC ACGCGAAAC AAGCGTCAT GAGCCGAAG
 6024 TGGGAGGCC GATCTTCCCC ATCGGTGATG TCGGCGATAT
 6064 AGGCGCCAGC AACCACCTT GTGGGCCCGG TGATGCCGGC
 6104 CACGATCCGT CCGCGTAGA GGATCTGCTC ATGTTTGACA
 6144 GCTTATC

【図 8】

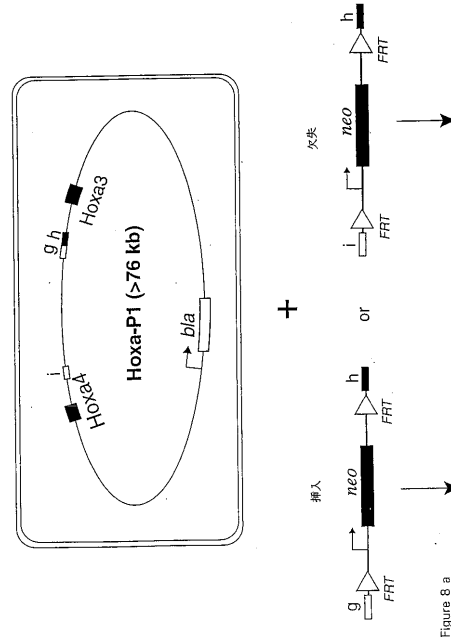


Figure 8 a

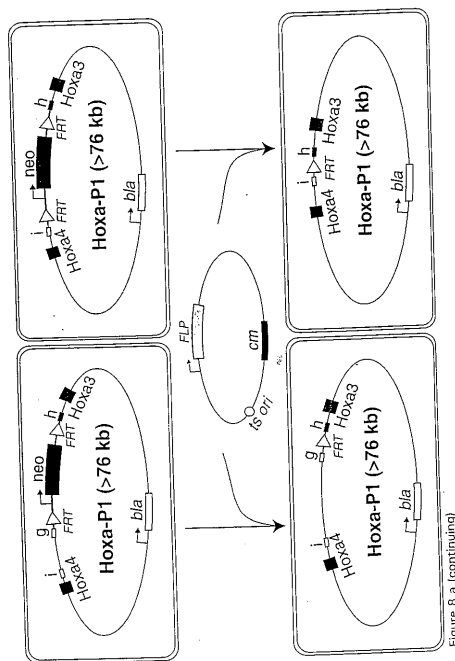
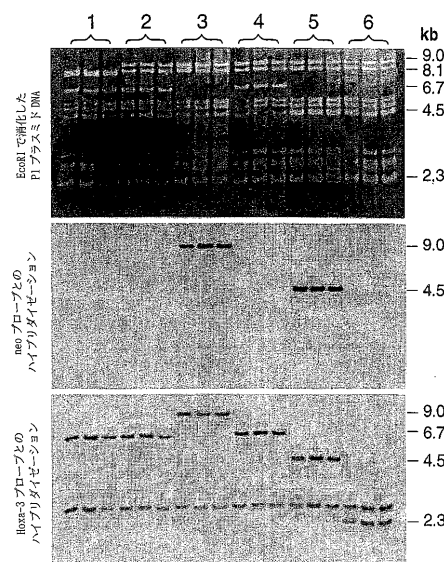


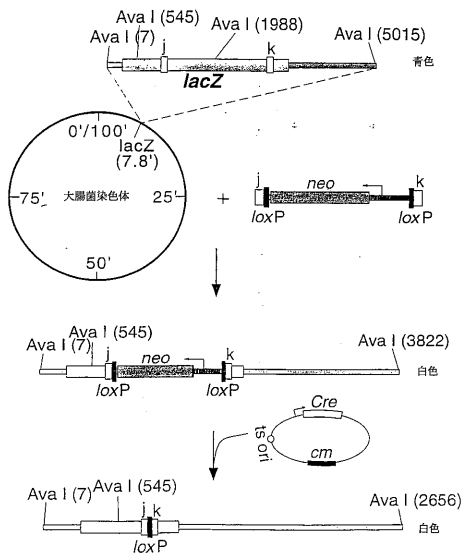
Figure 8 a (continuing)

Figure 8 b



【図 9】

Figure 9a



【図 10】

Figure 10a

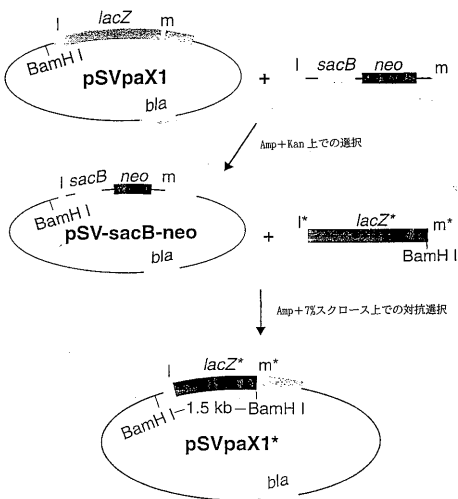


Figure 9

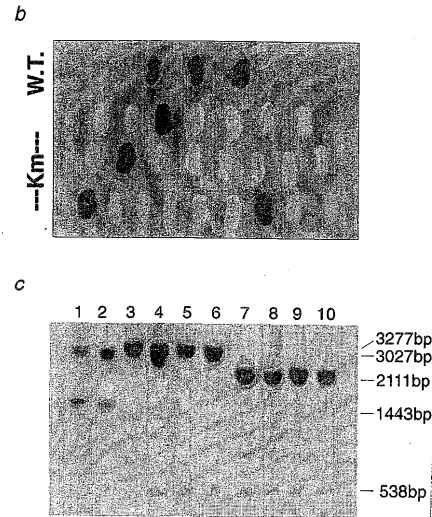
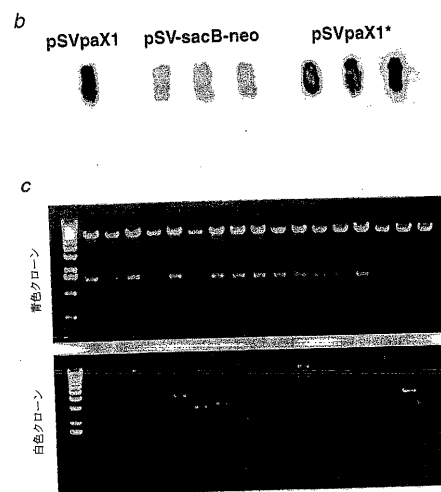


Figure 10



【図 11】

Figure 11a

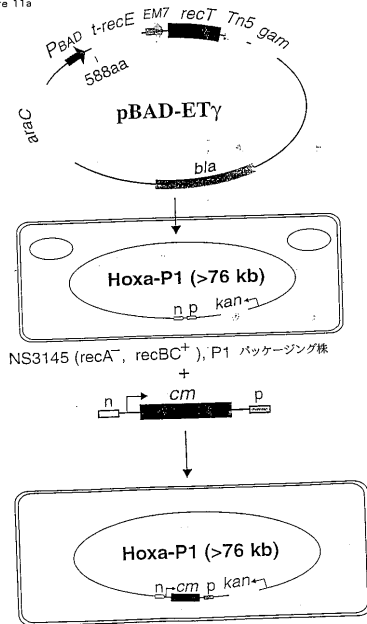
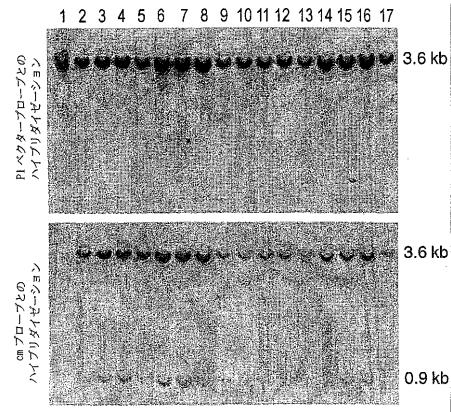
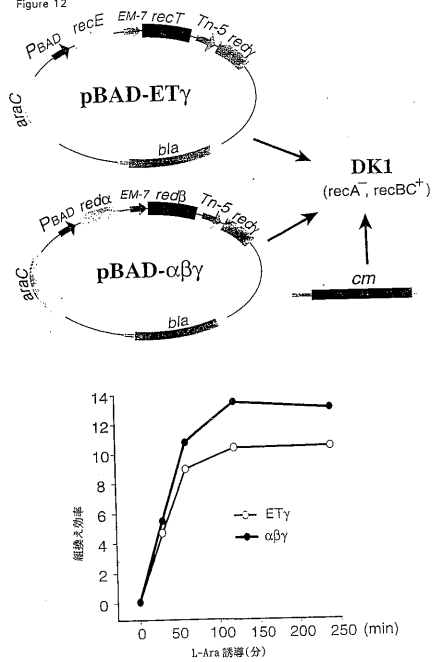


Figure 11 b



【図 12】

Figure 12



【図 13】

Figure 13 a

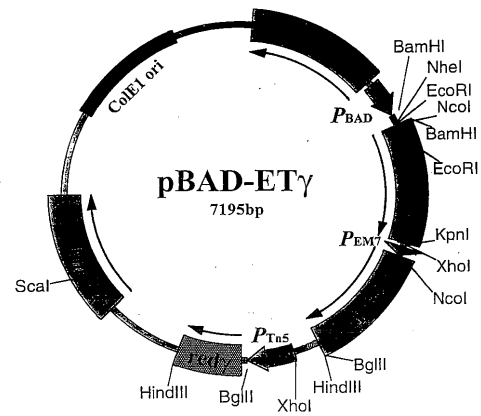


Figure 13b

1 ATCGATGCATAATGTGCTGTCAAAATGGACGAAGCAGGG
 40 ATCTCTGCAAAACCCCTATGCTACTCCGTCAAGCCGTCAATT
 79 GTCTGATTCGTTTACCAA TTA TGA CAA CTG GAC
 293♦♦♦ Ser Leu Lys Val
 111 GGC TAC ATC ATT CAC TTT TTC TTC ACA ACC
 288♦Ala Val Asp Asn Val Lys Glu Glu Cys Gly
 141 GGC ACG GAA CTC GCT CGG GCT GGC CCC GGT
 278♦Ala Arg Phe Glu Ser Pro Ser Ala Gly Thr
 171 GCA TTT TTT AAA TAC CCG CGA GAA ATA GAG
 268♦Cys Lys Lys Phe Val Arg Ser Phe Tyr Leu
 201 TTG ATC GTC AAA ACC AAC ATT GCG ACC GAC
 258♦Gln Asp Asp Phe Gly Val Asn Arg Gly Val
 231 GGT GGC GAT AGG CAT CCG GGT GGT GCT CAA
 248♦Thr Ala Ile Pro Met Arg Thr Thr Ser Leu
 261 AAG CAG CTT CGC CTG GCT GAT ACG TTG GTC
 238♦Leu Leu Lys Ala Gln Ser Ile Arg Gln Asp
 291 CTC GCG CCA GCT TAA GAC GCT AAT CCC TAA
 228♦Glu Arg Trp Ser Leu Val Ser Ile Gly Leu
 321 CTG CTG GCG GAA AAG ATG TGA CAG ACG CGA
 218♦Gln Gln Arg Phe Leu His Ser Leu Arg Ser
 351 CGG CGA CAA GCA AAC ATG CTG TGC GAC GCT
 208♦Pro Ser Leu Cys Val His Gln Ala Val Ser
 381 GGC GAT ATC AAA ATT GCT GTC TGC CAG GTG
 198♦Ala Ile Asp Phe Asn Ser Asp Ala Leu His
 411 ATC GCT GAT GTA CTG ACA AGC CTC GCG TAC

Figure 13b (続き)

188♦Asp Ser Ile Tyr Gln Cys Ala Glu Arg Val
 441 CCG ATT ATC CAT CGG TGG ATG GAG CGA CTC
 178♦Arg Asn Asp Met Pro Pro His Leu Ser Glu
 471 GTT AAT CGC TTC CAT GCG CCG CAG TAA CAA
 168♦Asn Ile Ala Glu Met Arg Arg Leu Leu Leu
 501 TTG CTC AAG CAG ATT TAT CGC CAG CAG CTC
 158♦Gln Glu Leu Leu Asn Ile Ala Leu Leu Glu
 531 CGA ATA GCG CCC TTC CCC TTG CCC GGC GTT
 148♦Ser Tyr Arg Gly Glu Gly Gln Gly Ala Asn
 561 AAT GAT TTG CCC AAA CAG GTC GCT GAA ATG
 138♦Ile Ile Gln Gly Phe Leu Asp Ser Phe His
 591 CCG CTG GTG CCG TTC ATC CGG GCG AAA GAA
 128♦Pro Gln His Ala Glu Asp Pro Arg Phe Phe
 621 CCC CGT ATT GGC AAA TAT TGA CGG CCA GTT
 118♦Gly Thr Asn Ala Phe Ile Ser Pro Trp Asn
 651 AAG CCA TTC ATG CCA GTA GCG GCG CGG ACG
 108♦Leu Trp Glu His Trp Tyr Ala Arg Pro Arg
 681 AAA GTA AAC CCA CTG GTG ATA CCA TTC GCG
 98♦Phe Tyr Val Trp Gln His Tyr Trp Glu Arg
 711 AGC CTC CGG ATG ACG ACC GTA GTG ATG AAT
 88♦Ala Glu Pro His Arg Gly Tyr His His Ile
 741 CTC TCC TGG CCG GAA CAG CAA AAT ATC ACC
 78♦Glu Gly Pro Pro Phe Leu Leu Ile Asp Gly
 771 CGG TCG GCA AAC AAA TTC TCG TCC CTG ATT
 68♦Pro Arg Cys Val Phe Glu Arg Gly Gln Asn

Figure 13b (続き)

801 TTT CAC CAC CCC CTG ACC GCG AAT GGT GAG
 58♦Lys Val Val Gly Gln Gly Arg Ile Thr Leu
 831 ATT GAG AAT ATA ACC TTT CAT TCC CAG CGG
 48♦Asn Leu Ile Tyr Gly Lys Met Gly Leu Pro
 861 TCG GTC GAT AAA AAA ATC GAG ATA ACC GTT
 38♦Arg Asp Ile Phe Phe Asp Leu Tyr Gly Asn
 891 GGC CTC AAT CCG CGT TAA ACC CGC CAC CAG
 28♦Ala Glu Ile Pro Thr Leu Gly Ala Val Leu
 921 ATG GGC ATT AAA CGA GTA TCC CGG CAG CAG
 18♦His Ala Asn Phe Ser Tyr Gly Pro Leu Leu
 951 GGG ATC ATT TTG CCG TTC AGC CAT ACITTTTC
 8♦Pro Asp Asn Gln Ala Glu Ala Met
 982 ATACTCCCGCCATTCAGAGAAGAAACCAATTGTCCATAT
 1021 TGCATCAGACATTGCGCTCACTGCGTCTTTTACTGGCTC
 1060 TTCTCGCTAACCAACCGGTAACCCCGCTTATTAAAGC
 1099 ATTCTGTAACAAAGCGGGACCAAGCCATGACAAAAAGC
 1138 CGTAACAAAAGTGCTATAATCAACGGCAGAAAAGTCCAC
 1177 ATTGATTATTTGACCGGCTCACACTTTGCTATGCCATA
 1216 GCATTTTATCCATAAGATTAGCGGATCCTACCTGACGC
 BamHI
 1216 GCATTTTATCCATAAGATTAGCGGATCCTACCTGACGC

Figure 13b (続き)

1255 TTTTATTCGCAACTCTCTACTGTTTCTCCATACCCGTTT
 NheI EcoRI NcoI BamHI
 1294 TTTTGGGCTAGCAGGAGGAAT TCACC AATG GAT CCC
 1♦Met Asp Pro
 1329 GTA ATC GTA GAA GAC ATA GAG CCA GGT ATT
 4♦Val Ile Val Glu Asp Ile Glu Pro Gly Ile
 1359 TAT TAC GGA ATT TCG AAT GAG AAT TAC CAC
 14♦Tyr Tyr Gly Ile Ser Asn Glu Asn Tyr His
 1389 GCG GGT CCC GGT ATC AGT AAG TCT CAG CTC
 24♦Ala Gly Pro Gly Ile Ser Lys Ser Gln Leu
 1419 GAT GAC ATT GCT GAT ACT CCG GCA CTA TAT
 34♦Asp Asp Ile Ala Asp Thr Pro Ala Leu Tyr
 1449 TTG TGG CGT AAA AAT GCC CCC GTG GAC ACC
 44♦Leu Trp Arg Lys Asn Ala Pro Val Asp Thr
 1479 ACA AAG ACA AAA ACG CTC GAT TTA GGA ACT
 54♦Thr Lys Thr Lys Thr Leu Asp Leu Gly Thr
 1509 GCT TTC CAC TGC CCG GTA CTT GAA CCG GAA
 64♦Ala Phe His Cys Arg Val Leu Glu Pro Glu
 EcoRI
 1539 GAA TTC AGT AAC CGC TTT ATC GTA GCA CCT
 74♦Glu Phe Ser Asn Arg Phe Ile Val Ala Pro
 1569 GAA TTT AAC CGC CGT ACA AAC GCC GGA AAA
 84♦Glu Phe Asn Arg Arg Thr Asn Ala Gly Lys
 1599 GAA GAA GAG AAA GCG TTT CTG ATG GAA TGC
 94♦Glu Glu Glu Lys Ala Phe Leu Met Glu Cys
 1629 GCA AGC ACA GGA AAA ACG GTT ATC ACT GCG
 104♦Ala Ser Thr Gly Lys Thr Val Ile Thr Ala

Figure 13b (続き)

1659 GAA GAA GGC CGG AAA ATT GAA CTC ATG TAT
 114▶Glu Glu Gly Arg Lys Ile Glu Leu Met Tyr
 1689 CAA AGC GTT ATG GCT TTG CCG CTG GGG CAA
 124▶Gln Ser Val Met Ala Leu Pro Leu Gly Gln
 1719 TGG CTT GTT GAA AGC GCC GGA CAC GCT GAA
 134▶Trp Leu Val Glu Ser Ala Gly His Ala Glu
 1749 TCA TCA ATT TAC TGG GAA GAT CCT GAA ACA
 144▶Ser Ser Ile Tyr Trp Glu Asp Pro Glu Thr
 1779 GGA ATT TTG TGT CGG TGC CGT CCG GAC AAA
 154▶Gly Ile Leu Cys Arg Cys Arg Pro Asp Lys
 1809 ATT ATC CCT GAA TTT CAC TGG ATC ATG GAC
 164▶Ile Ile Pro Glu Phe His Trp Ile Met Asp
 1839 CTG AAA ACT ACG GCG GAT ATT CAA CGA TTC
 174▶Val Lys Thr Thr Ala Asp Ile Gln Arg Phe
 1869 AAA ACC GCT TAT TAC GAC TAC CGC TAT CAC
 184▶Lys Thr Ala Tyr Tyr Asp Tyr Arg Tyr His
 1899 GTT CAG GAT GCA TTC TAC AGT GAC GGT TAT
 194▶Val Gln Asp Ala Phe Tyr Ser Asp Gly Tyr
 1929 GAA GCA CAG TTT GGA GTG CAG CCA ACT TTC
 204▶Glu Ala Glu Phe Gly Val Gln Pro Thr Phe
 1959 GTT TTT CTG GTT GCC AGC ACA ACT ATT GAA
 214▶Val Phe Leu Val Ala Ser Thr Thr Ile Glu
 1989 TGC GGA CGT TAT CCG GTT GAA ATT TTC ATG
 224▶Cys Gly Arg Tyr Pro Val Glu Ile Phe Met
 2019 ATG GGC GAA GAA GCA AAA CTG GCA GGT CAA
 234▶Met Gly Glu Glu Ala Lys Leu Ala Gly Gln

Figure 13b (続き)

2049 CAG GAA TAT CAC CGC AAT CTG CGA ACC CTG
 244▶Gln Glu Tyr His Arg Asn Leu Arg Thr Leu
 2079 TCT GAC TGC CTG AAT ACC GAT GAA TGG CCA
 254▶Ser Asp Cys Leu Asn Thr Asp Glu Trp Pro
 2109 GCT ATT AAG ACA TTA TCA CTG CCC CGC TGG
 264▶Ala Ile Lys Thr Leu Ser Leu Pro Arg Trp
 XhoI KpnI
 2139 GCT AAG GAA TAT GCA AAT GAC TAGATCTCGAG
 274▶Ala Lys Glu Tyr Ala Asn Asp
 2171 GTACCCGAGCAGCTGTGTGACAATTAATCAATCGGCATAGT
 2210. ATATCGGCATAGTATAATACGACAAGGTGAGGAACATAAA
 NcoI
 2249 CC ATG GCT AAG CAA CCA CCA ATC GCA AAA
 1▶Met Ala Lys Gln Pro Pro Ile Ala Lys
 2278 GCC GAT CTG CAA AAA ACT CAG GGA AAT CGT
 10▶Ala Asp Leu Gln Lys Thr Gln Gly Asn Arg
 2308 GCA CCA GCA GCA GTT AAA AAT AGC GAC GTG
 20▶Ala Pro Ala Ala Val Lys Asn Ser Asp Val
 2338 ATT AGT TTT ATT AAC CAG CCA TCA ATG AAA
 30▶Ile Ser Phe Ile Asn Gln Pro Ser Met Lys
 2368 GAG CAA CTG GCA GCA GCT CTT CCA CGC CAT
 40▶Glu Gln Leu Ala Ala Leu Pro Arg His
 2398 ATG ACG GCT GAA CGT ATG ATC CGT ATC GCC
 50▶Met Thr Ala Glu Arg Met Ile Arg Ile Ala
 2428 ACC ACA GAA ATT CGT AAA GTT CCG GCG TTA
 60▶Thr Thr Glu Ile Arg Lys Val Pro Ala Leu

Figure 13b (続き)

2458 GGA AAC TGT GAC ACT ATG AGT TTT GTC AGT
 70▶Gly Asn Cys Asp Thr Met Ser Phe Val Ser
 2488 GCG ATC GTA CAG TGT TCA CAG CTC GGA CTT
 80▶Ala Ile Val Gln Cys Ser Gln Leu Gly Leu
 2518 GAG CCA GGT AGC GCC CTC GGT CAT GCA TAT
 90▶Glu Pro Gly Ser Ala Leu Gly His Ala Tyr
 2548 TTA CTG CCT TTT GGT AAT AAA AAC GAA AAG
 100▶Leu Leu Pro Phe Gly Asn Lys Asn Glu Lys
 2578 AGC GGT AAA AAG AAC GTT CAG CTA ATC ATT
 110▶Ser Gly Lys Lys Asn Val Gln Leu Ile Ile
 2608 GGC TAT CGC GGC ATG ATT GAT CTG GCT CGC
 120▶Gly Tyr Arg Gly Met Ile Asp Leu Ala Arg
 2638 CGT TCT GGT CAA ATC GCC AGC CTG TCA GCC
 130▶Arg Ser Gly Gln Ile Ala Ser Leu Ser Ala
 2668 CGT GTT GTC CGT GAA GGT GAC GAG TTT AGC
 140▶Arg Val Val Arg Glu Gly Asp Glu Phe Ser
 2698 TTC GAA TTT GGC CTT GAT GAA AAG TTA ATA
 150▶Phe Glu Phe Gly Leu Asp Glu Lys Leu Ile
 2728 CAC CGC CCG GGA GAA AAC GAA GAT GCC CCG
 160▶His Arg Pro Gly Glu Asn Glu Asp Ala Pro
 2758 GTT ACC CAC GTC TAT GCT GTC GCA AGA CTG
 170▶Val Thr His Val Tyr Ala Val Ala Arg Leu
 2788 AAA GAC GGA GGT ACT CAG TTT GAA GTT ATG
 180▶Lys Asp Gly Gly Thr Gln Phe Glu Val Met
 2818 ACG CGC AAA CAG ATT GAG CTG GTG CGC AGC
 190▶Thr Arg Lys Gln Ile Glu Leu Val Arg Ser

Figure 13b (続き)

2848 CTG AGT AAA GCT GGT AAT AAC GGG CCG TGG
 200▶Leu Ser Lys Ala Gly Asn Asn Gly Pro Trp
 2878 GTA ACT CAC TGG GAA GAA ATG GCA AAG AAA
 210▶Val Thr His Trp Glu Glu Met Ala Lys Lys
 2908 ACG GCT ATT CGT CGC CTG TTC AAA TAT TTG
 220▶Thr Ala Ile Arg Arg Leu Phe Lys Tyr Leu
 2938 CCC GTA TCA ATT GAG ATC CAG CGT GCA GTA
 230▶Pro Val Ser Ile Glu Ile Gln Arg Ala Val
 2968 TCA ATG GAT GAA AAG GAA CCA CTG ACA ATC
 240▶Ser Met Asp Glu Lys Glu Pro Leu Thr Ile
 2998 GAT CCT GCA GAT TCC TCT GTA TTA ACC GGG
 250▶Asp Pro Ala Asp Ser Ser Val Leu Thr Gly
 3028 GAA TAC AGT GTA ATC GAT AAT TCA GAG GAA
 260▶Glu Tyr Ser Val Ile Asp Asn Ser Glu Glu
 BglII HindIII
 3058 TAG ATCTAAGCTTCTCTGCTGAACATCAAGGCAAGAAA
 270▶...
 3096 ACATCTGTGTCAAAGACAGCATCCCTTGAAACAAGGACAA
 3135 TTAACAGTTAACAATAAAACGCAAAAGAAAATGCCGA
 3174 TATCCTATTGGCATTTTCTTTTATTTCCTATCAACATAA
 XhoI
 3213 AGGTGAATCCCATACCTCGAGCTTCACGCTGCCGCAAGC
 3252 ACTCAGGGCGCAAGGGCTGCTAAAGGAAGCGGAACACG
 3291 TAGAAAGCCAGTCCGCAAGAACGGTGCTGACCCCGGATG
 3330 AATGTCAGCTACTGGGCTATCTGGACAAGGAAAACGCA
 3369 AGCGCAAAGAGAAAGCAGGTAGCTTCAGTGGGCTTACA

Figure 13b (続き)

3897 CGT CGG TTT GAT GAC GAT GTT GAG TTT CAG
104▶ Arg Ala Phe Asp Asp Val Glu Phe Gl n
3924 GAG CGC ATG GCA GAA CAC ATC CGG TAC ATG
114▶ Glu Arg Met Ala Glu His Ile Arg Tyr Met
3957 GTT GAA ACC ATT GCT CAC CAC CAG GTT GAG
124▶ Val Glu Thr Ile Ala His His Gl n Val Asp
HindIII
3987 ATT GAT TCA GAG GTA TAA AACGAGTAGA AGCT
134▶ Ile Asp Ser Glu Val * * *
4019 TGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGAT
4058 ACAGATTAAATCAGAAACGCAAGACGGTCTGATAAACA
4097 GAATTTCCCTGACCGCGCAGTAGCCCGGTGGTCCCACTGA
4136 CCCCATGCCAACTCAGAAGTGAACCCCGTAGCCGCGGA
4175 TGGTAGTGTGGGGTCTCCCATCGCAGAGTAGGGAACTG
4214 CCAGGCATCAAATAAACGAAAGGCTCAGTCGAAAGACT
4253 GGGCCTTTCTGTTTATCTGTGTGTTGTGCGTGAACGCTC
4292 TCCTGAGTAGGACAAATCCGCGGGAGCGGATTGTGAAGC
4331 TTGCGAAGCAACGGCCCGGAGGTTGGCGGAGGAGCGCC
4370 CGCCATAAACTGCCCAAGTCAAATTAACGAGAGGCCA
4409 TCTTCGACGATGGGCTCTTTGCGTTTATGCAAACTCTTT
4448 TGTTTATTTTCTAAATACATTCAAATATGTATCCGCTC
4487 ATGAGACAATAACCCCTGATAAATGCTTCAATAATATTGA
4526 AAAAGGAGAGT ATG AGT ATT CAA CAT TTC
1▶ Met Ser Ile Gl n His Phe

Figure 13b (続き)

4946	CTG	ACA	ACG	ATC	GGA	GGA	CCG	AAG	GAG	CTA
137▶	Leu	Thr	Thr	Ile	Gly	Gly	Pro	Lys	Glu	Leu
4976	ACC	GCT	TTT	TTG	CAC	AAC	ATG	GGG	GAT	CAT
147▶	Thr	Ala	Phe	Leu	His	Asn	Met	Gly	Asp	His
5006	GTA	ACT	CGC	CTT	GAT	CGT	TGG	GAA	CCG	GAT
157▶	Val	Thr	Arg	Leu	Leu	Asp	Arg	Glu	Pro	Glu
5036	CTG	AAT	GAA	GCC	ATA	CCA	AAC	GAC	GAG	CGT
167▶	Leu	Asn	Glu	Ala	Ile	Pro	Asn	Asp	Glu	Arg
5066	GAC	ACC	ACG	ATG	CCT	GTA	GCA	ATG	GCA	ACA
177▶	Asp	Thr	Met	Pro	Met	Val	Ala	Met	Ala	Thr
5096	ACG	TTG	CGC	AAA	CTA	TTA	ACT	GGC	GAA	CTA
187▶	Thr	Leu	Arg	Lys	Leu	Leu	Gly	Glu	Glu	Thr
5126	CTT	ACT	CTA	GCT	TCC	CGC	CAA	CAA	TTA	ATA
197▶	Leu	Thr	Leu	Ala	Ser	Arg	Gln	Gln	Leu	Ile
5156	GAC	TGG	ATG	GAG	GCG	GAT	AAA	GTT	GCA	GGA
207▶	Asp	Trp	Met	Glu	Ala	Asp	Lys	Val	Ala	Gly
5186	CCA	CTT	CTG	CGC	TCG	GCC	CTT	CCG	GCT	GGC
217▶	Pro	Leu	Leu	Arg	Ser	Ala	Leu	Pro	Ala	Gly
5216	TGG	TTT	ATT	GCT	GAT	AAA	TCT	GGA	GCC	GGT
227▶	Trp	Phe	Ile	Ala	Asp	Lys	Gly	Gly	Ala	Gly
5246	GAG	CGT	GGG	TCT	CGC	GGT	ATC	ATT	GCA	GCA
237▶	Glu	Arg	Gly	Ser	Arg	Gly	Ile	Ile	Ala	Ala
5276	CTG	GGG	GCA	GAT	GGT	AGC	CCC	TCC	CGT	ATC
247▶	Leu	Gly	Pro	Asp	Gly	Lys	Pro	Ser	Arg	Ile
5306	GTA	GTT	ATC	TAC	ACG	AGC	GGG	AGT	CAG	GCA
257▶	Val	Val	Ile	Tyr	Thr	Thr	Gly	Ser	Gln	Ala

Figure 13b (続き)

5336 ACT ATG GAT GAA CGA AAT AGA CAG ATC GCT
 267 Thr Met Asp Glu Arg Asn Arg Gln Ile Ala
 5366 GAG ATA GGT GCC TCA CTG AIT AAG CAT TGG
 277 Glu Ile Gly Ala Ser Leu Ile Lys His Trp
 5396 TAA CTGTACAGACCAAGTTTACTCATATATACCTTAGAT
 287 ...
 5434 TGATTACGCGCCCTGTAGCGGCGCATTAAGCGCGCGG
 5473 GTGTGGTGGTTACGCGAGCGTGACCGCTACACTTGGCA
 5512 GCGCCCTAGCGCCCGCTCTTTCGCTTTCTTCCCTTCCT
 5551 TTCTCGCCACGTTTCGCGGCTTCCCGTCAAGCTCTAA
 5590 ATCGGGGGCTCCCTTTAGGGTTCGATTTAGTGCTTTAC
 5629 GGCACCTCGACCCCAAAAACCTTGATTTGGGTGATGGTT
 5668 CACGTAGTGGCCATCGCCCTGATAGACGGTTTTCGCC
 5707 CTTTGACCTTGGAGTCCACGTTCTTTAATAGTGGACTCT
 5746 TGTTCACAACTTGAACAACACTCAACCTATCTCGGGCT
 5785 ATTCTTTTGGATTATTAAGGGATTTTGCCGATTTTCGGCT
 5824 ATTGGTTAAAAAATGAGCTGATTTAACAAAATTTAACG
 5863 CGAATTTTAACAAAATATTACGTTTACAATTTAAAGG
 5902 ATCTAGGTGAAGATCCCTTTTGATATCTCATGACCAA
 5941 ATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGAC
 5980 CCCGTAGAAAAGATCAAGGATCTTCTTGAGATCCCTTT
 6019 TTTCTGCGCGTAATCTGCTGCTTGCAACAAAAAACCA
 6058 CCGCTACCAGCGGTGGTTTGTGTCGCGATCAAGAGCTA
 6097 CCAACTCTTTTTCGAAGGTAAGTGGCTTCAGCAGAGCG

Figure 13b (続き)

6136 CAGATACCAAATCTGTCTCTTAGTGTAGCCGTAGTTA
 6175 GGCCACCACCTTCAAGAACTCTGTAGCACCGCTACATAC
 6214 CTCGCTCTGCTAATCTGTACCAGTGGCTGCTGCCAGT
 6253 GGCATAAGTCTGTCTTACCAGGTGGACTCAAGACGA
 6292 TAGTTACCGGATAAGGCGCAGCGGTGCGGCTGAACGGGG
 6331 GGTTCGTGCACACAGCCAGCTTGGAGCGAACGACCTAC
 6370 ACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGC
 6409 GCCACGCTTCCGAAAGGAGAAAGCGGACAGGTATCCG
 6448 GTAAGCGCGAGGGTCGGAACAGGAGAGCGCACGAGGAG
 6487 CTTCCAGGGGAAACGCTTGGTATCTTTATAGTCTCTGTC
 6526 GGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGA
 6565 TGCTCGTCAGGGGGCGGAGCCTATGGAACACGCCAGC
 6604 AACGCGGCTTTTTCAGGTTCTTGGCTTTTGTGCGCT
 6643 TTTGCTCACAIGTCTTTCCTGCGTTATCCCTGATTTCT
 6682 GTGGATAACCGTATTACCGCTTTTGAGTGAGCTGATACC
 6721 GCTCGCCGCGAGCGAAGCAGCGAGCGCAGCGAGTCACTG
 6760 AGCGAGGAAGCGGAGAGCGCCTGATGCGGTATTTCTC
 6799 CTTACGCATCTGTGCGGTATTTACACCGCATAGGGTCA
 6838 TGGCTGCGCCCGACACCCGCCAACCCGCTGACGCGC
 6877 CCTGACGGGCTTGTCTGCTCCGCGATCCGCTTACAGAC
 6916 AAGCTGTGACCGTCTCCGGGAGCTGATGTGTGAGAGT
 6955 TTTACCGCTCATCACCGAAACGCGCGAGGCAGCAAGGAG

Figure 13b (続き)

6994 ATGGCGCCCAACAGTCCCCCGGCCACGGGCGCTGCCACC
 7033 ATACCCACGCGGAACAAGCGCTCATGAGCCGAAAGTGG
 7072 CGAGCCCGATCTTCCCATCGGTGATGTGCGGATATAG
 7111 GCGCCAGCAACCGCACCTGTGGCGCGGTGATGCCGGCC
 7150 ACGATGCGTCCGGCTAGAGGATCTGCTCATGTTTGACA
 7189 GCTTATC

【図 14】

Figure 14 a

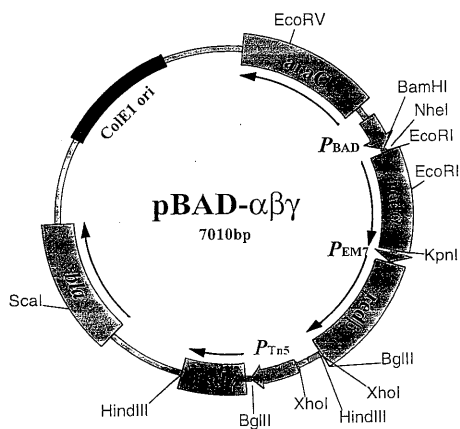


Figure 14b

NsiI

1 ATCGATGCATAATGTGCCTGTCAAATGGACGAAGCAGGG
 40 ATTCTGCAAAACCTATGCTACTCCGTCAAGCCGTCATTT
 79 GTCTGATTCGTTACCAA TTA TGA CAA CTT GAC
 293 Ser Leu Lys Val
 111 GGC TAC ATC ATT CAC TTT TTC TTC ACA ACC
 288 Ala Val Asp Asn Val Lys Glu Glu Cys Gly
 141 GGC ACG GAA CTC GCT CGG GCT GGC CCC GGT
 278 Ala Arg Phe Glu Ser Pro Ser Ala Gly Thr
 171 GCA TTT TTT AAA TAC CCG CGA GAA ATA GAG
 268 Cys Lys Lys Phe Val Arg Ser Phe Tyr Leu
 201 TTG ATC GTC AAA ACC AAC ATT GCG ACC GAC
 258 Gln Asp Asp Phe Gly Val Asn Arg Gly Val
 231 GGT GGC GAT AGG CAT CCG GGT GGT GCT CAA
 248 Thr Ala Ile Pro Met Arg Thr Thr Ser Leu
 261 AAG CAG CTT CGC CTG GCT GAT ACG TTG GTC
 238 Leu Leu Lys Ala Gln Ser Ile Arg Gln Asp
 291 CTC GCG CCA GCT TAA GAC GCT AAT CCC TAA
 228 Glu Arg Trp Ser Leu Val Ser Ile Gly Leu
 321 CTG CTG GCG GAA AAG ATG TGA CAG ACG CGA
 218 Gln Gln Arg Phe Leu His Ser Leu Arg Ser
 351 CGG CGA CAA GCA AAC ATG CTG TGC GAC GCT
 208 Pro Ser Leu Cys Val His Gln Ala Val Ser
 EcoRV
 381 GGC GAT ATC AAA ATT GCT GTC TGC CAG GTG
 198 Ala Ile Asp Phe Asn Ser Asp Ala Leu His

Figure 14b (続き)

411 ATC GCT GAT GTA CTG ACA AGC CTC GCG TAC
 188 Asp Ser Ile Tyr Gln Cys Ala Glu Arg Val
 441 CCG ATT ATC CAT CCG TGG ATG GAG CGA CTC
 178 Arg Asn Asp Met Pro Pro His Leu Ser Glu
 471 GTT AAT CCG TTC CAT GCG CCG CAG TAA CAA
 168 Asn Ile Ala Glu Met Arg Arg Leu Leu Leu
 501 TTG CTC AAG CAG ATT TAT CGC CAG CAG CTC
 158 Gln Glu Leu Leu Asn Ile Ala Leu Leu Glu
 531 CGA ATA GCG CCC TTC CCC TTG CCC GCG GTT
 148 Ser Tyr Arg Gly Glu Gly Gln Gly Ala Asn
 561 AAT GAT TTG CCC AAA CAG GTC GCT GAA ATG
 138 Ile Ile Gln Gly Phe Leu Asp Ser Phe His
 591 CCG CTG GTG CCG TTC ATC CCG GCG AAA GAA
 128 Pro Gln His Ala Glu Asp Pro Arg Phe Phe
 621 CCC CGT ATT GGC AAA TAT TGA CCG CCA GTT
 118 Gly Thr Asn Ala Phe Ile Ser Pro Trp Asn
 651 AAG CCA TTC ATG CCA GTA GGC GCG CCG ACG
 108 Leu Trp Glu His Trp Tyr Ala Arg Pro Arg
 681 AAA GTA AAC CCA CTG GTG ATA CCA TTC GCG
 98 Phe Tyr Val Trp Gln His Tyr Trp Glu Arg
 711 AGC CTC CCG ATG ACG ACC GTA GTG ATG AAT
 88 Ala Glu Pro His Arg Gly Tyr His His Ile
 741 CTC TCC TGG CCG GAA CAG CAA AAT ATC ACC
 78 Glu Gly Pro Pro Phe Leu Leu Ile Asp Gly
 771 CCG TCG GCA AAC AAA TTC TCG TCC CTG ATT
 68 Pro Arg Cys Val Phe Glu Arg Gly Gln Asn

Figure 14b (続き)

4 Asp Ile Ile Leu Gln Arg Thr Gly Ile Asp
 1359 GTG AGA GCT GTC GAA CAG GGG GAT GAT GCG
 14 Val Arg Ala Val Glu Gln Gly Asp Asp Ala
 1389 TGG CAC AAA TTA CCG CTC GGC GTC ATC ACC
 24 Trp His Lys Leu Arg Leu Gly Val Ile Thr
 1419 GCT TCA GAA GTT CAC AAC GTG ATA GCA AAA
 34 Ala Ser Glu Val His Asn Val Ile Ala Lys
 1449 CCC CCG TCC GGA AAG AAG TGG CCT GAC ATG
 44 Pro Arg Ser Gly Lys Lys Trp Pro Asp Met
 1479 AAA ATG TCC TAC TTC CAC ACC CTG CTT GCT
 54 Lys Met Ser Tyr Phe His Thr Leu Leu Ala
 1509 GAG GTT TGC ACC GGT GTG GCT CCG GAA GTT
 64 Glu Val Cys Thr Gly Val Ala Pro Glu Val
 1539 AAC GCT AAA GCA CTG GCC TGG GGA AAA CAG
 74 Asn Ala Lys Ala Leu Ala Trp Gly Lys Gln
 1569 TAC GAG AAC GAC GCC AGA ACC CTG TTT GAA
 84 Tyr Glu Asn Asp Ala Arg Thr Leu Phe Glu
 1599 TTC ACT TCC GGC GTG AAT GTT ACT GAA TCC
 94 Phe Thr Ser Gly Val Asn Val Thr Glu Ser
 1629 CCG ATC ATC TAT CCG GAC GAA AGT ATG CGT
 104 Pro Ile Ile Tyr Arg Asp Glu Ser Met Arg
 1659 ACC GCC TGC TCT CCC GAT GGT TTA TGC AGT
 114 Thr Ala Cys Ser Pro Asp Gly Leu Cys Ser
 1689 GAC GGC AAC GGC CTT GAA CTG AAA TGC CCG
 124 Asp Gly Asn Gly Leu Glu Leu Lys Cys Pro

Figure 14b (続き)

801 TTT CAC CAC CCC CTG ACC GCG AAT GGT GAG
 58 Lys Val Val Gly Gln Gly Arg Ile Thr Leu
 831 ATT GAG AAT ATA ACC TTT CAT TCC CAG CCG
 48 Asn Leu Ile Tyr Gly Lys Met Gly Leu Pro
 861 TCG GTC GAT AAA AAA ATC GAG ATA ACC GTT
 38 Arg Asp Ile Phe Phe Asp Leu Tyr Gly Asn
 891 GGC CTC AAT CCG CGT TAA ACC CCG CAC CAG
 28 Ala Glu Ile Pro Thr Leu Gly Ala Val Leu
 921 ATG GGC ATT AAA CGA GTA TCC CCG CAG CAG
 18 His Ala Asn Phe Ser Tyr Gly Pro Leu Leu
 951 GGG ATC ATT TTG CCG TTC AGC CAT ACTTTC
 8 Pro Asp Asn Gln Ala Glu Ala Met
 982 ATACTCCCGCCATTCAGAGAAGAAACCAATTGTCCATAT
 1021 TGCATCAGACATTGCCGTCACTGCGTCTTTTACTGGCTC
 1060 TTCTCGCTAACCAACCGGTAAACCCCGCTTATTAAGC
 1099 ATTCTGTACAAAGCGGACCAAGCCATGACAAAAAGC
 1138 CGTAACAAAAGTGCTATATACCGGAGAAAAGTCCAC
 1177 ATTGATTATTTGCACGGCGTCACACTTTGCTATGCCATA
 BamHI
 1216 GCATTTTATCCATAAGATTAGCGGATCCTACCTGACGC
 1255 TTTTATCGCAACTCTCTACTGTCTCTCCATACCCGTTC
 NheI EcoRI
 1294 TTTTGGCTAGCAGGAGGAATTCACC ATG ACA CCG
 1 Met Thr Pro
 PstI
 1329 GAC ATT ATC CTG CAG CGT ACC GGG ATC GAT

Figure 14b (続き)

1719 TTT ACC TCC CCG GAT TTC ATG AAG TTC CCG
 134 Phe Thr Ser Arg Asp Phe Met Lys Phe Arg
 1749 CTC GGT GGT TTC GAG GCC ATA AAG TCA GCT
 144 Leu Gly Gly Phe Glu Ala Ile Lys Ser Ala
 1779 TAC ATG GCC CAG GTG CAG TAC AGC ATG TGG
 154 Tyr Met Ala Gln Val Gln Tyr Ser Met Trp
 1809 GTG ACG CGA AAA AAT GCC TGG TAC TTT GCC
 164 Val Thr Arg Lys Asn Ala Trp Tyr Phe Ala
 1839 AAC TAT GAC CCG CGT ATG AAG CGT GAA GCG
 174 Asn Tyr Asp Pro Arg Met Lys Arg Glu Gly
 1869 CTG CAT TAT GTC GTG ATT GAG CCG GAT GAA
 184 Leu His Tyr Val Val Ile Glu Arg Asp Glu
 1899 AAG TAC ATG GCG AGT TTT GAC GAG ATC GTG
 194 Lys Tyr Met Ala Ser Phe Asp Glu Ile Val
 1929 CCG GAG TTC ATC GAA AAA ATG GAC GAG GCA
 204 Pro Glu Phe Ile Glu Lys Met Asp Glu Ala
 1959 CTG GCT GAA ATT GGT TTT GTA TTT GGG GAG
 214 Leu Ala Glu Ile Gly Phe Val Phe Gly Glu
 KpnI
 1989 CAA TGG CGA TAGATCCGGTACCCGAGCACGTTGA
 224 Gln Trp Arg ...
 2025 CAATTAATCATCGGCATAGTATATCGGCATAGTATAATA
 2064 CGACAAGGTGAGGAACATAACC ATG AGT ACT GCA
 1 Met Ser Thr Ala
 2098 CTC GCA ACG CTG GCT GGG AAG CTG GCT GAA
 5 Leu Ala Thr Leu Ala Gly Lys Leu Ala Glu

Figure 14b (続き)

Sall
 2128 CGT GTC GGC ATG GAT TCT GTC GAC CCA CAG
 15▶Arg Val Gly Met Asp Ser Val Asp Pro Gln
 2158 GAA CTG ATC ACC ACT CTT CGC CAG ACG GCA
 25▶Glu Leu Ile Thr Thr Leu Arg Gln Thr Ala
 2188 TTT AAA GGT GAT GCC AGC GAT GCG CAG TTC
 35▶Phe Lys Gly Asp Ala Ser Asp Ala Gln Phe
 2218 ATC GCA TTA CTG ATC GTT GCC AAC CAG TAC
 45▶Ile Ala Leu Leu Ile Val Ala Asn Gln Tyr
 2248 GGC CTT AAT CCG TGG ACG AAA GAA ATT TAC
 55▶Gly Leu Asn Pro Trp Thr Lys Glu Ile Tyr
 2278 GCC TTT CCT GAT AAG CAG AAT GGC ATC GTT
 65▶Ala Phe Pro Asp Lys Gln Asn Gly Ile Val
 2308 CCG GTG GTG GGC GTT GAT GGC TGG TCC CGC
 75▶Pro Val Val Gly Val Asp Gly Trp Ser Arg
 2338 ATC ATC AAT GAA AAC CAG CAG TTT GAT GGC
 85▶Ile Ile Asn Glu Asn Gln Gln Phe Asp Gly
 2368 ATG GAC TTT GAG CAG GAC AAT GAA TCC TGT
 95▶Met Asp Phe Glu Gln Asp Asn Glu Ser Cys
 2398 ACA TGC CGG ATT TAC CGC AAG GAC CGT AAT
 105▶Thr Cys Arg Ile Tyr Arg Lys Asp Arg Asn
 2428 CAT CCG ATC TGC GTT ACC GAA TGG ATG GAT
 115▶His Pro Ile Cys Val Thr Glu Trp Met Asp
 2458 GAA TGC CGC CGC GAA CCA TTC AAA ACT CGC
 125▶Glu Cys Arg Arg Glu Pro Phe Lys Thr Arg
 2488 GAA GGC AGA GAA ATC ACG GGC CCG TGG CAG
 135▶Glu Gly Arg Glu Ile Thr Gly Pro Trp Gln

Figure 14b (続き)

2518 TCG CAT CCC AAA CGG ATG TTA CGT CAT AAA
 145▶Ser His Pro Lys Arg Met Leu Arg His Lys
 2548 GCC ATG ATT CAG TGT GCC CGT CTG GCC TTC
 155▶Ala Met Ile Gln Cys Ala Arg Leu Ala Phe
 2578 GGA TTT GCT GGT ATC TAT GAC AAG GAT GAA
 165▶Gly Phe Ala Gly Ile Tyr Asp Lys Asp Glu
 2608 GCC GAG CGC ATT GTC GAA AAT ACT GCA TAC
 175▶Ala Glu Arg Ile Val Glu Asn Thr Ala Tyr
 PstI
 2638 ACT GCA GAA CGT CAG CCG GAA CGC GAC ATC
 185▶Thr Ala Glu Arg Gln Pro Glu Arg Asp Ile
 2668 ACT CCG GTT AAC GAT GAA ACC ATG CAG GAG
 195▶Thr Pro Val Asn Asp Glu Thr Met Gln Glu
 2698 ATT AAC ACT CTG CTG ATC GCC CTG GAT AAA
 205▶Ile Asn Thr Leu Leu Ile Ala Leu Lys
 2728 ACA TGG GAT GAC GAC TTA TTG CCG CTC TGT
 215▶Thr Trp Asp Asp Asp Leu Leu Pro Leu Cys
 2758 TCC CAG ATA TTT CGC CGC GAC ATT CGT GCA
 225▶Ser Gln Ile Phe Arg Arg Asp Ile Arg Ala
 2788 TCG TCA GAA CTG ACA CAG GCC GAA GCA GTA
 235▶Ser Ser Glu Leu Thr Gln Ala Glu Ala Val
 2818 AAA GCT CTT GGA TTC CTG AAA CAG AAA GCC
 245▶Lys Ala Leu Gly Phe Leu Lys Gln Lys Ala
 BglII XhoI
 2848 GCA GAG CAG AAG GTG GCA GCA TAGATCTCGAG
 255▶Ala Glu Gln Lys Val Ala Ala ...

Figure 14b (続き)

HindIII
 2880 AAGCTTCCTGCTGAACATCAAGGCAAGAAAACATCTGT
 2919 TGTCAAGACAGCATCTCTGAACAAGGACAAATTAACAGT
 2958 TAACAATAAAAAACGCAAAAGAAAATGCCGATATCTAT
 2997 TGGCATTTTCTTTTATTTCTTATCAACATAAAGGTGAAT
 XhoI
 3036 CCCATACCTCGAGCTTCAACGCTGCCGCAAGCACTCAGGG
 3075 CGCAAGGGCTGCTAAAAGGAAGCGGAACACGTAGAAAGC
 3114 CAGTCCGCAGAAACGGTGCTGACCCCGGATGAATGTGAG
 3153 CTACTGGGCTATCTGGACAAGGGAAACGCAAGCGCAAA
 3192 GAGAAAGCAGGTAGCTTGCAGTGGGCTTACATGGCGATA
 3231 GCTAGACTGGGCGGTTTATGGACAGCAAGCAACCGGA
 PvuII
 3270 ATTGCCAGCTGGGGGCCCTCTGGTAAGGTGGGAAGCC
 3309 CTGCAAGATAAAGTGGATGGCTTTCTTCCGCCAAGGAT
 BglII
 3348 CTGATGGCGCAGGGATCAAGATCTGATCAAGAGACAGG
 3387 ATGAGGATCGTTTCGC ATG GAT ATT AAT ACT
 1▶Met Asp Ile Asn Thr
 3418 GAA ACT GAG ATC AAG CAA AAG CAT TCA CTA
 6▶Glu Thr Glu Ile Lys Gln Lys His Ser Leu
 3448 ACC CCC TTT CCT GTT TTC CTA ATC AGC CCG
 16▶Thr Pro Phe Pro Val Phe Leu Ile Ser Pro
 3478 GCA TTT CGC GGG CGA TAT TTT CAC AGC TAT
 26▶Ala Phe Arg Gly Arg Tyr Phe His Ser Tyr
 3508 TTC AGG AGT TCA GCC ATG AAC GCT TAT TAC
 36▶Phe Arg Ser Ser Ala Met Asn Ala Tyr Tyr

Figure 14b (続き)

3538 ATT CAG GAT CGT CTT GAG GCT CAG AGC TGG
 46▶Ile Gln Asp Arg Leu Glu Ala Gln Ser Trp
 3568 GCG CGT CAC TAC CAG CAG CTC GCC CGT GAA
 56▶Ala Arg His Tyr Gln Gln Leu Ala Arg Glu
 3598 GAG AAA GAG GCA GAA CTG GCA GAC GAC ATG
 66▶Glu Lys Glu Ala Glu Leu Ala Asp Asp Met
 3628 GAA AAA GGC CTG CCC CAG CAC CTG TTT GAA
 76▶Glu Lys Gly Leu Pro Gln His Leu Phe Glu
 3658 TCG CTA TGC ATC GAT CAT TTG CAA CGC CAC
 86▶Ser Leu Cys Ile Asp His Leu Gln Arg His
 3688 GGG GCC AGC AAA AAA TCC ATT ACC CGT GCG
 96▶Gly Ala Ser Lys Lys Ser Ile Thr Arg Ala
 3718 TTT GAT GAC GAT GTT GAG TTT CAG GAG CGC
 106▶Phe Asp Asp Asp Val Glu Phe Gln Glu Arg
 3748 ATG GCA GAA CAC ATC CGG TAC ATG GTT GAA
 116▶Met Ala Glu His Ile Arg Tyr Met Val Glu
 3778 ACC ATT GCT CAC CAC CAG GTT GAT ATT GAT
 126▶Thr Ile Ala His His Gln Val Asp Ile Asp
 HindIII
 3808 TCA GAG GTA TAA AACGAGTAGA AGC TTG GCT
 136▶Ser Glu Val ...
 3839 GTT TTG GCG GAT GAG AGA AGA TTT TCA GCC
 3869 TGA TACAGATTAAATCAGAACGCAAGCGGTCTGATA
 3907 AAACAGAAATTGCGTGGCGGAGTAGCGGGTGGTCCCA
 3946 COTGACCCCATGCCGAACCTCAGAAAGTGAACCCGTAGC
 3985 GCCGATGGTAGTGTGGGCTCTCCCCATGCGAGTAGGG

Figure 14b (続き)

Figure 14b (続)

4578	GCG	GCG	GTA	TTA	TCC	CGT	GTT	GAC	GCC	GGG
76	Gly	Ala	Val	Leu	Ser	Arg	Val	Asp	Ala	Gly
4608	CAA	GAG	CAA	CTC	GGT	CGC	CGC	ATA	CAC	TAT
86	Gly	Ala	Gln	Leu	Gly	Arg	Arg	Ile	His	Tyr
Scale										
4638	TCT	CAG	AAT	GAC	TTG	GTT	GAG	TAC	TCA	CCA
96	Ser	Gln	Asn	Asp	Leu	Val	Glu	Tyr	Ser	Pro
4668	GTC	ACA	GAA	AAG	CAT	CTT	ACG	GAT	GGC	ATG
106	Val	Thr	Gly	Lys	His	Leu	Thr	Asp	Gly	Met
4698	ACA	GTA	AGA	GAA	TTA	TGC	AGT	GCT	GCC	ATA
116	Thr	Val	Arg	Glu	Leu	Cys	Ser	Ala	Ala	Ile
4728	ACC	ATT	AGT	GAT	AAC	ACT	GCG	GCC	AAC	TTA
126	Thr	Met	Ser	Asp	Asn	Thr	Ala	Ala	Asn	Leu
4758	CTT	CTG	ACA	ACG	ATC	GGA	GGA	CCG	AAG	GAG
136	Leu	Leu	Thr	Thr	Ile	Gly	Gly	Pro	Lys	Glu
4788	CTA	ACC	GTT	TTT	TTG	CAC	AAC	ATG	GGG	GAT
146	Thr	Thr	Ala	Phe	Leu	His	Asn	Met	Gly	Asp
4818	CAT	GTA	ACT	CGC	CTT	GAT	CGT	TGG	GAA	CCG
156	His	Val	Thr	Arg	Leu	Asp	Arg	Trp	Glu	Pro
4848	GAG	CTG	AAT	GAA	GCC	ATA	CCA	AAC	GAC	GAG
166	Glu	Leu	Asn	Glu	Ala	Ile	Pro	Asn	Asp	Glu
4878	CGT	GAC	ACC	ACG	ATG	CCT	GTA	GCA	ATG	GCA
176	Arg	Asp	Thr	Thr	Met	Pro	Val	Ala	Met	Ala
4908	ACA	ACG	TTG	CGC	AAA	CTA	TTA	ACT	GGC	GAA
186	Thr	Thr	Leu	Arg	Gly	Leu	Leu	Thr	Gly	Glu
4938	CTA	CYT	ACT	CTA	GCT	TCC	CGG	CAA	CAN	TTA
196	Leu	Leu	Thr	Leu	Ala	Ser	Arg	Gln	Gln	Leu

Figure 14b (続き)

5518 GCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGAA
5557 CTCTTGTGTCCAAACTTGAACAACACTCAACCCCTATCTCG
5596 GGCATATCTCTTTTGATTTATAGAAGGATTTTGGCGATTTCG
5635 GCCTATTGSGTTAAAAAATGAGCTGATTTAAACAAAATTTC
5674 AACCGGAATTTTAAACAAATATTAACGCTTTACAATTTAA
5713 AAGGATCTAGTGGAGAAATCCCTTTTGATGATCTCATGAC
5752 CAAAATCCCTTAAACGTGAGTTTTCGTGTCCACTGAGCGTC
5791 AGACCCCGTAGAAAAGATCAAAGGATCTCTCTGAGATCC
5830 TTTTTCCTTGGCGGTAATCTGCTGCTTGCAACAACAAAAA
5869 ACCACCGCTACCCGCGGTGGTTTGTGTGGCCGCAAAAAA
5908 GCTACCAACTCTTTTCCGAAGTAACTGGCTTCAGAG
5947 AGCGCAGATACCAAAATACCTGCTCTCTAGTGTAGCCGTA
5986 GTTAGGCCACCACCTTCAAGAACTCTGTAGACCCGCCATC
6025 ATACCTCGCTCTGCTAATCTGTTACCAAGTGGCTGCTCG
6064 CAGTGGCGGATACCGCTGTGCTTACCGGGTGGGCTCAAG
6103 ACGTAGGTTAAGGATAGGCGAGCGGTGCGGCTGAAC
6142 GGGGGTGTCTGTGCACACAGCCGACTTGGAGCGCAACGAC
6181 CTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGA
6220 AAGCGCAGCGTTCCTCCGAAGGAGAAAGGGCGGACAGGTA
6259 TCCGGTAAACGCGAGGCTGGCAACAGGAGAGCGCACGAC
6298 GGAGCTTCCAGGGGGAACCGCTGGTATCTTTATAGTCC
6337 TGTGGGTTTCGCGACCTCTGACTCTGAGCGCTGCGATTTT

Figure 14b (続き)

6376 GTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGC
6415 CAGCAACGCGGCCCTTTTACGGTTCCCTGGCCTTTTGCTG
6454 GCCTTTGTGCTCACATGTTCTTTCCCTGCGTTATCCCCTGA
6493 TTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGA
6532 TACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTC
6571 AGTGAGCGAGGAAGCGGAAGAGCGCCTGATCGGTATTT
6610 TCTCCTTACGCATCTGTGCGGTATTTACACCGCATAGG
6649 GTCATGSGCTGCGCCCCGACACCCGCCAACACCCGCTGAC
6688 GCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTAC
6727 AGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTGAG
6766 AGGTTTTCACCGTCATCACGAAACGCGGAGGCAGCAA
6805 GGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGC
6844 CACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAA
6883 GTGGCGAGCCCGATCTTCCCATCGGTGATGTGCGCGAT
6922 ATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCC
6961 GGCCACGATGCGTCCGGGTAGAGGATCTGCTCATGTTT
7000 GACAGCTTATC

フロントページの続き

(72)発明者 ツァング、ヨウミング

ドイツ連邦共和国 ディー - 6 9 1 1 7 ハイデルベルク、フリードリッヒ - エバート - アンラゲ
5 1 エー

(72)発明者 ブーフホルツ、フランク

ドイツ連邦共和国 ディー - 2 8 7 7 9 ブレメン、ニューエンキルシェナー ヴェク 4 4 アー

審査官 深草 亜子

(56)参考文献 Nucleic Acids Res. , 1 9 9 3 年, Vol.21, p.5192-5197

Mol. Microbiol., , , , 1 9 9 4 年, Vol.11, p.23-30

Genetics, 1 9 9 2 年, Vol.130, p.37-49

Proc.Natl.Acad.Sci.USA , 1 9 9 4 年, Vol.91, p.3205-3209

(58)調査した分野(Int.Cl., D B 名)

C12N 15/09

PubMed