(19) INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL PORTUGAL

(11) *Número de Publicação:* PT 92886 B

(51) *Classificação Internacional:* (Ed. 6)
C07K014/00 A C12N015/62 B
C12N015/70 B

(12) *FASCÍCULO DE PATENTE DE INVENÇÃO*

(22) *Data de depósito:* 1990.01.18

(30) *Prioridade:* 1989.01.21 DE 3901681

(43) *Data de publicação do pedido:* 1990.07.31

(45) *Data e BPI da concessão:* 08/95 1995.08.24

(73) *Titular(es):*
BEHRINGWERKE AKTIENGESELLSCHAFT
- D-3550 MARBURG/LAHN DE

(72) *Inventor(es):*
STEFAN KNAPP DE
EGON AMANN DE
KARL-JOSEF ABEL DE

(74) *Mandatário(s):*
JOÃO DE ARANTES E OLIVEIRA
RUA DO PATROCÍNIO 94 1350 LISBOA PT

(54) *Epígrafe:* PROCESSO PARA A PREPARAÇÃO DE UM PÉPTIDO DE SINAL E PARA A SECREÇÃO DE PÉPTIDOS EM ESCHERICHIA COLI POR MEIO DESSE PÉPTIDO DE SINAL

(57) *Resumo:*

[Fig.]

MEMÓRIA DESCRITIVA

DA

PATENTE DE INVENÇÃO

Nº 92886

NOME: BEHRINGWERKE AKTIENGESELLSCHAFT

EPÍGRAFE: "PROCESSO PARA A PREPARAÇÃO DE UM PÉPTIDO DE SINAL E PARA A SECREÇÃO DE PÉPTIDOS EM ESCHERICHIA COLI POR MEIO DESSE PÉPTIDO DE SINAL"

INVENTORES: Dr. Stefan Knapp, Dr. Egon Amann e Dr. Karl-Josef Abel, residentes na Alemanha Ocidental.

Reivindicação do direito de prioridade ao abrigo do artigo 4º da Convenção da União de Paris de 20 de Março de 1883. República Federal Alemã em 21 de Janeiro de 1989, sob o Nº. P 39 01 681.1.

Descrição referente à patente de invenção de BEHRINGWERKE AKTIENGESELLSCHAFT, alemã, industrial e comercial, com sede em D-3550 Marburg, Alemanha Ocidental, (inventores: Dr. Stefan Knapp, Dr. Egon Amann e Dr. Karl-Josef Abel, residentes na Alemanha Ocidental), para "PROCESSO PARA A PREPARAÇÃO DE UM PÉPTIDO DE SINAL E PARA A SECREÇÃO DE PÉPTIDOS EM ESCHERICHIA COLI POR MEIO DESSE PÉPTIDO DE SINAL".

DESCRIÇÃO

A invenção refere-se ao péptido de sinal de uma proteína de Bordetella pertussis, que é capaz de dirigir proteínas heterólogas no espaço periplasmático entre as membranas interna e externa de espécies bacterianas gran-negativas. A invenção refere-se ainda a sequências de DNA que codificam para este péptido de sinal, a plasmídeos que contêm uma tal estrutura genética, bem como a organismos hospedeiros com esses plasmídeos.

A invenção refere-se também a vectores plasmídeos, com os quais se pode determinar e comparar a eficiência de sequências de sinal conhecidas e desconhecidas. Em consequência destes estudos comparativos podem identificar-se e clonar-se sequências de sinal particularmente eficientes e uti-

PA

lizar-se em todas as três ordens de leitura de trnaslacção possíveis para a expressão de proteínas heterólogas.

Basicamente podem distinguir-se dois tipos diferentes de sequências de sinal: um tipo "hidrofóbico" e um tipo "hidrofílico". O grupo hidrofóbico de sequências de sinal abrange em regra cerca de 13-30 aminoácidos, enquanto que o grupo hidrofílico compreende, pelo contrário, cerca de 12-70 aminoácidos. A sequência de sinal do tipo hidrofóbico pode sub-dividir-se em três elementos estruturais. É constituída por um terminal-NH_2 relativamente hidrofílico com um ou dois aminoácidos básicos, por um bloco apolar na maior parte dos casos hidrofóbico, de sete ou oito aminoácidos e ainda por um terminal-COOH relativamente hidrofílico, que termina por um aminoácido de pequena cadeia lateral. Tais sequências de sinal "hidrofóbicas" conduzem as proteínas através da membrana do retículo endoplasmático (ER) e através das membranas das bactérias. Embora as sequências de sinal bacterianas e de ER pouco se distingam uma da outra, são funcionalmente permutáveis. O tipo hidrofílico distingue-se claramente na sua constituição do tipo hidrofóbico acima mencionado: não existem no tipo hidrofílico troços longos não interrompidos de aminoácidos hidrofóbicos, encontrando-se sim numerosos aminoácidos básicos e hidroxilados e poucos ou nenhuns aminoácidos acídicos. O tipo hidrofílico de sequências de sinal conduz as proteínas em mitocôndrias, cloroplastos e possívelmente também em peroxisonas. Este tipo não é significativo para a presente invenção.

Embora, como acima mencionado, os tipos hidrofóbicos das sequências de sinal de origem procarió-tica e ensariótica apresentem caracteristicas comuns e se possam permutar funcionalmente, podem contudo observar-se algumas diferenças: deste modo a maior parte das sequências de sinal procarióticas até agora conhecidas apresentam, em comparação com o tipo hidrofóbico (tipo ER) das sequências de sinal encarióticas, uma menor hidrofobicidade na região apolar, bem como, na maior parte dos casos, possuem um aminoácido básico adicional na região -NH_2. Possivelmente esta é a razão para que, em geral, a sequência de sinal natural de uma proteína heteróloga em microorganismos seja reconhecida e processada menos eficaz-

mente do que uma sequência de sinal bacteriana anteposta a esta proteína.

A secreção de uma proteína heteróloga processa-se em geral na E. coli como transporte através da membrana interna para o espaço periplasmático; conhecem-se apenas poucas excepções nas quais as proteínas heterólogas são segregadas para o meio envolvente. O transporte de uma proteína para o espaço periplasmático em E. coli corresponde em grande medida funcionalmente ao transporte de uma proteína para o lumen do retículo endoplasmático das células encarióticas. Como consequência deste processo podem dobrar-se correctamente as proteínas e formar-se pontes di-sulfureto intramoleculares, também correctamente, em E. coli. Por meio de peptidases de sinal especificos pode cindir-se de forma proteolítica a sequência de sinal e desse modo sintetizar-se a proteína heteróloga madura "processada", em E. coli.

Muitas proteínas são instáveis após expressão citoplasmática em bactérias, por exemplo Escherichia coli sendo rapidamente destruídas por proteases. Pode impedir-se esta destruição segregando estas proteínas rapidamente para o espaço periplasmático devido a uma sequência de sinal prévia muito eficiente. Nasceu assim a necessidade de isolar sequências de sinal particularmente eficientes e de conceber para tal processos adequados.

Hoffman e Wright (Proc. Acad. Natl. Sci. USA, (1985) 82, 5107-5111) descrevem plasmídeos que codificam para a fosfatase alcalina periplasmática de E. coli (Phoa EC 3.1.3.1.) sem sequências de sinal própria. Em fusões in vitro segrega-se então fosfastose alcalina activa na forma de uma proteína de fusão, por parceiros de fusão com sequência de sinal própria, enquanto que faltando uma sequência de sinal fundida para a fosfatose alcalina libertada para o citoplasma não se observa qualquer actividade. Manoil e Beckwith (Proc. Natl. Acad. Sci USA (1985) 82, 8129-8133) continuaram este trabalho, colocando na posição 3' antes de transposão Tn 5 (loc-cit) o cDNA codificante para a PhoA sem sequência de sinal e para os cinco aminoácidos seguintes, demonstrando assim que só as fusões com proteínas segregadas mas também com proteínas de mem-

brana podem conduzir a PhoA activa. A construção designada por "TnPhoA" é por isso adequada para a identificação de sequências de sinal ou de estruturas análogas a sequências de sinal.

S. Knapp e J. Mekalanos (J. Bacteriology (1988) 170, 5059-5066) prepararam então mutantes de Bordetella pertassis, por mutagénese com TnPhoA, mutantes esses influenciáveis por sinais de modulação (aqui ácido nicotínico e $MgSO_4$), sendo a maior parte desses mutantes reprimidos e alguns deles activados, o que leva a pelo menos dois genes reguladores "trans-acting".

Descobriu-se que no mutante designado por SK 6 está contida uma sequência de sinal muito eficiente.

Esta nova sequência de sinal pertence a uma proteína segregada por Bordetella pertassis e possui a sequência seguinte (vide Tab. 2 e 3).

M K K W F V A A G I G A G L L M L S S A A.

Encontram-se ainda descritos plasmídeos contendo PhoA que, por um lado, são bem adequados como "vectores de clonagem de sequência de sinal e, por outro lado, possibilitam uma comparação quantitativa entre várias sequências de sinal em relação à sua "eficiência de secreção". É particularmente útil para estes dois fins o vector pTrc99C-PhoA (Fig. 1, Tab. 1 e exemplo 1). Este vector foi composto a partir de pTrc99C (Amann et al. Gene 69(1988) 301-315) e de um DNA de PhoA modificado isento de sequência peptídica de sinal, de modo que o gene estrutural para a PhoA se encontre na ordem de leitura correcta em relação ao codão de início de translacção do ptrc99C e que se produzisse imediatamente a seguir ao extremo 5' do gene estrutural de PhoA (sem sequência de sinal) uma posição de corte NcoI.

A invenção refere-se sucessivamente:

a) à sequência de sinal

M K K W F V A A G I G A G L L M L S S A A

b) a plasmideos que contenham essa sequência

c) à sua utilização para a secreção de proteínas, bem como

d) a plasmídeos que sejam particularmente adequados para a clonagem e para a valorização quantitativa de sequências de sinal, nos quais, após um promotor fortemente regulável com trc, se segue a posição de ligação de ribossonas (RBS) LacZ bem como um codão de iniciação de translacção codificado por vector, numa distância em relação ao RBS optimizada para uma elevada expressão, e nos quais se encontra presente uma posição de corte NcoI imediatamente a seguir ao extremo 5' do gene estrutural da PhoA isento de sequência de sinal, mas em que se removeu essa posição no interior de PhoA por mutação, preferindo-se o pTrc99C-PhoA.

A invenção encontra-se ainda especificada nos exemplos e nas reivindicações da patente.

Exemplo 1

Identificação e isolamento da sequência de sinal de Bordetella Pentussis.

O transposão TnPhoA a seguir utilizado trata-se de um derivado de transposão Tn5. O TnPhoA contém no elemento de inserção IS50 enquanto um derivado isento de sequência de sinal do gene estrutural do PhoA de E. coli. Este foi construido por Manoil e Beckwith (vide acima) de modo a que, por transposição de TnPhoA num gene cromossonal ou codificado por plasmídeo, resulte apenas uma fusão genética PhoA positiva, quando o código de leitura do gene estrutural PhoA de E. coli de TnPhoA e a sequência de sinal do gene estrutural atingido pela transposição coincidem. Estas colónias PhoA positivas podem identificar-se facilmente por meio do indicador corante 5-bromo-4-cloroindoxil-fosfato-toluidina (XP). Através da técnica descrita efectuou-se uma mutagénese TnPhoA na estirpe nativa 18323 de Bordetella pentassis (Knapp e Mekalanos (1988), vide acima). Produziu-se assim, entre outros, o mutante SK6 de TnPhoA PhoA positivo, cuja fusão genética TnPhoA se designou por vrg-6. A fusão genética vrg6 clonou-se num fragmento BamHI de 20k6 no vector plasmídeo pBR322, da maneira seguinte. Cindiu-se o DNA genómico do mutante SK6 com BamHI, ligou-se com o DNA de pBR322 cortado com BamHI e transformou-se na estir



pe de E. coli CC118 (= PhoA negativa). Os clonos que contêm o fragmento genómico com a fusão genética TnPhoA, seleccionaram-se em placas de agar com canamicina/ampicilina (o TnPhoA codifica, como o Tn5, para um gene resistente à canamicina, que está localizado entre a parte 5' PhoA de TnPhoA e a posição de corte Bam HI interna singular de TnPhoA). Um fragmento BamHI genómico de um mutante TnPhoA, que apresenta uma resistência à canamicina, deve por isso conter também o gene estrutural PhoA e o DNA genómico de B. pertussis colocado no sentido ascendente da cadeia até à próxima posição de corte BamHI genómica. No caso do fragmento Bam HI de 20 kb, que contém a fusão genética vrg6, cerca de 14 kb correspondem ao DNA de B.pertassis e cerca de 6 kb ao DNA codificante de TnPhoA. Limitaram-se ainda mais as sequências de regulação transcripcionais e translaccionais da fusão genética vrg6. Para tal submeteu-se o fragmento BamHI de 2o kb a uma análise de restrição, e clonaram-se os subfragmentos que contêm a sequência PhoA total de TnPhoA, mas que contêm zonas de DNA de B.pertassis encuratadas em comparação com o fragmento de 20 kb, nos vectores pBR322 ou pUC18. Os derivados de delecção assim obtidos clonaram-se no plasmídeo pLAFR2 com capacidade de replicação em B. pentassis (Friedmann et al. (1982), Gene 18, 289-196) e, por meio de transferência conjugativa em B. pentassis, testaram-se em relação à actividade de phoA modulável. Identificou-se assim um fragmento PstI de cerca de 3,2 kb e subclonou-se em pUC18 (designado a seguir por pUC-PI) que contém agora apenas cerca de 500 pares de bases do DNA de B. pertassis no sentido da corrente da posição de inserção TnphoA da fusão genética vrg6, e que é positivo em B. pertassis após indução por PhoA. Uma vez que a actividade PhoA dos derivados de B. pertassis, que contêm o fragmento BamHI clonado de 20 kb ou o fragmento PstI de 3,2 kb, não se distinguem efectivamente pela sua acetividade de posfatase, as sequências de regulação trnascripcionais e trnaslaccionais da fusão genética vrg6 devem estar completamente contidas no último fragmento. Partindo do pUC-PI introduziram-se deleccções na região de DNA colocada na posição de inserção TnphoA de 500 pares de bases, no sentido da corrente, por meio dos enzimas exomuclease III e 51 nuclease, de acordo com o método de Henikoff ((1984) Gene,

28, 351-359). Obtiveram-se assim, entre outros, os dois derivados de pUC-PI, vrg96-delta 12 e vrg6-delta 11. O vrg6-delta 12 contém ainda cerca de 200 pares do DNA específico de B. pertassis no sentido da corrente da posição de inversão TnphoA e é também PhoA-positivo. Por meio de sequenciação de DNA determinou-se a sequência de sinal de B. pertassis nestes dois plasmídeos recombinantes.

A sequência de sinal é a seguinte (vide também Tab. 2)

M K K W F V A A G I G A G L L M L S S A A

A sequência de sinal de B. pertassis caracterizada deste modo abrange 21 aminoácidos e preparou-se em seguida, como descrito no exemplo 3, clonou-se e verificou-se ser adequada para a secreção de proteínas heterólogas.

O vrg6-delta 11 contém simplesmente quatro nucleótidos específicos de B. pertassis, no sentido da corrente da posição de inserção TnphoA, seguidos por uma posição de corte SacI especificada por pUC18 (Tab. 1). Por meio de cisão com PstI/SacI do DNA de vrg6-delta 11, obtém-se o gene estrutural PhoA completo, isento de sequência de sinal, do TnPhoA, na forma de um fragmento de cerca de 2,6 kb, o qual serve como fonte para o gene estrutural phoA, isento de sequência de sinal, no exemplo 2.

Exemplo 2

Construção de um vector plasmídeo (pTrc 99C-phoA) para a clonagem e medição de eficiência comparativa de sequências de sinal

Descreve-se em seguida a construção do vector plasmídeo pTrc99C-phoA. Este vector plasmídeo contém como elemento efectivo o gene estrutural phoA isento de sequência de sinal, já acima descrito, isolado a partir de TnphoA. O gene estrutural phoA contém uma posição de corte NcoI interna. Eliminou-se esta posição de corte através do método de mutagénese dirigida, mantendo a sequência de aminoácidos.

Para tal, cindiu-se em primeiro lugar o plasmídeo recombinante PhoA-negativo pvrg6-delta 11 (vide

exemplo 1) com EcoRI e isolou-se o fragmento de 330 pares de bases da região interna do gene estrutural phoA. Este fragmento que contem a posição de corte NcoI a matagenizar, ligou-se na posição EcoRI do vector de mutagénese pMa5-8 (Fig. 3). Isolou-se o plasmídeo resultante pMa5-8-EcoRI330 e utilizou-se para preparar uma cadeia simples. A cadeia simples assim obtida, com o fragmento EcoRI clonado, isolou-se então de acordo com métodos conhecidos e submeteu-se ao protocolo de mutagénese "gapped-duplex" publicado (Kramer et al. (1984) Nucl. Acids Res. 12, 9441-9456), tendo-se utilizado o seguinte oligodesoxinuleótido.

5' ATCGATATTGCCGTGGTACGTTGCTTTC 3'

Identificou-se um plasmídeo, que apresenta a mutação NcoI desejada, por meio de análise de restrição adequada, sequenciou-se a região relevante e confirmou-se ser a correcta. A partir deste plasmídeo reisolou-se em seguida o fragmento EcoRI de 330 pares de bases e utilizou-se no lugar do fragmento correspondeu-se do plasmídeo purg-6-delta 11. Para tal, digeriu-se parcialmente o purg-6-delta 11 com EcoRI, e comparado com o plasmídeo de partida pvrg-delta 11 (cerca de 6700 bp) linearizado por digestão parcial com EcoRI, isolou-se um fragmento menor com menos cerca de 330 pares de bases. O fragmento EcoRI com este tamanho (cerca de 6400 bp) tratou-se com fosfatase alcalina e ligou-se com o fragmento EcoRI de 330 pares de bases mutagenizado e transformou-se a mistura de ligação em E. coli. Os plasmídeos recombinantes, que contêm um gene estrutural phoA restaurado com o fragmento EcoRI de 330 pares de bases correctamente inserido, foram identificados por meio de análise de restrição e sequenciação do DNA. Multiplicou-se um tal plasmídeo recombinante, pvrg6-delta 11-delta NcoI, e utilizou-se para a construção do plasmídeo híbrido pTrc 99c-phoA. Para isso, isolou-se um fragmento SacI-ScaI com cerca de 2600 pares de bases a partir do prvg6-deta 11-delta NCoI. No passo seguinte, substituiu-se o fragmento SacI-ScaI com cerca de 900 pares de bases de pTrc 99C (Amann et al. (1988), Gene 69, 301-315) por este fragmento SacI-ScaI de cerca de 2600 pares de bases. O plasmídeo recombinante resultante pTrc99C-phoA contém, em consequência das manipulações anteriores, uma posição de corte NcoI

singular imediatamente junto do extremo 5' do gene estrutural phoA isento de sequência de sinal e pode, como se mostra no exemplo seguinte, utilizar-se para a clonagem de sequências de sinal naturais ou sintéticas pretendidas. O pTrc99C-phoA contém o gene estrutural de phoA na ordem de leitura correcta em relação ao codão de iniciação de translacção do vector de expressão pTrc99c, mas não é capaz, devido à ausência da sequência de sinal phoA, de promover a síntese de uma fosfatase alcalina enzimaticamente activa, em células de Escherichia coli transformadas, sendo por isso apropriado como "vector de clonagem de sequência de sinal". Além disso o pTrc99C-PhoA contém, no sentido contrário da corrente do promotor trc híbrido (Amann e Brosins (1985) Gene 40, 183-190) a posição de ligação ribossomal lacZ (RBS) bem como um codão de iniciação de translacção numa distância em relação a lacZ RBS optimizada para uma elevada expressão. As células de E. coli, que contêm o plasmídeo recombinante pTrc99C-phoA não têm qualquer actividade de fosfatase alcalina biologicamente activa codificada pelo plasmídeo, uma vez que ao gene estrutural phoA deste plasmídeo falta a sequência de sinal. Podem então preparar-se colónias phoA-positivas colocando um fragmento de DNA codificante para uma sequência de sinal antes do gene estrutural phoA, na ordem de leitura correcta. Pode acontecer deste modo que o pTrc 99C-phoA seja cortado com NcoI e que se inseriram neste DNA vector fragmentos de DNA sintéticos que codificam para sequências de sinal. As colónias de bactérias, que contêm plasmídeos híbridos com esta manipulação, podem agora identificar-se facilmente pelos seus novos fenotipos PhoA-positivos através do indicador corante XP já caima descrito. O principio apresentado será adiante melhor esclarecido na forma dos exemplos de execução. Clonando-se sequências de sinal de vários proteínas de secreção no vector pTrc 99C-phoA, obtêm-se plasmídeos recombinantes isógenos, que se distinguem apenas pela sequência de sinal. Nesta base, a actividade de phoA das células de E. coli que contêm uma tal construção, da uma medida da eficiência das respectivas sequências de sinal clonadas.

Uma outra possibilidade de utilização do vector pTrc99C-phoA consiste em clonar fragmentos de DNA

sintéticos que não codificam para uma sequência de sinal claramente definida, mas que são degenerados de modo que para cada posição da sequência de sinal é possível um grande número de aminoácidos. Trata-se neste caso de certo modo de uma clonagem "shotgun", em que a actividade de phoA determinada pelo vector representa uma medida da eficiência da sequência de sinal artificial. Com este método podem preparar-se e valorizar-se novas sequências de sinal a utilizar para a expressão heteróloga de genes clonadas.

O princípio de construção do pTrc99c--phoA está ilustrado na Fig. 1. As abreviaturas têm os seguintes significados: N = NcoI, S = SacI, P = PstI, /N/ = posição NcoI não regenerada após ligação, "phoA = gene estrutural phoA sem sequência de sinal; as setas indicam a orientação de transcripção ou a orientação $NH_2 \rightarrow COOH$ das regiões translaccionadas. Oligo significa uma sequência de oigonucleótido sintética. A Tab. 1 mostra as região relevante de inicação de clonagem e de trnaslacção pTrc99c-phoA.

Exemplo 3

Síntese do DNA e clonagem da sequência de sinal de Bordetella pertussis bem como de outras cinco sequências de sinal microbianas de origem natural de proteínas de secreção.

Por meio do vector pTrc99C-phoA clonaram-se seis sequências de sinal diferentes, cujas sequências de aminoácidos se apresentam na Tab. 2. Para além da nova sequência de sinal de Bordetella pertussis escolheram-se mais cinco sequências de sinal, de acordo com os critérios seguintes:

a) sequência de sinal de uma proteína periplasmática
   - fosfatase alcalina (phoA) de E. coli (kituchi et al. (1981) Nucleic Acids Res. 9, 5671-5678)

b) sequência de sinal de uma proteína de membrana externa
   - "outer membrane protein" (ompA) de E. coli (Hovva et al. (1980) J. Biol. Chem. 255, 27-29)

c) Sequências de sinal de três proteínas segregadas para o meio
   - "Heat stable toxin I" (STI) de E. coli (So e McCarthy

(1980), Proc. Natl. Acad. Sci. USA 77, 4011-4015)

- "Heat stable toxin II" (STII) de E. coli (Lee et al. (1983), Infect. Immun. 42, 264-268)
- amilase de Bacillus subtitis (Yang et al. (1983), Nucleic Acids Res. 11, 237-249).

Para s síntese e clonagem destas sequências de sinal utilizou-se a seguinte nomenclatura simplificada:

Sequência de final de Bordetella pertussis vrg-6 = Seq 1
" " " de phoA = Seq 2
" " " de OmpA = Seq 3
" " " de STI = Seq 4
" " " de STII = Seq 5
" " " de amilase de Bacillus subtitis = Seq 6

Prepararam-se todas as seis sequências de sinal mencionadas por meio da síntese de DNA. Clonaram-se os fragmentos de DNA para tal sintetizados (representados na Tab. 3), utilizando a selecção de fosfastase alcalina descrita no exemplo 2, no vector leste pTrc99C-phoA, e identificaram-se. Os fragmentos de DNA codificantes para sequências de sinal sintéticos foram concebidos de modo a que, após inserção na orientação correcta no vector pTrc99C-phoA, apenas se regenere uma posição NcoI, nomeadamente no sentido da corrente da região codificante para a sequência de sinal (vide também as Fig. 1, Tab 3 e Tab. 4). Deste modo, esta posição NcoI, como se indicará a seguir no exemplo 4, pode continuar a ser utilizada como posição de clonagem para a inserção de genes heterólogos nos vectores pSEC (pSEC = secreção).

A síntese dos doze fragmentos de DNA apresentados na Tab. 3 efectuou-se de acordo com métodos conhecidos (Sinha et al. (1984) Nucl. Acids Res. 12, 4539-4557) utilizando β-cianoetilamiditeno. As sínteses efectuaram-se num sintetizador da firma Biosearch segundo o método de fosfite-triéster (Lectringer (1975) J. Arner. Chem. Soc. 97, 3278; Letringer (1976) J. Amer. Chem. Soc. 98, 3655). Após a cisão do suporte (CPG) com amoníaco concentrado à temperatura ambiente, durante 5-8 horas, e a remoção dos grupos protectores nas bases da mesma solução a 55°C, durante cerca de 12 horas, purificaram

-se os oligodesoximucleótidos por electroforesse de gel ou por HPLC de fase revessa. Tomaram-se os oligodesoxinucleótidos em tampão de "anncaling" (100 mMNacl, 10 mM Tris-cl (pH 7,8) 0,1 mM EDTA), adicionaram-se-lhes quantidades molares das duas cadeias, incubou-se durante 5 minutos a 95°C e deixou-se arrefecer lentamente até à temperatura ambiente. Os fragmentos de DNA de cadeia dupla apresentam nos extremos 5' regiões de cadeia simples com quatro bases de comprimento complementares a uma posição de reconhecimento NcoI. Linearizou-se o vector teste pTrc 99C-phoA com NcoI e ligou-se em várias misturas reaccionais com fragmentos de DNA hibridizados. Transformaram-se células adequadas de E. coli, de acordo com métodos conhecidos, com as misturas reaccionais de ligação, colocaram-se em placas de agar LB/Amp e incubaram-se a 37° C durante a noite. Inocolaram-se as colónias por meio do método das réplicas em placas de indicador LB/Amp/XP/IPTG e incubaram-se novamente a 37° C. As colónias phoA-positivas aparecem coradas de azul nesta placa de indicador. Isolou-se o DNA plasmídeo destas colónias e sequenciou-se, tendo-se conseguido confirmar a orientação correcta dos fragmentos de DNA sintéticas bem como a sequência de sinal correcto eperada por os seis exemplos acima mencionados. Os plasmídeos assim obtidos, com as respectivas sequências de sinal confirmadas como correctas por sequenciação, designaram-se de acordo com a Tabela anterior como Seql-, -2, -3, -4, -5 e -6 de pTrc99C-phoA, respectivamente. Em condições padrão podem agora, por meio de extinção (medição do corante libertado), comparar-se e valorizar-se estas sequências de sinal, pertencendo as descobertas de B. pertussis às relativamente mais fortes.

Exemplo 4

Construção dos vectores de secreção pSEC-Bp-1, pSEC-Bp-2 e pSEC-Bp-3

Digeriu-se o DNA plasmídeo do clono pTrc99C-phoA-Seql com SacI e ScaI e isolou-se o fragmento com cerca de 3,1 kb. Este fragmento contém para além da sequência de sinal de B. pertussis exclusivamente sequências específicas pTrc 99C (vide também Fig. 1). Ligou-se este fragmento, em

três misturas reaccionais separadas, contendo cada uma um fragmento SaCI/ScaI de cerca de 0,9 kb dos plasmídeos pTrc97A, pTrc97B e pTrc97C (Amann et al; vide acima) e designaram-se os plasmídeos resultantes como pSEC-Bp-1, pSEC-Bp-2 e pSEC-Bp-3. Por meio desta manipulação utilizou-se a longa região de polilincagem dos plasmídeos pTrc97A, pTrc97B e pTrc97C, para introduzir em todas as três códigos de leitura um grande número de posições de restrição no sentido da corrente da região codificante da sequência de sinal de Bordetella pertussis (Tab. 4). De modo análogo a estas construções podem preparar-se vectores de secreção semelhantes para a expressão e secreção de proteínas heterólogas utilizando os plasmídeos pTrc99C-phoA-Seq2, -3, -4, -5 e -6. Os vectores de secreção assim preparados distinguem-se pela sua eficiência relativa bem como na localização celular dos produtos expressos, conforme a origem da sequência de sinal em causa utilizada. A Fig. 2 mostra, como exemplo, a estrutura de plasmídeo do pSEC-BP1 e a Tab. 5 mostra a sequência de DNA total do pSEC-BP1, representando XXX um codão de início ou de fim.

Legenda da Fig. 1:

Carta dos plasmídeos pMAC5-8(=pMA5-8 e pMC5-8).
F1-ORI: Origem de replicação do fogo f1;
ORI: Origem de replicação do ColE1-Typ;
CAT: região codificante para a cloranfenicol-acetil-transferase
AMP: região codificante para a B-lactamase.
OpMA5-8 tem uma mutação em CAT (A na posição 3409) e opMC5-8 uma mutação em AMP (C na posição 2238).

pTrc99C-phoA

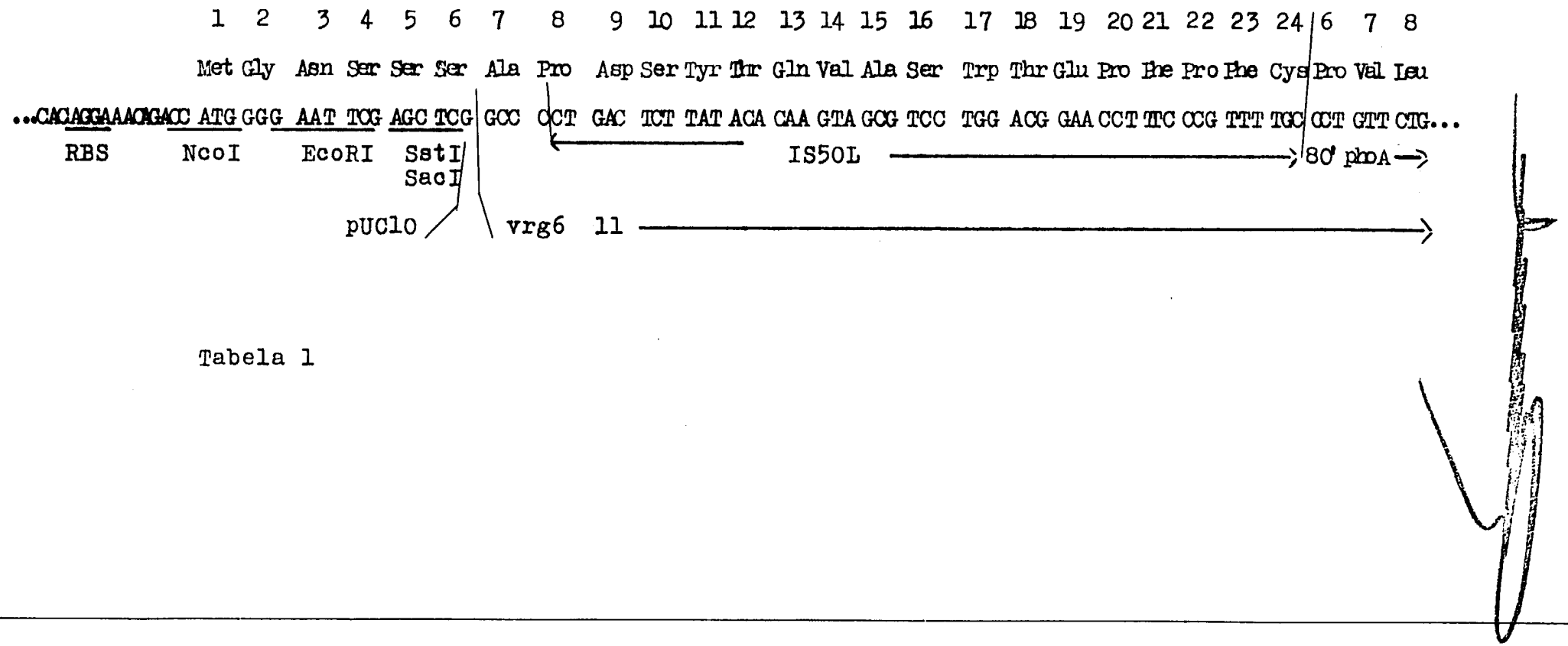


Tabela 1

-30	-20	-10	-1	Aminoácidos	
		M K K W F V A A G I G A G L L M L S S A A		B.p.	21
		M K Q S T I A L A L L P L L F T P V T K A		E.c. phoA	21
		M K K T A I A I A V A L A G F A T V A Q A		E.c.ompA	21
		M K K L M L A I F F S V L S F P S F S Q S		E.c. ST II	21
	M K K N I A F L L A S M F V F S I A T N A Y A			E.C. ST II	23
M F A K R F K T S L L P L F A G F L L L F H L V L A G P A A A S				B.S. amilase	32

Tab. 2

Tab. 3

Sequência de sinal de Bordetella pertussis

5′ C ATG AAA AAG TGG TTC GTT GCT GCC GGC ATC GGC GCT GCC GGA CTC ATG CTC TCC AGC GCC GC
TTT TTC ACC AAG CAA CGA CGG CCG TAG CCG CGA CGG CCT GAG TAC GAG AGG TCG CGG CGG TAC 5′

Sequência de sinal phoA de E. coli

5′ C ATG AAA CAA AGC ACT ATT GCA CTG GCA CTC TTA CCG TTA CTG TTT ACC CCT GTG ACA AAA GC
TTT GTT TCG TGA TAA CGT GAC CGT GAG AAT GGC AAT GAC AAA TGG GGA CAC TGT TTT CGG TAC 5′

Sequência de sinal OmpA de E. coli

5′ C ATG AAA AAG ACA GCT ATC GCG ATT GCA GTG GCA CTG GCT GGT TTC GCT ACC GTA GCG CAG GC
TTT TTC TGT CGA TAG CGC TAA CGT CAC CGT GAC CGA CCA AAG CGA TGG CAT CGC GTC CGG TAC 5′

Sequência de sinal da "heat-stable" toxinaIde E. coli

5′ C ATG AAA AAG CTA ATG TTG GCA ATT TTT ATT TCT GTA TTA TCT TTC CCC TCT TTT AGT CAG TCA CC
TTT TTC GAT TAC AAC CGT TAA AAA TAA AGA CAT AAT AGA AAG GGG AGA AAA TCA GTC AGT GGG TAC 5′

Sequência de sinal de "heat stable" toxina II de E. coli

5′ C ATG AAA AAG AAT ATC GCA TTT CTT CTT GCA TCT ATG TTC GTT TTT TCT ATT GCT ACA AAT GCC TAT GC
TTT TTC TTA TAG CGT AAA GCC GAA CGT AGA TAC AAG CAA AAA AGA TAA CGA TGT TTA CGG ATA CGG TAC 5′

Sequência de sinal de amilase de Bacillus subtilis

5′ C ATG TTT GCA AAA CGA TTC AAA ACC TCT TTA CTG CCG TTA TTC GCT GGA TTT TTA TTG CTG TTT CAT TTG GTT
AAA CGT TTT GCT AAG TTT TGG AGA AAT GAC GGC AAT AAG CGA CCT AAA AAT AAC GAC AAA GTA AAC CAA

CTG GCA GGA CCG GCG GCT GCG AGT CC
GAC CGT CCT GGC CGC CGA CGC TCA GGG TAC 5′

Tab. 4

péptido de sinal B. p.

pSEC-Bp1

17	18	19	20	21	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Leu	Ser	Ser	Ala	Ala	Met	Glu	Phe	Glu	Leu	Gly	Thr	Arg	Gly	Ser	Ser	Arg	Val	Asp	Leu	Gln	Ala	Cys	Lys	Leu	rrnB
CTC	TCC	AGC	GCC	GCC	ATG	GAA	TTC	GAG	CTC	GGT	ACC	CGG	GGA	TCC	TCT	AGA	GTC	GAC	CTG	CAG	GCA	TGC	AAG	CTT	GGCTG
					NcoI		EcoRI		SstI		KpnI	SmaI XmaI		BamHI		XbaI		SalI AccI HincII		PstI		SphI		HindIII	

péptido de sinal B. p.

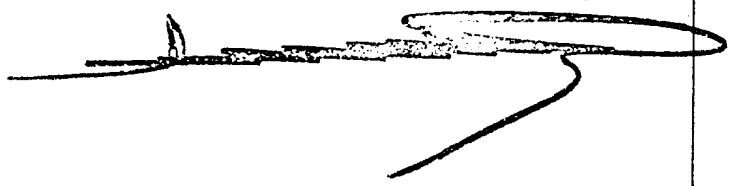
pSEC-Bp2

17	18	19	20	21	1	2	3	4	5	6	7	8	9	10	11										
Leu	Ser	Ser	Ala	Ala	Met	Gly	Ile	Arg	Ala	Arg	Tyr	Pro	Gly	Ile	Leu										rrnB
CTC	TCC	AGC	GCC	GCC	ATG	GGA	ATT	CGA	GCT	CGG	TAC	CCG	GGG	ATC	CTC	TAG	AGTCGAC	CTGCAG	GCATGC	AAGCTT	GGCTG				
					NcoI		EcoRI		SstI		KpnI	SmaI XmaI		BamHI		XbaI		SalI AccI HincII		PstI		SphI		HindIII	

péptido de sinal B. p.

pSEC-Bp3

17	18	19	20	21	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
Leu	Ser	Ser	Ala	Ala	Met	Gly	Asn	Ser	Ser	Val	Val	Pro	Gly	Asp	Pro	Leu	Glu	Ser	Thr	Cys	Arg	His	Ala	Ser	Leu	rrnB
CTC	TCC	AGC	GCC	GCC	ATG	GGG	AAT	TCG	AGC	TCG	GTA	CCC	GGG	GAT	CCT	CTA	GAG	TCG	ACC	TGC	AGG	CAT	GCA	AGC	TTG	GCTG
					NcoI		EcoRI		SstI		KpnI	SmaI XmaI		BamHI		XbaI		SalI AccI HincII		PstI		SphI		HindIII		



```
   1  GTTTGACAGC TTATCATCGA CTGCACGGTG CACCAATGCT TCTGGCGTCA

  51  GGCAGCCATC GGAAGCTGTG GTATGGCTGT GCAGGTCGTA AATCACTGCA

 101  TAATTCGTGT CGCTCAAGGC GCACTCCCGT TCTGGATAAT GTTTTTTGCG
                                                      -35
 151  CCGACATCAT AACGGTTCTG GCAAATATTC TGAAATGAGC TGTTGACAAT
           trcP          -10
 201  TAATCATCCG GCTCGTATAA TGTGTGGAAT TGTGAGCGGA TAACAATTTC

                        M   K   K   W   F     V   A   A     G   I   G
 251  ACACAGGAAA CAGACCATGA AAAAGTGGTT CGTTGCTGCC GGCATCGGCG
                         ***

      A   G   L   L     M   L   S     S   A   A
 301  CTGCCGGACT CATGCTCTCC AGCGCCGCCA TGGAATTCGA GCTCGGTACC
                                       ----------------------
                                        NcoI  EcoRI  SstI  KpnI

 351  CGGGGATCCT CTAGAGTCGA CCTGCAGGCA TGCAAGCTTG GCTGTTTTGG
      ------------------------------------------
      SmaI BamHI XbaI   SalI  PstI  SphI HindIII

 401  CGGATGAGAG AAGATTTTCA GCCTGATACA GATTAAATCA GAACGCAGAA
       ***                   ***         ***

 451  GCGGTCTGAT AAAACAGAAT TTGCCTGGCG GCAGTAGCGC GGTGGTCCCA

 501  CCTGACCCCA TGCCGAACTC AGAAGTGAAA CGCCGTAGCG CCGATGGTAG

 551  TGTGGGGTCT CCCCATGCGA GAGTAGGGAA CTGCCAGGCA TCAAATAAAA

 601  CGAAAGGCTC AGTCGAAAGA CTGGGCCTTT CGTTTTATCT GTTGTTTGTC

 651  GGTGAACGCT CTCCTGAGTA GGACAAATCC GCCGGGAGCG GATTTGAACG

 701  TTGCGAAGCA ACGGCCCGGA GGGTGGCGGG CAGGACGCCC GCCATAAACT

 751  GCCAGGCATC AAATTAAGCA GAAGGCCATC CTGACGGATG GCCTTTTTGC

 801  GTTTCTACAA ACTCTTTTTG TTTATTTTTC TAAATACATT CAAATATGTA

 851  TCCGCTCATG AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA

 901  GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT TCCCTTTTTT

 951  GCGGCATTTT GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT

1001  AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG

1051  ATCTCAACAG CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT

1101  CCAATGATGA GCACTTTTAA AGTTCTGCTA TGTGGCGCGG TATTATCCCG

1151  TGTTGACGCC GGGCAAGAGC AACTCGGTCG CCGCATACAC TATTCTCAGA
```



```
1201  ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC
1251  ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC
1301  TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG
1351  CTTTTTTGCA CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA
1401  CCGGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGACA CCACGATGCC
1451  TACAGCAATG GCAACAACGT TGCGCAAACT ATTAACTGGC GAACTACTTA
1501  CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC GGATAAAGTT
1551  GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA
1601  TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG
1651  GGCCAGATGG TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT
1701  CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA TAGGTGCCTC
1751  ACTGATTAAG CATTGGTAAC TGTCAGACCA AGTTTACTCA TATATACTTT
1801  AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA GGTGAAGATC
1851  CTTTTTGATA ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA
1901  CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT
1951  TTTTTCTGCG CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA
2001  GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA CCAACTCTTT TTCCGAAGGT
2051  AACTGGCTTC AGCAGAGCGC AGATACCAAA TACTGTCCTT CTAGTGTAGC
2101  CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC TACATACCTC
2151  GCTCTGCTAA TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG
2201  TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG GCGCAGCGGT
2251  CGGGCTGAAC GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC
2301  TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA GCGCCACGCT
2351  TCCCGAAGGG AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA
2401  CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG GTATCTTTAT
2451  AGTCCTGTCG GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG
2501  CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT
2551  TACGGTTCCT GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG
```



```
2601  TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGCCTTTG AGTGAGCTGA
2651  TACCGCTCGC CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG
2701  AAGCGGAAGA GCGCCTGATG CGGTATTTTC TCCTTACGCA TCTGTGCGGT
2751  ATTTCACACC GCATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC
2801  ATAGTTAAGC CAGTATACAC TCCGCTATCG CTACGTGACT GGGTCATGGC
2851  TGCGCCCCGA CACCCGCCAA CACCCGCTGA CGCGCCCTGA CGGGCTTGTC
2901  TGCTCCCGGC ATCCGCTTAC AGACAAGCTG TGACCGTCTC CGGGAGCTGC
2951  ATGTGTCAGA GGTTTTCACC GTCATCACCG AAACGCGCGA GGCAGCAGAT
3001  CAATTCGCGC GCGAAGGCGA AGCGGCATGC ATTTACGTTG ACACCATCGA
3051  ATGGTGCAAA ACCTTTCGCG GTATGGCATG ATAGCGCCCG GAAGAGAGTC
3101  AATTCAGGGT GGTGAATGTG AAACCAGTAA CGTTATACGA TGTCGCAGAG
3151  TATGCCGGTG TCTCTTATCA GACCGTTTCC CGCGTGGTGA ACCAGGCCAG
3201  CCACGTTTCT GCGAAAACGC GGGAAAAAGT GGAAGCGGCG ATGGCGGAGC
3251  TGAATTACAT TCCCAACCGC GTGGCACAAC AACTGGCGGG CAAACAGTCG
3301  TTGCTGATTG GCGTTGCCAC CTCCAGTCTG GCCCTGCACG CGCCGTCGCA
3351  AATTGTCGCG GCGATTAAAT CTCGCGCCGA TCAACTGGGT GCCAGCGTGG
3401  TGGTGTCGAT GGTAGAACGA AGCGGCGTCG AAGCCTGTAA AGCGGCGGTG
3451  CACAATCTTC TCGCGCAACG CGTCAGTGGG CTGATCATTA ACTATCCGCT
3501  GGATGACCAG GATGCCATTG CTGTGGAAGC TGCCTGCACT AATGTTCCGG
3551  CGTTATTTCT TGATGTCTCT GACCAGACAC CCATCAACAG TATTATTTTC
3601  TCCCATGAAG ACGGTACGCG ACTGGGCGTG GAGCATCTGG TCGCATTGGG
3651  TCACCAGCAA ATCGCGCTGT TAGCGGGCCC ATTAAGTTCT GTCTCGGCGC
3701  GTCTGCGTCT GGCTGGCTGG CATAAATATC TCACTCGCAA TCAAATTCAG
3751  CCGATAGCGG AACGGGAAGG CGACTGGAGT GCCATGTCCG GTTTTCAACA
3801  AACCATGCAA ATGCTGAATG AGGGCATCGT TCCCACTGCG ATGCTGGTTG
3851  CCAACGATCA GATGGCGCTG GGCGCAATGC GCGCCATTAC CGAGTCCGGG
3901  CTGCGCGTTG GTGCGGATAT CTCGGTAGTG GGATACGACG ATACCGAAGA
3951  CAGCTCATGT TATATCCCGC CGTCAACCAC CATCAAACAG GATTTTCGCC
```



4001 TGCTGGGGCA AACCAGCGTG GACCGCTTGC TGCAACTCTC TCAGGGCCAG

4051 GCGGTGAAGG GCAATCAGCT GTTGCCCGTC TCACTGGTGA AAAGAAAAAC

4101 CACCCTGGCG CCCAATACGC AAACCGCCTC TCCCCGCGCG TTGGCCGATT

4151 CATTAATGCA GCTGGCACGA CAGGTTTCCC GACTGGAAAG CGGGCAGTGA

4201 GCGCAACGCA ATTAATGTGA GTTAGCGCGA ATTGATCTG

R E I V I N D I C A Ç Õ E S

- 1ª -

Processo para a secreção de proteínas em bactérias gran-negativas, caracterizado por se posicionar um péptido de sinal, derivado de Bordetella pertussis, com a sequência de aminoácidos seguintes

M K K V F V A A G I G A G L L M L S S A A

antes da estrutura genética pretendida.

- 2ª -

Processo para o isolamento de sequência de sinal, caracterizado por se empregarem vectores de expressão, que possuem sucessivamente:

(a) um promotor forte, regulável, de preferência trc,

(b) a posição de ligação de ribossoma lacz e

(c) o gene estrutural de fosfatase alcalina periplasmética (phoA) isenta de sequência de sinal, na ordem de leitura correcta em relação ao vector, no qual a posição de corte NcoI contida na phoA se encontra afastada e se encontra presente imediatamente a seguir ao extremo 5' do gene estrutural da phoA isento de se-

quência de sinal.

- 3ª -

Processo de acordo com a reivindicação 2, caracterizado por se empregar como vector de expressão o plasmídeo pTrc 99C - phoA.

- 4ª -

Processo para a reprodução quantitativa de sequências de sinal, caracterizado por se empregarem vectores de expressão, que possuem sucessivamente:

(a) um promotor forte, regulável, de preferência trc,
(b) a posição de ligação de ribossoma lacz e
(c) o gene estrutral da fosfatase alcalina periplasmética (phoA) isenta de sequência de sinal, no ordem de leitura correcta em relação ao vector, no qual a posição de corte NcoI contida na phoA se encontra afastada e se encontra presente imediatamente a seguir ao extremo 5' do gene estrutural da phoA isento de sequência de sinal.

A requerente reivindica a prioridade do pedido alemão apresentado em 21 de Janeiro de 1989, sob o nº P 39 01 681.1.

Lisboa, 18 de Janeiro de 1990
O AGENTE OFICIAL DA PROPRIEDADE INDUSTRIAL



RESUMO

"PROCESSO PARA A PREPARAÇÃO DE UM PÉPTIDO DE SINAL E PARA A SECREÇÃO DE PÉPTIDOS EM ESCHERICHIA COLI POR MEIO DESSE PÉPTIDO DE SINAL"

A invenção refere-se a um processo para a secreção de proteínas em bactérias gram-negativas, que compreende posicionar-se um péptido de sinal, derivado de Bordetella pertussis, com a sequência de aminoácidos seguintes

M K K W F V A A G I G A G L L M L S S A A

antes da estrutura genética pretendida.

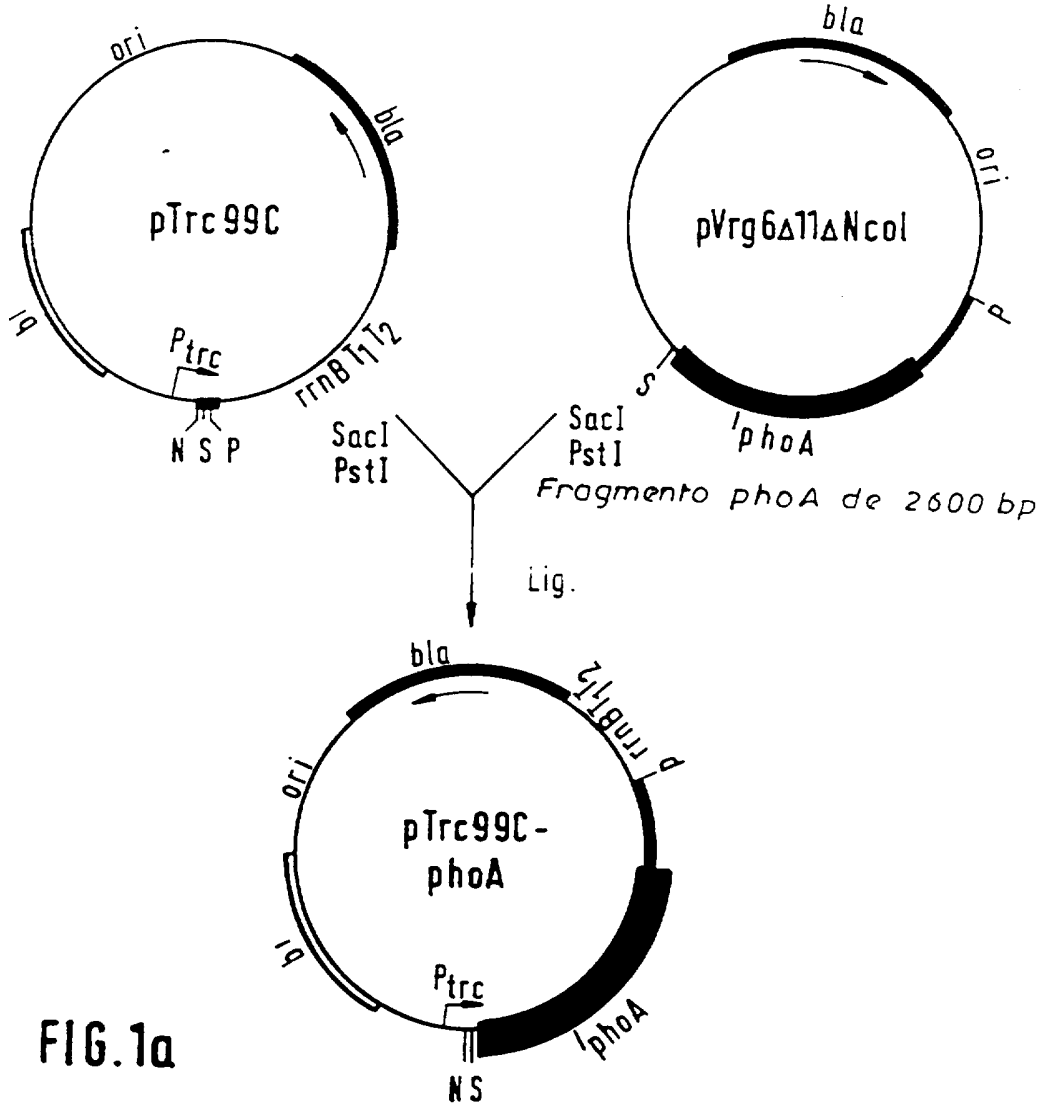

ori
bla
pTrc 99C
Ptrc
rrnBT1T2
N S P
SacI
PstI
bla
ori
pVrg6Δ11ΔNcol
P
S
IphoA
SacI
PstI
Fragmento phoA de 2600 bp
Lig.
bla
P rrnBT1T2
ori
pTrc99C-
phoA
Ptrc
IphoA
N S

FIG.1a

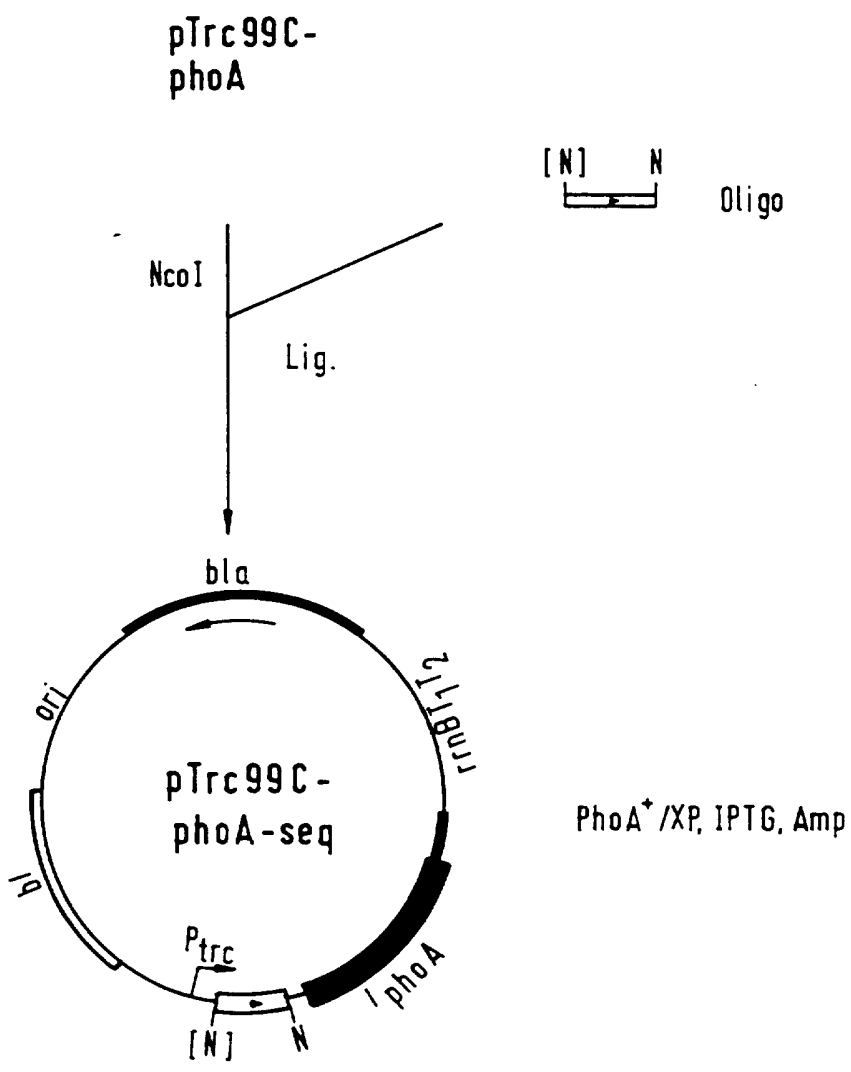

pTrc99C-
phoA
[N]
N
Oligo
NcoI
Lig.
bla
ori
rrnBT1T2
pTrc99C-
phoA-seq
bl
Ptrc
phoA
[N]
N
PhoA+/XP, IPTG, Amp

FIG. 1b

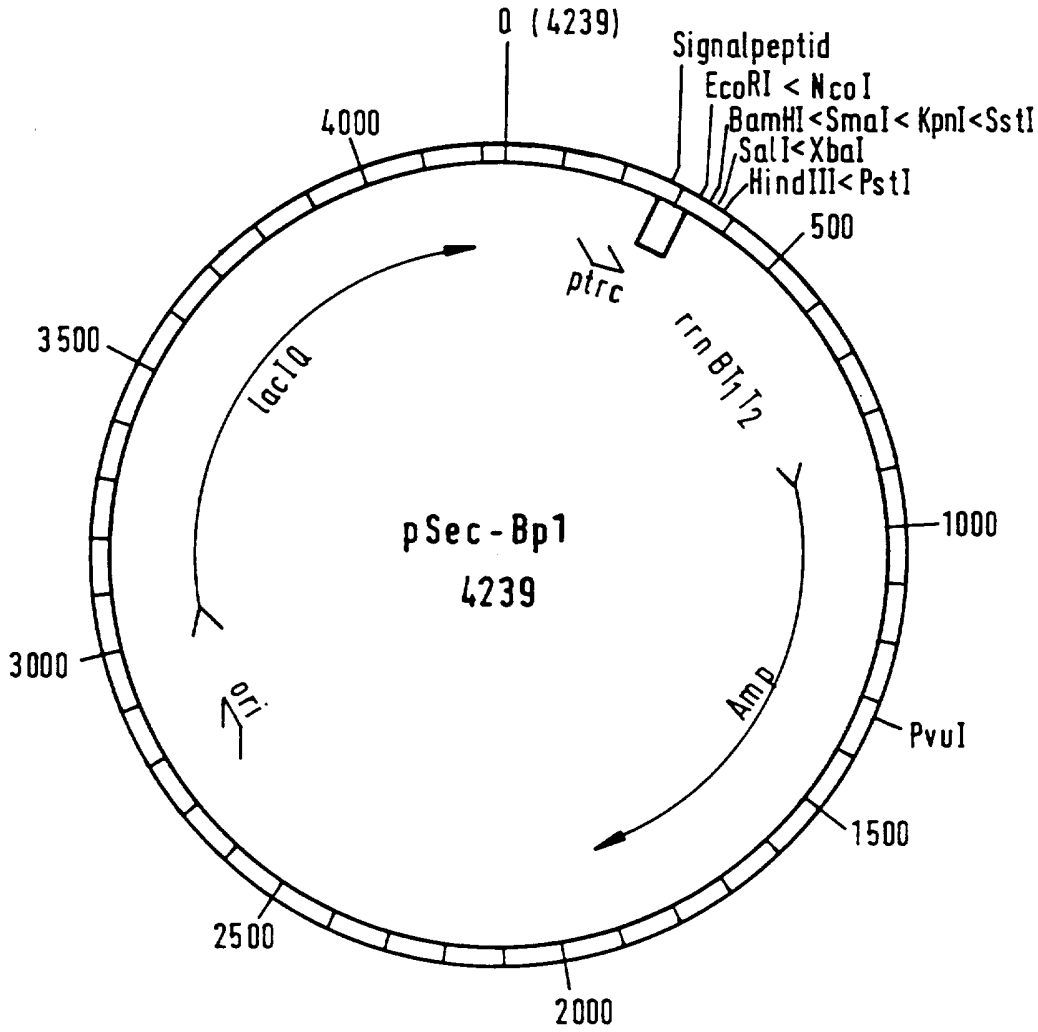

0 (4239)
Signalpeptid
EcoRI < NcoI
BamHI<SmaI< KpnI<SstI
SalI<XbaI
HindIII< PstI
500
1000
PvuI
1500
2000
2500
3000
3500
4000
ptrc
rrnB T1 T2
lacIq
pSec-Bp1
4239
Amp
ori

FIG. 2

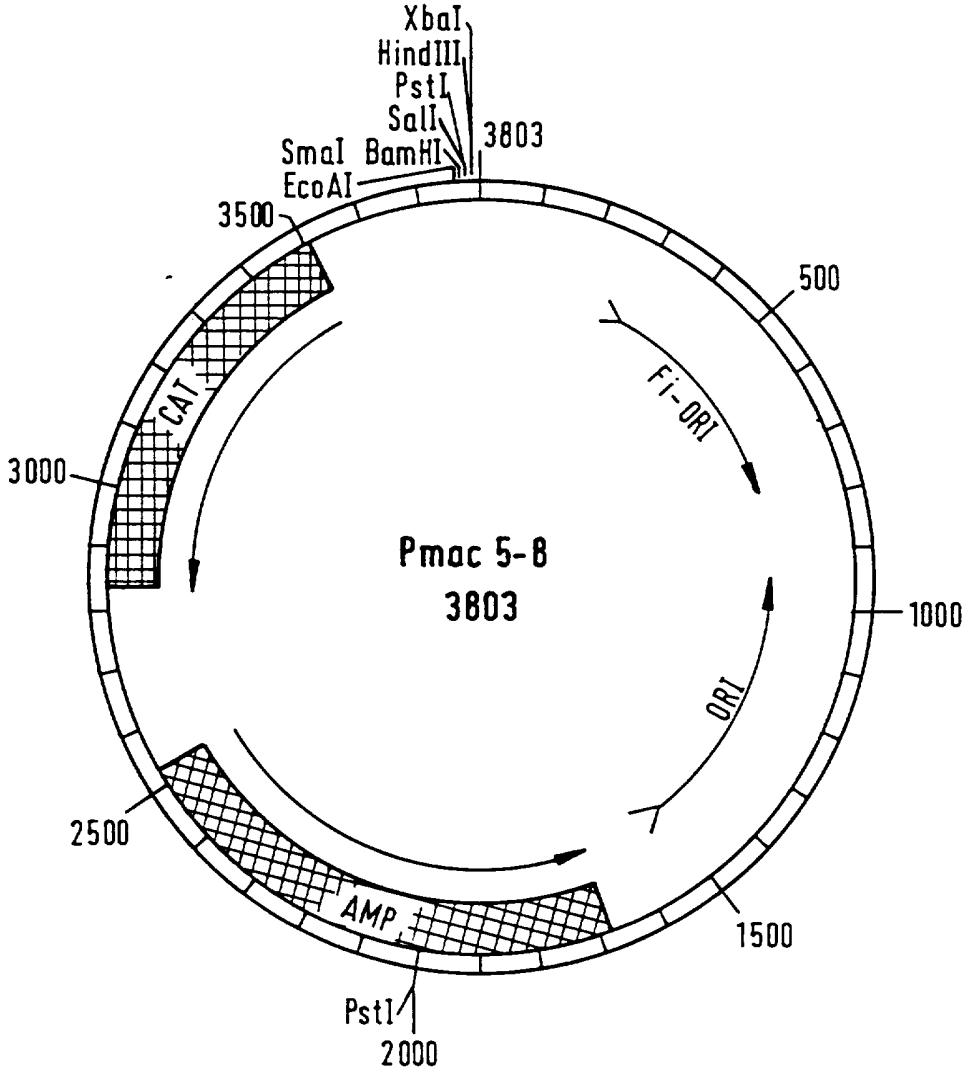

XbaI
HindIII
PstI
SalI
SmaI
BamHI
EcoAI
3803
3500
500
CAT
Fi-ORI
3000
Pmac 5-8
3803
1000
ORI
2500
AMP
1500
PstI
2000

FIG. 3