



( I P ) INSTITUTO NACIONAL  
DA PROPRIEDADE INDUSTRIAL  
PORTUGAL

(11) *Número de Publicação:* PT 91929 B

(51) *Classificação Internacional:* (Ed. 6)

C12N009/72 A

C12N015/58 B

**(12) FASCÍCULO DE PATENTE DE INVENÇÃO**

|                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                          |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>(22) <i>Data de depósito:</i> 1989.10.09</p> <p>(30) <i>Prioridade:</i> 1988.10.11 GB 8823833</p> <p>(43) <i>Data de publicação do pedido:</i><br/>1990.04.30</p> <p>(45) <i>Data e BPI da concessão:</i><br/>11/94 1994.11.18</p> | <p>(73) <i>Titular(es):</i><br/>FARMITALIA CARLO ERBA, S.P.A.<br/>VIA CARLO IMBONATI, 24 MILÃO 20159 IT</p> <p>(72) <i>Inventor(es):</i><br/>PAOLO SARMIENTOS IT<br/>ANNA BRANDAZZA IT<br/>GAETANO ORSINI IT</p> <p>(74) <i>Mandatário(s):</i><br/>ANTÓNIO LUÍS LOPES VIEIRA DE SAMPAIO<br/>RUA DE MIGUEL LUPI 16 R/C 1200 LISBOA PT</p> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

(54) *Epígrafe:* PROCESSO PARA A PREPARAÇÃO DE PRÓ-UROQUINASE HUMANA

(57) *Resumo:*

[Fig.]

MEMÓRIA DESCRITIVA

DA

PATENTE DE INVENÇÃO

Nº 91.929

NOME: Farmitalia Carlo Erba S.R.L., italiana, com sede em  
Via Carlo Imbonati, 24 - 20159, Milano, Itália,

EPIGRAFE: "Processo para a preparação de pró-uroqui-  
nase humana",

INVENTORES: Anna Brandanza,  
Paolo Sarmientos,  
Gaetano Orsini,

Reivindicação do direito de prioridade ao abrigo do  
artigo 4º da Convenção da União de Paris de 20 de  
Março de 1883.

Grã-Bretanha, 11.10.1988, sob o Nº 8823833.2,

4.

FARMITALIA CARLO ERBA S.R.L.

"PROCESSO PARA A PREPARAÇÃO DE PRÓ-UROQUINASE HUMANA"

A presente invenção refere-se a um método de ADN re combinante para a preparação de pró-uroquinase de cadeia simples não glicosilada (adiante designada por proUK). Mais particularmente, refere-se a um método para a preparação de proUK não glicosilada, o qual consiste em recuperar o mARN a partir duma linha de células contínua, em preparar o cADN com base no referido mARN, em inserir o cADN num vector, em introduzir o plasmídeo resultante numa célula bacteriana para desse modo produzir um transformante e em recuperar proUK não glicosilada a partir da referida célula bacteriana.

A presente invenção diz respeito também a diversos plasmídeos de expressão utilizados no método anterior.

INTRODUÇÃO

O conhecimento crescente das interacções moleculares que regulam a fibrinólise fisiológica tem implicações importantes no domínio da compreensão dos mecanismos que dissolvem os coágulos sanguíneos, e no desenvolvimento de novos agentes trombolíticos.

No sistema fibrinolítico humano, existe uma proenzima, o plasminogénio, que pode ser activada para proporcionar a enzima activa, a plasmina, através de diversos tipos de activadores de plasminogénio (Collen, D., e Lijnen, H.R., CRC Critical Reviews in oncology/hematology, 4, n. 3, 1986, p. 249; Verstraete,

M., e Collen, D., Blood, 67, n. 6, 1986, p. 1529). A plasmina é a protease principal responsável pela degradação do componente fibrina de um coágulo sanguíneo (Rakoczi, I., Wiman, B., e Collen, D., Biochim. Biophys. Acta, 540, 1978, p. 295; Robbins, K.C. Summaria, L., Hsieh, B.; e Shah, R.S., J. Biol.Chem., 242, 1967, p.2333; Wiman, B., Eur. J. Biochem, 76 1977, p. 129).

Todavia a plasmina pode exercer também o seu efeito proteolítico sobre diversas proteínas do plasma entre as quais se faz referência aos componentes do fibronogéneo da sequência de coagulação, ao factor V e ao factor VIII (Collen, D., e Lijnen, H.R., CRC Critical Reviews in oncology/hematology, 4, n. 3, 1986, p. 249; Verstraete, M., e Collen, D., Blood, 67, n. 6, 1986, p. 1529; Wiman, B., Lijnen, H.R., e Collen, D., Biochim. Biophys. Acta, 579, 1979, p. 142).

A activação do plasminogéneo pode ocorrer ao nível sistémico, originando plasmina circulante que é rapidamente neutralizada por alfa2-antiplasmina e que por isso não se encontra disponível para a fibrinólise (Collen, D., e Lijnen, H.R., CRC Critical Reviews in oncology/hematology, 4, n. 3, 1986, p. 249; Verstraete, M., e Collen, D., Blood, 67, n. 6, 1986, p. 1529).

Quando o nível de alfa2-antiplasmina for acentuadamente reduzido, a plasmina é neutralizada menos rapidamente e pode exercer os seus efeitos proteolíticos não só sobre a fibrina mas também sobre as proteínas de coagulação sanguínea conforme anteriormente descrito.

A diminuição excessiva das concentrações de fibrinogéneo, de factor V e de factor VIII no plasma, juntamente com os efeitos inibidores exercidos por alguns dos produtos de degradação do fibrinogéneo sobre o processo hemostático, sobre a agregação das plaquetas e sobre a polimerização da fibrina originam

4.

uma deficiência hemostática e subsequentemente elevado risco de hemorragia (Latallo, Z.S. e Lopaciuk, S., Thrombos. Diath. Haemouh., 56, 1973, p. 253; Totty, W.G., Gilula, L.A., Mc. Clennman, M., Ahmed, P., e Sherman, L., Radiology, 143, 1982, p. 59).

Por outro lado, a activação do plasminogénio pode ocorrer ao nível da fibrina (activação do plasminogénio ligado à fibrina) originando plasmina ligada à fibrina (Collen, D. e Lijnen, H.R., CRC Critical Reviews in oncology/hematology, 4, n. 3, 1986, P. 249; Verstraete, M. e Collen, D., Blood, 67, n. 6, 1986, p. 1529), a qual, contrariamente, não é afectada pela alfa2-antiplasmina e não pode induzir a fibrinogenólise sistémica.

A uroquinase e a estreptoquinase, os activadores de plasminogénio mais vulgarmente utilizados na terapêutica trombolítica convencional nos seres humanos não possuem actividade específica para a fibrina. Os dois compostos activam de modo relativamente indiscriminado o plasminogénio circulante ou o plasminogénio ligado à fibrina (Zamarron, C., Lijnen, H.R., Van Hoef, B., e Collen, D., Thromb. Haemostas., 52, 1984, p. 19; Samama, M., e Kher, A., Sem. Hop. Paris., 61, n. 20, 1985, p. 1423).

Por conseguinte, o colapso hemostático sistémico observado frequentemente durante o tratamento com estreptoquinase e com uroquinase e, consequentemente, o elevado risco de hemorragia impediram frequentemente a utilização clínica vulgarizada destes agentes trombolíticos apesar de demonstrada a sua eficácia clínica (Samama, M., e Kher, A., Sem. Hop. Paris, 61, n. 20, 1985, p. 1423; Maizel, A.S., e Bookstein, J.J., Cardiovasc. Intervent. Radiol., 9, 1986, p. 236; Bell, W.R., Thromb. Haemostas., 35, 1976, p. 57; Acar, J., Vahanian, A., Michel, P.L., Slama, M., Cormier, B. e Roger, V., Seminars in Thromb. and Haemost., 13, n. 2, 1987, p. 186;

Gruppo Italiano per lo studio della Streptochinase nell'infarto miocardico (GISSI), Lancet, 1, 1986, p. 397).

Pelo contrário, o activador de plasminogénio de tipo tecidual (t-PA) (Hoylaerts, M., Ryken, D.C., Lijnen, H.R. e Collen, D., J.Biol. Chem., 257, n. 6, 1982, p. 2912), e mais recentemente a pró-uroquinase (proUK) (Husain, S.S., e Gurewich, V., Arch. Biochem. Biophys., 220, 1983, p. 31), ambas proteínas naturais, demonstraram ser fracos activadores do plasminogénio circulante e, inversamente, fortes activadores do plasminogénio ligado à fibrina, sem que haja risco de colapso hemostático sistémico ou de consumo de alfa2-antiplasmina e de plasminogénio, pelo que a sua utilização clínica pode originar menor risco de hemorragia.

A actividade de t-PA trombolítica específica da fibrina tem sido explicada pela sua capacidade para ligar a fibrina através de sítios específicos de ligação da lisina, localizados nos "domínios de Kringle" da molécula de ligação tripla dissulfureto.

Em consequência, o plasminogénio ligado à fibrina poderia ser activado sem significativo colapso hemostático (Collen, D., e Lijnen, H.R., Haemostasis, 16, n. 3, 1986, p. 25). Por outro lado, a proUK (também designada por activador de plasminogénio do tipo uroquinase de cadeia simples, scu-PA) não se liga à fibrina, mas existe actividade trombolítica específica da fibrina sem consumo hemostático sistémico (Pannell, R. e Gurewich, V., Blood, 67, 1986, p. 1215; Gurewich, V., e Pannell, R.; Seminars in Thromb. e Haemost., 13, n. 2, 1987, p. 146; Lijnen, H.R., Zamarron, C., Blaber, M., Winbler, M.E., e Collen, D., J. Biol. Chem., 261, 1986, p. 1253).

O t-PA recombinante foi submetido a ensaios clínicos

4.

em diversas instituições em pacientes com enfartes do miocárdio agudos tendo-se verificado que é significativamente mais eficaz que a estreptoquinase para a recanalização de artérias coronárias obstruídas (The European Cooperative Study Group for Recombinant Tissue-type Plasminogen Activator, Lancet, 1, 1985, p. 842; Sheehan, F.H. Braunwald E., Canner, P., Doodge, H.T., Gore, J., Van Natta P., Passamani, E.R. Williams, D.O., Zaret, B., Circulation, 75, 4, 1987, p.817).

Presentemente a pró-uroquinase encontra-se na fase de ensaios clínicos avançados e admite-se que seja, pelo menos, tão eficaz quanto a t-PA em termos de segurança e de actividade trombolítica (Van de Werf, F., Nobuhara, M., e Collen, D., Annals of Internal Medicine, 104, 1986, p. 345; Van de Werf, F., Vanhaecke, J., De Geest, H., Verstraete, M., e Collen, D., Circulation, 74, n.5, 1986, p. 1066).

#### ANTECEDENTES DA INVENÇÃO

Os activadores de plasminogénio do tipo uroquinase (u-PAs) encontram-se em pelo menos três formas diferentes na urina humana, no plasma e em meio de cultura condicionada a partir de uma diversidade de linhas de células. A primeira forma caracterizada como u-PA é constituída por um polipéptido fibrinoliticamente activo constituído por 410 aminoácidos, com um peso molecular aparente de 54000 daltons, que tem duas cadeias ligadas por dissulfureto (Gunzler, W.A., Steffens, G.J., Oetting, F., Kim, S.M.A. Frankus, E., e Flohé, L., Hoppe - Seyler's Z. Physiol.Chem., 363, 1982, p.1155).

A cadeia A, ou cadeia leve, contém 157 aminoácidos e uma estrutura de "Kringle" de ligação tripla dissulfureto. Esta

cadeia contém também um domínio de ligação ao receptor para células normais e neoplásicas (monócitos, células semelhantes a monócitos e células epidermóides A 431). A cadeia B ou cadeia pesada (30000 daltons) é constituída por 253 aminoácidos e contém o domínio catalítico.

Esta forma molecular de u-PA é geralmente designada por uroquinase (UK), por uroquinase de cadeia dupla (TC-UK) ou por uroquinase de elevado peso molecular (HMW-UK) (Gunzler, W. A., Steffens, G.J., Oetting, F., Buze, G., e Flohé, L., Hoppe-Seyler's Z. Physiol. Chem., 363, 1982, p.133).

A segunda forma de u-PA tem um peso molecular de 33000 daltons e resulta da degradação proteolítica da forma HMW, pela via da plasmina ou da tripsina. É designada por uroquinase de baixo peso molecular (LMW-UK). As determinações da sequência de proteínas revelaram que a LMW-UK é idêntica à HMW-UK, excepto no que diz respeito à ausência dos 135 aminoácidos do terminal NH<sub>2</sub> que são removidos especificamente pela acção da plasmina ou da tripsina (Steffens, G.J., Gunzler, W.A. Oetting, F., Frankus, E., e Flohé, L., Hoppe-Seyler's Z. Physiol. Chem., 363, 1982, p. 1043). A pró-uroquinase nativa (proUK) é uma forma de uroquinase de cadeia simples (54000 daltons) e é também designada por activador de plasminogénio do tipo uroquinase de cadeia simples (scu-PA). Conforme anteriormente especificado, a proUK apresenta uma actividade trombolítica específica para a fibrina e consequentemente constitui um agente trombolítico melhor por comparação com as uroquinases de elevado ou de baixo peso molecular presentemente utilizadas.

No sentido de se produzir pró-uroquinase, os inventores da presente invenção desenvolveram um processo de ADN recom-

4.

binante, o qual permite a preparação de grandes quantidades de proUK de cadeia polipeptídica.

Existem diversos métodos descritos na literatura científica e de patentes de invenção para a produção de proUK (Holmes, W.E., Pennica, D., Blaber, M., Rey, M.W., Gunzler, W.A., Steffens, G.J. e Heynecker, H.L., Biotechnology, 3, 1985, p. 923; pedido de patente europeia nº 0092182). Todavia, o método descrito no âmbito da presente invenção explora parâmetros conhecidos como sendo importantes para a expressão de proteínas heterólogas em E. coli, mas cuja combinação nunca foi aplicada antes à produção de proUK recombinante.

Os parâmetros principais, cuja combinação contribui para o estabelecimento de uma estirpe recombinante de E. coli, susceptíveis de produzir proUK, e os quais constituem o objectivo da presente invenção, são o promotor Ptrp de E. coli, a sequência MS-2 de Shine-Dalgarno do fago MS-2, e as estirpes do tipo B de E. coli como hospedeiros para a expressão do gene de proUK humana (ver a descrição que se segue). Essa combinação é crucial. A substituição de um destes parâmetros por um sinal de expressão alternativo pode não proporcionar tanta proUK.

Em consequência, o objectivo da presente invenção é um processo para a preparação de proUK não glicosilada, caracterizado pelo facto de a proUK não glicosilada ser expressa sob o controlo do promotor Ptrp de E. coli e da sequência MS-2 de Shine-Dalgarno por E. coli B.

#### DESCRIÇÃO DO PROCEDIMENTO DE PRODUÇÃO

A presente invenção diz respeito à construção, por téc-



nicas de engenharia genética, de estirpes de E.coli susceptíveis de fazer a expressão do gene proUK humano a níveis elevados. Em consequência, estas estirpes recombinantes são susceptíveis de sintetizar grandes quantidades de proUK de cadeia polipeptídica.

No sentido de se isolar as referidas estirpes recombinantes da bactéria E.coli, é necessário efectuar diversos passos, incluindo :

- o isolamento do gene cADN humano que codifica para a proUK desejada,
- a inserção do referido gene num plasmídeo de expressão apropriado,
- a transformação da estirpe seleccionada de E.coli com o plasmídeo modificado pela engenharia e a cultura de transformantes em condições apropriadas.

1) Clonagem do gene cADN humano que codifica a proUK

Para se obter o clone cADN que codifica para a pró-uroquinase humana, os inventores utilizaram os dados sequenciais sobre proteínas publicados na literatura (Gunzler, W.A., Steffens, G.J., Oetting, F., Kim, S.M.A., Frankus, E., e Flohé, L., Hoppe - Seyler's Z. Physiol.Chem., 363, 1982, p. 1155; Gunzler, W.A., Steffens, G.J., Oetting, F., Buze, G., e Flohé, L., Hoppe-Seyler's Z. Physiol. Chem., 363, 1982, p. 133; Steffens, G.J., Gunzler, W.A., Oetting, F., Frankus, E., e Flohé, L., Hoppe-Seyler's Z. Physiol. Chem., 363, 1982, p. 1043). Em consequência, procedeu-se à preparação de sondas específicas e fez-se o rastreio de uma biblioteca de cADN apropriada.

Fez-se a síntese química de oligonucleótidos que codificam para os péptidos seleccionados do activador de plasminogénio.



neo do tipo uroquinase de cadeia simples (proUK) (Caruthers, M.H., Gassen, H.G. e Lang, J.A. (eds) Verlag-chimie, Weinheim, Deefield Beach, Basel, (1982), p.71) para servirem como sondas específicas para controlar o enriquecimento do mRNA de proUK e para seleccionar os clones que contêm cADN de pró-uroquinase, a partir de uma biblioteca de cADN enriquecida. Os oligómeros tinham um comprimento de 14 a 17 unidades mero, tendo cada um deles sido sintetizado ou como sequência única (designada p7) ou em agregados que continham dois (designados p1, p2, p3) ou 16 (designados p6) oligonucleótidos, conforme indicado na Figura 1. Testou-se a especificidade dos oligómeros relativamente a proUK segundo a hibridação de Northern. Para esta análise extraiu-se ARN que continha poli-A a partir de carcinoma epidermóide HEP-3 (Miskin, R., Haemostasis (Switzerland), 11, nº supl. 1, (1982) p. 63). Para cada oligómero, após a reacção de hibridação, ajustou-se a temperatura de lavagem de modo a ser 2 a 5 graus Celsius inferior à temperatura mínima de fusão, conforme calculado por Suggs e outros para a hibridação em coágulos de Southern blots (Suggs, S.V., Hirose, T., Miyake, T., Kawashima, E.G., Johnson M.Y., Itakura, K. e Wallach, R.B., Developmental Biology Using Purified Genes, Brown D.D. e Fox, C.F. (eds), New York, Academic Press, 1981 p. 638). No presente trabalho as cinco sondas de proUK representada na Figura 1 reagiram com uma banda de mRNA de carcinoma principal comum de aproximadamente 2,3kb sendo estas dimensões as esperadas para o mRNA de proUK.

A clonagem ocorre utilizando fracções de mRNA enriquecido, a partir de carcinoma epidermóide de HEP-3. As preparações de ARN foram extraídas e enriquecidas aproximadamente três vezes em dois gradientes de sacarose sucessivos. Fez-se a síntese de

4.

cADN de acordo com procedimentos conhecidos (Efstratiadis, A., Kafatos, F.C., Maxam, A.M. e Maniatis, T., Cell, 7, 1976, p. 279; Buell, G.N., Wickens, M.P., Payvar, F. e Schimke, R.T., J. Biol. Chem., 253, (1978), p. 2471), utilizando oligo-dT como origem iniciadora. Fez-se o isolamento das moléculas mais compridas utilizando a electroforese em gel de poliacrilamida seguindo-se a electro-eluição das fracções de gel apropriadas. Depois, extraiu-se o cADN utilizando a técnica de extracção com fenol/clorofórmio normalizada seguindo-se a precipitação com etanol.

As moléculas deste cADN foram ligadas primeiro aos ligantes de EcoRI e depois clonadas no vector  $\lambda$ gt10 do fago de acordo com uma modificação da técnica de Davis (Maniatis, T., Fritsch, E. F., Sambrook, J., Molecular cloning. A laboratory manual., Cold Spring Harbour, NY, Cold Spring Harbour, Laboratory, 1982). Procedendo deste modo, construiu-se uma biblioteca que continha  $2 \times 10^5$  u.f.p. (unidades de formação de placa).

Metade da biblioteca foi filtrada em filtros duplicados, um filtro com a sonda p1 marcada com  $^{32}\text{P}$  e o filtro complementar com uma mistura de sondas p3 e p6. Obteve-se um total de 36 clones positivos, sete dos quais foram positivos nos filtros complementares, indicando assim inserções cDNA correspondentes a uma grande porção da sequência que codifica proUK.

Os fagos recombinantes que hibridaram com as três sondas foram purificados de placas utilizando a sonda p1 e ainda caracterizados por cartografia de restrição com EcoRI e por sequenciação de ADN. A fracção dos clones positivos na biblioteca de cADN total indicou que a frequência de mRNA de pró-uroquinase no carcinoma HEP-3 é de aproximadamente 0,01%.

A análise sequencial de quatro clones de cADN de proUK

revelou que três dos clones apresentavam anulações ou sequências não consistentes com as sequências de aminoácidos da enzima. Apenas um clone, o  $\lambda$ Uc17, apresentou uma sequência completamente concordante com a sequência de aminoácidos conhecida.

Todavia, o  $\lambda$ Uc-17 não incluía a extremidade não codificadora 3' completa do mRNA e verificou-se que faltavam 30 nucleótidos da sequência codificadora. Construiu-se um clone cADN de pré-pró-uroquinase de comprimento completo por ligação de um fragmento SmaI-BamHI de  $\lambda$ Uc-17 de 1325 pb, que continha a região não codificadora 5' e a maior parte da sequência codificadora, com um fragmento BamHI-EcoRI que continha a restante região 3' em falta a partir de outro clone  $\lambda$ Uc-6 (Figura 2).

Esta construção foi ligada ao sítio SmaI-EcoRI do vector plasmídeo pUN121 (Nilsson B., Uhlen M., Josephson S., Gatenbeck S. e Philipson L., Nucleic Acid Research, 11, (1983), p. 8019), eliminando-se, deste modo, a maior parte do gene cI, e designou-se por pcUK176 (Figura 3).

A sequência de ADN do clone de cADN completo está representada na Figura 4. É constituída por 2296 nucleótidos, incluindo 69 nucleótidos não codificadores na extremidade 5', 1296 nucleótidos codificadores e 931 nucleótidos não codificadores na extremidade 3', seguida por uma cauda poli(A) de mais de 80 resíduos.

A sequência codificadora começa com 60 pb que codificam para 20 aminoácidos que constituem a "pré-pró-uroquinase" (Heyneker, H.L., Holmes, W.E. e Vehar, G.A. (1983), publicação do pedido de patente de invenção europeia nº 0092182), seguindo-se-lhe a sequência que codifica para a proteína de pró-uroqui-

nase completa, a qual se encontra em total concordância com a sequência de aminoácidos.

O fragmento completo foi verificado por análise sequencial e de restrição. A sequência que codifica para a pró-uroquinase madura foi inserida no vector de expressão para a produção.

## 2) Construção do plasmídeo de expressão de proUK

Utilizou-se o cADN de comprimento completo original, presente no pcUK176 para se construir um plasmídeo de expressão de pró-uroquinase designado por pFC44, no qual o gene proUK se encontra sob controlo transcripcional e translaccional do promotor P<sub>trp</sub> e da sequência MS-2 de "Shine-Dalgarno", respectivamente. Na Figura 7 representa-se o plasmídeo pFC44.

No sentido de se obter pFC44 foram construídos diversos plasmídeos intermédios. Partindo de pDS20 (Figura 5) (Duester, G., Helfard, R.M. e Holmes, W.M., Cell, 30, 1982, p. 855), substituiu-se primeiro o fragmento EcoRI-HindIII, que codifica o promotor P<sub>gal</sub> operação de galactose, pela sequência poli-ligante EcoRI-HindIII obtida a partir do vector M13mp8 (Vieira, J. e Messing, I., Gene, 19, 1982, p. 259), obtendo-se um novo plasmídeo designado por pAB1 (Figura 5).

O promotor P<sub>trp</sub> foi obtido a partir do plasmídeo pDR720 (comprado a Pharmacia) como fragmento de restrição EcoRI-SalI. Este fragmento foi inserido na região poli-ligante do pAB1 entre os sítios do EcoRI e do SalI. Procedendo deste modo, obteve-se um novo plasmídeo designado por pFC10 (Figura 5).

O pFC10 pode ser considerado como o vector base no qual se inseriu o gene de proUK, assim como a sequência de "Shine



-Dalgarno" proveniente do fago MS-2.

Para se conseguir a expressão da pró-uroquinase madura, é necessário fundir a sequência que codifica proUK, a partir do primeiro codão da proteína madura do triploto iniciador ATG. Esta fusão deve ser, por isso, precedida pela sequência de "Shine-Dalgarno".

O sítio de ligação do ribossoma (RBS) do fago bacteriano MS-2 era conhecido e a sua sequência de nucleótidos já havia sido publicada (Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G. e Ysebaert, M., Nature, 260, 1976, p. 500).

Admite-se que seja um sinal forte para uma tradução eficiente do mRNA. Em consequência, escolheu-se esta região como sinal de tradução para a produção de proUK. No sentido de se obter a correcta fusão do nucleótido com o gene proUK sintetizou-se uma região de ADN de cordão duplo do RBS do MS-2 unida directamente ao início do gene proUK. Existe um sítio TaqI no 25º nucleótido da sequência de proUK madura. Aproveitou-se este sítio e, por síntese química, isolou-se a seguinte sequência de ADN:

HindIII

MetSer

5'-AGCTTTAATAGACGCCGGCCATTCAAACATGAGGATTACCCATGAGCA

3'-AATTATCTGCGGCCGGTAAGTTTGTACTCCTAATGGGTACTTCGT

TaqI

ATGAACTTCATCAAGTTCCAT-3'

TACTTGAAGTAGTTCAAGGTAGC-5'

a qual é flanqueada a montante por um sítio HindIII e a jusante

por um sítio *TaqI*. O codão iniciador ATG encontra-se representado em caracteres a negrito. A sequência que codifica para o início da sequência de proUK madura está sublinhada.

Utilizou-se o fragmento sintético numa reacção de ligação com os dois fragmentos de restrição seguintes:

- o fragmento *TaqI*-*BglII* proveniente de pcUK176 (Figura 3), o qual suporta a sequência de proUK desde o nucleótido 155 até ao nucleótido 392 (ver Figura 4);

- o fragmento *BamHI*-*HindIII* grande proveniente de PFC10 (Figura 5), o qual suporta a resistência antibiótica à ampicilina assim como o promotor *P<sub>trp</sub>*.

Através desta construção fez-se o isolamento de um novo plasmídeo designado por pAB8, cujo mapa esquemático se representa na Figura 6. Neste plasmídeo, o promotor *P<sub>trp</sub>* e o RBS da MS-2 encontram-se fundidos com os primeiros 260 nucleótidos do gene proUK maduro (correspondente aos nucleótidos 131-391 na Figura 4). Além disso, o pAB8 tem um sítio *NcoI* único no qual se inseriu o resto da sequência de proUK através de um fragmento de restrição *NcoI*-*NcoI* proveniente de pcUK176. Esta ligação provocou a duplicação de um fragmento de *NcoI*-*BglII* a jusante do gene proUK na região de não codificação. Contudo, esta duplicação não afecta a estabilidade do plasmídeo. Através desta construção os sinais podem orientar agora a síntese da sequência de proUK completa (ver Figura 6).

Todos os plasmídeos até agora descritos foram seleccionados na estirpe hospedeira C-600 galK de *E.coli* K-12 (ATCC 33955), na base da resistência à ampicilina. Na realidade, comportam o gene que codifica a  $\beta$ -lactamase, a enzima responsável pela degradação da ampicilina antibiótica no meio de cultura.

Experiências recentes demonstraram que o pFC16 pode ser inserido com sucesso nas estirpes de E.coli do tipo B e originar a produção de proUK recombinante em níveis elevados.

Todavia, para obedecer às directrizes internacionais para a produção de produtos derivados de ADN recombinante, procedeu-se à modificação do pFC16 para criar um novo plasmídeo resistente à tetraciclina susceptível de fazer a expressão do gene proUK em níveis elevados. Em particular, a partir do bem conhecido plasmídeo pBR322 (Maniatis, T., Fritsch, E.F. Sambrook, J., Molecular cloning. A laboratory manual, Cold Spring Harbour, NY, Cold Spring Harbour Laboratory, 1982) (Figura 6), isolou-se o fragmento EcoRI-AvaI em que as extremidades aderentes foram preenchidas utilizando o fragmento de Klenow de ADN polimerase I "Klenow" fragment of ADN polimerase I (Perbal, B., A Wiley-Interscience Publication John Wiley and Sons, 1984, p. 231). Este fragmento foi ligado ao fragmento AatII-PvuI maior a partir de pFC16, cujas extremidades foram embotadas ("blunt") por ADN polimerase I (Perbal, B., A Wiley-Interscience Publication John Wiley And Sons, 1984, p. 231). Ao proceder-se deste modo, substituiu-se a porção amino-terminal do gene de  $\beta$ -lactamase e as suas sequências de controlo, pelo gene da resistência à tetraciclina. Seguiu-se a ligação do gene de resistência à tetraciclina com a mesma orientação que o gene de proUK.

Além disso criou-se um novo sítio EcoRI na junção entre os sítios PvuI e EcoRI, previamente preenchidos. O novo plasmídeo pFC44 (ver Figura 7) constitui a construção final que se utilizou para a produção de pró-uroquinase recombinante.

O plasmídeo pFC44 (resistente à tetraciclina) e o plasmídeo pFC16 (resistente à ampicilina) constituem um dos obje<sub>c</sub>

tos da presente invenção. Os sinais de expressão, presentes nestes dois plasmídeos, designadamente o promotor P<sub>trp</sub> e a sequência "MS-2" de Shine-Dalgarno, haviam sido já descritos na bibliografia da expressão de proteínas heterólogas (Remaut E., Stranssens P. e Fiers W., Nucl. Acid. Res., 11, 1983, p.4677). Contudo, nunca antes a sua combinação havia sido aplicada à expressão do gene proUK.

### 3) Transformação de estirpes de E. coli do tipo B

Constitui o segundo objecto principal da presente invenção a utilização de estirpes de E.coli do tipo B para a expressão e produção de pró-uroquinase. Na realidade, os inventores da presente invenção descobriram que a inserção de plasmídeos pFC16 ou pFC44 na estirpe do tipo B da bactéria E.coli proporciona níveis elevados de produção de proUK de cadeia polipeptídica. É interessante referir que a inserção dos plasmídeos pFC16 ou pFC44 noutras estirpes de E.coli (tipo K-12, tipo C, tipo W, etc.) não proporciona tanta proUK. Em consequência, o tipo de estirpe hospedeira parece ser crucial para o bom sucesso da produção de proUK.

Encontram-se disponíveis diversas estirpes do tipo B de E.coli as quais podem ser utilizadas para a expressão bem sucedida do gene de proUK. As estirpes preferidas são: ATCC 12407, ATCC 11303, NCTC 10537. Apresenta-se a seguir um exemplo de transformação da estirpe NCTC 10537 com o plasmídeo pFC44 e subsequente cultura do transformante.

Procedeu-se à preparação de células competentes da estirpe NCTC 10537 utilizando o processo do cloreto de cálcio de Mandel e Higa (Mandel, M. e Higa, A., J. Mol. Biol., 53, 1970,



p. 154). Procedeu-se à transformação de aproximadamente 200  $\mu$ l de uma preparação destas células na concentração de  $1 \times 10^9$  células por mililitro utilizando 2  $\mu$ l de ADN do plasmídeo (concentração aproximada de 5  $\mu$ g/ml). Fez-se a selecção dos transformantes sobre placas de ágar do tipo L que continham 12,5  $\mu$ g/ml de tetraciclina. Duas colónias pequenas foram riscadas com palitos de madeira (cada uma com três riscas de aproximadamente 1 cm de comprimento) sobre ágar do tipo L que continha o mesmo antibiótico. Após 12 horas de incubação à temperatura de 37°C, fez-se o teste às porções entre as riscas para se determinar a produção de pró-uroquinase humana por inoculação sobre 10 ml de meio LB (contendo tetraciclina numa concentração de 2,5  $\mu$ g/ml) e fez-se a incubação durante a noite à temperatura de 37°C. No dia seguinte fez-se a diluição das culturas a 1:100 em Meio M9 que continha a mesma concentração de tetraciclina e fez-se a incubação durante 6 horas à temperatura de 37°C. Fez-se a análise das proteínas celulares totais de alíquotas de 250  $\mu$ l de cultura D.O.<sub>550</sub> = 1-1,5) utilizando o processo da electroforese em gel de poliacrilamida/dodecil-sulfato de sódio conforme descrito por Laemmli (Laemmli, U.K. Nature, 227, (1970), p. 680). Para as duas amostras independentes (Figura 8) observou-se uma banda proteica principal com um peso molecular correspondente ao da pró-uroquinase humana não glicosilada (45000 daltons).

O conjunto de riscas correspondente à colónia nº 2 (clone 2) foi escolhido arbitrariamente para caracterização adicional e depois seleccionou-se como estirpe produtora de proUK.

## MATERIAIS E METODOS

Meio de desenvolvimento: Fez-se a preparação dos meios utilizando os récipes descritos por Maniatis e outros (Maniatis, T., Fritsch, E.F. Sambrook, J., Molecular cloning. A laboratory manual. Cold Spring Harbour, NY, Cold Spring Harbour Laboratory, 1982). Preparou-se Meio LB, ágar do tipo LB e ágar do tipo MacConkey utilizando produtos obtidos em "Difcobacto Products". O meio M9 é constituído pelos componentes seguintes:  $\text{Na}_2\text{HPO}_4$ , 6 gramas/l;  $\text{KH}_2\text{PO}_4$ , 3 g/l;  $\text{NaCl}$ , 0,5 g/l;  $\text{NH}_4\text{Cl}$ , 1 g/l.

Após esterilização dos componentes anteriores por tratamento em autoclave (1 atmosfera;  $120^\circ\text{C}$  durante 20 minutos), adicionou-se a cada litro 1 ml de  $\text{MgSO}_4$  1M; 0,1 ml de  $\text{CaCl}_2$  1M; 16 ml de glucose a 25%, 20 ml de Thiamine na proporção de 0,5 mg/ml (SIGMA) e 20 ml de ácidos Casamino a 20% (DIFCO) e esterilizou-se a solução anterior por filtração.

## UTILIZAÇÃO DAS ENDONUCLEASES DE RESTRIÇÃO E DE OUTRAS

### ENZIMAS:

As endonucleases de restrição, a T4 ADN Ligase e a ADN Polimerase I (Fragmento de Klenow) foram obtidas na New England Biolabs e na Boehringer Mannheim e foram utilizadas nas condições recomendadas pelo fabricante.

### PREPARAÇÃO DE ADN PLASMIDICO:

Preparou-se o ADN plasmídico utilizando um método que implica a centrifugação por densidade de flutuação do corante baseado no procedimento de Birnboim e Doly (Birnboim, H.C. e Doly. J.,

Nucleic Acid Res., 7, (1979) p. 1513).

Análise da sequência de ADN: Os dados sequenciais foram obtidos utilizando o conjunto de sequenciação Amersham M13 de acordo com as instruções do fabricante. Resumidamente, esta técnica baseada na técnica de Sanger (Sanger, F., Science, 214, (1981), p. 1205) consiste na subclonagem de diversos fragmentos de restrição em vectores M13 convenientes (a "família mp") os quais podem ser obtidos na sua configuração de cordão simples. Após o tratamento de formas de cordão simples com "origens iniciadoras universais", é possível copiar a matriz de ADN com DNA polimerase I (Fragmento de Klenow). Copiando a matriz na presença de quatro didesoxinucleótidos é possível provocar terminações aleatórias do prolongamento da cadeia. Depois faz-se a separação dos fragmentos truncados sobre géis de poliacrilamida de desnaturação sendo o perfil electroforético evidenciado por auto-radiografia.

Oligonucleótidos: Os oligonucleótidos sintéticos utilizados nas construções plasmídicas e como origens iniciadoras na sequenciação de ADN foram sintetizados utilizando o sintetizador 380B de ADN da Applied Biosystem (ABI) de acordo com o manual ABI.

Outros procedimentos: Os procedimentos para a electroforese em agarose e gel de poliacrilamida de ácidos nucleicos foram os descritos por Maniatis e outros (Maniatis, T., Fritsch, E.F. Sambrook, J., Molecular cloning. A laboratory manual, Cold Spring Harbour, NY, Cold Spring Harbour Laboratory, 1982). As proteínas foram separadas por electroforese em gel de poliacrilamida conforme descrito por Laemmli (Laemmli, U.K. Nature, 227, 1970, p.680). Extraiu-se o ADN dos géis de agarose por electroforese em sacos de diálise e concentrou-se por precipitação com etanol.

### Legendas das Figuras

Fig. 1: representa a sequência de nucleótidos das sondas P1, P2, P3, P6 e P7, a sequência do mRNA complementar e os péptidos proUK codificados pelas sondas.

Fig. 2: Avaliou-se a dimensão e localização dos produtos de clivagem da endonuclease de restrição por electroforese e fez-se a confirmação por análise sequencial de ADN. A região a negro indica a sequência de codificação da proteína proUK madura, a região tracejada representa a sequência de codificação peptídica "pre-pro" e as regiões a branco indicam as sequências não traduzidas das extremidades 5' e 3'. A extremidade 5' do mRNA é a da esquerda. As linhas sob o mapa de restrição indicam a contribuição dos dois clones parciais  $\lambda$  Ucl7 e  $\lambda$  Uc6.

Fig. 3: O fragmento EcoRI-SmaI que suporta o clone cADN foi inserido no pUN121 substituindo a maior parte do gene cI. O plasmídeo pcUK176 é também resistente à tetraciclina e à ampicilina. \*cI representa uma proteína cI desactivada.

Fig. 4: Representa a sequência de cADN completa do clone pcUK176 com a correspondente sequência de aminoácidos traduzida. Os sítios de restrição que se utilizou para as construções plasmídicas encontram-se sublinhados. Também se encontram sublinhados dois sítios de poliadenilação nas posições 2264 e 2277 e o resíduo da Serina na posição 1 da sequência de proUK madura.

Fig. 5: Esta figura representa quatro construções intermédias. O plasmídeo de partida, o pDS20, suporta os antecedentes gerais, os quais, através de diferentes plasmídeos intermédios, perde o promotor Pgal, o gene galK e o gene da  $\beta$ -lactamase.

Fig. 6: Representa três construções intermédias adicionais e o plasmídeo pFC16, o qual faz a expressão de alto nível de proUK. Para pormenores vide o texto.

Fig. 7: No pFC44 a sequência que codifica a proUK madura está sob o controlo do promotor Ptrp e do "sítio de ligação do ribossoma" do fago MS-2. O gene da tetraciclina foi inserido no lugar do gene da  $\beta$ -lactamase. Em consequência, o pFC44 é sensível à ampicilina.

Fig. 8: Fez-se a preparação das amostras e a sua análise conforme descrito no texto, procedendo-se à carga de um gel de poliacrilamida-SDS a 12,5% (relação de 40:1 entre a acrilamida e a bis-acrilamida). As pistas ("lanes") 1 e 2 contêm material derivado de duas culturas da estirpe NCTC 10537 transformada com pFC44. A posição da proteína pró-uroquinase recombinante está indicada pela seta. A pista 3 apresenta material derivado da estirpe hospedeira de controlo NCTC 10537. Na pista 4 apresenta-se um padrão de peso molecular.

#### DISCUSSÃO E CONCLUSÕES

A presente invenção descreve um método de ADN recombinante para a produção de pró-uroquinase não glicosilada. O método baseia-se na inserção do gene humano que codifica a proUK em estirpes bacterianas de E.coli e na cultura subsequente das referidas estirpes transformadas.

A produção de proteínas heterólogas em E.coli constitui um domínio bem estudado da moderna biotecnologia (Harris T.J.R. e Emtage J.S., Microbiological Sciences, 3, 1986, pp. 28-31).

Actualmente, os biólogos moleculares dispõem de diversos sinais de expressão tais como promotores, sequências de Shine-Dalgarno, terminadores, etc., que podem ser utilizados para a preparação da proteína pretendida. O promotor é responsável pela síntese do mensageiro ARN ao passo que a sequência de Shine-Dalgarno pode garantir uma tradução eficaz do mRNA na cadeia polipeptídica.

Contudo a combinação destes parâmetros constitui um aspecto importante na expressão de genes heterólogos. Por exemplo, a fusão de uma sequência de Shine-Dalgarno eficaz em diferentes regiões do promotor pode proporcionar diferentes níveis de expressão. Além disso, o comprimento dos fragmentos de restrição que suportam os sinais de expressão afecta frequentemente os níveis de produção (McCarthy J.E.G., Sebald W., Gross G. e Lammers R., Gene, 41, 1986, p. 201).

A escolha da estirpe hospedeira constitui também um passo crítico no desenvolvimento de um método eficiente de produção. Na realidade, sabe-se que a inserção do mesmo plasmídeo de expressão em diferentes estirpes pode originar eficiências de expressão bastante diferentes (Harris T.J.R. e Emtage J.S., Microbiological Sciences, 3, 1986, pp. 28-31).

Apesar de os sinais de expressão descritos na presente invenção serem já conhecidos da literatura científica, a sua combinação nunca antes havia sido explorada para a expressão específica da pró-uroquinase humana. Mais particularmente, os plasmídeos pFC16 e pFC44 suportam o gene proUK sob o controlo do promotor Ptrp de E.coli e da sequência MS-2 do fago de Shine-Dalgarno.

Em consequência, o método de produção descrito no âm-



bito da presente invenção baseia-se em plasmídeos de expressão essencialmente diferentes de outros plasmídeos de expressão anteriormente descritos. Estes plasmídeos, pFC16 e pFC44, representam consequentemente um dos novos aspectos da presente invenção e, conforme já referido, constituem um dos seus objectivos.

Além disso, o método agora descrito retira vantagem das estirpes de E.coli do tipo B. A maior parte dos métodos de expressão descritos na literatura científica baseia-se nas estirpes de E.coli do tipo K-12. Assim, a produção de proUK em estirpes de E.coli do tipo B representa outro aspecto novo da presente invenção.

Este segundo aspecto é extremamente importante. Na realidade, a escolha do organismo hospedeiro pode afectar o processo de produção global em diversos passos.

Por exemplo, as fermentações em grandes biomassas pode ser dramaticamente influenciada pelo tipo de hospedeiro. Os inventores da presente invenção e bem assim outros grupos de investigadores descobriram consistentemente que as estirpes de E.coli do tipo B podem crescer mais facilmente do que as estirpes do tipo K-12, por exemplo. A inserção dos mesmos plasmídeos de expressão, pFC16 ou pFC44, em estirpes K-12 tais como as do tipo C600 gera estirpes recombinantes, as quais não podem desenvolver-se, em fermentadores tão eficientemente quanto as estirpes do tipo B recombinantes. Por outras palavras, as produções de pró-UK não glicosilada recombinante são maiores a partir das estirpes B, quando se utilizam os mesmos plasmídeos de expressão.

Outro aspecto importante relacionado com a escolha da

estirpe hospedeira reside na natureza diferente dos contaminantes bacterianos durante o processo de produção de pró-UK.

Os contaminantes indesejáveis, tais como as proteases, podem afectar gravemente as produções de produto recombinante. É interessante referir que em 1986, Winkler e Blaber (Winkler, M.E., Blaber, M., Biochemistry, 25, n. 14, 1986, p.4041) descreveram um processo de produção de pró-UK baseado na estirpe 294 de K-12 (ATCC 31446).

Neste processo, os inventores tiveram de tomar diversas precauções para evitar a digestão proteolítica da pró-UK. De acordo com os inventores esta actividade proteolítica ficou a dever-se às proteases bacterianas da estirpe hospedeira. Em contraste, a utilização de estirpes do tipo B, de acordo com a presente invenção, proporciona extractos celulares com actividade proteolítica muito menor. Em particular, descobriu-se que a pró-UK extraída da estirpe C600 de K-12 é contaminada pela uroquinase com uma intensidade maior do que a pró-UK derivada das estirpes do tipo B.

Em conclusão, os inventores da presente invenção admitem que os rendimentos superiores observados na produção de pró-UK recombinante com a utilização do procedimento agora descrito, por comparação com o que sucedia com a técnica anterior, representam um resultado imprevisível e um aperfeiçoamento em relação aos processos conhecidos.

R E I V I N D I C A Ç Õ E S

1.- Processo para a preparação de pró-UK não glicosilada, ca racterizado pelo facto de se expressar a pró-UK não glicosilada sob o controlo do promotor de E. coli Ptrp e da sequência de Shine-Dalgarno MS-2 por E. coli B.

2.- Processo de acordo com a reivindicação 1, caracterizado pelo facto de a pró-uroquinase de cadeia simples não glicosilada ter um peso molecular de 45 000 daltons aproximadamente.

3.- Processo de acordo com a reivindicação 1, caracterizado pelo facto de o E. coli B exprimir fundamentalmente a sequência de pró-UK.

4.- Processo de acordo com a reivindicação 1, caracterizado

pelo facto de se obter a sequência cADN para pró-UK a partir de mARN das células de carcinoma epidermóide HEP-3.

5.- Processo de acordo com a reivindicação 1, caracterizado pelo facto de o promotor Ptrp ser constituído por um fragmento de restrição EcoRI-SalI obtido a partir do plasmídeo pDR-720.

6.- Processo de acordo com a reivindicação 1, caracterizado pelo facto de a sequência que incorpora a sequência de Shine-Dalgarno MS-2, o codão de partida ATG e o início do gene pró-UK, flanqueado a montante para um sítio HindIII e a jusante por um sítio TaqI ser como segue:

HindIII

5'-AGCTTTAATAGACGCCGCCATTCAAACATGAGGATTACCCATGAGC

3'-AATTATCTGCGGCCGGTAAGTTTGTACTCCTAATGGGTACTCG

TaqI

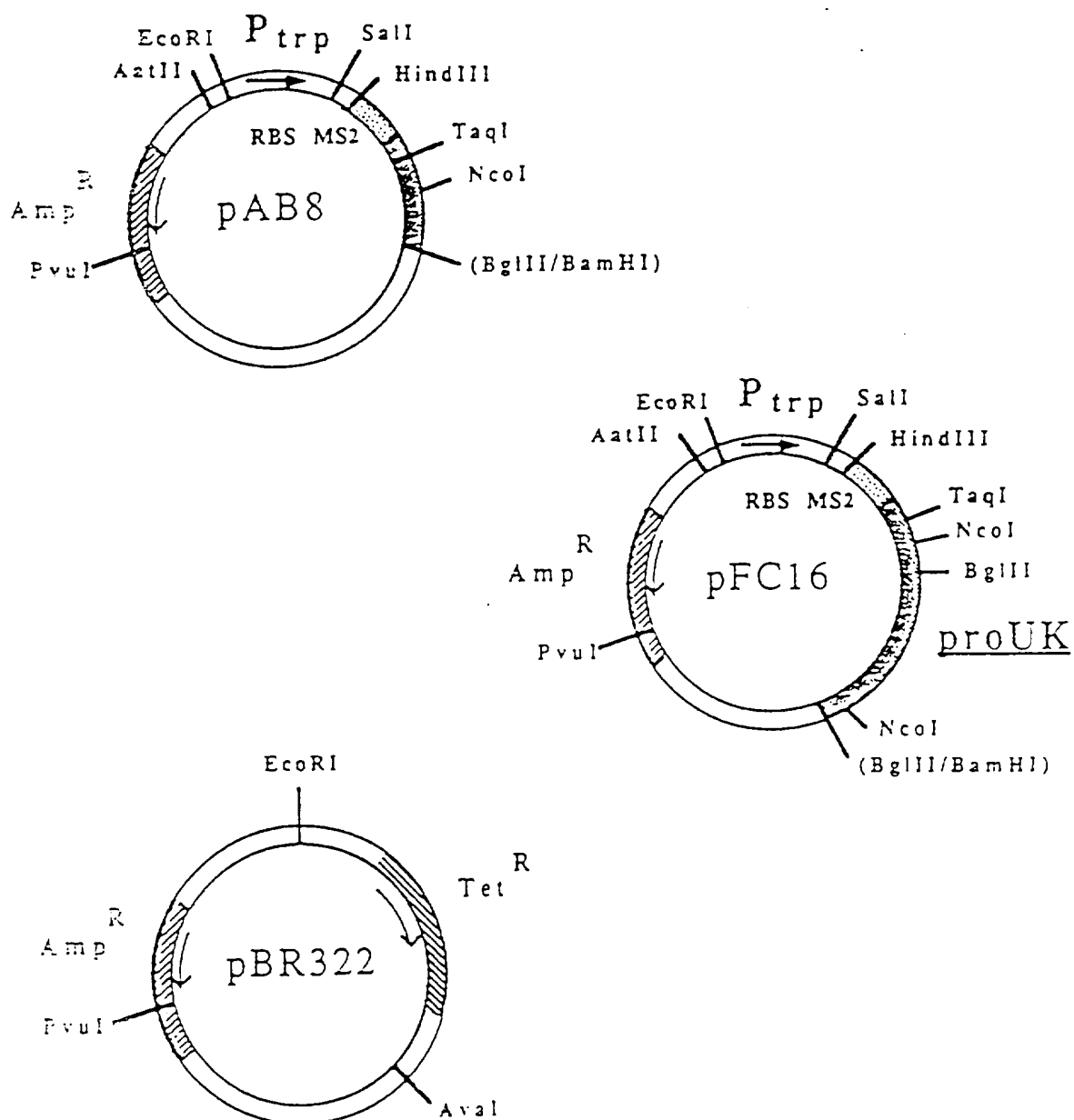
AATGAACTTCATCAAGTTCCAT-3'

TTACTTGAAGTAGTTCAAGGTAGC-5'

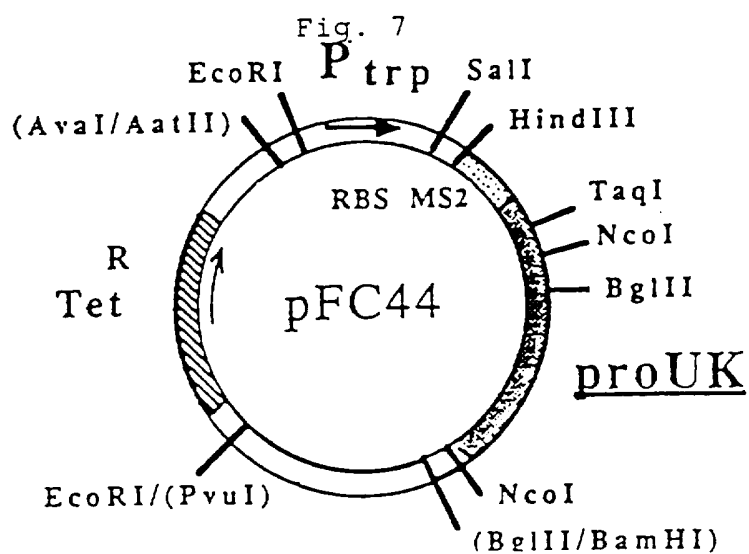
7.- Processo de acordo com a reivindicação 1 caracterizado pelo facto de o plasmídeo FC-16 de expressão ser como segue:

4.

FIG. 6



8.- Processo de acordo com a reivindicação 1, caracterizado pelo facto de o plasmídeo FC-44 de expressão ser como segue:



Lisboa, 09 de Outubro de 1989

O Agente Oficial da Propriedade Industrial

*[Handwritten signature]*

R E S U M O

"PROCESSO PARA A PREPARAÇÃO DE PRÓ-UROQUINASE HUMANA"

A presente invenção refere-se a um processo para preparar pró-uroquinase de cadeia simples não glicosilada (pró-UK). O processo consiste em cultivar estirpes bacterianas de E. coli que foram transformadas com plasmídeos com a sequência de cADN que codifica pró-UK.

Lisboa, 09 de Outubro de 1989

O Agente Oficial da Propriedade Industrial

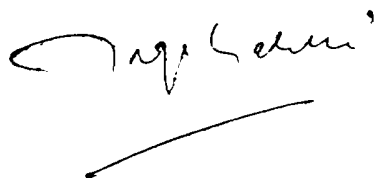


FIG. 1

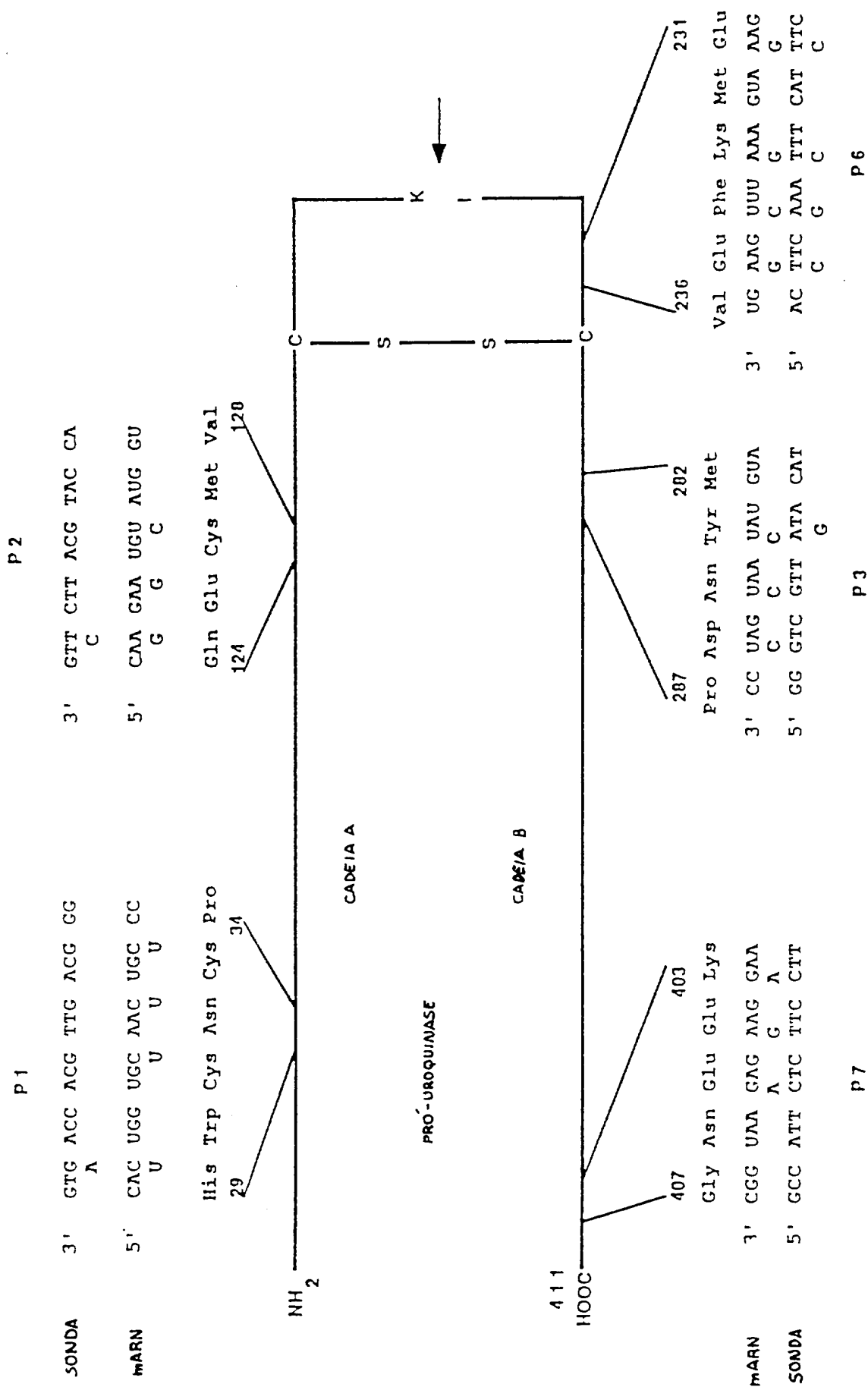


FIG. 2

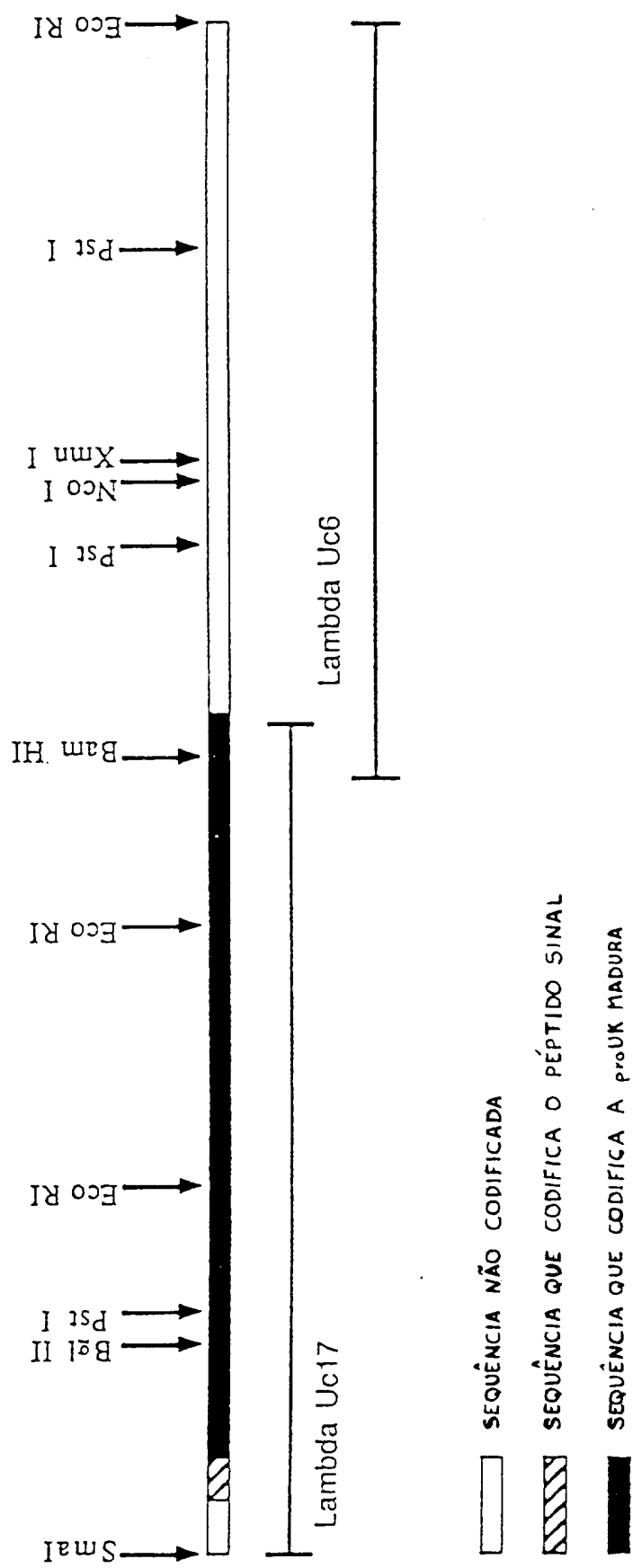


FIG. 3

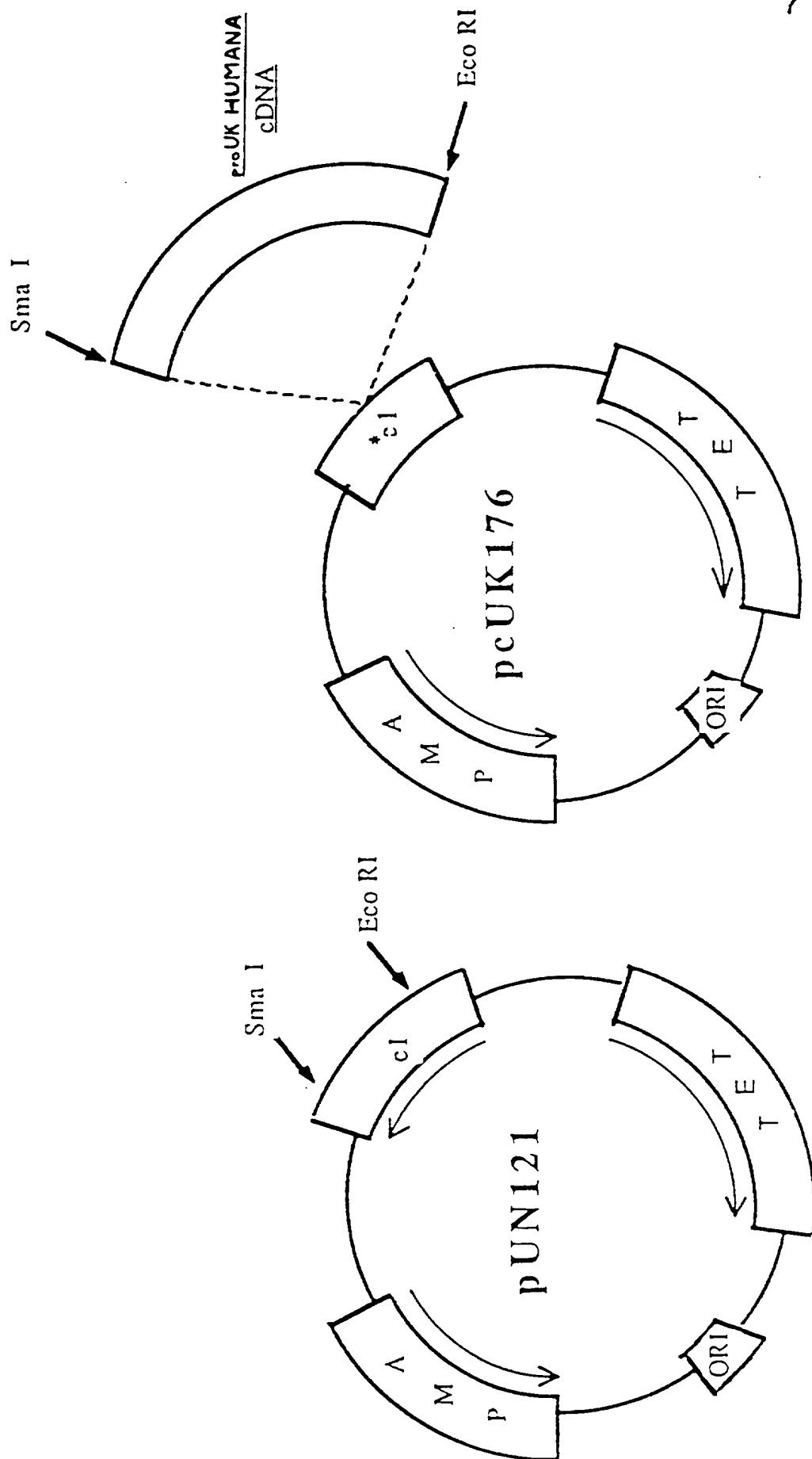


FIG. 4

1SmaI  
 CCCGGGCTCCGGGCTGCGGTCTCCTGCCGACGCCACCGAGCCGCGCTCTAGCGCCCCGACCTC

50

100  
 GCCACC ATG AGA GCC CTG CTG GCG CGC CTG CTT CTC TGC GTC CTG GTC  
 Met Arg Ala Leu Leu Ala Arg Leu Leu Leu Cys Val Leu Val

150 TaqI  
 GTG AGC GAC TCC AAA GGC AGC AAT GAA CTT CAT CAA GTT CCA TCG AAC  
 Val Ser Asp Ser Lys Gly Ser Asn Glu Leu His Gln Val Pro Ser Asn

200  
 TGT GAC TGT CTA AAT GGA GGA ACA TGT GTG TCC AAC AAG TAC TTC TCC  
 Cys Asp Cys Leu Asn Gly Gly Thr Cys Val Ser Asn Lys Tyr Phe Ser

250  
 AAC ATT CAC TGG TGC AAC TGC CCA AAG AAA TTC GGA GGG CAG CAC TGT  
 Asn Ile His Trp Cys Asn Cys Pro Lys Lys Phe Gly Gly Gln His Cys

300  
 GAA ATA GAT AAG TCA AAA ACC TGC TAT GAG GGG AAT GGT CAC TTT TAC  
 Glu Ile Asp Lys Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr

NcoI 350  
 CGA GGA AAG GCC AGC ACT GAC ACC ATG GGC CGG CCC TGC CTG CCC TGG  
 Arg Gly Lys Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp

BglII  
 AAC TCT GCC ACT GTC CTT CAG CAA ACG TAC CAT GCC CAC AGA TCT GAT  
 Asn Ser Ala Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp

400  
 GCT CTT CAG CTG GGC CTG GGG AAA CAT AAT TAC TGC AGG AAC CCA GAC  
 Ala Leu Gln Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp

450  
 AAC CGG AGG CGA CCC TGG TGC TAT GTG CAG GTG GGC CTA AAG CCG CTT  
 Asn Arg Arg Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu

500  
 GTC CAA GAG TGC ATG GTG CAT GAC TGC GCA GAT GGA AAA AAG CCC TCC  
 Val Gln Glu Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser

550  
 TCT CCT CCA GAA GAA TTA AAA TTT CAG TGT GGC CAA AAG ACT CTG AGG  
 Ser Pro Pro Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg

600  
 CCC CGC TTT AAG ATT ATT GGG GGA GAA TTC ACC ACC ATC GAG AAC CAG  
 Pro Arg Phe Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln

650  
 CCC TGG TTT GCG GCC ATC TAC AGG AGG CAC CGG GGG GGC TCT GTC ACC  
 Pro Trp Phe Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr

FIG. 4 (cont'd)

700  
TAC GTG TGT GGA GGC AGC CTC ATC AGC CCT TGC TGG GTG ATC AGC GCC  
Tyr Val Cys Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala

750  
ACA CAC TGC TTC ATT GAT TAC CCA AAG AAG GAG GAC TAC ATC GTC TAC  
Thr His Cys Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr

800  
CTG GGT CGC TCA AGG CTT AAC TCC AAC ACG CAA GGG GAG ATG AAG TTT  
Leu Gly Arg Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe

850  
GAG GTG GAA AAC CTC ATC CTA CAC AAG GAC TAC AGC GCT GAC ACG CTT  
Glu Val Glu Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu

900  
GCT CAC CAC AAC GAC ATT GCC TTG CTG AAG ATC CGT TCC AAG GAG GGC  
Ala His His Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly

950  
AGG TGT GCG CAG CCA TCC CGG ACT ATA CAG ACC ATC TGC CTG CCC TCG  
Arg Cys Ala Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser

1000  
ATG TAT AAC GAT CCC CAG TTT GGC ACA AGC TGT GAG ATC ACT GGC TTT  
Met Tyr Asn Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe

1050  
GGA AAA GAG AAT TCT ACC GAC TAT CTC TAT CCG GAG CAG CTG AAA ATG  
Gly Lys Glu Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met

1100  
ACT GTT GTG AAG CTG ATT TCC CAC CGG GAG TGT CAG CAG CCC CAC TAC  
Thr Val Val Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr

1150  
TAC GGC TCT GAA GTC ACC ACC AAA ATG CTA TGT GCT GCT GAC CCA CAA  
Tyr Gly Ser Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln

1200  
TGG AAA ACA GAT TCC TGC CAG GGA GAC TCA GGG GGA CCC CTC GTC TGT  
Trp Lys Thr Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys

1250  
TCC CTC CAA GGC CGC ATG ACT TTG ACT GGA ATT GTG AGC TGG GGC CGT  
Ser Leu Gln Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg

1300  
GGA TGT GCC CTG AAG GAC AAG CCA GGC GTC TAC ACG AGA GTC TCA CAC  
Gly Cys Ala Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His

BamHI 1350  
TTC TTA CCC TGG ATC CGC AGT CAC ACC AAG GAA GAG AAT GGC CTG GCC  
Phe Leu Pro Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala

FIG. 4 (cont'd)

1400

CTC TGA GGGTCCCCAGGGAGGAAACGGGCACCACCCGCTTTCTTGCTGGTTGTCATTTTT  
Leu end

1450

GCAGTAGAGTCATCTCCATCAGCTGTAAGAAGAGACTGGGAAGATAGGCTCTGCACAGATGGA

1500

TTTGCTGTGCCACCCACCAGGGTGAACGACAATAGCTTTACCCTCAGGCATAGGCCTGGGTG

1550

CTGGCTGCCCAGACCCCTCTGGCCAGGATGGAGGGGTGGTCCTGACTCAACATGTTACTGACC

1600

1650

AGCAACTTGTCTTTTTCTGGACTGAAGCCTGCAGGAGTTAAAAAGGGCAGGGCATCTCCTGTG

1700

CATGGGTGAAGGGAGAGCCAGCTCCCCCGACGGTGGGCATTTGTGAGGCCCATGGTTGAGAAA

NcoI

1750

TGAATAATTTCCCAATTAGGAAGTGTAACAGCTGAGGTCTCTTGAGGGAGCTTAGCCAATGTG

1800

GGAGCAGCGGTTTGGGGGAGCAGAGACACTAACGACTTCAGGGCAGGGCTCTGATATTCCATG

1850

1900

AATGTATCAGGAAATATATATGTGTGTGTATGTTGCACACTTGTGTGTGGGCTGTGAGTGTA

1950

AGTGTGAGTAAGAGCTGGTGTCTGATTGTTAAGTCTAAATATTTTCCTTAACTGTGTGGACTG

2000

TGATGCCACACAGAGTGGTCTTTCTGGAGAGGTTATAGGTCACTCCTGGGGCCTCTTGGGTCC

2050

2100

CCCACGTGACAGTGCCTGGGAATGTATTATTCTGCAGCATGACCTGTGACCAGCACTGTCTCA

2150

GTTTCACTTTCACATAGATGTCCCTTTCTTGGCCAGTTATCCCTTCCTTTTAGCCTAGTTCAT

2200

CCAATCCTCACTGGGTGGGGTGAGGACCACTCCTTACACTGAATATTTATATTTCACTATTTT

2250

TATTTATATTTTTGTAATTTTAAATAAAAGTGATCAATAAAATGTGATTTTTCTG (A)<sub>n</sub>GAATT

EcoRI

FIG. 5

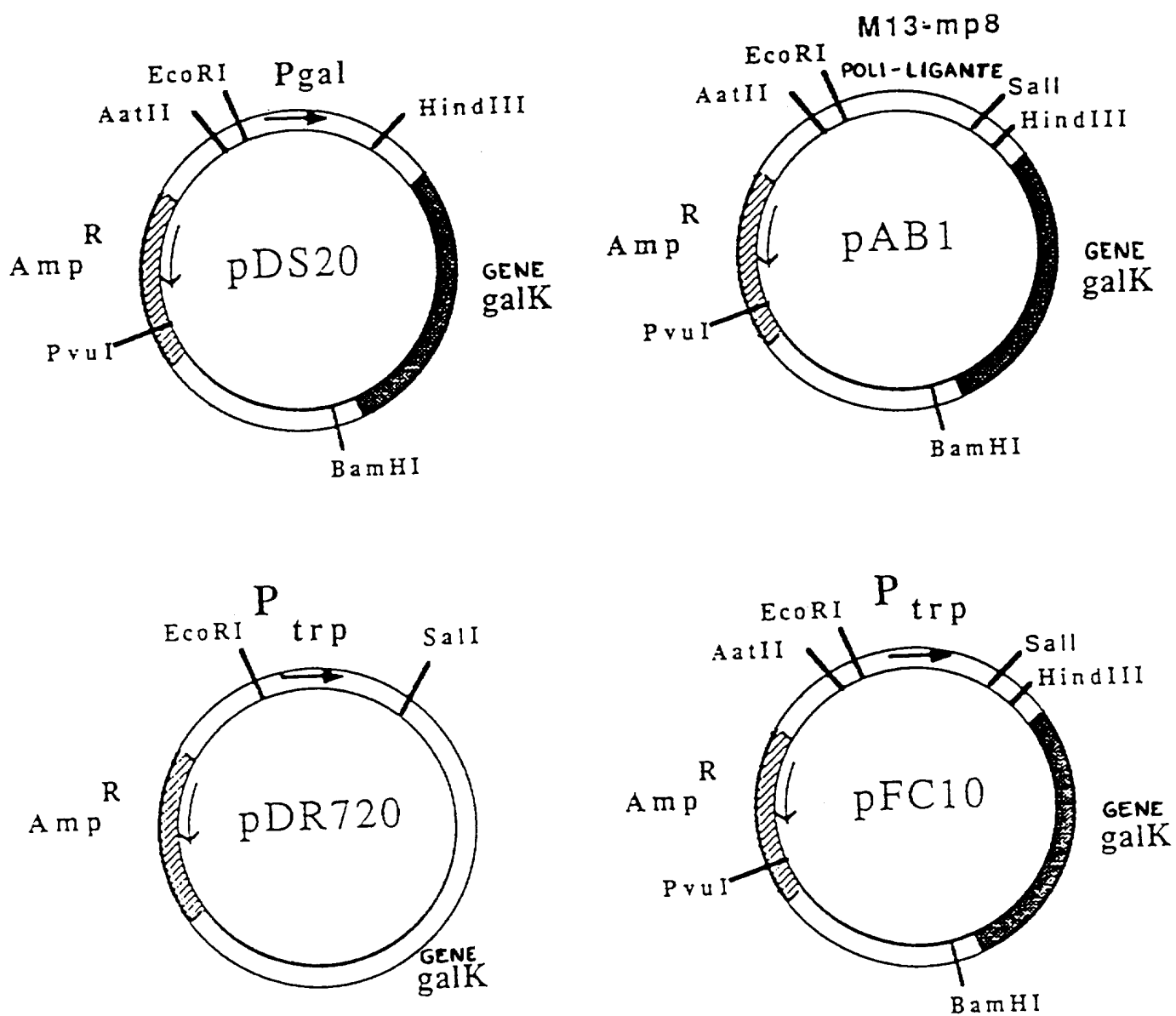


FIG. 6

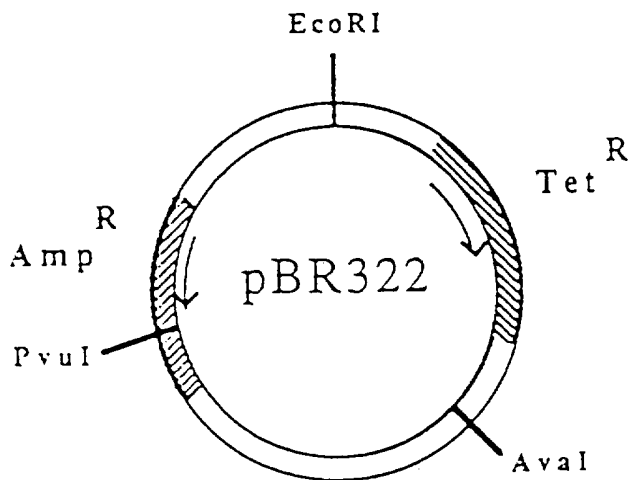
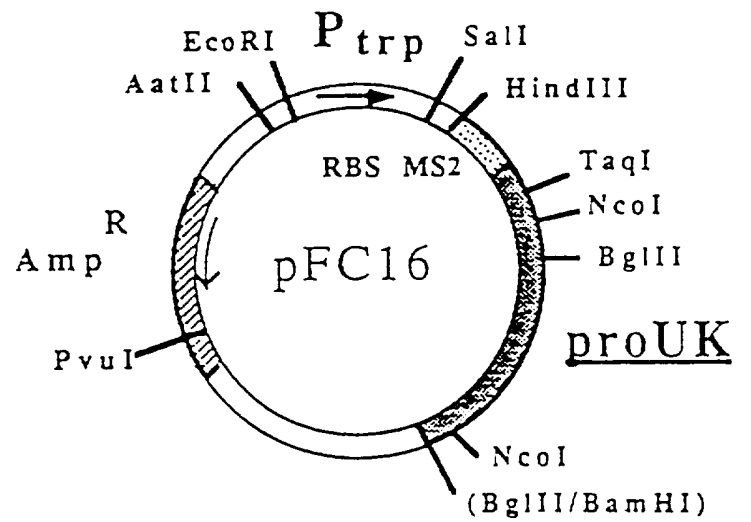
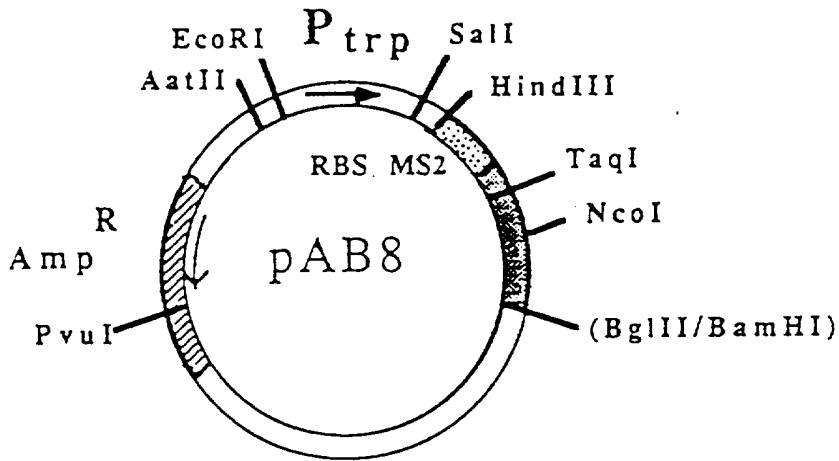
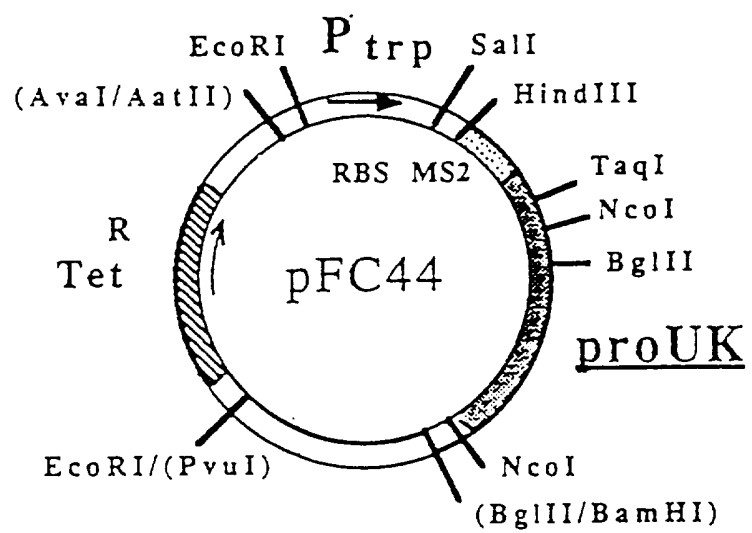


FIG. 7



4.

FIG. 8

