

DESCRIÇÃO

“Variantes de uroquinase bifuncionais com propriedades fibrinolíticas melhoradas e actividade inibidora de trombina”

O presente invento refere-se a variantes de uroquinase bifuncionais com propriedades fibrinolíticas melhoradas e actividade inibidora de trombina, a plasmídeos para a obtenção destes polipéptidos, assim como a agentes trombolíticos contendo, como substância activa, uma variante de uroquinase bifuncional.

Uma característica importante do sangue humano é a sua capacidade de fechar lesões dos vasos sanguíneos através da formação de trombos. A coagulação do sangue é provocada por uma série de enzimas presentes no sangue as quais são responsáveis, em última análise, na chamada "cascata de coagulação", pela transformação proteolítica, da proteína precursora, fibrinogénio, em fibrina, pela enzima trombina. A fibrina polimeriza-se incluindo trombócitos, eritrócitos e outros componentes do sangue no local da lesão, formando um trombo.

Além disso, o sangue contém também uma série de enzimas que têm um efeito contrário à coagulação e asseguram o fluxo de sangue depois da regeneração das paredes dos vasos. A enzima mais importante para a trombólise é a plasmina, que ataca o tecido de fibrina proteoliticamente e provoca, desta maneira, a dissolução dos trombos. A plasmina é formada através de cisão proteolítica da proteína precursora inactiva, o plasminogénio. Esta activação é provocada por activadores de plasminogénio, através de cisão proteolítica do plasminogénio. Conhecem-se dois activadores de plasminogénio endógenos humanos: a uroquinase, um activador de plasminogénio existente na urina, e o activador de plasminogénio tissular.

O enfarte cardíaco e o enfarte cerebral estão estreitamente ligados à formação patológica de trombos. Em ambas as formas de enfarte formam-se, na maior parte das vezes em consequência de alterações arterioscleróticas das artérias, sob determinadas condições, trombos nas paredes dos vasos. Estes trombos estorvam o fluxo sanguíneo nas artérias, sendo a consequência disso

que o tecido já não pode ser alimentado suficientemente com oxigénio. Isto provoca, no caso do enfarte cardíaco, a necrose parcial ou total do músculo cardíaco. De modo análogo o bloqueio de artérias cerebrais provoca lesões graves no tecido cerebral.

Para a terapia de pacientes de enfarte são utilizados activadores de plasminogénio como agentes trombolíticos que dão início à dissolução dos trombos através da plasmina. Actualmente dispõe-se, para estas terapias, de estreptoquinase, APSAC (*Anisolated Plasminogen Streptokinase Activator Complex*), uroquinase de duas cadeias (UK), uroquinase recombinante de uma só cadeia (pro-uroquinase recombinante) e do activador de plasminogénio tissular (tPA) (Collen e Lijnen, *Blood* 78, 3114-3124 (1991)). A estreptoquinase é uma proteína de estreptococos hemolíticos. A estreptoquinase activa o plasminogénio, formando um complexo com o plasminogénio e transformando assim o plasminogénio numa estrutura activa. Este próprio complexo transforma plasminogénio livre em plasmina a qual, por sua vez, cinde o plasminogénio complexado com a estreptoquinase. Um aperfeiçoamento da estreptoquinase é o APSAC, um complexo fabricado *in vitro*, de estreptoquinase e plasminogénio humano. Devido a uma modificação química do local activo do plasminogénio, o APSAC possui uma semi-vida biológica superior à da estreptoquinase.

A uroquinase é uma proteína humana que pode ser obtida como proteína com actividade proteolítica, a partir da urina, sob duas formas diferentes: a uroquinase de elevado peso molecular (HUK) e a uroquinase de baixo peso molecular (LUK) (Stump *et al.*, *J. Biol. Chem.* 261, 1267-1273 (1986)). HUK e LUK são moléculas de duas cadeias. A uroquinase é formada sob a forma de uroquinase de uma só cadeia (pro-uroquinase) em tecidos diferentes e pode ser detectada, sob a forma proenzima, em quantidades pequenas no sangue humano (Wun *et al.*, *J. Biol. Chem.* 257, 3276-3283 (1982)). A forma activada da pro-uroquinase tem, sob a forma de HUK, um peso molecular de 54 quilodalton e consiste em 3 domínios: o domínio amino-terminal de *growth factor*, o "Kringle" e um domínio de serina-protease (Günzler *et al.*, *Hoppe-Seyler's Z. Physiol. Chem.* 363, 1155-1165 (1982); Steffens *et al.*, *Hoppe-Seyler's Z. Physiol. Chem.* 363, 1043-1058 (1982)). Embora a pro-uroquinase e o plasminogénio estejam presentes sob a forma de proenzimas, a pro-uroquinase consegue, devido a uma actividade intrínseca, transformar

plasminogénio em plasmina activa. Este activador de plasminogénio, no entanto, apenas fica com a sua actividade completa depois da plasmina formada, por sua vez, ter cindido a pro-uroquinase entre ¹⁵⁸lisina e ¹⁵⁹isoleucina (Lijnen *et al.*, J. Biol. Chem. 261, 1253-1258 (1986)). A obtenção de uroquinase através de métodos da tecnologia genética, em *Escherichia coli* foi descrita, pela primeira vez, por Heyneker (Proceedings of the IVth International Symposium on Genetics of Industrial Microorganisms 1982). Pro-uroquinase não glicosilada (*Saruplase*) é fabricada utilizando um gene sintético (Brigelius-Flohé *et al.*, Appl. Microbiol. Biotech. 36, 640-649 (1992)).

O activador de plasminogénio tissular é uma proteína que existe no sangue e nos tecidos, com um peso molecular de 72 quilodalton. Este activador de plasminogénio consiste em 5 domínios: o domínio amino-terminal "*finger*", o domínio de *growth factor*, o "*Kringle*" 1, o "*Kringle*" 2 e um domínio de serina-protease. Ao contrário da pro-uroquinase, o tPA é capaz de cindir o plasminogénio só depois de ligação à fibrina. Como a pro-uroquinase, o tPA é transformado na forma activa através de uma cisão catalisada por plasmina, entre o "*Kringle*" 2 e o domínio de serina-protease. O activador de plasminogénio tissular liga-se, neste caso, à fibrina mas não ao fibrinogénio, sendo desta maneira o plasminogénio activado de forma específica do trombo. Ao contrário da uroquinase de duas cadeias, uma activação geral de plasminogénio é amplamente evitada (Collen e Lijnen, Blood 78, 3114-3124 (1991)).

Desde o início dos anos 80, o tratamento activo do enfarte do miocárdio com medicamentos trombolíticos tem-se mostrado eficaz e eficiente. Numa série de estudos foi demonstrado que o tratamento de pacientes de enfarte cardíaco com estreptoquinase, APSAC, UK, pro-uroquinase recombinante ou tPA, reduz significativamente a mortalidade em comparação com pacientes não tratados. Para melhorar a eficácia destas substâncias na terapia, foram fabricados, utilizando métodos de tecnologia genética, uma série de derivados do activador de plasminogénio tissular e da pro-uroquinase. No centro do interesse está, além de um aumento da actividade fibrinolítica e da redução dos efeitos secundários, o desenvolvimento de formas que sejam apropriadas para aplicação de *bolus*. Um resumo das tentativas de melhoramento dos activadores



de plasminogénio encontra-se em *Thrombosis and Haemostasis* 66, 88-110 (1991), assim como em *Trends in Biotech.* 9, 86-90 (1991).

Para melhorar a eficácia dos activadores de plasminogénio na terapia de lise e, especialmente, para aumentar a sua semi-vida biológica, foram fabricadas variantes, por deleção ou substituição, do activador de plasminogénio tissular, no quais foram removidos, por exemplo, o domínio "finger" e o domínio de *growth factor* ou em que o domínio de serina-protease foi substituído pelo domínio de serina-protease da uroquinase (Collen *et al.*, *Thromb. Haemostasis* 65, 174-180 (1991); Fromage *et al.*, *Fibrinolysis* 5, 187-190 (1991); Lu *et al.*, *Blood* 78, 125-131 (1991)). Mostrou-se, efectivamente, que a deleção dos domínios "finger" e de *growth factor* aumentava a semi-vida biológica das variantes de tPA (Lijnen e Collen, *Thromb. Haemostasis* 66, 94-95 (1991)). Uma variante obtida por deleção e substituição consistindo em ambos os domínios "Kringle" do tPA e no domínio da serina-protease da UK, era, na trombólise, superior aos activadores de plasminogénio originais, isto é não modificados, devido a uma semi-vida nitidamente superior. Estas variantes de activadores de plasminogénio, no entanto, possuíam apenas uma reduzida especificidade para fibrina (Lu *et al.*, *Blood* 78, 125-131 (1991)).

Foram realizadas várias tentativas para fabricar activadores de plasminogénio com maior especificidade para a fibrina. Para reduzir o risco de hemorragias, as substâncias activas deste género deveriam activar, se possível, plasminogénio exclusivamente na proximidade do trombo, sem provocarem uma activação sistémica do plasminogénio. Assim, por exemplo, conhece-se uma variante de tPA em que o "Kringle" 1 foi substituído pelo "Kringle" 2 da molécula original. É certo que esta variante possui uma maior afinidade para os resíduos lisina N-terminais, mas não para a fibrina. Relativamente à trombólise, esta variante não foi mais eficaz em modelos animais do que o activador de plasminogénio tissular original (Collen *et al.*, *Thromb. Haemostasis* 65, 174-180 (1991)). Outras variantes conhecidas, variantes essas que consistem em fusões entre um anticorpo específico para trombos e um activador de plasminogénio, são mais eficazes em modelos animais do que os activadores de plasminogénio originais (Lijnen e Collen, *Thromb. Haemostasis* 66, 88-110 (1991)). Um activador de plasminogénio que foi isolado do morcego *Desmodus rotundus* possui uma especificidade muito elevada para a fibrina (Gardell *et al.*, *J. Biol.*



Chem. 264, 17947-17952 (1989)). Este activador de plasminogénio apresenta, no modelo animal, uma trombólise melhorada em relação ao tPA, com uma semi-vida maior e uma activação sistémica de plasminogénio reduzida (Gardell *et al.*, *Circulation* 84, 244 a 253 (1991); Mellott *et al.*, *Arterioscl. Thromb.* 12, 212 a 221 (1992)).

O êxito de um tratamento de pacientes de enfarte com activadores de plasminogénio, no entanto, depende não só da trombólise como também do grau com que se consegue impedir uma nova obstrução dos vasos sanguíneos que foram desentupidos. Vários diagnósticos indicam que, na trombólise, a trombina ligada nos trombos é outra vez libertada sob a forma de enzima activa e pode provocar novamente a obstrução de vasos (Szczeklik *et al.*, *Arterioscl. Thromb.* 12, 548 a 553 (1992); Eisenberg, *Circulation* 84, 2601-2603 (1991)). Com efeito, o efeito de agentes trombolíticos é melhorado através da administração simultânea ou anterior do inibidor de trombina heparina. Também através da administração de *Argatroban*, *Hirugen* ou proteína C é possível reduzir a recorrência de obstruções na terapia de lise (Schneider, *Thromb. Res.* 64, 677-689 (1990); Yao *et al.*, *Am. Physiol.* 262 (Heart Circ. Physiol. 31, H 374-H 379 (1992); Gruber *et al.*, *Circulation* 84, 2454-2462 (1991)). Além disso é sabido que a mortalidade de pacientes de enfarte é reduzida significativamente através da administração prévia de heparina e aplicação subsequente de pro-uroquinase, em relação a um grupo de controlo (aplicação de pro-uroquinase sem administração prévia de heparina) (Tebbe *et al.*, *Z. Kardiol.* 80, Suppl 3, 32 (1991)).

Um dos inibidores de trombina mais potentes é a hirudina, isolada da sanguessuga *Hirudo medicinalis*, a qual se liga com a sua parte carboxi-terminal especificamente ao local chamado local de ligação de aniões da trombina. Determinados aminoácidos da parte amino-terminal da molécula de hirudina bloqueiam o acesso à bolsa de ligação ao substrato da trombina (Rydel *et al.*, *Science* 249, 277-280 (1990)). Além disso sabe-se que a trombina pode também ser inibida por derivados mais pequenos da hirudina, devendo sublinhar-se especialmente as moléculas *Hirulog* (Krstenansky *et al.*, *J. Med. Chem.* 30, 1688-1691 (1987); Yue *et al.*, *Protein Engineering* 5, 77-85 (1992)).

A utilização de hirudina em combinação com um activador de plasminogénio, para o tratamento de doenças vasculares provocadas por trombozes é descrita nos pedidos de patente europeia EP 328 957 e EP 365 468. Do pedido internacional de patente WO 91/01142 conhece-se a utilização terapêutica de derivados da hirudina em combinação com um agente trombolítico.

No pedido internacional de patente WO 92/18139 são descritas moléculas quiméricas com uma parte activadora de plasminogénio que não se liga à fibrina e uma parte com afinidade em relação a componentes de placas arterioscleróticas que não de fibrina, por exemplo uma parte inibidora de trombina. Ambas as partes podem estar ligadas uma à outra através de um 12-mero consistindo em alanina ou contendo principalmente alanina. Num 12-mero rico em alanina podem estar substituídos 2 ou 3 resíduos de alanina por lisina ou arginina.

Do pedido internacional de patente WO 91/09125 conhecem-se proteínas de fusão, com um local de cisão entre a primeira e a segunda sequência, que possuem uma actividade fibrinolítica e/ou antitrombótica. A actividade fibrinolítica e/ou antitrombótica completa é obtida apenas depois da cisão, podendo esta ser provocada especialmente através de trombina. Isto tem, no entanto, a consequência de o inibidor de trombina libertado através da cisão inibir a trombina, o que impede uma continuação da cisão provocada por trombina.

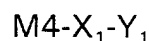
A trombina pode também ser inibida por um péptido derivado da sequência amino-terminal do receptor de trombina humano (Vu *et al.*, Nature 253, 674-677 (1991)). O receptor de trombina contém, na região amino-terminal, uma sequência que se liga à trombina, próxima de um local de cisão para trombina. A região do receptor que se liga à trombina, assemelha-se muito, na sua estrutura, ao domínio carboxi-terminal da hirudina. O receptor é activado por trombina através da cisão da sequência do receptor. Imitando a interacção que existe entre o receptor e a trombina, um fragmento do receptor com a região de ligação e um local de cisão modificado, actua como inibidor de trombina.

Da mesma maneira, a trombina pode ser inibida por um péptido derivado dos aminoácidos 41 a 57 da hemodina (Strube *et al.*, J. Biol. Chem. 268, 8590-8595 (1993)).

O problema em que o presente invento assenta consistia no desenvolvimento de substâncias activas para o tratamento de obstruções vasculares causadas por trombozes, que conseguissem num espaço de tempo muito curto uma trombólise completa e impedissem ao mesmo tempo uma nova obstrução dos vasos após uma trombólise inicialmente bem sucedida. Além disso pretende-se evitar, com estas substâncias activas, uma activação sistémica do plasminogénio.

Verificou-se agora que estas rigorosas exigências para novas substâncias activas deste género, são cumpridas por determinadas variantes de uroquinase bifuncionais.

Em conformidade, o presente invento refere-se a variantes de uroquinase bifuncionais, de fórmula geral I



em que

M4 significa a sequência de aminoácidos ⁴⁷Ser a ⁴¹¹Leu da pro-uroquinase não glicosilada, de acordo com a figura 1 (SEQ ID NO: 21 e 22),

X₁ é uma ligação directa entre M4 e Y₁ ou representa um péptido de sequência

Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe

ou

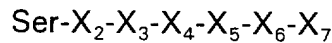
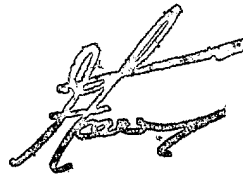
Ser-Pro-Pro-Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe

ou

Ser-Pro-Pro-Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe-Gly

ou

uma sequência peptídica de fórmula geral II



em que X_2 significa Pro ou Leu, X_3 significa Val ou Pro, X_4 significa Lys, Val, Arg, Gly ou Glu, X_5 significa Ala, Val, Gly, Leu ou Ile, X_6 significa Phe, Trp, Tyr ou Val, e X_7 significa Gly ou uma ligação directa entre X_6 e Y_1 ,

e

Y_1 representa um péptido de sequência

$Y_2\text{-Arg-Pro-}Y_3\text{-Gly-Gly-Gly-Gly-Asn-Gly-Asp-Phe-Glu-Glu-Ile-Pro-Glu-Glu-Tyr-Leu-}Y_4$

ou

$Y_2\text{-Arg-Pro-Phe-Leu-Leu-Arg-Asn-Pro-Asn-Asp-Lys-Tyr-Glu-Pro-Phe-Trp-Glu-Asp-Glu-Glu-Lys-Asn-Glu}$

ou

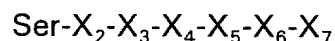
$Y_2\text{-Arg-Pro-Ser-Ser-Glu-Phe-Glu-Glu-Phe-Glu-Ile-Asp-Glu-Glu-Glu-Lys,}$

onde Y_2 é Pro ou Val, Y_3 é Leu ou uma ligação directa entre Pro e Gly, e Y_4 é Gln ou um grupo hidroxilo.

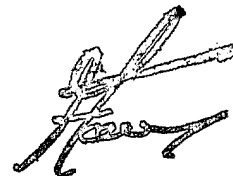
Em variantes de uroquinase bifuncionais de fórmula geral I em que Y_1 representa um péptido com a sequência

$Y_2\text{-Arg-Pro-}Y_3\text{-Gly-Gly-Gly-Gly-Asn-Gly-Asp-Phe-Glu-Glu-Ile-Pro-Glu-Glu-Tyr-Leu-}Y_4$

em que Y_2 significa Pro ou Val, Y_3 significa Leu ou uma ligação directa entre Pro e Gly, e Y_4 significa Gln ou um grupo hidroxilo, X_1 é, de preferência, uma sequência peptídica de fórmula geral II



em que X_2 significa Pro ou Leu, X_3 significa Val, X_4 significa Lys, Val ou Arg, X_5 significa Ala, Val ou Gly, X_6 significa Phe, Trp, Tyr ou Val, e X_7 significa Gly ou uma ligação directa entre X_6 e Y_1 . Nestas variantes de uroquinase bifuncionais preferem-se particularmente sequências peptídicas de fórmula geral II em que X_2 significa Pro ou Leu, X_3 significa Val, X_4 significa Lys ou Val, X_5 significa Ala ou



Val, X_6 significa Phe, Trp ou Tyr, e X_7 significa Gly ou uma ligação directa entre X_6 e Y_1 , especialmente em que X_7 significa uma ligação directa entre X_6 e Y_1 .

Em variantes de uroquinase bifuncionais de fórmula geral I, em que Y_1 representa um péptido com a sequência

Y_2 -Arg-Pro-Phe-Leu-Leu-Arg-Asn-Pro-Asn-Asp-Lys-Tyr-Glu-Pro-Phe-Trp-Glu-Asp-Glu-Glu-Lys-Asn-Glu,

Onde Y_2 é Pro ou Val, X_1 é, de preferência, uma sequência peptídica de fórmula geral II,

Ser- X_2 - X_3 - X_4 - X_5 - X_6 - X_7 ,

em que X_2 significa Pro ou Leu, X_3 significa Val, X_4 significa Lys ou Val, X_5 significa Ala ou Val, X_6 significa Phe ou Trp, e X_7 significa uma ligação directa entre X_6 e Y_1 .

Em comparação com os activadores de plasminogénio que se conhecem e com as misturas, de um activador de plasminogénio e de um inibidor de trombina, que se conhecem, as variantes de uroquinase bifuncionais de acordo com o invento distinguem-se por uma actividade fibrinolítica mais forte e, simultaneamente, boas características, não previsíveis, de inibição de trombina. Além disso, os polipéptidos de acordo com o invento consomem, inesperadamente, fibrinogénio plasmático em quantidades nitidamente mais baixas. A especificidade para fibrina significativamente mais elevada que daí resulta, especialmente também em comparação com as misturas de um activador de plasminogénio e de um inibidor de trombina que se conhecem, tem como efeito a capacidade de coagulação do sangue ser pouco influenciada e o risco de hemorragias incontroladas, como complicação potencial de uma degradação sistémica do fibrinogénio, ser minimizado. A elevada especificidade para a fibrina das variantes de uroquinase de acordo com o invento torna portanto possíveis aplicações de *bolus* com um risco de ocorrência de hemorragias nitidamente reduzido em comparação com as aplicações de *bolus* de agentes trombolíticos conhecidos.

As variantes de uroquinase bifuncionais de fórmula geral I são toxicologicamente inofensivas, podendo por isso ser administradas como tal, em preparações farmacêuticas adequadas, a pacientes com obstruções vasculares causadas por trombos.

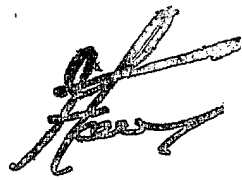
Em conformidade, um outro objectivo do invento consiste em agentes trombolíticos contendo como substância activa uma variante de uroquinase bifuncional de fórmula geral I.

Para o tratamento de obstruções vasculares causadas por trombos, por exemplo enfarte cardíaco, enfarte cerebral, obstrução aguda de artérias periféricas, embolia pulmonar e trombose das veias profundas da perna e da bacia, são necessários 0,1 a 1 mg/kg de um polipéptido de acordo com o invento. As variantes de uroquinase bifuncionais podem ser administradas de forma intravenosa e, particularmente, através de injeção de *bolus*.

Os agentes trombolíticos de acordo com o invento contêm além de, pelo menos, uma variante de uroquinase bifuncional, aditivos auxiliares, como por exemplo materiais de transportadores, solventes, diluentes, corantes e aglutinantes. A escolha dos aditivos auxiliares, assim como das quantidades a aplicar, depende da forma prevista para a aplicação do medicamento, e não constitui nenhum problema para o perito.

A produção das variantes de uroquinase bifuncionais é realizada através de métodos de tecnologia genética. Em conformidade, são também um objectivo do invento, plasmídeos para serem utilizados na obtenção de variantes de uroquinase bifuncionais de fórmula geral I contendo um operão com um promotor regulável, uma sequência de Shine-Dalgarno que actua como local de ligação ao ribossoma, um códon de iniciação, um gene estrutural sintético para uma variante de uroquinase bifuncional de fórmula geral I e, a jusante do gene estrutural, um ou dois terminadores.

Um promotor regulável particularmente apropriado é o promotor *trp* ou o promotor *tac*. Como terminador utiliza-se, de preferência, o terminador *trp A* e/ou o terminador *tet A/orf L* de Tn 10.



Na região de controlo dos plasmídeos de acordo com o invento, a distância entre a sequência de Shine-Dalgarno e o códon de iniciação é de 6-12, de preferência 8-10 nucleótidos.

A expressão dos plasmídeos de acordo com o invento é realizada em estirpes de *Escherichia coli*, especialmente em estirpes de *Escherichia coli* do grupo K12, por exemplo *E.coli* K12 JM 101 (ATCC 33876), *E.coli* K12 JM 103 (ATCC 39403), *E.coli* K12 JM 105 (DSM 4162) e *E.coli* K12 DH 1 (ATCC 33849). As variantes de uroquinase bifuncionais de fórmula geral I de acordo com o invento são obtidas na célula bacteriana com elevado rendimento, em corpos de inclusão nos quais a proteína existe sob a forma desnaturada. Após o isolamento dos corpos de inclusão, a proteína desnaturada é redobrada na estrutura terciária desejada utilizando técnicas químicas usuais para proteínas, envolvendo um sistema redox,.

Exemplos

1) Preparação, isolamento e purificação de variantes de uroquinase bifuncionais de acordo com o invento

a) Métodos de clonagem

Os plasmídeos de expressão para a produção, através de métodos de tecnologia genética, dos polipéptidos de acordo com o invento em *Escherichia coli*, foram produzidos de modo conhecido. A ordem sequencial dos passos individuais da produção está representada nas figuras 2 e 2a a 2p. Os produtos de partida para a produção dos plasmídeos foram os plasmídeos pBluescriptKSII+ (Stratagene, Heidelberg), pUC8 (Pharmacia, Freiburg) e pGR201. O pGR201 é idêntico ao plasmídeo pBF160 descrito em EP 408 945 e Appl. Microbiol. Biotechn. 36, 640-649 (1992). As endonucleases de restrição BanII, BamHI, ClaI, HindIII, NcoI, NdeI, NheI e NotI, assim como as enzimas modificadoras de ADN, como a fosfatase alcalina, a ligase de T4, a quinase de T4 e a polimerase de T4, foram compradas a Pharmacia, Stratagene, Boehringer (Mannheim) e Gibco (Eggenstein). As modificações dos plasmídeos durante a sua fabricação foram controladas através de análise de restrição e sequenciação do ADN. A sequenciação do ADN foi realizada de acordo com as

recomendações do fabricante, com um conjunto de reagentes de Pharmacia. Na produção dos plasmídeos utilizaram-se diferentes oligodesoxirribonucleótidos (oligos) cujas sequências estão indicadas, juntamente com as respectivas designações, na tabela 1 seguinte.

Tabela 1:

Designação do oligo	Sequência do oligo (de 5' para 3')
0105 SEQ ID NO:49	TATGAGCAAACCTTGCTACGAAGGTAACGGTCACTTCTA CCGTGGTAAGGCTTCTACCGACAC
0106 SEQ ID NO:50	CATGGTGTCTGGTAGAAGCCTTACCACGGTAGAAGTGACC GTTACCTTCGTAGCAAGTTTTGCTCA
0220 SEQ ID NO:51	CGGTTAAGGCTTTCCCGAGGCCTGGTGGTGGTGGTAACG GTGACTTCGAAGAAATCCCGAAGAGTACCTGTGATAGG ATCAA
0221 SEQ ID NO:52	CTAGTTGATCCTATCACAGGTA CTCTTCCGGGATTTCTT CGAAGTCACCGTTACCACCACCACCAGGCCTCGGGAAAG CCTTAACCGGGCT
0222 SEQ ID NO:53	CGCCGAGCCCGCCGAGCCCGCCGGGTGGTTTTCCCGAGGC CTGGTGGTGGTGGTAACGGTGA CTTCGAAGAAATCCCGG AAGAGTACCTGTGATAGGATCAA
0223 SEQ ID NO:54	CTAGTTGATCCTATCACAGGTA CTCTTCCGGGATTTCTT CGAAGTCACCGTTACCACCACCACCAGGCCTCGGGAAAC CACCCGGCGGGCTCGGCGGGCTCGGCGGCT
0224 SEQ ID NO:55	CGCCGGGTGGTTTTCCCGAGGCCTGGTGGTGGTGGTAACG GTGACTTCGAAGAAATCCCGAAGAGTACCTGTGATAGG ATCAA
0225 SEQ ID NO:56	CTAGTTGATCCTATCACAGGTA CTCTTCCGGGATTTCTT CGAAGTCACCGTTACCACCACCACCAGGCCTCGGGAAAC CACCCGGCGGGCT
0226 SEQ ID NO:57	CGCCGAGCCCGCCGAGCCCGCCGGGTGGTTTTCGGTCCGA GGCCTGGTGGTGGTGGTAACGGTGA CTTCGAAGAAATCC CGGAAGAGTACCTGTGATAGGATCAA

Designação do oligo	Sequência do oligo (de 5' para 3')
0227 SEQ ID NO:58	CTAGTTGATCCTATCACAGGTA CTCTCCGGGATTTCTT CGAAGTCACCGTTACCACCACCACCAGGCCTCGGACCGA AACCACCCGGCGGGCTCGGCGGGCTCGGCGGGCT
0265 SEQ ID NO:59	CACCCGGCGGAGACGGCGGGCTCAGAGCCAGACCGTTTT CTTCTTTGGTGTGAGAACG
0281 SEQ ID NO:60	CGTCCGGGTGGTGGTGGTAACGGTGACTTCGAAGAAATC CCGGAAGAATACCTGTAAG
0282 SEQ ID NO:61	GATCCGTTCTCACACCAAAGAAGAAAACGGTCTGGCTCT GAGCCCGCCGTCTCCGCCGGGTGGTTTCCCG
0283 SEQ ID NO:62	CTAGCTTACAGGTATTCTTCCGGGATTTCTTCGAAGTCA CCGTTACCACCACCACCCGGACGCGGGAAAC
0329 SEQ ID NO:63	AAGAAATCCCGGAAGAATACCTGCAATAAG
0330 SEQ ID NO:64	CGGTAAAGGCTTGGGGACCGCGGCCGCTGGGTGGTGGTG GTAACGGTGACTTCG
0331 SEQ ID NO:65	ACCACCACCCAGCGGCCGCGGTCCCCAAGCCTTAACCGG GCT
0332 SEQ ID NO:66	CTAGCTTATTGCAGGTATTCTTCCGGGATTTCTTCGAAG TCACCGTTACC
0333 SEQ ID NO:67	CGGTAAAGGCTTT6CGGACCGC
0334 SEQ ID NO:68	GGCCGCGGTCCGAAAGCCTTAACCGGGCT
0335 SEQ ID NO:69	CGGTTCGGGCTTTCGGTCCGC
0336 SEQ ID NO:70	GGCCGCGGACCGAAAGCCCGAACCGGGCT
0337 SEQ ID NO:71	CGGTAAAGGCTTACGGACCGC
0338 SEQ ID NO:72	GGCCGCGGTCCGTAAGCCTTAACCGGGCT

Designação do oligo	Sequência do oligo (de 5' para 3')
0339 SEQ ID NO:73	CGGTTGTTGCTTTTCGGTCCGC
0340 SEQ ID NO:74	GGCCGCGGACCGAAAGCAACAACCGGGCT
0341 SEQ ID NO:75	CGGTTGTTGCTTTTCGGTCCGC
0342 SEQ ID NO:76	GGCCGCGGGAAAGCCCGAACCGGGCT
0343 SEQ ID NO:77	CGGTTAAGGCTTACCCGC
0344 SEQ ID NO:78	GGCCGCGGGTAAGCCTTAACCGGGCT
0347 SEQ ID NO:79	CGGTTGTTGCTTTTCGGTCCGC
0348 SEQ ID NO:80	GGCCGCGGGAAAGCAACAACCGGGCT
0381 SEQ ID NO:81	CGGTTAAGGCTTGGCCGC
0383 SEQ ID NO:82	GGCCGCGGCCAAGCCTTAACCGGGCT
0384 SEQ ID NO:83	CGGTTAAGGCTTTTCGGTCCGC
0385 SEQ ID NO:84	GGCCGCGGGAAAGCCTTAACCGGGCT
0386 SEQ ID NO:85	CGGTTGTAGTTTTTCGGTCCGC
0387 SEQ ID NO:86	CGGTTGAAGTTTTTCGGTCCGC
0388 SEQ ID NO:87	GGCCGCACTACAACCTACAACCGGGCT

Designação do oligo	Sequência do oligo (de 5' para 3')
0389 SEQ ID NO:88	CGGTTGTAGTTGTAGTGC
0390 SEQ ID NO:89	GGCCGCGGGAAAACCTTCAACCGGGCT
0391 SEQ ID NO:90	GGCCGCGGGAAAACCTACAACCGGGCT
0392 SEQ ID NO:91	CTAGCTTATTCGTTTTTTTCTTCGTCTTCCCAGAACGGT TCGTATTTGTCGTTCCGGTTCGCAGCAGGAAC
0393 SEQ ID NO:92	GGCCGTTCTTGCTGCGGAACCCGAACGACAAATACGAAC CGTTCTGGGAAGACGAAGAAAAAACGAATAAG
0453 SEQ ID NO:93	TGGTTAAAGCTTTCCCGC
0454 SEQ ID NO:94	GGCCGCGGGAAAGCTTTAACCAGGCT
0455 SEQ ID NO:95	TGGTTGTTGCTTTCCCGC
0456 SEQ ID NO:96	GGCCGCGGGAAAGCAACAACCAGGCT
0465 SEQ ID NO:97	GGCCGCGGGAACAGAGCCAGACCGTTTTCTTCTTTGGTG TGAGAACG
0466 SEQ ID NO:98	GATCCGTTCTCACACCAAAGAAGAAAACGGTCTGGCTCT GTTCCCGC
0467 SEQ ID NO:99	CGGTAAAGCTTTCCCGCGGCCGTTCTGCTGCGGAAC
0468 SEQ ID NO:100	TTTGTGTTCCGGTTCGCAGCAGGAACGGCCGCGGGAA AGCCTTAACCGGGCT
0469 SEQ ID NO:101	CTAGCTTATTCGTTTTTTTCTTCGTCTTCCCAGAACGGT TCGTA
0470 SEQ ID NO:102	CCGAACGACAAATACGAACCGTTCTGGGAAGACGAAGAA AAAAACGAATAAG

Os oligodesoxirribonucleótidos foram fabricados sob a forma destrilada à escala de 0,1 μ mole num sintetizador (Modelo 391) de Applied Biosystems (Weiterstadt), de acordo com as indicações do fabricante, utilizando di-isopropilaminofosfoamidatos protegidos com b-cianoetilo. Fosforilaram-se 100 pmol de oligodesoxirribonucleótido de cada vez, em tri-(hidroximetil)-aminometano/HCl (tris/HCl) 50 mM, cloreto de magnésio 10 mM e ditioneitol 5 mM, a um valor de pH 7,5, com uma unidade de enzima quinase de T4 na presença de adenosina-trifosfato 10 mM e a seguir, no mesmo tampão, transformaram-se em moléculas de ADN de cadeia dupla. As moléculas sintéticas de ADN de cadeia dupla obtidas foram purificadas através de electroforese num gel de poliacrilamida (poliacrilamida a 5%) e, a seguir, foram utilizadas para ligação com os plasmídeos adequadamente preparados. A preparação dos plasmídeos através de digestão com enzimas de restrição, isolamento dos respectivos fragmentos de restrição e desfosforilação dos terminais 5', a ligação subsequente e a transformação em *E.coli* K12 JM103, assim como todas as operações de tecnologia genética seguintes foram realizadas de modo convencional e estão descritas em *Sambrook et al. "Molecular Cloning: A Laboratory Manual", 2a. edição, Cold Spring Harbor Laboratory Press, Cold Spring Harbour, USA, 1989.*

b) Produção de culturas permanentes e fermentação

Os plasmídeos de expressão recombinantes pSJ69, pSJ76, pSJ77, pSJ78, pSJ79, pSJ81, pSJ83, pSJ90, pSJ91, pSJ92, pSJ93, pSJ94, pSJ95, pSJ101, pSJ102, pSJ103, pSJ104, pSJ105, pSJ106, pSJ109, pSJ111, pSJ114 e pSJ113 foram inseridos em *E.coli* K12 JM103 (ATCC39403) e espalhados em ágar padrão I (150 mg/l de ampicilina) (*Sambrook et al. "Molecular Cloning: A Laboratory Manual"*). Colónias individuais de cada transformação, respectivamente, foi cultivada no meio padrão I (pH 7,0; 150 mg/l de ampicilina) a 20°C até à densidade óptica (DO) de 1 a 578 nm, e congelaram-se 5 porções de 2 ml, sob a forma de cultura permanente com adição de dimetilsulfóxido (DMSO) (concentração final de 7,5%) em azoto líquido e conservada a -70°C. Para obter as variantes de uroquinase bifuncionais suspendeu-se 1 ml, respectivamente, de cada cultura permanente

em 20 ml de meio padrão I (pH 7,0; 150 mg/l de ampicilina) e cultivou-se a 37°C até à DO de 1 a 578 nm.

A seguir, suspendeu-se a quantidade total da cultura obtida em 1 l de meio padrão I (pH 7,0; 150 mg/l de ampicilina) e fermentou-se em balões de agitação a 37°C. A indução foi realizada por adição de 2 ml de solução de ácido indoloacrílico (60 mg em 2 ml de etanol) a uma DO de 0,5 a 1 a 578 nm.

c) Teste de expressão

Para verificar a taxa de expressão (unidades *Ploug* por DO por ml) separaram-se por centrifugação, imediatamente antes da indução e de hora a hora depois da indução (6 horas no total) células correspondendo a 1 ml de uma suspensão celular com uma DO de 1 a 578 nm. As células centrifugadas foram lisadas com lisozima (1 mg de lisozima por ml em tampão tris/HCl 50 mM, pH 8,0, ácido etilenodiaminotetra-acético (EDTA) 50 mM e sacarose a 15% em volume). As células lisadas foram solubilizadas em solução 4-5 M de cloridrato de guanidina e, após diluição até 1,2 M de cloridrato de guanidina, submetidas, por adição de um agente redutor (glutathiona ou cisteína) durante 2-5 horas, à reacção de redobragem (Winkler *et al.*, *Biochemistry* 25 4041 a 4045 (1986)). As variantes de uroquinase bifuncionais de uma só cadeia obtidas foram transformadas, através da adição de plasmina, nas respectivas variantes de uroquinase de duas cadeias, cuja actividade foi determinada com o substrato cromogénico piro-Glu-Gly-Arg-p-nitroanilida, que é cindido só por uroquinases activas de duas cadeias. A activação das variantes de uroquinase bifuncionais de acordo com o invento, com plasmina, foi realizada em tampão tris/HCl 50 mM, cloreto de sódio 12 mM, *Tween 80* a 0,02%, a um pH 7,4 e 37°C. A relação entre a variante de uroquinase bifuncional e a plasmina era cerca de 100-1500 para 1, relação molar, ou cerca de 8 000-36 000 para 1 relação de unidades de enzima. A incubação de teste foi realizada em tampão tris/HCl 50 mM e cloreto de sódio 38 mM, a um pH 8,8, na presença de aprotinina 0,36 µM (para inibição da plasmina) e substrato piro-Glu-Gly-Arg-p-nitroanilida 0,27 mM, a 37°C. Em função da concentração de variante de uroquinase bifuncional, a reacção foi feita parar após uma incubação de 5 a 60 minutos, através da adição de ácido acético a 50%, e foi medida a extinção a 405 nm. Ao proceder desta maneira, uma alteração da extinção de 0,05 por minuto a 405 nm

corresponde, de acordo com as indicações do fabricante do substrato (Kabi Vitrum, Suécia), a uma actividade de uroquinase de 25 unidades *Ploug* por ml de solução de teste. As variantes de uroquinase bifuncionais de acordo com o invento apresentavam actividades específicas entre 120 000 e 155 000 unidades *Ploug* por mg de proteína purificada. O teor de proteína das soluções foi determinado através do ensaio BCA da Pierce.

d) Isolamento e purificação

5 a 6 horas após a indução foi terminada a fermentação realizada sob as condições descritas em 1b) (densidade 5-6 DO a 578 nm). As células foram separadas por centrifugação. As células precipitadas foram ressuspensas em 200 ml de água e lisadas no homogeneizador de alta pressão. A seguir a uma nova centrifugação o precipitado, que continha toda a quantidade da variante de uroquinase bifuncional de uma só cadeia, foi dissolvido em 500 ml de cloridrato de guanidina 5 M, cisteína 40 mM, EDTA 1 mM, a um valor de pH 8,0, e diluído com 2000 ml de tris/HCl 25 mM com um valor de pH 9,0. A reacção de redobragem ficou concluída ao fim de cerca de 12 horas.

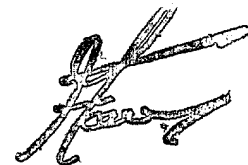
Após a adição de 8 g de gel de sílica, as variantes de uroquinase bifuncionais obtidas foram ligadas totalmente ao gel de sílica através de uma agitação de 2 horas. O gel de sílica carregado foi separado e lavado com tampão acetato (pH 4,0). As variantes de uroquinase foram eluídas com cloreto de tetrametilamónio (TMAC) 0,5 M em tampão de acetato 0,1 M (pH 4). Após duas separações cromatográficas (coluna de quelato de cobre e permuta iónica), as variantes de uroquinase foram obtidas sob a forma pura. Através de análise de sequenciação do terminal N provou-se, por um lado, a existência de uma só cadeia e, por outro lado, a que a sequência amino-terminal era a desejada. A caracterização através de métodos químicos usual para as proteínas, do domínio carboxi-terminal modificado das variantes individuais, foi obtida depois de cisão da proteína com CNBr (dissolvida em 1 ml de ácido fórmico a 90% e 1 ml de ácido heptafluorobutírico), ocorrendo a cisão da cadeia peptídica a seguir a resíduos triptofano. O péptido carboxi-terminal foi separado e purificado, antes da análise de sequenciação, através de HPLC (*High Pressure Liquid Chromatography*).



Todas as variantes de uroquinase bifuncionais isoladas e listadas na tabela 2 não apresentavam, num teste de actividade directo com o substrato cromogénico para uroquinase, nenhuma actividade ou apenas uma actividade muito reduzida (inferior a 1200 unidades *Ploug* por mg de proteína purificada). Só depois de cisão com plasmina (as condições estão indicadas no ponto 1c)), foi obtida uma actividade enzimática entre as 120 000 e 155 000 unidades *Ploug* por mg de proteína purificada. De acordo com isto, todas as variantes de uroquinase foram expressas em *E.coli* K12 JM103 sob a forma de proteínas de uma só cadeia.

Tabela 2: Variantes de uroquinase bifuncionais (bU) de acordo com o invento de fórmula geral I, M4-X₁-Y₁

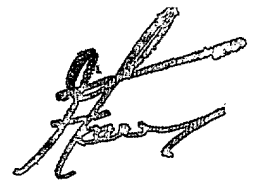
bU	X ₁	Y ₁
M11	Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe	A ¹⁾ com Y ₂ = Pro, Y ₃ = ligação directa entre Pro e Gly, Y ₄ = OH
M12	Ser-Pro-Val-Lys-Ala-Phe	A ¹⁾ com Y ₂ = Pro, Y ₃ = ligação directa entre Pro e Gly, Y ₄ = OH
M13	Ser-Pro-Pro-Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe	A ¹⁾ com Y ₂ = Pro, Y ₃ = ligação directa entre Pro e Gly, Y ₄ = OH
M14	Ser-Pro-Pro-Gly-Gly-Phe	A ¹⁾ com Y ₂ = Pro, Y ₃ = ligação directa entre Pro e Gly, Y ₄ = OH
M15	Ser-Pro-Pro-Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe-Gly	A ¹⁾ com Y ₂ = Pro, Y ₃ = ligação directa entre Pro e Gly, Y ₄ = OH
M16	Ser-Pro-Val-Lys-Ala-Trp-Gly	A ¹⁾ com Y ₂ = Pro, Y ₃ = Leu, Y ₄ = Gln
M17	Ser-Pro-Val-Lys-Ala-Phe-Gly	A ¹⁾ com Y ₂ = Pro, Y ₃ = Leu, Y ₄ = Gln
M18	Ser-Pro-Val-Arg-Ala-Phe-Gly	A ¹⁾ com Y ₂ = Pro, Y ₃ = Leu, Y ₄ = Gln
M19	Ser-Pro-Val-Lys-Ala-Tyr-Gly	A ¹⁾ com Y ₂ = Pro, Y ₃ = Leu, Y ₄ = Gln



bU	X ₁	Y ₁
M20	Ser-Pro-Val-Val-Ala-Phe-Gly	A ¹⁾ com Y ₂ = Pro, Y ₃ = Leu, Y ₄ = Gln
M21	Ser-Pro-Val-Arg-Ala-Phe	A ¹⁾ com Y ₂ = Pro, Y ₃ = Leu, Y ₄ = Gln
M22	Ser-Pro-Val-Lys-Ala-Tyr	A ¹⁾ com Y ₂ = Pro, Y ₃ = Leu, Y ₄ = Gln
M23	Ser-Pro-Val-Val-Ala-Phe	A ¹⁾ com Y ₂ = Pro, Y ₃ = Leu, Y ₄ = Gln
M24	Ser-Pro-Val-Lys-Ala-Trp	A ¹⁾ com Y ₂ = Pro, Y ₃ = Leu, Y ₄ = Gln
M25	Ser-Pro-Val-Lys-Ala-Phe	A ¹⁾ com Y ₂ = Pro, Y ₃ = Leu, Y ₄ = Gln
M26	Ser-Pro-Val-Val-Val-Phe	A ¹⁾ com Y ₂ = Pro, Y ₃ = Leu, Y ₄ = Gln
M27	Ser-Pro-Val-Glu-Val-Phe	A ¹⁾ com Y ₂ = Pro, Y ₃ = Leu, Y ₄ = Gln
M28	Ser-Pro-Val-Val-Val-Val	A ¹⁾ com Y ₂ = Val, Y ₃ = Leu, Y ₄ = Gln
M29	Ser-Pro-Val-Val-Ala-Phe	B ²⁾ com Y ₂ = Pro
M30	Ser-Leu-Val-Val-Ala-Phe	A ¹⁾ com Y ₂ = Pro, Y ₃ = Leu, Y ₄ = Gln
M31	Ser-Leu-Val-Lys-Ala-Phe	A ¹⁾ com Y ₂ = Pro, Y ₃ = Leu, Y ₄ = Gln
M32	Ser-Pro-Val-Lys-Ala-Phe	B ²⁾ com Y ₂ = Pro
M33	ligação directa entre M4 e Y ₁	A ¹⁾ com Y ₂ = Pro, Y ₃ = Leu, Y ₄ = Gln

¹⁾ A = Y₂-Arg-Pro-Y₃-Gly-Gly-Gly-Gly-Asn-Gly-Asp-Phe-Glu-Glu-Ile-Pro-Glu-Glu-Tyr-Leu-Y₄

²⁾ B = Y₂-Arg-Pro-Phe-Leu-Leu-Arg-Asn-Pro-Asn-Asp-Lys-Tyr-Glu-Pro-Phe-Trp-Glu-Asp-Glu-Glu-Lys-Asn-Glu



2) Análises farmacológicas

Determinação da actividade inibidora de trombina

A actividade inibidora das variantes de uroquinase bifuncionais de acordo com o invento foi determinada através de medição do tempo de trombina, misturando 200 µl de uma diluição de 1:10 de plasma de citrato humano num tampão de veronal, com 50 µl de solução de trombina (0,2 unidades) e 50 µl de uma solução aquosa contendo 0,5-50 µg de uma variante de uroquinase bifuncional. A seguir mediu-se o tempo decorrido até à formação de um tecido de fibrina. Na tabela 3 estão indicados os factores de inibição medidos, os quais indicam, em cada caso, o aumento do tempo de trombina na presença de 10 µg de uma variante de uroquinase bifuncional de acordo com o invento. O aumento do tempo de trombina em função da concentração foi também determinado e está representado graficamente na figura 3, para as variantes de uroquinase bifuncionais M12, M23, M29, M32, M33 e, a título de comparação, M4. Ao contrário das variantes de uroquinase bifuncionais de acordo com o invento, o tempo decorrido até à coagulação não foi prolongado nem por M4, isto é, a sequência de aminoácidos ⁴⁷Ser a ⁴¹¹Leu da pro-uroquinase não glicosilada de acordo com a figura 1, nem pela pro-uroquinase não glicosilada (*Saruplase*), nem pela LUK em doses de 1 mg.

(Segue Tabela)

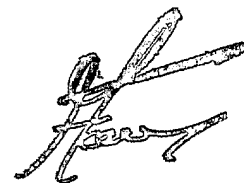


Tabela 3: Aumento do tempo de trombina por de variantes de uroquinase bifuncionais de acordo com o invento de fórmula geral I M4-X₁-Y₁

Variante de uroquinase bifuncional	Factor de inibição ¹⁾
M11 (SEQ ID NO: 23 e 24)	1,8
M12 (SEQ ID NO: 25 e 26)	4,6
M13 (SEQ ID NO: 27 e 28)	1,7
M14 (SEQ ID NO: 29 e 30)	1,8
M15 (SEQ ID NO: 31 e 32)	2,5
M16 (SEQ ID NO: 33 e 34)	3,2
M17 (SEQ ID NO: 35 e 36)	3,1
M18 (SEQ ID NO: 37 e 38)	2,9
M19 (SEQ ID NO: 39 e 40)	2,0
M20 (SEQ ID NO: 41 e 42)	2,2
M21 (SEQ ID NO: 43 e 44)	2,3
M22 (SEQ ID NO: 45 e 46)	3,7
M23 (SEQ ID NO: 47 e 48)	5,3
M24 (SEQ ID NO: 1 e 2)	6,2
M25 (SEQ ID NO: 3 e 4)	2,9
M26 (SEQ ID NO: 5 e 6)	3,2
M27 (SEQ ID NO: 7 e 8)	2,0
M28 (SEQ ID NO: 9 e 10)	2,1
M29 (SEQ ID NO: 11 e 12)	2,6
M30 (SEQ ID NO: 13 e 14)	3,4
M31 (SEQ ID NO: 15 e 16)	2,0
M32 (SEQ ID NO: 19 e 20)	3,0
M33 (SEQ ID NO: 17 e 18)	2,0

¹⁾ em relação ao efeito de 10 µg de proteína;

Factor de inibição = quociente entre o tempo de trombina na presença de um inibidor, e o tempo de trombina na ausência de um inibidor

Características farmacológicas das variantes de uroquinase bifuncionais M12 e M23 num ensaio em animais

Num modelo farmacológico *in vivo* foi ensaiado o efeito das variantes de uroquinase bifuncionais M12 e M23 na trombólise de obstruções vasculares arteriais, em comparação com *Saruplase* (pro-uroquinase não glicosilada). Para este efeito injectou-se localmente, em coelhos anestesiados, num segmento da artéria femural temporariamente isolado, de cerca de 1 cm, através de um ramo lateral, trombina e fibrinogénio humano marcados com ^{125}I . Em consequência, formou-se um trombo que deu origem a uma obstrução vascular completa. O tamanho do trombo formado foi determinado através da radioactividade incorporada na fibrina humana, utilizando um detector de raios gama extracorporal. A medição electromagnética do fluxo de sangue e o registo da radioactividade do trombo foram realizados continuamente durante todo o período do ensaio. A actividade fibrinolítica foi portanto quantificada não apenas como reperfusão do vaso obstruído pelo trombo, como também através da degradação da fibrina incorporada, marcada radioactivamente, do trombo. Antes da aplicação das variantes de uroquinase bifuncionais de acordo com o invento, assim como 30, 60 e 90 minutos depois da aplicação das variantes de uroquinase, foram recolhidas amostras de sangue nas quais se determinaram as concentrações de fibrinogénio no plasma. Foram aplicados, sob a forma de injeções de *bolus* intravenosas, respectivamente, 6 mg/kg de M12, M23 e *Saruplase*. Dado que M12 e M23 possuem além disso, e ao contrário da *Saruplase*, um efeito anticoagulante, combinou-se num 4º grupo de ensaio, *Saruplase* com o anticoagulante heparina (150 U/kg de *bolus* i.v. Cada grupo continha 6 animais.

Durante os 90 minutos de duração do ensaio, a trombólise da fibrina marcada do trombo foi de $46\pm 11\%$ para M12, $43\pm 12\%$ para M23, $22\pm 5\%$ para *Saruplase* e $39\pm 15\%$ para a combinação de *Saruplase*-heparina. As aplicações de *bolus* de M12 e M23 resultaram, em todos os 6 animais, na desobstrução do vaso obstruído pelo trombo; através da aplicação de *Saruplase* foi possível a reperfusão do vaso em 5 de 6 animais e, através da aplicação de *Saruplase* e heparina, em 4 de 6 animais. O valor máximo do fluxo de reperfusão (em % do valor inicial) foi de $95\pm 10\%$ para M12 e $82\pm 9\%$ para M23, e diferiu significativamente do valor máximo do fluxo de reperfusão, de $43\pm 12\%$, para

Saruplase. O valor máximo do fluxo de reperfusão, de $58 \pm 8\%$, no tratamento com *Saruplase* e heparina, encontrava-se entre os resultados de M12 e M23, por um lado, e *Saruplase*, por outro lado, e não difere significativamente em relação aos dois lados. A actividade fibrinolítica total foi determinada sob a forma da área do fluxo de reperfusão (como % do fluxo inicial) ao longo dos 90 minutos de duração do ensaio. Esta actividade total foi de $4502 \pm 1127\% \cdot \text{min}$ para M12 e $4270 \pm 885\% \cdot \text{min}$ para M23 e, para ambas as variantes de uroquinase de acordo com o invento, foi significativamente superior ao valor de $1519 \pm 643\% \cdot \text{min}$ para *Saruplase*. Com o tratamento combinado com *Saruplase*-heparina foi medida uma actividade total de $2217 \pm 761\% \cdot \text{min}$, que não foi significativamente melhor do que o valor obtido no tratamento exclusivamente com *Saruplase* e se encontrava nitidamente abaixo dos resultados obtidos com M12 e M23. Os resultados estão resumidos na tabela 4.

Tabela 4: Actividade trombolítica após uma aplicação de *bolus* i.v.; trombose arterial femural, coelho anestesiado.

Polipéptido	Dose	^{125}I -fibrinólise	Fluxo de reperfusão máximo (% do valor prévio)	Fluxo de reperfusão cumulativo (% · min)
M12	6 mg/kg	46 ± 11	$95 \pm 10^*$	4502 ± 1127
M23	6 mg/kg	43 ± 12	$82 \pm 9^*$	$4270 \pm 885^*$
<i>Saruplase</i>	6 mg/kg	22 ± 5	43 ± 12	1519 ± 643
<i>Saruplase</i> + heparina	6 mg/kg 150 U/kg	39 ± 15	58 ± 8	2217 ± 761

* $p < 0.05$ vs *Saruplase*

Inesperadamente verificou-se que não apenas após a aplicação de *bolus* de M12 como também após a aplicação de *bolus* de M23, as concentrações de fibrinogénio no plasma diminuíram significativamente menos do que após a aplicação de *bolus* de *Saruplase*. Os resultados encontram-se resumidos na tabela 5.

Tabela 5: Efeito das aplicações de *bolus* de M12 e M23 em comparação com a *Saruplase*, sem e com heparina, na diminuição da concentração de fibrinogénio no plasma; coelhos anestesiados

Polipéptido	Dose	Diminuição de fibrinogénio no plasma (alteração em % em relação ao valor inicial)		
		Tempo após a aplicação		
		30 min	60 min	90 min
M12	6 mg/kg	-19 ± 9	-20 ± 9*	-19 ± 9*
M23	6 mg/kg	-20 ± 11*	-21 ± 11*	-20 ± 11*
<i>Saruplase</i>	6 mg/kg	-64 ± 7	-66 ± 6	-67 ± 6
<i>Saruplase</i> + heparina	6 mg/kg 150 U/kg	n.d. ¹⁾	-46 ± 8	-45 ± 9

* $p < 0,05$ vs *Saruplase*

¹⁾ n.d. = não determinada

Os resultados mostram que as variantes de uroquinase bifuncionais de acordo com o invento, M12 e M23, dissolvem trombos arteriais que provocam uma obstrução vascular completa, restaurando o fluxo sanguíneo nos vasos obstruídos pelo trombo. Este efeito foi obtido através de uma aplicação única de *bolus* de M12 ou M23, respectivamente, em animais não heparinizados. Inesperadamente, esta actividade fibrinolítica mais forte de M12 e M23 em comparação com *Saruplase* foi acompanhado por um consumo mais baixo de fibrinogénio do plasma. Isto significa que M12 e M23 apresentam uma especificidade para a fibrina significativamente mais elevada em comparação com a *Saruplase*.

A melhor preservação de fibrinogénio do plasma de M12 e M23 em comparação com a *Saruplase*, significa que a capacidade de coagulação do sangue é mantida em melhores condições, o que reduz o risco de hemorragias incontroladas como complicações possíveis de uma degradação sistémica de fibrinogénio. M12 e M23 devem portanto ser classificadas como mais seguros do que *Saruplase* em relação ao risco de efeitos secundários hemostaseológicos.

LISTAGEM DE SEQUÊNCIAS

(1) INFORMAÇÃO GERAL:

(i) REQUERENTE:

- (A) NOME: Gruenenthal GmbH
- (B) RUA: Zieglerstrasse 6
- (C) CIDADE: Aachen
- (E) PAÍS: Alemanha
- (F) CÓDIGO POSTAL (ZIP): 52078

(ii) TÍTULO DO INVENTO: Variantes de uroquinase bifuncionais com propriedades fibrinolíticas melhoradas e actividade inibidora de trombina

(iii) NÚMERO DE SEQUÊNCIAS: 102

(iv) FORMA LEGÍVEL EM COMPUTADOR:

- (A) TIPO DE MEIO: Disquete, 3,5", DS, 1,44 MB
- (B) COMPUTADOR: Compatível com PC da IBM
- (C) SISTEMA OPERATIVO: PC-DOS/MS-DOS
- (D) SUPORTE LÓGICO: PatentIn Release #1.0, Versão #1.30

(v) DADOS DO PEDIDO ACTUAL:

NÚMERO DO PEDIDO: EP 94 108 872.6-1270

(2) INFORMAÇÃO PARA SEQ ID NO:1:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1182 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: CDS

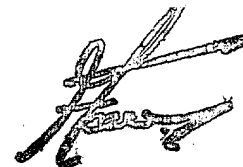
(B) LOCALIZAÇÃO: 1..1182

(D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada proteoliticamente em E.coli"

/product = "gene sintético para a proteína M24 (pSJ101)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 1:

ATG	AGC	AAA	ACT	TGC	TAC	GAA	GGT	AAC	GGT	CAC	TTC	TAC	CGT	GGT	AAG	48
Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys	
1				5					10					15		
GCT	TCT	ACC	GAC	ACC	ATG	GGT	CGT	CCG	TGC	CTG	CCG	TGG	AAC	TCT	GCT	96
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala	
			20					25					30			
ACC	GTT	CTG	CAG	CAG	ACC	TAC	CAC	GCT	CAC	CGT	TCT	GAT	GCA	TTG	CAG	144
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln	
		35					40					45				
CTG	GGT	CTG	GGT	AAA	CAC	AAC	TAC	TGC	CGT	AAC	CCG	GAC	AAC	CGT	CGT	192
Leu	Gly	Leu	Gly	Lys	His	Asn	Tyr	Cys	Arg	Asn	Pro	Asp	Asn	Arg	Arg	
	50					55					60					
CGT	CCG	TGG	TGC	TAC	GTT	CAG	GTT	GGT	CTG	AAA	CCG	CTA	GTT	CAG	GAA	240
Arg	Pro	Trp	Cys	Tyr	Val	Gln	Val	Gly	Leu	Lys	Pro	Leu	Val	Gln	Glu	
	65				70					75					80	
TGC	ATG	GTT	CAC	GAC	TGC	GCT	GAC	GGT	AAA	AAA	CCG	TCT	TCT	CCG	CCG	288
Cys	Met	Val	His	Asp	Cys	Ala	Asp	Gly	Lys	Lys	Pro	Ser	Ser	Pro	Pro	
				85					90					95		
GAA	GAG	CTC	AAA	TTC	CAG	TGC	GGT	CAA	AAA	ACC	CTA	CGT	CCG	CGT	TTT	336
Glu	Glu	Leu	Lys	Phe	Gln	Cys	Gly	Gln	Lys	Thr	Leu	Arg	Pro	Arg	Phe	
			100					105					110			
AAA	ATC	ATC	GGT	GGT	GAG	TTC	ACC	ACC	ATC	GAA	AAC	CAG	CCG	TGG	TTC	384
Lys	Ile	Ile	Gly	Gly	Glu	Phe	Thr	Thr	Ile	Glu	Asn	Gln	Pro	Trp	Phe	
		115					120					125				
GCT	GCT	ATC	TAC	CGT	CGT	CAC	CGT	GGT	GGT	TCT	GTT	ACC	TAC	GTT	TGC	432
Ala	Ala	Ile	Tyr	Arg	Arg	His	Arg	Gly	Gly	Ser	Val	Thr	Tyr	Val	Cys	
	130					135					140					
GGT	GGT	TCT	CTG	ATC	TCT	CCG	TGC	TGG	GTT	ATC	TCT	GCT	ACC	CAC	TGC	480
Gly	Gly	Ser	Leu	Ile	Ser	Pro	Cys	Trp	Val	Ile	Ser	Ala	Thr	His	Cys	
	145				150					155					160	
TTC	ATC	GAC	TAC	CCG	AAA	AAA	GAA	GAC	TAC	ATC	GTT	TAC	CTC	GGC	CGT	528
Phe	Ile	Asp	Tyr	Pro	Lys	Lys	Glu	Asp	Tyr	Ile	Val	Tyr	Leu	Gly	Arg	
				165					170					175		



(2) INFORMAÇÃO PARA SEQ ID NO: 2:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 394 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 2:

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

Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

Val Lys Ala Trp Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
370 375 380

Glu Glu Ile Pro Glu Glu Tyr Leu Gln *
385 390

(2) INFORMAÇÃO PARA SEQ ID NO: 3:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1182 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1182
- (D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada proteoliticamente em E.coli"
/product = "gene sintético para a proteína M25 (pSJ102)"



(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 3:

ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys 1 5 10 15	48
GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala 20 25 30	96
ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln 35 40 45	144
CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg 50 55 60	192
CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu 65 70 75 80	240
TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro 85 90 95	288
GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe 100 105 110	336
AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe 115 120 125	384
GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys 130 135 140	432
GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys 145 150 155 160	480
TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg 165 170 175	528
TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu 180 185 190	576
AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His 195 200 205	624

AAC Asn	GAC Asp	ATC Ile	GCT Ala	CTG Leu	CTA Leu	AAA Lys	ATC Ile	CGT Arg	TCT Ser	AAA Lys	GAA Glu	GGT Gly	CGT Arg	TGC Cys	GCT Ala	672
210						215					220					
CAG Gln	CCG Pro	TCT Ser	CGT Arg	ACC Thr	ATC Ile	CAG Gln	ACC Thr	ATC Ile	TGC Cys	CTG Leu	CCG Pro	TCT Ser	ATG Met	TAC Tyr	AAC Asn	720
225					230					235					240	
GAC Asp	CCG Pro	CAG Gln	TTC Phe	GGT Gly	ACC Thr	TCT Ser	TGC Cys	GAA Glu	ATC Ile	ACC Thr	GGT Gly	TTC Phe	GGT Gly	AAA Lys	GAA Glu	768
				245					250					255		
AAC Asn	TCT Ser	ACC Thr	GAC Asp	TAC Tyr	CTG Leu	TAC Tyr	CCG Pro	GAA Glu	CAG Gln	CTG Leu	AAA Lys	ATG Met	ACC Thr	GTT Val	GTT Val	816
			260					265					270			
AAA Lys	CTG Leu	ATC Ile	TCT Ser	CAC His	CGT Arg	GAA Glu	TGC Cys	CAG Gln	CAG Gln	CCG Pro	CAC His	TAC Tyr	TAC Tyr	GGT Gly	TCT Ser	864
		275					280					285				
GAA Glu	GTT Val	ACC Thr	ACC Thr	AAA Lys	ATG Met	CTG Leu	TGC Cys	GCT Ala	GCT Ala	GAC Asp	CCG Pro	CAG Gln	TGG Trp	AAA Lys	ACC Thr	912
	290					295					300					
GAC Asp	TCT Ser	TGC Cys	CAA Gln	GGT Gly	GAC Asp	TCT Ser	GGT Gly	GGT Gly	CCA Pro	CTA Leu	GTT Val	TGC Cys	TCT Ser	CTC Leu	CAG Gln	960
305					310					315					320	
GGT Gly	CGT Arg	ATG Met	ACC Thr	CTG Leu	ACC Thr	GGT Gly	ATT Ile	GTT Val	TCT Ser	TGG Trp	GGT Gly	CGT Arg	GGT Gly	TGC Cys	GCT Ala	1009
				325				330						335		
CTG Leu	AAA Lys	GAC Asp	AAA Lys	CCG Pro	GGT Gly	GTT Val	TAC Tyr	ACC Thr	CGT Arg	GTT Val	TCT Ser	CAC His	TTC Phe	CTG Leu	CCG Pro	1056
			340				345						350			
TGG Trp	ATC Ile	CGT Arg	TCT Ser	CAC His	ACC Thr	AAA Lys	GAA Glu	GAA Glu	AAC Asn	GGT Gly	CTG Leu	GCT Ala	CTG Leu	AGC Ser	CCG Pro	1104
		355					360					365				
GTT Val	AAG Lys	GCT Ala	TTC Phe	CCG Pro	CGG Arg	CCG Pro	CTG Leu	GGT Gly	GGT Gly	GGT Gly	GGT Gly	AAC Asn	GGT Gly	GAC Asp	TTC Phe	1152
	370					375					380					
GAA Glu	GAA Glu	ATC Ile	CCG Pro	GAA Glu	GAA Glu	TAC Tyr	CTG Leu	CAA Gln	TAA *							1182
385					390											

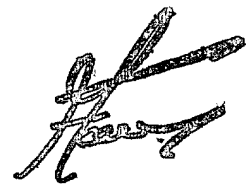
(2) INFORMAÇÃO PARA SEQ ID NO: 4:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

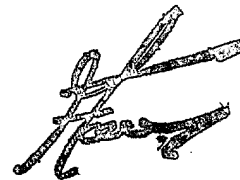
- (A) COMPRIMENTO: 394 aminoácidos
- (B) TIPO: aminoácido
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 4:



Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
1 5 10 15
Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
20 25 30
Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
35 40 45
Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
50 55 60
Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
65 70 75 80
Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
85 90 95
Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
100 105 110
Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
115 120 125
Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
130 135 140
Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
145 150 155 160
Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
165 170 175
Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190
Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205
Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220
Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240
Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255
Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270
Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285
Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300
Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320



Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

Val Lys Ala Phe Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
370 375 380

Glu Glu Ile Pro Glu Glu Tyr Leu Gin *
385 390

(2) INFORMAÇÃO PARA SEQ ID NO: 5:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1182 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

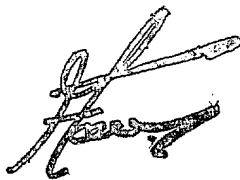
(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1182
- (D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada proteoliticamente em E.coli"
/product = "gene sintético para a proteína M26 (pSJ103)"

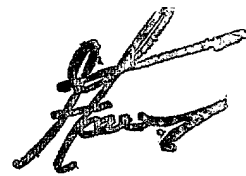
(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 5:

ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
1 5 10 15

GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT
Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
20 25 30



ACC	GTT	CTG	CAG	CAG	ACC	TAC	CAC	GCT	CAC	CGT	TCT	GAT	GCA	TTG	CAG	144
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln	
		35					40					45				
CTG	GGT	CTG	GGT	AAA	CAC	AAC	TAC	TGC	CGT	AAC	CCG	GAC	AAC	CGT	CGT	192
Leu	Gly	Leu	Gly	Lys	His	Asn	Tyr	Cys	Arg	Asn	Pro	Asp	Asn	Arg	Arg	
	50					55					60					
CGT	CCG	TGG	TGC	TAC	GTT	CAG	GTT	GGT	CTG	AAA	CCG	CTA	GTT	CAG	GAA	240
Arg	Pro	Trp	Cys	Tyr	Val	Gln	Val	Gly	Leu	Lys	Pro	Leu	Val	Gln	Glu	
	65				70					75					80	
TGC	ATG	GTT	CAC	GAC	TGC	GCT	GAC	GGT	AAA	AAA	CCG	TCT	TCT	CCG	CCG	288
Cys	Met	Val	His	Asp	Cys	Ala	Asp	Gly	Lys	Lys	Pro	Ser	Ser	Pro	Pro	
				85					90					95		
GAA	GAG	CTC	AAA	TTC	CAG	TGC	GGT	CAA	AAA	ACC	CTA	CGT	CCG	CGT	TTT	336
Glu	Glu	Leu	Lys	Phe	Gln	Cys	Gly	Gln	Lys	Thr	Leu	Arg	Pro	Arg	Phe	
			100					105					110			
AAA	ATC	ATC	GGT	GGT	GAG	TTC	ACC	ACC	ATC	GAA	AAC	CAG	CCG	TGG	TTC	384
Lys	Ile	Ile	Gly	Gly	Glu	Phe	Thr	Thr	Ile	Glu	Asn	Gln	Pro	Trp	Phe	
			115				120					125				
GCT	GCT	ATC	TAC	CGT	CGT	CAC	CGT	GGT	GGT	TCT	GTT	ACC	TAC	GTT	TGC	432
Ala	Ala	Ile	Tyr	Arg	Arg	His	Arg	Gly	Gly	Ser	Val	Thr	Tyr	Val	Cys	
		130				135						140				
GGT	GGT	TCT	CTG	ATC	TCT	CCG	TGC	TGG	GTT	ATC	TCT	GCT	ACC	CAC	TGC	480
Gly	Gly	Ser	Leu	Ile	Ser	Pro	Cys	Trp	Val	Ile	Ser	Ala	Thr	His	Cys	
	145				150					155					160	
TTC	ATC	GAC	TAC	CCG	AAA	AAA	GAA	GAC	TAC	ATC	GTT	TAC	CTC	GGC	CGT	528
Phe	Ile	Asp	Tyr	Pro	Lys	Lys	Glu	Asp	Tyr	Ile	Val	Tyr	Leu	Gly	Arg	
				165					170					175		
TCT	CGT	TTA	AAC	TCT	AAC	ACC	CAG	GGT	GAA	ATG	AAA	TTC	GAA	GTT	GAA	576
Ser	Arg	Leu	Asn	Ser	Asn	Thr	Gln	Gly	Glu	Met	Lys	Phe	Glu	Val	Glu	
			180					185					190			
AAC	CTG	ATC	CTG	CAC	AAA	GAC	TAC	TCT	GCT	GAC	ACC	CTG	GCT	CAC	CAC	624
Asn	Leu	Ile	Leu	His	Lys	Asp	Tyr	Ser	Ala	Asp	Thr	Leu	Ala	His	His	
			195				200					205				
AAC	GAC	ATC	GCT	CTG	CTA	AAA	ATC	CGT	TCT	AAA	GAA	GGT	CGT	TGC	GCT	672
Asn	Asp	Ile	Ala	Leu	Leu	Lys	Ile	Arg	Ser	Lys	Glu	Gly	Arg	Cys	Ala	
	210					215					220					
CAG	CCG	TCT	CGT	ACC	ATC	CAG	ACC	ATC	TGC	CTG	CCG	TCT	ATG	TAC	AAC	720
Gln	Pro	Ser	Arg	Thr	Ile	Gln	Thr	Ile	Cys	Leu	Pro	Ser	Met	Tyr	Asn	
					230					235					240	
GAC	CCG	CAG	TTC	GGT	ACC	TCT	TGC	GAA	ATC	ACC	GGT	TTC	GGT	AAA	GAA	768
Asp	Pro	Gln	Phe	Gly	Thr	Ser	Cys	Glu	Ile	Thr	Gly	Phe	Gly	Lys	Glu	
				245					250					255		
AAC	TCT	ACC	GAC	TAC	CTG	TAC	CCG	GAA	CAG	CTG	AAA	ATG	ACC	GTT	GTT	816
Asn	Ser	Thr	Asp	Tyr	Leu	Tyr	Pro	Glu	Gln	Leu	Lys	Met	Thr	Val	Val	
			260				265						270			
AAA	CTG	ATC	TCT	CAC	CGT	GAA	TGC	CAG	CAG	CCG	CAC	TAC	TAC	GGT	TCT	864
Lys	Leu	Ile	Ser	His	Arg	Glu	Cys	Gln	Gln	Pro	His	Tyr	Tyr	Gly	Ser	
		275					280						285			



36

GAA	GTT	ACC	ACC	AAA	ATG	CTG	TGC	GCT	GCT	GAC	CCG	CAG	TGG	AAA	ACC	912
Glu	Val	Thr	Thr	Lys	Met	Leu	Cys	Ala	Ala	Asp	Pro	Gln	Trp	Lys	Thr	
290						295					300					
GAC	TCT	TGC	CAA	GGT	GAC	TCT	GGT	GGT	CCA	CTA	GTT	TGC	TCT	CTC	CAG	960
Asp	Ser	Cys	Gln	Gly	Asp	Ser	Gly	Gly	Pro	Leu	Val	Cys	Ser	Leu	Gln	
305				310					315						320	
GGT	CGT	ATG	ACC	CTG	ACC	GGT	ATT	GTT	TCT	TGG	GGT	CGT	GGT	TGC	GCT	1008
Gly	Arg	Met	Thr	Leu	Thr	Gly	Ile	Val	Ser	Trp	Gly	Arg	Gly	Cys	Ala	
				325					330					335		
CTG	AAA	GAC	AAA	CCG	GGT	GTT	TAC	ACC	CGT	GTT	TCT	CAC	TTC	CTG	CCG	1056
Leu	Lys	Asp	Lys	Pro	Gly	Val	Tyr	Thr	Arg	Val	Ser	His	Phe	Leu	Pro	
			340					345					350			
TGG	ATC	CGT	TCT	CAC	ACC	AAA	GAA	GAA	AAC	GGT	CTG	GCT	CTG	AGC	CCG	1104
Trp	Ile	Arg	Ser	His	Thr	Lys	Glu	Glu	Asn	Gly	Leu	Ala	Leu	Ser	Pro	
		355					360					365				
GTT	GTA	GTT	TTC	CCG	CGG	CCG	CTG	GGT	GGT	GGT	GGT	AAC	GGT	GAC	TTC	1152
Val	Val	Val	Phe	Pro	Arg	Pro	Leu	Gly	Gly	Gly	Gly	Asn	Gly	Asp	Phe	
		370				375					380					
GAA	GAA	ATC	CCG	GAA	GAA	TAC	CTG	CAA	TAA							1182
Glu	Glu	Ile	Pro	Glu	Glu	Tyr	Leu	Gln	*							
385						390										

(2) INFORMAÇÃO PARA SEQ ID NO: 6:

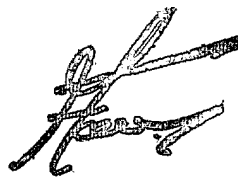
(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 394 aminoácidos
- (B) TIPO: aminoácido
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 6:

Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys
1				5					10					15	
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala
			20					25					30		
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln
		35					40					45			
Leu	Gly	Leu	Gly	Lys	His	Asn	Tyr	Cys	Arg	Asn	Pro	Asp	Asn	Arg	Arg
	50					55					60				



Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
65 70 75 80
Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
85 90 95
Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
100 105 110
Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
115 120 125
Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
130 135 140
Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
145 150 155 160
Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
165 170 175
Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190
Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205
Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220
Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240
Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255
Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270
Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285
Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300
Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320
Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335
Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350
Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365
Val Val Val Phe Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
370 375 380
Glu Glu Ile Pro Glu Glu Tyr Leu Gln *
385 390



(2) INFORMAÇÃO PARA SEQ ID NO: 7:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1182 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1182
- (D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada proteoliticamente em E.coli"
/product = "gene sintético para a proteína M27 (pSJ104)"

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1182

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 7:

ATG	AGC	AAA	ACT	TGC	TAC	GAA	GGT	AAC	GGT	CAC	TTC	TAC	CGT	GGT	AAG	43
Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys	
1				5			10						15			
GCT	TCT	ACC	GAC	ACC	ATG	GGT	CGT	CCG	TGC	CTG	CCG	TGG	AAC	TCT	GCT	96
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala	
			20				25						30			

ACC	GTT	CTG	CAG	CAG	ACC	TAC	CAC	GCT	CAC	CGT	TCT	GAT	GCA	TTG	CAG	144
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln	
		35					40					45				
CTG	GGT	CTG	GGT	AAA	CAC	AAC	TAC	TGC	CGT	AAC	CCG	GAC	AAC	CGT	CGT	192
Leu	Gly	Leu	Gly	Lys	His	Asn	Tyr	Cys	Arg	Asn	Pro	Asp	Asn	Arg	Arg	
	50					55					60					
CGT	CCG	TGG	TGC	TAC	GTT	CAG	GTT	GGT	CTG	AAA	CCG	CTA	GTT	CAG	GAA	240
Arg	Pro	Trp	Cys	Tyr	Val	Gln	Val	Gly	Leu	Lys	Pro	Leu	Val	Gln	Glu	
	65				70					75					80	
TGC	ATG	GTT	CAC	GAC	TGC	GCT	GAC	GGT	AAA	AAA	CCG	TCT	TCT	CCG	CCG	288
Cys	Met	Val	His	Asp	Cys	Ala	Asp	Gly	Lys	Lys	Pro	Ser	Ser	Pro	Pro	
				85					90					95		
GAA	GAG	CTC	AAA	TTC	CAG	TGC	GGT	CAA	AAA	ACC	CTA	CGT	CCG	CGT	TTT	336
Glu	Glu	Leu	Lys	Phe	Gln	Cys	Gly	Gln	Lys	Thr	Leu	Arg	Pro	Arg	Phe	
			100					105					110			
AAA	ATC	ATC	GGT	GGT	GAG	TTC	ACC	ACC	ATC	GAA	AAC	CAG	CCG	TGG	TTC	384
Lys	Ile	Ile	Gly	Gly	Glu	Phe	Thr	Thr	Ile	Glu	Asn	Gln	Pro	Trp	Phe	
			115				120					125				
GCT	GCT	ATC	TAC	CGT	CGT	CAC	CGT	GGT	GGT	TCT	GTT	ACC	TAC	GTT	TGC	432
Ala	Ala	Ile	Tyr	Arg	Arg	His	Arg	Gly	Gly	Ser	Val	Thr	Tyr	Val	Cys	
		130				135						140				
GGT	GGT	TCT	CTG	ATC	TCT	CCG	TGC	TGG	GTT	ATC	TCT	GCT	ACC	CAC	TGC	480
Gly	Gly	Ser	Leu	Ile	Ser	Pro	Cys	Trp	Val	Ile	Ser	Ala	Thr	His	Cys	
	145				150					155					160	
TTC	ATC	GAC	TAC	CCG	AAA	AAA	GAA	GAC	TAC	ATC	GTT	TAC	CTC	GGC	CGT	528
Phe	Ile	Asp	Tyr	Pro	Lys	Lys	Glu	Asp	Tyr	Ile	Val	Tyr	Leu	Gly	Arg	
				165					170					175		
TCT	CGT	TTA	AAC	TCT	AAC	ACC	CAG	GGT	GAA	ATG	AAA	TTC	GAA	GTT	GAA	576
Ser	Arg	Leu	Asn	Ser	Asn	Thr	Gln	Gly	Glu	Met	Lys	Phe	Glu	Val	Glu	
			180					185					190			
AAC	CTG	ATC	CTG	CAC	AAA	GAC	TAC	TCT	GCT	GAC	ACC	CTG	GCT	CAC	CAC	624
Asn	Leu	Ile	Leu	His	Lys	Asp	Tyr	Ser	Ala	Asp	Thr	Leu	Ala	His	His	
			195				200					205				
AAC	GAC	ATC	GCT	CTG	CTA	AAA	ATC	CGT	TCT	AAA	GAA	GGT	CGT	TGC	GCT	672
Asn	Asp	Ile	Ala	Leu	Leu	Lys	Ile	Arg	Ser	Lys	Glu	Gly	Arg	Cys	Ala	
	210					215					220					
CAG	CCG	TCT	CGT	ACC	ATC	CAG	ACC	ATC	TGC	CTG	CCG	TCT	ATG	TAC	AAC	720
Gln	Pro	Ser	Arg	Thr	Ile	Gln	Thr	Ile	Cys	Leu	Pro	Ser	Met	Tyr	Asn	
	225				230					235					240	
GAC	CCG	CAG	TTC	GGT	ACC	TCT	TGC	GAA	ATC	ACC	GGT	TTC	GGT	AAA	GAA	768
Asp	Pro	Gln	Phe	Gly	Thr	Ser	Cys	Glu	Ile	Thr	Gly	Phe	Gly	Lys	Glu	
				245					250					255		
AAC	TCT	ACC	GAC	TAC	CTG	TAC	CCG	GAA	CAG	CTG	AAA	ATG	ACC	GTT	GTT	816
Asn	Ser	Thr	Asp	Tyr	Leu	Tyr	Pro	Glu	Gln	Leu	Lys	Met	Thr	Val	Val	
			260					265					270			
AAA	CTG	ATC	TCT	CAC	CGT	GAA	TGC	CAG	CAG	CCG	CAC	TAC	TAC	GGT	TCT	864
Lys	Leu	Ile	Ser	His	Arg	Glu	Cys	Gln	Gln	Pro	His	Tyr	Tyr	Gly	Ser	
		275					280						285			

GAA Glu 290	GTT Val	ACC Thr	ACC Thr	AAA Lys	ATG Met	CTG Leu	TGC Cys	GCT Ala	GCT Ala	GAC Asp	CCG Pro	CAG Gln	TGG Trp	AAA Lys	ACC Thr	912
GAC Asp 305	TCT Ser	TGC Cys	CAA Gln	GGT Gly	GAC Asp	TCT Ser	GGT Gly	GGT Gly	CCA Pro	CTA Leu	GTT Val	TGC Cys	TCT Ser	CTC Leu	CAG Gln	960
GGT Gly	CGT Arg	ATG Met	ACC Thr	CTG Leu	ACC Thr	GGT Gly	ATT Ile	GTT Val	TCT Ser	TGG Trp	GGT Gly	CGT Arg	GGT Gly	TGC Cys	GCT Ala	1008
CTG Leu	AAA Lys	GAC Asp	AAA Lys	CCG Pro	GGT Gly	GTT Val	TAC Tyr	ACC Thr	CGT Arg	GTT Val	TCT Ser	CAC His	TTC Phe	CTG Leu	CCG Pro	1056
TGG Trp	ATC Ile	CGT Arg	TCT Ser	CAC His	ACC Thr	AAA Lys	GAA Glu	GAA Glu	AAC Asn	GGT Gly	CTG Leu	GCT Ala	CTG Leu	AGC Ser	CCG Pro	1104
GTT Val	GAA Glu	GTT Val	TTC Phe	CCG Pro	CGG Arg	CCG Pro	CTG Leu	GGT Gly	GGT Gly	GGT Gly	GGT Gly	AAC Asn	GGT Gly	GAC Asp	TTC Phe	1152
GAA Glu	GAA Glu	ATC Ile	CCG Pro	GAA Glu	GAA Glu	TAC Tyr	CTG Leu	CAA Gln	TAA *							1182

(2) INFORMAÇÃO PARA SEQ ID NO: 8:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 394 aminoácidos
- (B) TIPO: aminoácido
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 8:

Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys	1	5	10	15
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala	20	25	30	
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln	35	40	45	
Leu	Gly	Leu	Gly	Lys	His	Asn	Tyr	Cys	Arg	Asn	Pro	Asp	Asn	Arg	Arg	50	55	60	
Arg	Pro	Trp	Cys	Tyr	Val	Gln	Val	Gly	Leu	Lys	Pro	Leu	Val	Gln	Glu	65	70	75	80

Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
85 90 95

Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
100 105 110

Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
115 120 125

Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
130 135 140

Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
145 150 155 160

Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
165 170 175

Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190

Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205

Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220

Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

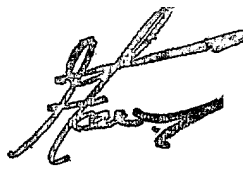
Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

Val Glu Val Phe Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
370 375 380

Glu Glu Ile Pro Glu Glu Tyr Leu Gln *
385 390



(2) INFORMAÇÃO PARA SEQ ID NO: 9:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1182 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1182
- (D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada proteoliticamente em E.coli"
/product = "gene sintético para a proteína M28 (pSJ105)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 9:

ATG	AGC	AAA	ACT	TGC	TAC	GAA	GGT	AAC	GGT	CAC	TTC	TAC	CGT	GGT	AAG	48
Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys	
1				5					10					15		
GCT	TCT	ACC	GAC	ACC	ATG	GGT	CGT	CCG	TGC	CTG	CCG	TGG	AAC	TCT	GCT	95
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala	
			20					25					30			
ACC	GTT	CTG	CAG	CAG	ACC	TAC	CAC	GCT	CAC	CGT	TCT	GAT	GCA	TTG	CAG	144
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln	
		35					40					45				
CTG	GGT	CTG	GGT	AAA	CAC	AAC	TAC	TGC	CGT	AAC	CCG	GAC	AAC	CGT	CGT	192
Leu	Gly	Leu	Gly	Lys	His	Asn	Tyr	Cys	Arg	Asn	Pro	Asp	Asn	Arg	Arg	
	50					55					60					
CGT	CCG	TGG	TGC	TAC	GTT	CAG	GTT	GGT	CTG	AAA	CCG	CTA	GTT	CAG	GAA	240
Arg	Pro	Trp	Cys	Tyr	Val	Gln	Val	Gly	Leu	Lys	Pro	Leu	Val	Gln	Glu	
65					70					75					80	
TGC	ATG	GTT	CAC	GAC	TGC	GCT	GAC	GGT	AAA	AAA	CCG	TCT	TCT	CCG	CCG	288
Cys	Met	Val	His	Asp	Cys	Ala	Asp	Gly	Lys	Lys	Pro	Ser	Ser	Pro	Pro	
				85					90					95		



GAA Glu	GAG Glu	CTC Leu	AAA Lys 100	TTC Phe	CAG Gln	TGC Cys	GGT Gly	CAA Gln 105	AAA Lys	ACC Thr	CTA Leu	CGT Arg	CCG Pro 110	CGT Arg	TTT Phe	336
AAA Lys	ATC Ile	ATC Ile 115	GGT Gly	GGT Gly	GAG Glu	TTC Phe	ACC Thr 120	ACC Thr	ATC Ile	GAA Glu	AAC Asn	CAG Gln 125	CCG Pro	TGG Trp	TTC Phe	384
GCT Ala	GCT Ala	ATC Ile	TAC Tyr	CGT Arg	CGT Arg	CAC His 135	CGT Arg	GGT Gly	GGT Gly	TCT Ser	GTT Val 140	ACC Thr	TAC Tyr	GTT Val	TGC Cys	432
GGT Gly	GGT Gly	TCT Ser	CTG Leu	ATC Ile	TCT Ser 150	CCG Pro	TGC Cys	TGG Trp	GTT Val	ATC Ile 155	TCT Ser	GCT Ala	ACC Thr	CAC His	TGC Cys 160	480
TTC Phe	ATC Ile	GAC Asp	TAC Tyr	CCG Pro 165	AAA Lys	AAA Lys	GAA Glu	GAC Asp	TAC Tyr 170	ATC Ile	GTT Val	TAC Tyr	CTC Leu	GGC Gly 175	CGT Arg	528
TCT Ser	CGT Arg	TTA Leu 180	AAC Asn	TCT Ser	AAC Asn	ACC Thr	CAG Gln	GGT Gly 185	GAA Glu	ATG Met	AAA Lys	TTC Phe	GAA Glu 190	GTT Val	GAA Glu	576
AAC Asn	CTG Leu 195	ATC Ile	CTG Leu	CAC His	AAA Lys	GAC Asp	TAC Tyr 200	TCT Ser	GCT Ala	GAC Asp	ACC Thr 205	CTG Leu	GCT Ala	CAC His	CAC His	624
AAC Asn 210	GAC Asp	ATC Ile	GCT Ala	CTG Leu	CTA Leu	AAA Lys 215	ATC Ile	CGT Arg	TCT Ser	AAA Lys	GAA Glu 220	GGT Gly	CGT Arg	TGC Cys	GCT Ala	572
CAG Gln 225	CCG Pro	TCT Ser	CGT Arg	ACC Thr 230	ATC Ile	CAG Gln	ACC Thr	ATC Ile	TGC Cys	CTG Leu 235	CCG Pro	TCT Ser	ATG Met	TAC Tyr	AAC Asn 240	720
GAC Asp	CCG Pro	CAG Gln	TTC Phe	GGT Gly 245	ACC Thr	TCT Ser	TGC Cys	GAA Glu	ATC Ile	ACC Thr	GGT Gly	TTC Phe	GGT Gly	AAA Lys 255	GAA Glu	768
AAC Asn	TCT Ser	ACC Thr 260	GAC Asp	TAC Tyr	CTG Leu	TAC Tyr	CCG Pro	GAA Glu 265	CAG Gln	CTG Leu	AAA Lys	ATG Met	ACC Thr 270	GTT Val	GTT Val	816
AAA Lys	CTG Leu	ATC Ile	TCT Ser	CAC His	CGT Arg	GAA Glu	TGC Cys	CAG Gln 280	CAG Gln	CCG Pro	CAC His	TAC Tyr 285	TAC Tyr	GGT Gly	TCT Ser	864
GAA Glu	GTT Val	ACC Thr	ACC Thr	AAA Lys	ATG Met	CTG Leu 295	TGC Cys	GCT Ala	GCT Ala	GAC Asp	CCG Pro 300	CAG Gln	TGG Trp	AAA Lys	ACC Thr	912
GAC Asp 305	TCT Ser	TGC Cys	CAA Gln	GGT Gly	GAC Asp	TCT Ser	GGT Gly	GGT Gly	CCA Pro	CTA Leu	GTT Val	TGC Cys	TCT Ser	CTC Leu	CAG Gln 320	960
GGT Gly	CGT Arg	ATG Met	ACC Thr	CTG Leu 325	ACC Thr	GGT Gly	ATT Ile	GTT Val	TCT Ser	TGG Trp 330	GGT Gly	CGT Arg	GGT Gly	TGC Cys 335	GCT Ala	1008
CTG Leu	AAA Lys	GAC Asp	AAA Lys	CCG Pro	GGT Gly	GTT Val	TAC Tyr	ACC Thr 345	CGT Arg	GTT Val	TCT Ser	CAC His	TTC Phe	CTG Leu	CCG Pro	1056

TGG	ATC	CGT	TCT	CAC	ACC	AAA	GAA	GAA	AAC	GGT	CTG	GCT	CTG	AGC	CCG	1104
Trp	Ile	Arg	Ser	His	Thr	Lys	Glu	Glu	Asn	Gly	Leu	Ala	Leu	Ser	Pro	
		355					360					365				
GTT	GTA	GTT	GTA	GTG	CGG	CCG	CTG	GGT	GGT	GGT	GGT	AAC	GGT	GAC	TTC	1152
Val	Val	Val	Val	Val	Arg	Pro	Leu	Gly	Gly	Gly	Gly	Asn	Gly	Asp	Phe	
		370				375					380					
GAA	GAA	ATC	CCG	GAA	GAA	TAC	CTG	CAA	TAA							1132
Glu	Glu	Ile	Pro	Glu	Glu	Tyr	Leu	Gln	*							
385						390										

(2) INFORMAÇÃO PARA SEQ ID NO: 10:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 394 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 10:

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

Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
100 105 110

Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
115 120 125

Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
130 135 140

Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
145 150 155

Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
165 170 175

Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190

Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205

Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220

Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

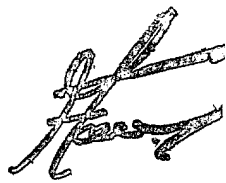
Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

Val Val Val Val Val Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
370 375 380

Glu Glu Ile Pro Glu Glu Tyr Leu Gln *
385 390



(2) INFORMAÇÃO PARA SEQ ID NO: 11:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1191 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

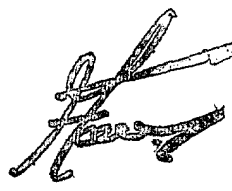
(iv) ANTI-SENTIDO: não

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1191
- (D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada proteoliticamente em E.coli"
/product = "gene sintético para a proteína M29 (pSJ106)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 11:

ATG	AGC	AAA	ACT	TGC	TAC	GAA	GGT	AAC	GGT	CAC	TTC	TAC	CGT	GGT	AAG	48
Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys	
1				5				10						15		
GCT	TCT	ACC	GAC	ACC	ATG	GGT	CGT	CCG	TGC	CTG	CCG	TGG	AAC	TCT	GCT	96
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala	
			20					25					30			
ACC	GTT	CTG	CAG	CAG	ACC	TAC	CAC	GCT	CAC	CGT	TCT	GAT	GCA	TTG	CAG	144
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln	
		35					40					45				
CTG	GGT	CTG	GGT	AAA	CAC	AAC	TAC	TGC	CGT	AAC	CCG	GAC	AAC	CGT	CGT	192
Leu	Gly	Leu	Gly	Lys	His	Asn	Tyr	Cys	Arg	Asn	Pro	Asp	Asn	Arg	Arg	
	50					55					60					
CGT	CCG	TGG	TGC	TAC	GTT	CAG	GTT	GGT	CTG	AAA	CCG	CTA	GTT	CAG	GAA	240
Arg	Pro	Trp	Cys	Tyr	Val	Gln	Val	Gly	Leu	Lys	Pro	Leu	Val	Gln	Glu	
	65				70					75				80		
TGC	ATG	GTT	CAC	GAC	TGC	GCT	GAC	GGT	AAA	AAA	CCG	TCT	TCT	CCG	CCG	288
Cys	Met	Val	His	Asp	Cys	Ala	Asp	Gly	Lys	Lys	Pro	Ser	Ser	Pro	Pro	
				85						90				95		



GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe 100 105 110	336
AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe 115 120 125	384
GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys 130 135 140	432
GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys 145 150 155 160	480
TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg 165 170 175	528
TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu 180 185 190	576
AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His 195 200 205	624
AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala 210 215 220	672
CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn 225 230 235 240	720
GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu 245 250 255	768
AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val 260 265 270	816
AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser 275 280 285	864
GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr 290 295 300	912
GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln 305 310 315 320	960
GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala 325 330 335	1008

CTG	AAA	GAC	AAA	CCG	GGT	GTT	TAC	ACC	CGT	GTT	TCT	CAC	TTC	CTG	CCG	1056
Leu	Lys	Asp	Lys	Pro	Gly	Val	Tyr	Thr	Arg	Val	Ser	His	Phe	Leu	Pro	
			340					345					350			
TGG	ATC	CGT	TCT	CAC	ACC	AAA	GAA	GAA	AAC	GGT	CTG	GCT	CTG	AGC	CCG	1104
Trp	Ile	Arg	Ser	His	Thr	Lys	Glu	Glu	Asn	Gly	Leu	Ala	Leu	Ser	Pro	
		355					360					365				
GTT	GTT	GCT	TTC	CCG	CGG	CCG	TTC	CTG	CTG	CGG	AAC	CCG	AAC	GAC	AAA	1152
Val	Val	Ala	Phe	Pro	Arg	Pro	Phe	Leu	Leu	Arg	Asn	Pro	Asn	Asp	Lys	
	370					375					380					
TAC	GAA	CCG	TTC	TGG	GAA	GAC	GAA	GAA	AAA	AAC	GAA	TAA				1191
Tyr	Glu	Pro	Phe	Trp	Glu	Asp	Glu	Glu	Lys	Asn	Glu	*				
385					390					395						

(2) INFORMAÇÃO PARA SEQ ID NO: 12:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 397 aminoácidos
- (B) TIPO: aminoácido
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 12:

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

Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
165 170 175

Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190

Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205

Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220

Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285 295

Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

Val Val Ala Phe Pro Arg Pro Phe Leu Leu Arg Asn Pro Asn Asp Lys
370 375 380

Tyr Glu Pro Phe Trp Glu Asp Glu Glu Lys Asn Glu *
385 390 395

(2) INFORMAÇÃO PARA SEQ ID NO: 13:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1182 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico



(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: CDS

(B) LOCALIZAÇÃO: 1..1182

(D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada proteoliticamente em E.coli"

/product = "gene sintético para a proteína M30 (pSJ109)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 13:

ATG	AGC	AAA	ACT	TGC	TAC	GAA	GGT	AAC	GGT	CAC	TTC	TAC	CGT	GGT	AAG	48
Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys	
1				5					10					15		
GCT	TCT	ACC	GAC	ACC	ATG	GGT	CGT	CCG	TGC	CTG	CCG	TGG	AAC	TCT	GCT	96
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala	
			20					25					30			
ACC	GTT	CTG	CAG	CAG	ACC	TAC	CAC	GCT	CAC	CGT	TCT	GAT	GCA	TTG	CAG	144
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln	
		35					40					45				
CTG	GGT	CTG	GGT	AAA	CAC	AAC	TAC	TGC	CGT	AAC	CCG	GAC	AAC	CGT	CGT	192
Leu	Gly	Leu	Gly	Lys	His	Asn	Tyr	Cys	Arg	Asn	Pro	Asp	Asn	Arg	Arg	
	50					55					60					
CGT	CCG	TGG	TGC	TAC	GTT	CAG	GTT	GGT	CTG	AAA	CCG	CTA	GTT	CAG	GAA	240
Arg	Pro	Trp	Cys	Tyr	Val	Gln	Val	Gly	Leu	Lys	Pro	Leu	Val	Gln	Glu	
65					70					75					80	
TGC	ATG	GTT	CAC	GAC	TGC	GCT	GAC	GGT	AAA	AAA	CCG	TCT	TCT	CCG	CCG	288
Cys	Met	Val	His	Asp	Cys	Ala	Asp	Gly	Lys	Lys	Pro	Ser	Ser	Pro	Pro	
				85					90					95		
GAA	GAG	CTC	AAA	TTC	CAG	TGC	GGT	CAA	AAA	ACC	CTA	CGT	CCG	CGT	TTT	336
Glu	Glu	Leu	Lys	Phe	Gln	Cys	Gly	Gln	Lys	Thr	Leu	Arg	Pro	Arg	Phe	
			100					105					110			
AAA	ATC	ATC	GGT	GGT	GAG	TTC	ACC	ACC	ATC	GAA	AAC	CAG	CCG	TGG	TTC	384
Lys	Ile	Ile	Gly	Gly	Glu	Phe	Thr	Thr	Ile	Glu	Asn	Gln	Pro	Trp	Phe	
			115				120					125				
GCT	GCT	ATC	TAC	CGT	CGT	CAC	CGT	GGT	GGT	TCT	GTT	ACC	TAC	GTT	TGC	432
Ala	Ala	Ile	Tyr	Arg	Arg	His	Arg	Gly	Gly	Ser	Val	Thr	Tyr	Val	Cys	
	130					135					140					
GGT	GGT	TCT	CTG	ATC	TCT	CCG	TGC	TGG	GTT	ATC	TCT	GCT	ACC	CAC	TGC	480
Gly	Gly	Ser	Leu	Ile	Ser	Pro	Cys	Trp	Val	Ile	Ser	Ala	Thr	His	Cys	
145					150					155					160	

(2) INFORMAÇÃO PARA SEQ ID NO: 14:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 394 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 14:

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

Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
130 135 140

Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
145 150 155 160

Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
165 170 175

Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190

Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205

Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220

Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Val
355 360 365

Val Val Ala Phe Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
370 375 380

Glu Glu Ile Pro Glu Glu Tyr Leu Gln *
385 390

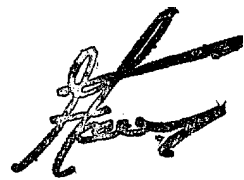
(2) INFORMAÇÃO PARA SEQ ID NO: 15:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1182 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não



(iv) ANTI-SENTIDO: não

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: CDS

(B) LOCALIZAÇÃO: 1..1182

(D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada proteoliticamente em E.coli"

/product = "gene sintético para a proteína M31 (pSJ111)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 15:

ATG	AGC	AAA	ACT	TGC	TAC	GAA	GGT	AAC	GGT	CAC	TTC	TAC	CGT	GGT	AAG	48
Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys	
1				5					10					15		
GCT	TCT	ACC	GAC	ACC	ATG	GGT	CGT	CCG	TGC	CTG	CCG	TGG	AAC	TCT	GCT	96
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala	
			20					25					30			
ACC	GTT	CTG	CAG	CAG	ACC	TAC	CAC	GCT	CAC	CGT	TCT	GAT	GCA	TTG	CAG	144
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln	
		35					40					45				
CTG	GGT	CTG	GGT	AAA	CAC	AAC	TAC	TGC	CGT	AAC	CCG	GAC	AAC	CGT	CGT	192
Leu	Gly	Leu	Gly	Lys	His	Asn	Tyr	Cys	Arg	Asn	Pro	Asp	Asn	Arg	Arg	
	50					55					60					
CGT	CCG	TGG	TGC	TAC	GTT	CAG	GTT	GGT	CTG	AAA	CCG	CTA	GTT	CAG	GAA	240
Arg	Pro	Trp	Cys	Tyr	Val	Gln	Val	Gly	Leu	Lys	Pro	Leu	Val	Gln	Glu	
65					70					75					80	
TGC	ATG	GTT	CAC	GAC	TGC	GCT	GAC	GGT	AAA	AAA	CCG	TCT	TCT	CCG	CCG	288
Cys	Met	Val	His	Asp	Cys	Ala	Asp	Gly	Lys	Lys	Pro	Ser	Ser	Pro	Pro	
				85					90					95		
GAA	GAG	CTC	AAA	TTC	CAG	TGC	GGT	CAA	AAA	ACC	CTA	CGT	CCG	CGT	TTT	336
Glu	Glu	Leu	Lys	Phe	Gln	Cys	Gly	Gln	Lys	Thr	Leu	Arg	Pro	Arg	Phe	
			100					105					110			
AAA	ATC	ATC	GGT	GGT	GAG	TTC	ACC	ACC	ATC	GAA	AAC	CAG	CCG	TGG	TTC	384
Lys	Ile	Ile	Gly	Gly	Glu	Phe	Thr	Thr	Ile	Glu	Asn	Gln	Pro	Trp	Phe	
			115			120						125				
GCT	GCT	ATC	TAC	CGT	CGT	CAC	CGT	GGT	GGT	TCT	GTT	ACC	TAC	GTT	TGC	432
Ala	Ala	Ile	Tyr	Arg	Arg	His	Arg	Gly	Gly	Ser	Val	Thr	Tyr	Val	Cys	
	130					135					140					
GGT	GGT	TCT	CTG	ATC	TCT	CCG	TGC	TGG	GTT	ATC	TCT	GCT	ACC	CAC	TGC	480
Gly	Gly	Ser	Leu	Ile	Ser	Pro	Cys	Trp	Val	Ile	Ser	Ala	Thr	His	Cys	
145					150					155					160	
TTC	ATC	GAC	TAC	CCG	AAA	AAA	GAA	GAC	TAC	ATC	GTT	TAC	CTC	GGC	CGT	528
Phe	Ile	Asp	Tyr	Pro	Lys	Lys	Glu	Asp	Tyr	Ile	Val	Tyr	Leu	Gly	Arg	
				165					170					175		
TCT	CGT	TTA	AAC	TCT	AAC	ACC	CAG	GGT	GAA	ATG	AAA	TTC	GAA	GTT	GAA	576
Ser	Arg	Leu	Asn	Ser	Asn	Thr	Gln	Gly	Glu	Met	Lys	Phe	Glu	Val	Glu	
			180					185					190			

AAC	CTG	ATC	CTG	CAC	AAA	GAC	TAC	TCT	GCT	GAC	ACC	CTG	GCT	CAC	CAC	624
Asn	Leu	Ile	Leu	His	Lys	Asp	Tyr	Ser	Ala	Asp	Thr	Leu	Ala	His	His	
		195					200					205				
AAC	GAC	ATC	GCT	CTG	CTA	AAA	ATC	CGT	TCT	AAA	GAA	GGT	CGT	TGC	GCT	672
Asn	Asp	Ile	Ala	Leu	Leu	Lys	Ile	Arg	Ser	Lys	Glu	Gly	Arg	Cys	Ala	
	210					215					220					
CAG	CCG	TCT	CGT	ACC	ATC	CAG	ACC	ATC	TGC	CTG	CCG	TCT	ATG	TAC	AAC	720
Gln	Pro	Ser	Arg	Thr	Ile	Gln	Thr	Ile	Cys	Leu	Pro	Ser	Met	Tyr	Asn	
	225				230					235					240	
GAC	CCG	CAG	TTC	GGT	ACC	TCT	TGC	GAA	ATC	ACC	GGT	TTC	GGT	AAA	GAA	768
Asp	Pro	Gln	Phe	Gly	Thr	Ser	Cys	Glu	Ile	Thr	Gly	Phe	Gly	Lys	Glu	
				245					250					255		
AAC	TCT	ACC	GAC	TAC	CTG	TAC	CCG	GAA	CAG	CTG	AAA	ATG	ACC	GTT	GTT	816
Asn	Ser	Thr	Asp	Tyr	Leu	Tyr	Pro	Glu	Gln	Leu	Lys	Met	Thr	Val	Val	
			260					265					270			
AAA	CTG	ATC	TCT	CAC	CGT	GAA	TGC	CAG	CAG	CCG	CAC	TAC	TAC	GGT	TCT	864
Lys	Leu	Ile	Ser	His	Arg	Glu	Cys	Gln	Gln	Pro	His	Tyr	Tyr	Gly	Ser	
		275					280					285				
GAA	GTT	ACC	ACC	AAA	ATG	CTG	TGC	GCT	GCT	GAC	CCG	CAG	TGG	AAA	ACC	912
Glu	Val	Thr	Thr	Lys	Met	Leu	Cys	Ala	Ala	Asp	Pro	Gln	Trp	Lys	Thr	
	290					295					300					
GAC	TCT	TGC	CAA	GGT	GAC	TCT	GGT	GGT	CCA	CTA	GTT	TGC	TCT	CTC	CAG	960
Asp	Ser	Cys	Gln	Gly	Asp	Ser	Gly	Gly	Pro	Leu	Val	Cys	Ser	Leu	Gln	
	305				310					315					320	
GGT	CGT	ATG	ACC	CTG	ACC	GGT	ATT	GTT	TCT	TGG	GGT	CGT	GGT	TGC	GCT	1008
Gly	Arg	Met	Thr	Leu	Thr	Gly	Ile	Val	Ser	Trp	Gly	Arg	Gly	Cys	Ala	
				325					330					335		
CTG	AAA	GAC	AAA	CCG	GGT	GTT	TAC	ACC	CGT	GTT	TCT	CAC	TTC	CTG	CCG	1056
Leu	Lys	Asp	Lys	Pro	Gly	Val	Tyr	Thr	Arg	Val	Ser	His	Phe	Leu	Pro	
			340					345					350			
TGG	ATC	CGT	TCT	CAC	ACC	AAA	GAA	GAA	AAC	GGT	CTG	GCT	CTG	AGC	CTG	1104
Trp	Ile	Arg	Ser	His	Thr	Lys	Glu	Glu	Asn	Gly	Leu	Ala	Leu	Ser	Leu	
		355					360					365				
GTT	AAA	GCT	TTC	CCG	CGG	CCG	CTG	GGT	GGT	GGT	GGT	AAC	GGT	GAC	TTC	1152
Val	Lys	Ala	Phe	Pro	Arg	Pro	Leu	Gly	Gly	Gly	Gly	Asn	Gly	Asp	Phe	
	370					375					380					
GAA	GAA	ATC	CCG	GAA	GAA	TAC	CTG	CAA	TAA							1182
Glu	Glu	Ile	Pro	Glu	Glu	Tyr	Leu	Gln	*							
	385				390											

(2) INFORMAÇÃO PARA SEQ ID NO: 16:

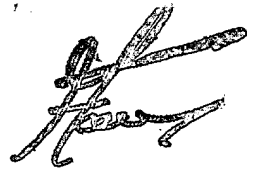
(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 394 aminoácidos
- (B) TIPO: aminoácido
- (D) TOPOLOGIA: linear

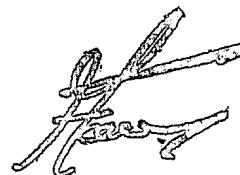
(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 16:

Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
1 5 10 15
Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
20 25 30
Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
35 40 45
Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
50 55 60
Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
65 70 75 80
Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
85 90 95
Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
100 105 110
Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
115 120 125
Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
130 135 140
Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
145 150 155 160
Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
165 170 175
Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190
Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205
Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220
Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240
Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255
Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270
Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285
Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300
Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320
Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335



GAA	GAG	CTC	AAA	TTC	CAG	TGC	GGT	CAA	AAA	ACC	CTA	CGT	CCG	CGT	TTT	336
Glu	Glu	Leu	Lys	Phe	Gln	Cys	Gly	Gln	Lys	Thr	Leu	Arg	Pro	Arg	Phe	
			100					105					110			
AAA	ATC	ATC	GGT	GGT	GAG	TTC	ACC	ACC	ATC	GAA	AAC	CAG	CCG	TGG	TTC	384
Lys	Ile	Ile	Gly	Gly	Glu	Phe	Thr	Thr	Ile	Glu	Asn	Gln	Pro	Trp	Phe	
			115				120					125				
GCT	GCT	ATC	TAC	CGT	CGT	CAC	CGT	GGT	GGT	TCT	GTT	ACC	TAC	GTT	TGC	432
Ala	Ala	Ile	Tyr	Arg	Arg	His	Arg	Gly	Gly	Ser	Val	Thr	Tyr	Val	Cys	
			130			135					140					
GGT	GGT	TCT	CTG	ATC	TCT	CCG	TGC	TGG	GTT	ATC	TCT	GCT	ACC	CAC	TGC	480
Gly	Gly	Ser	Leu	Ile	Ser	Pro	Cys	Trp	Val	Ile	Ser	Ala	Thr	His	Cys	
145					150					155					160	
TTC	ATC	GAC	TAC	CCG	AAA	AAA	GAA	GAC	TAC	ATC	GTT	TAC	CTC	GGC	CGT	528
Phe	Ile	Asp	Tyr	Pro	Lys	Lys	Glu	Asp	Tyr	Ile	Val	Tyr	Leu	Gly	Arg	
				165					170					175		
TCT	CGT	TTA	AAC	TCT	AAC	ACC	CAG	GGT	GAA	ATG	AAA	TTC	GAA	GTT	GAA	576
Ser	Arg	Leu	Asn	Ser	Asn	Thr	Gln	Gly	Glu	Met	Lys	Phe	Glu	Val	Glu	
			180					185					190			
AAC	CTG	ATC	CTG	CAC	AAA	GAC	TAC	TCT	GCT	GAC	ACC	CTG	GCT	CAC	CAC	624
Asn	Leu	Ile	Leu	His	Lys	Asp	Tyr	Ser	Ala	Asp	Thr	Leu	Ala	His	His	
			195			200						205				
AAC	GAC	ATC	GCT	CTG	CTA	AAA	ATC	CGT	TCT	AAA	GAA	GGT	CGT	TGC	GCT	672
Asn	Asp	Ile	Ala	Leu	Leu	Lys	Ile	Arg	Ser	Lys	Glu	Gly	Arg	Cys	Ala	
	210					215					220					
CAG	CCG	TCT	CGT	ACC	ATC	CAG	ACC	ATC	TGC	CTG	CCG	TCT	ATG	TAC	AAC	720
Gln	Pro	Ser	Arg	Thr	Ile	Gln	Thr	Ile	Cys	Leu	Pro	Ser	Met	Tyr	Asn	
225					230					235					240	
GAC	CCG	CAG	TTC	GGT	ACC	TCT	TGC	GAA	ATC	ACC	GGT	TTC	GGT	AAA	GAA	768
Asp	Pro	Gln	Phe	Gly	Thr	Ser	Cys	Glu	Ile	Thr	Gly	Phe	Gly	Lys	Glu	
				245					250					255		
AAC	TCT	ACC	GAC	TAC	CTG	TAC	CCG	GAA	CAG	CTG	AAA	ATG	ACC	GTT	GTT	816
Asn	Ser	Thr	Asp	Tyr	Leu	Tyr	Pro	Glu	Gln	Leu	Lys	Met	Thr	Val	Val	
			260					265					270			
AAA	CTG	ATC	TCT	CAC	CGT	GAA	TGC	CAG	CAG	CCG	CAC	TAC	TAC	GGT	TCT	864
Lys	Leu	Ile	Ser	His	Arg	Glu	Cys	Gln	Gln	Pro	His	Tyr	Tyr	Gly	Ser	
		275				280						285				
GAA	GTT	ACC	ACC	AAA	ATG	CTG	TGC	GCT	GCT	GAC	CCG	CAG	TGG	AAA	ACC	912
Glu	Val	Thr	Thr	Lys	Met	Leu	Cys	Ala	Ala	Asp	Pro	Gln	Trp	Lys	Thr	
	290					295					300					
GAC	TCT	TGC	CAA	GGT	GAC	TCT	GGT	GGT	CCA	CTA	GTT	TGC	TCT	CTC	CAG	960
Asp	Ser	Cys	Gln	Gly	Asp	Ser	Gly	Gly	Pro	Leu	Val	Cys	Ser	Leu	Gln	
305				310						315					320	
GGT	CGT	ATG	ACC	CTG	ACC	GGT	ATT	GTT	TCT	TGG	GGT	CGT	GGT	TGC	GCT	1008
Gly	Arg	Met	Thr	Leu	Thr	Gly	Ile	Val	Ser	Trp	Gly	Arg	Gly	Cys	Ala	
				325				330						335		
CTG	AAA	GAC	AAA	CCG	GGT	GTT	TAC	ACC	CGT	GTT	TCT	CAC	TTC	CTG	CCG	1056
Leu	Lys	Asp	Lys	Pro	Gly	Val	Tyr	Thr	Arg	Val	Ser	His	Phe	Leu	Pro	
			340					345					350			



TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG TTC CCG	1104
Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Phe Pro	
355 360 365	
CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC TTC GAA GAA ATC CCG GAA	1152
Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe Glu Glu Ile Pro Glu	
370 375 380	
GAA TAC CTG CAA TAA	1167
Glu Tyr Leu Gln *	
385	

(2) INFORMAÇÃO PARA SEQ ID NO: 18:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 389 aminoácidos

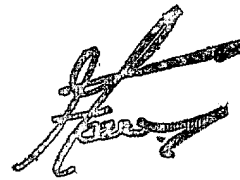
(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 18:

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



Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Phe Pro
355 360 365

Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe Glu Glu Ile Pro Glu
370 375 380

Glu Tyr Leu Gln *
385

2) INFORMAÇÃO PARA SEQ ID NO: 19:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1191 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1191
- (D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada proteoliticamente em E.coli"
/product = "gene sintético para a proteína M32 (pSJ114)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 19:

ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG
Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
1 5 10 15



61																
GCT	TCT	ACC	GAC	ACC	ATG	GGT	CGT	CCG	TGC	CTG	CCG	TGG	AAC	TCT	GCT	96
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala	
			20					25					30			
ACC	GTT	CTG	CAG	CAG	ACC	TAC	CAC	GCT	CAC	CGT	TCT	GAT	GCA	TTG	CAG	144
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln	
		35					40					45				
CTG	GGT	CTG	GGT	AAA	CAC	AAC	TAC	TGC	CGT	AAC	CCG	GAC	AAC	CGT	CGT	192
Leu	Gly	Leu	Gly	Lys	His	Asn	Tyr	Cys	Arg	Asn	Pro	Asp	Asn	Arg	Arg	
	50					55					60					
CGT	CCG	TGG	TGC	TAC	GTT	CAG	GTT	GGT	CTG	AAA	CCG	CTA	GTT	CAG	GAA	240
Arg	Pro	Trp	Cys	Tyr	Val	Gln	Val	Gly	Leu	Lys	Pro	Leu	Val	Gln	Glu	
	65				70					75					80	
TGC	ATG	GTT	CAC	GAC	TGC	GCT	GAC	GGT	AAA	AAA	CCG	TCT	TCT	CCG	CCG	288
Cys	Met	Val	His	Asp	Cys	Ala	Asp	Gly	Lys	Lys	Pro	Ser	Ser	Pro	Pro	
				85					90					95		
GAA	GAG	CTC	AAA	TTC	CAG	TGC	GGT	CAA	AAA	ACC	CTA	CGT	CCG	CGT	TTT	336
Glu	Glu	Leu	Lys	Phe	Gln	Cys	Gly	Gln	Lys	Thr	Leu	Arg	Pro	Arg	Phe	
			100					105					110			
AAA	ATC	ATC	GGT	GGT	GAG	TTC	ACC	ACC	ATC	GAA	AAC	CAG	CCG	TGG	TTC	384
Lys	Ile	Ile	Gly	Gly	Glu	Phe	Thr	Thr	Ile	Glu	Asn	Gln	Pro	Trp	Phe	
		115					120					125				
GCT	GCT	ATC	TAC	CGT	CGT	CAC	CGT	GGT	GGT	TCT	GTT	ACC	TAC	GTT	TGC	432
Ala	Ala	Ile	Tyr	Arg	Arg	His	Arg	Gly	Gly	Ser	Val	Thr	Tyr	Val	Cys	
	130					135					140					
GGT	GGT	TCT	CTG	ATC	TCT	CCG	TGC	TGG	GTT	ATC	TCT	GCT	ACC	CAC	TGC	480
Gly	Gly	Ser	Leu	Ile	Ser	Pro	Cys	Trp	Val	Ile	Ser	Ala	Thr	His	Cys	
145					150					155					160	
TTC	ATC	GAC	TAC	CCG	AAA	AAA	GAA	GAC	TAC	ATC	GTT	TAC	CTC	GGC	CGT	528
Phe	Ile	Asp	Tyr	Pro	Lys	Lys	Glu	Asp	Tyr	Ile	Val	Tyr	Leu	Gly	Arg	
				165					170					175		
TCT	CGT	TTA	AAC	TCT	AAC	ACC	CAG	GGT	GAA	ATG	AAA	TTC	GAA	GTT	GAA	576
Ser	Arg	Leu	Asn	Ser	Asn	Thr	Gln	Gly	Glu	Met	Lys	Phe	Glu	Val	Glu	
			180					185					190			
AAC	CTG	ATC	CTG	CAC	AAA	GAC	TAC	TCT	GCT	GAC	ACC	CTG	GCT	CAC	CAC	624
Asn	Leu	Ile	Leu	His	Lys	Asp	Tyr	Ser	Ala	Asp	Thr	Leu	Ala	His	His	
		195					200					205				
AAC	GAC	ATC	GCT	CTG	CTA	AAA	ATC	CGT	TCT	AAA	GAA	GGT	CGT	TGC	GCT	672
Asn	Asp	Ile	Ala	Leu	Leu	Lys	Ile	Arg	Ser	Lys	Glu	Gly	Arg	Cys	Ala	
	210					215					220					
CAG	CCG	TCT	CGT	ACC	ATC	CAG	ACC	ATC	TGC	CTG	CCG	TCT	ATG	TAC	AAC	720
Gln	Pro	Ser	Arg	Thr	Ile	Gln	Thr	Ile	Cys	Leu	Pro	Ser	Met	Tyr	Asn	
225					230					235					240	
GAC	CCG	CAG	TTC	GGT	ACC	TCT	TGC	GAA	ATC	ACC	GGT	TTC	GGT	AAA	GAA	768
Asp	Pro	Gln	Phe	Gly	Thr	Ser	Cys	Glu	Ile	Thr	Gly	Phe	Gly	Lys	Glu	
				245				250						255		
AAC	TCT	ACC	GAC	TAC	CTG	TAC	CCG	GAA	CAG	CTG	AAA	ATG	ACC	GTT	GTT	816
Asn	Ser	Thr	Asp	Tyr	Leu	Tyr	Pro	Glu	Gln	Leu	Lys	Met	Thr	Val	Val	
			260					265						270		



AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT	864
Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser	
275 280 285	
GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC	912
Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr	
290 295 300	
GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG	960
Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln	
305 310 315 320	
GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT	1008
Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala	
325 330 335	
CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG	1056
Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro	
340 345 350	
TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG	1104
Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro	
355 360 365	
GTT AAG GCT TTC CCG CGG CCG TTC CTG CTG CCG AAC CCG AAC GAC AAA	1152
Val Lys Ala Phe Pro Arg Pro Phe Leu Leu Arg Asn Pro Asn Asp Lys	
370 375 380	
TAC GAA CCG TTC TGG GAA GAC GAA GAA AAA AAC GAA TAA	1191
Tyr Glu Pro Phe Trp Glu Asp Glu Glu Lys Asn Glu *	
385 390 395	

(2) INFORMAÇÃO PARA SEQ ID NO: 20:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 397 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 20:

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

Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
115 120 125

Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
130 135 140

Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
145 150 155 160

Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
165 170 175

Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190

Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205

Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220

Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

Val Lys Ala Phe Pro Arg Pro Phe Leu Leu Arg Asn Pro Asn Asp Lys
370 375 380

Tyr Glu Pro Phe Trp Glu Asp Glu Glu Lys Asn Glu *
385 390 395

(2) INFORMAÇÃO PARA SEQ ID NO: 21:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1098 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: CDS

(B) LOCALIZAÇÃO: 1..1098

(D) OUTRA INFORMAÇÃO: /function= "metionina N-terminal é separada proteoliticamente em E.coli"

/product= "gene sintético para a proteína M4 (pSJ41)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 21:

ATG	AGC	AAA	ACT	TGC	TAC	GAA	GGT	AAC	GGT	CAC	TTC	TAC	CGT	GGT	AAG	48
Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys	
1				5					10					15		
GCT	TCT	ACC	GAC	ACC	ATG	GGT	CGT	CCG	TGC	CTG	CCG	TGG	AAC	TCT	GCT	96
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala	
			20					25					30			
ACC	GTT	CTG	CAG	CAG	ACC	TAC	CAC	GCT	CAC	CGT	TCT	GAT	GCA	TTG	CAG	144
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln	
		35					40					45				
CTG	GGT	CTG	GGT	AAA	CAC	AAC	TAC	TGC	CGT	AAC	CCG	GAC	AAC	CGT	CGT	192
Leu	Gly	Leu	Gly	Lys	His	Asn	Tyr	Cys	Arg	Asn	Pro	Asp	Asn	Arg	Arg	
	50					55					60					
CGT	CCG	TGG	TGC	TAC	GTT	CAG	GTT	GGT	CTG	AAA	CCG	CTA	GTT	CAG	GAA	240
Arg	Pro	Trp	Cys	Tyr	Val	Gln	Val	Gly	Leu	Lys	Pro	Leu	Val	Gln	Glu	
65					70					75					80	
TGC	ATG	GTT	CAC	GAC	TGC	GCT	GAC	GGT	AAA	AAA	CCG	TCT	TCT	CCG	CCG	288
Cys	Met	Val	His	Asp	Cys	Ala	Asp	Gly	Lys	Lys	Pro	Ser	Ser	Pro	Pro	
				85					90					95		
GAA	GAG	CTC	AAA	TTC	CAG	TGC	GGT	CAA	AAA	ACC	CTA	CGT	CCG	CGT	TTT	336
Glu	Glu	Leu	Lys	Phe	Gln	Cys	Gly	Gln	Lys	Thr	Leu	Arg	Pro	Arg	Phe	
			100					105					110			
AAA	ATC	ATC	GGT	GGT	GAG	TTC	ACC	ACC	ATC	GAA	AAC	CAG	CCG	TGG	TTT	384
Lys	Ile	Ile	Gly	Gly	Glu	Phe	Thr	Thr	Ile	Glu	Asn	Gln	Pro	Trp	Phe	
		115					120					125				
GCT	GCT	ATC	TAC	CGT	CGT	CAC	CGT	GGT	GGT	TCT	GTT	ACC	TAC	GTT	TGC	432
Ala	Ala	Ile	Tyr	Arg	Arg	His	Arg	Gly	Gly	Ser	Val	Thr	Tyr	Val	Cys	
	130					135					140					
GGT	GGT	TCT	CTG	ATC	TCT	CCG	TGC	TGG	GTT	ATC	TCT	GCT	ACC	CAC	TGC	480
Gly	Gly	Ser	Leu	Ile	Ser	Pro	Cys	Trp	Val	Ile	Ser	Ala	Thr	His	Cys	
145					150					155					160	
TTC	ATC	GAC	TAC	CCG	AAA	AAA	GAA	GAC	TAC	ATC	GTT	TAC	CTC	GGC	CGT	528
Phe	Ile	Asp	Tyr	Pro	Lys	Lys	Glu	Asp	Tyr	Ile	Val	Tyr	Leu	Gly	Arg	
				165					170					175		
TCT	CGT	TTA	AAC	TCT	AAC	ACC	CAG	GGT	GAA	ATG	AAA	TTC	GAA	GTT	GAA	576
Ser	Arg	Leu	Asn	Ser	Asn	Thr	Gln	Gly	Glu	Met	Lys	Phe	Glu	Val	Glu	
			180					185					190			



AAC	CTG	ATC	CTG	CAC	AAA	GAC	TAC	TCT	GCT	GAC	ACC	CTG	GCT	CAC	CAC	624
Asn	Leu	Ile	Leu	His	Lys	Asp	Tyr	Ser	Ala	Asp	Thr	Leu	Ala	His	His	
		195					200					205				
AAC	GAC	ATC	GCT	CTG	CTA	AAA	ATC	CGT	TCT	AAA	GAA	GGT	CGT	TGC	GCT	672
Asn	Asp	Ile	Ala	Leu	Leu	Lys	Ile	Arg	Ser	Lys	Glu	Gly	Arg	Cys	Ala	
	210					215					220					
CAG	CCG	TCT	CGT	ACC	ATC	CAG	ACC	ATC	TGC	CTG	CCG	TCT	ATG	TAC	AAC	720
Gln	Pro	Ser	Arg	Thr	Ile	Gln	Thr	Ile	Cys	Leu	Pro	Ser	Met	Tyr	Asn	
	225				230					235					240	
GAC	CCG	CAG	TTC	GGT	ACC	TCT	TGC	GAA	ATC	ACC	GGT	TTC	GGT	AAA	GAA	768
Asp	Pro	Gln	Phe	Gly	Thr	Ser	Cys	Glu	Ile	Thr	Gly	Phe	Gly	Lys	Glu	
				245					250					255		
AAC	TCT	ACC	GAC	TAC	CTG	TAC	CCG	GAA	CAG	CTG	AAA	ATG	ACC	GTT	GTT	816
Asn	Ser	Thr	Asp	Tyr	Leu	Tyr	Pro	Glu	Gln	Leu	Lys	Met	Thr	Val	Val	
			260					265					270			
AAA	CTG	ATC	TCT	CAC	CGT	GAA	TGC	CAG	CAG	CCG	CAC	TAC	TAC	GGT	TCT	864
Lys	Leu	Ile	Ser	His	Arg	Glu	Cys	Gln	Gln	Pro	His	Tyr	Tyr	Gly	Ser	
		275				280						285				
GAA	GTT	ACC	ACC	AAA	ATG	CTG	TGC	GCT	GCT	GAC	CCG	CAG	TGG	AAA	ACC	912
Glu	Val	Thr	Thr	Lys	Met	Leu	Cys	Ala	Ala	Asp	Pro	Gln	Trp	Lys	Thr	
	290					295					300					
GAC	TCT	TGC	CAA	GGT	GAC	TCT	GGT	GGT	CCA	CTA	GTT	TGC	TCT	CTC	CAG	960
Asp	Ser	Cys	Gln	Gly	Asp	Ser	Gly	Gly	Pro	Leu	Val	Cys	Ser	Leu	Gln	
	305				310					315					320	
GGT	CGT	ATG	ACC	CTG	ACC	GGT	ATT	GTT	TCT	TGG	GGT	CGT	GGT	TGC	GCT	1008
Gly	Arg	Met	Thr	Leu	Thr	Gly	Ile	Val	Ser	Trp	Gly	Arg	Gly	Cys	Ala	
				325					330					335		
CTG	AAA	GAC	AAA	CCG	GGT	GTT	TAC	ACC	CGT	GTT	TCT	CAC	TTC	CTG	CCG	1056
Leu	Lys	Asp	Lys	Pro	Gly	Val	Tyr	Thr	Arg	Val	Ser	His	Phe	Leu	Pro	
			340					345					350			
TGG	ATC	CGT	TCT	CAC	ACC	AAA	GAA	GAA	AAC	GGT	CTG	GCT	CTG			1098
Trp	Ile	Arg	Ser	His	Thr	Lys	Glu	Glu	Asn	Gly	Leu	Ala	Leu			
		355					360					365				

(2) INFORMAÇÃO PARA SEQ ID NO: 22:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 366 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 22:

Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
 1 5 10 15
 Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
 20 25 30

Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
35 40 45

Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
50 55 60

Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
65 70 75 80

Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
85 90 95

Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
100 105 110

Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
115 120 125

Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
130 135 140

Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
145 150 155

Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
165 170 175

Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190

Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205

Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220

Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

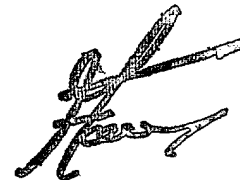
Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu
355 360 365



(2) INFORMAÇÃO PARA SEQ ID NO: 23:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1185 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

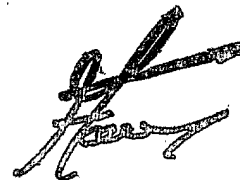
(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1185
- (D) OUTRA INFORMAÇÃO: /function= "metionina N-terminal é separada proteoliticamente em E.coli"
/product= "gene sintético para a proteína M11 (pSJ69)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 23:

ATG	AGC	AAA	ACT	TGC	TAC	GAA	GGT	AAC	GGT	CAC	TTC	TAC	CGT	GGT	AAG	43
Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys	
1				5					10					15		
GCT	TCT	ACC	GAC	ACC	ATG	GGT	CGT	CCG	TGC	CTG	CCG	TGG	AAC	TCT	GCT	95
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala	
			20					25					30			
ACC	GTT	CTG	CAG	CAG	ACC	TAC	CAC	GCT	CAC	CGT	TCT	GAT	GCA	TTG	CAG	144
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln	
		35					40					45				
CTG	GGT	CTG	GGT	AAA	CAC	AAC	TAC	TGC	CGT	AAC	CCG	GAC	AAC	CGT	CGT	192
Leu	Gly	Leu	Gly	Lys	His	Asn	Tyr	Cys	Arg	Asn	Pro	Asp	Asn	Arg	Arg	
	50					55					60					
CGT	CCG	TGG	TGC	TAC	GTT	CAG	GTT	GGT	CTG	AAA	CCG	CTA	GTT	CAG	GAA	240
Arg	Pro	Trp	Cys	Tyr	Val	Gln	Val	Gly	Leu	Lys	Pro	Leu	Val	Gln	Glu	
65					70					75					80	
TGC	ATG	GTT	CAC	GAC	TGC	GCT	GAC	GGT	AAA	AAA	CCG	TCT	TCT	CCG	CCG	288
Cys	Met	Val	His	Asp	Cys	Ala	Asp	Gly	Lys	Lys	Pro	Ser	Ser	Pro	Pro	
				85					90					95		
GAA	GAG	CTC	AAA	TTC	CAG	TGC	GGT	CAA	AAA	ACC	CTA	CGT	CCG	CGT	TTT	336
Glu	Glu	Leu	Lys	Phe	Gln	Cys	Gly	Gln	Lys	Thr	Leu	Arg	Pro	Arg	Phe	
			100					105					110			
AAA	ATC	ATC	GGT	GGT	GAG	TTC	ACC	ACC	ATC	GAA	AAC	CAG	CCG	TGG	TTC	384
Lys	Ile	Ile	Gly	Gly	Glu	Phe	Thr	Thr	Ile	Glu	Asn	Gln	Pro	Trp	Phe	
			115				120					125				
GCT	GCT	ATC	TAC	CGT	CGT	CAC	CGT	GGT	GGT	TCT	GTT	ACC	TAC	GTT	TGC	432
Ala	Ala	Ile	Tyr	Arg	Arg	His	Arg	Gly	Gly	Ser	Val	Thr	Tyr	Val	Cys	
			130				135				140					

GGT Gly 145	GGT Gly	TCT Ser	CTG Leu	ATC Ile	TCT Ser 150	CCG Pro	TGC Cys	TGG Trp	GTT Val	ATC Ile 155	TCT Ser	GCT Ala	ACC Thr	CAC His	TGC Cys 160	480
TTC Phe	ATC Ile	GAC Asp	TAC Tyr	CCG Pro 165	AAA Lys	AAA Lys	GAA Glu	GAC Asp	TAC Tyr 170	ATC Ile	GTT Val	TAC Tyr	CTC Leu	GGC Gly 175	CGT Arg	523
TCT Ser	CGT Arg	TTA Leu	AAC Asn 180	TCT Ser	AAC Asn	ACC Thr	CAG Gln	GGT Gly 185	GAA Glu	ATG Met	AAA Lys	TTC Phe	GAA Glu 190	GTT Val	GAA Glu	576
AAC Asn	CTG Leu	ATC Ile 195	CTG Leu	CAC His	AAA Lys	GAC Asp 200	TAC Tyr	TCT Ser	GCT Ala	GAC Asp	ACC Thr 205	CTG Leu	GCT Ala	CAC His	CAC His	624
AAC Asn 210	GAC Asp	ATC Ile	GCT Ala	CTG Leu	CTA Leu	AAA Lys 215	ATC Ile	CGT Arg	TCT Ser	AAA Lys	GAA Glu 220	GGT Gly	CGT Arg	TGC Cys	GCT Ala	672
CAG Gln 225	CCG Pro	TCT Ser	CGT Arg	ACC Thr	ATC Ile 230	CAG Gln	ACC Thr	ATC Ile	TGC Cys	CTG Leu 235	CCG Pro	TCT Ser	ATG Met	TAC Tyr	AAC Asn 240	720
GAC Asp	CCG Pro	CAG Gln	TTC Phe	GGT Gly 245	ACC Thr	TCT Ser	TGC Cys	GAA Glu	ATC Ile 250	ACC Thr	GGT Gly	TTC Phe	GGT Gly	AAA Lys 255	GAA Glu	768
AAC Asn	TCT Ser	ACC Thr 260	GAC Asp 260	TAC Tyr	CTG Leu	TAC Tyr	CCG Pro	GAA Glu 265	CAG Gln	CTG Leu	AAA Lys	ATG Met	ACC Thr 270	GTT Val	GTT Val	816
AAA Lys	CTG Leu 275	ATC Ile	TCT Ser	CAC His	CGT Arg	GAA Glu 280	TGC Cys	CAG Gln	CAG Gln	CCG Pro	CAC His	TAC Tyr 285	TAC Tyr	GGT Gly	TCT Ser	864
GAA Glu 290	GTT Val	ACC Thr	ACC Thr	AAA Lys	ATG Met	CTG Leu 295	TGC Cys	GCT Ala	GCT Ala	GAC Asp	CCG Pro 300	CAG Gln	TGG Trp	AAA Lys	ACC Thr	912
GAC Asp 305	TCT Ser	TGC Cys	CAA Gln	GGT Gly 310	GAC Asp 310	TCT Ser	GGT Gly	GGT Gly	CCA Pro	CTA Leu 315	GTT Val	TGC Cys	TCT Ser	CTC Leu	CAG Gln 320	960
GGT Gly	CGT Arg	ATG Met	ACC Thr 325	CTG Leu	ACC Thr	GGT Gly	ATT Ile	GTT Val	TCT Ser 330	TGG Trp	GGT Gly	CGT Arg	GGT Gly	TGC Cys 335	GCT Ala	1008
CTG Leu	AAA Lys	GAC Asp 340	AAA Lys	CCG Pro	GGT Gly	GTT Val	TAC Tyr	ACC Thr 345	CGT Arg	GTT Val	TCT Ser	CAC His	TTC Phe 350	CTG Leu	CCG Pro	1056
TGG Trp	ATC Ile	CGT Arg 355	TCT Ser	CAC His	ACC Thr	AAA Lys	GAA Glu 360	GAA Glu	AAC Asn	GGT Gly 365	CTG Leu	GCT Ala	CTG Leu	AGC Ser	CCG Pro	1104
CCG Pro 370	TCT Ser	CCG Pro	CCG Pro	GGT Gly	GGT Gly	TTC Phe 375	CCG Pro	CGT Arg	CCG Pro	GGT Gly 380	GGT Gly	GGT Gly	GGT Gly	AAC Asn	GGT Gly	1152
GAC Asp 385	TTC Phe	GAA Glu	GAA Glu	ATC Ile	CCG Pro 390	GAA Glu	GAA Glu	TAC Tyr	CTG Leu	TAA * 395						1185



Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

Pro Ser Pro Pro Gly Gly Phe Pro Arg Pro Gly Gly Gly Gly Asn Gly
370 375 380

Asp Phe Glu Glu Ile Pro Glu Glu Tyr Leu *
385 390 395

(2) INFORMAÇÃO PARA SEQ ID NO: 25:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1176 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1176
- (D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada proteoliticamente em E.coli"
- /product = "gene sintético para a proteína M12 (pSJ76)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 25:

ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG	48
Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys	
1 5 10 15	
GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT	96
Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala	
20 25 30	
ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG	144
Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln	
35 40 45	
CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT	192
Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg	
50 55 60	



CGT	CCG	TGG	TGC	TAC	GTT	CAG	GTT	GGT	CTG	AAA	CCG	CTA	GTT	CAG	GAA	240
Arg	Pro	Trp	Cys	Tyr	Val	Gln	Val	Gly	Leu	Lys	Pro	Leu	Val	Gln	Glu	
65					70					75					80	
TGC	ATG	GTT	CAC	GAC	TGC	GCT	GAC	GGT	AAA	AAA	CCG	TCT	TCT	CCG	CCG	288
Cys	Met	Val	His	Asp	Cys	Ala	Asp	Gly	Lys	Lys	Pro	Ser	Ser	Pro	Pro	
				85					90					95		
GAA	GAG	CTC	AAA	TTC	CAG	TGC	GGT	CAA	AAA	ACC	CTA	CGT	CCG	CGT	TTT	336
Glu	Glu	Leu	Lys	Phe	Gln	Cys	Gly	Gln	Lys	Thr	Leu	Arg	Pro	Arg	Phe	
			100					105					110			
AAA	ATC	ATC	GGT	GGT	GAG	TTC	ACC	ACC	ATC	GAA	AAC	CAG	CCG	TGG	TTC	384
Lys	Ile	Ile	Gly	Gly	Glu	Phe	Thr	Thr	Ile	Glu	Asn	Gln	Pro	Trp	Phe	
		115					120					125				
GCT	GCT	ATC	TAC	CGT	CGT	CAC	CGT	GGT	GGT	TCT	GTT	ACC	TAC	GTT	TGC	432
Ala	Ala	Ile	Tyr	Arg	Arg	His	Arg	Gly	Gly	Ser	Val	Thr	Tyr	Val	Cys	
	130					135					140					
GGT	GGT	TCT	CTG	ATC	TCT	CCG	TGC	TGG	GTT	ATC	TCT	GCT	ACC	CAC	TGC	480
Gly	Gly	Ser	Leu	Ile	Ser	Pro	Cys	Trp	Val	Ile	Ser	Ala	Thr	His	Cys	
145					150					155					160	
TTC	ATC	GAC	TAC	CCG	AAA	AAA	GAA	GAC	TAC	ATC	GTT	TAC	CTC	GGC	CGT	528
Phe	Ile	Asp	Tyr	Pro	Lys	Lys	Glu	Asp	Tyr	Ile	Val	Tyr	Leu	Gly	Arg	
				165					170					175		
TCT	CGT	TAA	AAC	TCT	AAC	ACC	CAG	GGT	GAA	ATG	AAA	TTC	GAA	GTT	GAA	576
Ser	Arg	Leu	Asn	Ser	Asn	Thr	Gln	Gly	Glu	Met	Lys	Phe	Glu	Val	Glu	
			180					185					190			
AAC	CTG	ATC	CTG	CAC	AAA	GAC	TAC	TCT	GCT	GAC	ACC	CTG	GCT	CAC	CAC	624
Asn	Leu	Ile	Leu	His	Lys	Asp	Tyr	Ser	Ala	Asp	Thr	Leu	Ala	His	His	
		195				200						205				
AAC	GAC	ATC	GCT	CTG	CTA	AAA	ATC	CGT	TCT	AAA	GAA	GGT	CGT	TGC	GCT	672
Asn	Asp	Ile	Ala	Leu	Leu	Lys	Ile	Arg	Ser	Lys	Glu	Gly	Arg	Cys	Ala	
	210					215					220					
CAG	CCG	TCT	CGT	ACC	ATC	CAG	ACC	ATC	TGC	CTG	CCG	TCT	ATG	TAC	AAC	720
Gln	Pro	Ser	Arg	Thr	Ile	Gln	Thr	Ile	Cys	Leu	Pro	Ser	Met	Tyr	Asn	
225					230					235					240	
GAC	CCG	CAG	TTC	GGT	ACC	TCT	TGC	GAA	ATC	ACC	GGT	TTC	GGT	AAA	GAA	768
Asp	Pro	Gln	Phe	Gly	Thr	Ser	Cys	Glu	Ile	Thr	Gly	Phe	Gly	Lys	Glu	
				245					250					255		
AAC	TCT	ACC	GAC	TAC	CTG	TAC	CCG	GAA	CAG	CTG	AAA	ATG	ACC	GTT	GTT	816
Asn	Ser	Thr	Asp	Tyr	Leu	Tyr	Pro	Glu	Gln	Leu	Lys	Met	Thr	Val	Val	
			260					265					270			
AAA	CTG	ATC	TCT	CAC	CGT	GAA	TGC	CAG	CAG	CCG	CAC	TAC	TAC	GGT	TCT	864
Lys	Leu	Ile	Ser	His	Arg	Glu	Cys	Gln	Gln	Pro	His	Tyr	Tyr	Gly	Ser	
		275					280					285				
GAA	GTT	ACC	ACC	AAA	ATG	CTG	TGC	GCT	GCT	GAC	CCG	CAG	TGG	AAA	ACC	912
Glu	Val	Thr	Thr	Lys	Met	Leu	Cys	Ala	Ala	Asp	Pro	Gln	Trp	Lys	Thr	
	290					295					300					
GAC	TCT	TGC	CAA	GGT	GAC	TCT	GGT	GGT	CCA	CTA	GTT	TGC	TCT	CTC	CAG	960
Asp	Ser	Cys	Gln	Gly	Asp	Ser	Gly	Gly	Pro	Leu	Val	Cys	Ser	Leu	Gln	
305					310					315					320	

GGT	CGT	ATG	ACC	CTG	ACC	GGT	ATT	GTT	TCT	TGG	GGT	CGT	GGT	TGC	GCT	1009
Gly	Arg	Met	Thr	Leu	Thr	Gly	Ile	Val	Ser	Trp	Gly	Arg	Gly	Cys	Ala	
				325					330					335		
CTG	AAA	GAC	AAA	CCG	GGT	GTT	TAC	ACC	CGT	GTT	TCT	CAC	TTC	CTG	CCG	1056
Leu	Lys	Asp	Lys	Pro	Gly	Val	Tyr	Thr	Arg	Val	Ser	His	Phe	Leu	Pro	
			340					345					350			
TGG	ATC	CGT	TCT	CAC	ACC	AAA	GAA	GAA	AAC	GGT	CTG	GCT	CTG	AGC	CCG	1104
Trp	Ile	Arg	Ser	His	Thr	Lys	Glu	Glu	Asn	Gly	Leu	Ala	Leu	Ser	Pro	
		355					360					365				
GTT	AAG	GCT	TTC	CCG	AGG	CCT	GGT	GGT	GGT	GGT	AAC	GGT	GAC	TTC	GAA	1152
Val	Lys	Ala	Phe	Pro	Arg	Pro	Gly	Gly	Gly	Gly	Asn	Gly	Asp	Phe	Glu	
	370					375					380					
GAA	ATC	CCG	GAA	GAA	TAC	CTG	TGA									1176
Glu	Ile	Pro	Glu	Glu	Tyr	Leu	*									
	385				390											

(2) INFORMAÇÃO PARA SEQ ID NO: 26:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 392 aminoácidos
- (B) TIPO: aminoácido
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 26:

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

Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190
Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205
Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220
Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240
Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255
Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270
Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285
Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300
Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320
Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335
Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350
Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365
Val Lys Ala Phe Pro Arg Pro Gly Gly Gly Gly Asn Gly Asp Phe Glu
370 375 380
Glu Ile Pro Glu Glu Tyr Leu *
385 390

(2) INFORMAÇÃO PARA SEQ.ID NO: 27:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1194 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

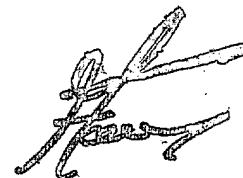
(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1194
- (D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada proteoliticamente em E.coli"
/product = "gene sintético para a proteína M13 (pSJ77)"



(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 27:

ATG Met 1	AGC Ser	AAA Lys	ACT Thr	TGC Cys 5	TAC Tyr	GAA Glu	GGT Gly	AAC Asn	GGT Gly 10	CAC His	TTC Phe	TAC Tyr	CGT Arg	GGT Gly 15	AAG Lys	43
GCT Ala	TCT Ser	ACC Thr	GAC Asp 20	ACC Thr	ATG Met	GGT Gly	CGT Arg	CCG Pro 25	TGC Cys	CTG Leu	CCG Pro	TGG Trp	AAC Asn 30	TCT Ser	GCT Ala	96
ACC Thr	GTT Val	CTG Leu 35	CAG Gln	CAG Gln	ACC Thr	TAC Tyr	CAC His 40	GCT Ala	CAC His	CGT Arg	TCT Ser	GAT Asp 45	GCA Ala	TTG Leu	CAG Gln	144
CTG Leu 50	GGT Gly	CTG Leu	GGT Gly	AAA Lys	CAC His	AAC Asn 55	TAC Tyr	TGC Cys	CGT Arg	AAC Asn 60	CCG Pro	GAC Asp	AAC Asn	CGT Arg	CGT Arg	192
CGT Arg 65	CCG Pro	TGG Trp	TGC Cys	TAC Tyr	GTT Val 70	CAG Gln	GTT Val	GGT Gly	CTG Leu	AAA Lys 75	CCG Pro	CTA Leu	GTT Val	CAG Gln	GAA Glu 80	240
TGC Cys	ATG Met	GTT Val	CAC His	GAC Asp 85	TGC Cys	GCT Ala	GAC Asp	GGT Gly	AAA Lys 90	AAA Lys	CCG Pro	TCT Ser	TCT Ser	CCG Pro 95	CCG Pro	298
GAA Glu	GAG Glu	CTC Leu	AAA Lys 100	TTC Phe	CAG Gln	TGC Cys	GGT Gly	CAA Gln 105	AAA Lys	ACC Thr	CTA Leu	CGT Arg	CCG Pro 110	CGT Arg	TTT Phe	336
AAA Lys	ATC Ile	ATC Ile 115	GGT Gly	GGT Gly	GAG Glu	TTC Phe	ACC Thr 120	ACC Thr	ATC Ile	GAA Glu	AAC Asn	CAG Gln 125	CCG Pro	TGG Trp	TTC Phe	384
GCT Ala 130	GCT Ala	ATC Ile	TAC Tyr	CGT Arg	CGT Arg	CAC His 135	CGT Arg	GGT Gly	GGT Gly	TCT Ser	GTT Val 140	ACC Thr	TAC Tyr	GTT Val	TGC Cys	432
GGT Gly 145	GGT Gly	TCT Ser	CTG Leu	ATC Ile	TCT Ser 150	CCG Pro	TGC Cys	TGG Trp	GTT Val	ATC Ile 155	TCT Ser	GCT Ala	ACC Thr	CAC His	TGC Cys 160	480
TTC Phe	ATC Ile	GAC Asp	TAC Tyr	CCG Pro 165	AAA Lys	AAA Lys	GAA Glu	GAC Asp	TAC Tyr 170	ATC Ile	GTT Val	TAC Tyr	CTC Leu	GGC Gly 175	CGT Arg	528
TCT Ser	CGT Arg	TTA Leu	AAC Asn 180	TCT Ser	AAC Asn	ACC Thr	CAG Gln	GGT Gly 185	GAA Glu	ATG Met	AAA Lys	TTC Phe	GAA Glu 190	GTT Val	GAA Glu	576
AAC Asn	CTG Leu	ATC Ile 195	CTG Leu	CAC His	AAA Lys	GAC Asp	TAC Tyr 200	TCT Ser	GCT Ala	GAC Asp	ACC Thr	CTG Leu 205	GCT Ala	CAC His	CAC His	624
AAC Asn 210	GAC Asp	ATC Ile	GCT Ala	CTG Leu	CTA Leu	AAA Lys 215	ATC Ile	CGT Arg	TCT Ser	AAA Lys	GAA Glu 220	GGT Gly	CGT Arg	TGC Cys	GCT Ala	672
CAG Gln 225	CCG Pro	TCT Ser	CGT Arg	ACC Thr	ATC Ile 230	CAG Gln	ACC Thr	ATC Ile	TGC Cys	CTG Leu 235	CCG Pro	TCT Ser	ATG Met	TAC Tyr	AAC Asn 240	720



GAC Asp	CCG Pro	CAG Gln	TTC Phe	GGT Gly	ACC Thr	TCT Ser	TGC Cys	GAA Glu	ATC Ile	ACC Thr	GGT Gly	TTC Phe	GGT Gly	AAA Lys	GAA Glu	768
				245				250						255		
AAC Asn	TCT Ser	ACC Thr	GAC Asp	TAC Tyr	CTG Leu	TAC Tyr	CCG Pro	GAA Glu	CAG Gln	CTG Leu	AAA Lys	ATG Met	ACC Thr	GTT Val	GTT Val	816
			260					265					270			
AAA Lys	CTG Leu	ATC Ile	TCT Ser	CAC His	CGT Arg	GAA Glu	TGC Cys	CAG Gln	CAG Gln	CCG Pro	CAC His	TAC Tyr	TAC Tyr	GGT Gly	TCT Ser	864
		275					280					285				
GAA Glu	GTT Val	ACC Thr	ACC Thr	AAA Lys	ATG Met	CTG Leu	TGC Cys	GCT Ala	GCT Ala	GAC Asp	CCG Pro	CAG Gln	TGG Trp	AAA Lys	ACC Thr	912
	290					295					300					
GAC Asp	TCT Ser	TGC Cys	CAA Gln	GGT Gly	GAC Asp	TCT Ser	GGT Gly	GGT Gly	CCA Pro	CTA Leu	GTT Val	TGC Cys	TCT Ser	CTC Leu	CAG Gln	960
305					310					315					320	
GGT Gly	CGT Arg	ATG Met	ACC Thr	CTG Leu	ACC Thr	GGT Gly	ATT Ile	GTT Val	TCT Ser	TGG Trp	GGT Gly	CGT Arg	GGT Gly	TGC Cys	GCT Ala	1008
				325					330					335		
CTG Leu	AAA Lys	GAC Asp	AAA Lys	CCG Pro	GGT Gly	GTT Val	TAC Tyr	ACC Thr	CGT Arg	GTT Val	TCT Ser	CAC His	TTC Phe	CTG Leu	CCG Pro	1056
			340				345						350			
TGG Trp	ATC Ile	CGT Arg	TCT Ser	CAC His	ACC Thr	AAA Lys	GAA Glu	GAA Glu	AAC Asn	GGT Gly	CTG Leu	GCT Ala	CTG Leu	AGC Ser	CCG Pro	1104
		355					360					365				
CCG Pro	AGC Ser	CCG Pro	CCG Pro	AGC Ser	CCG Pro	CCG Pro	GGT Gly	GGT Gly	TTC Phe	CCG Pro	AGG Arg	CCT Pro	GGT Gly	GGT Gly	GGT Gly	1152
	370					375					380					
GGT Gly	AAC Asn	GGT Gly	GAC Asp	TTC Phe	GAA Glu	GAA Glu	ATC Ile	CCG Pro	GAA Glu	GAG Glu	TAC Tyr	CTG Leu	TGA *			1194
385					390				395							

(2) INFORMAÇÃO PARA SEQ ID NO: 28:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 398 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 28:

Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys
1				5					10					15	
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala
			20					25					30		
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln
		35					40					45			
Leu	Gly	Leu	Gly	Lys	His	Asn	Tyr	Cys	Arg	Asn	Pro	Asp	Asn	Arg	Arg
	50					55					60				

76
Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
65 70 75 80
Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
85 90 95
Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
100 105 110
Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
115 120 125
Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
130 135 140
Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
145 150 155 160
Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
165 170 175
Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190
Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205
Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220
Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240
Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255
Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270
Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285
Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300
Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320
Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335
Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350
Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365
Pro Ser Pro Pro Ser Pro Pro Gly Gly Phe Pro Arg Pro Gly Gly Gly
370 375 380
Gly Asn Gly Asp Phe Glu Glu Ile Pro Glu Glu Tyr Leu *
385 390 395



(2) INFORMAÇÃO PARA SEQ ID NO: 29:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1176 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1176
- (D) OUTRA INFORMAÇÃO: /function= "metionina N-terminal é separada proteoliticamente em E.coli"
/product= "gene sintético para a proteína M14 (pSJ78)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 29:

ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG	48
Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys	
1 5 10 15	
GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT	96
Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala	
20 25 30	
ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG	144
Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln	
35 40 45	
CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT	192
Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg	
50 55 60	
CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA	240
Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu	
65 70 75 80	
TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG	288
Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro	
85 90 95	
GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT	336
Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe	
100 105 110	
AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC	384
Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe	
115 120 125	
GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC	432
Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys	
130 135 140	



(2) INFORMAÇÃO PARA SEQ ID NO: 30:

- (i) CARACTERÍSTICAS DA SEQUÊNCIA:
 (A) COMPRIMENTO: 392 aminoácidos
 (B) TIPO: aminoácido
 (D) TOPOLOGIA: linear

- (ii) TIPO DE MOLÉCULA: proteína
 (xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 30:

Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys
1				5					10					15	
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala
			20					25					30		
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln
		35					40					45			
Leu	Gly	Leu	Gly	Lys	His	Asn	Tyr	Cys	Arg	Asn	Pro	Asp	Asn	Arg	Arg
	50					55					60				
Arg	Pro	Trp	Cys	Tyr	Val	Gln	Val	Gly	Leu	Lys	Pro	Leu	Val	Gln	Glu
65					70					75					80
Cys	Met	Val	His	Asp	Cys	Ala	Asp	Gly	Lys	Lys	Pro	Ser	Ser	Pro	Pro
				85					90					95	
Glu	Glu	Leu	Lys	Phe	Gln	Cys	Gly	Gln	Lys	Thr	Leu	Arg	Pro	Arg	Phe
			100					105					110		
Lys	Ile	Ile	Gly	Gly	Glu	Phe	Thr	Thr	Ile	Glu	Asn	Gln	Pro	Trp	Phe
		115					120					125			
Ala	Ala	Ile	Tyr	Arg	Arg	His	Arg	Gly	Gly	Ser	Val	Thr	Tyr	Val	Cys
	130					135					140				
Gly	Gly	Ser	Leu	Ile	Ser	Pro	Cys	Trp	Val	Ile	Ser	Ala	Thr	His	Cys
145					150					155					160
Phe	Ile	Asp	Tyr	Pro	Lys	Lys	Glu	Asp	Tyr	Ile	Val	Tyr	Leu	Gly	Arg
			165						170					175	
Ser	Arg	Leu	Asn	Ser	Asn	Thr	Gln	Gly	Glu	Met	Lys	Phe	Glu	Val	Glu
			180					185					190		
Asn	Leu	Ile	Leu	His	Lys	Asp	Tyr	Ser	Ala	Asp	Thr	Leu	Ala	His	His
		195					200					205			
Asn	Asp	Ile	Ala	Leu	Leu	Lys	Ile	Arg	Ser	Lys	Glu	Gly	Arg	Cys	Ala
	210					215					220				
Gln	Pro	Ser	Arg	Thr	Ile	Gln	Thr	Ile	Cys	Leu	Pro	Ser	Met	Tyr	Asn
225					230					235					240
Asp	Pro	Gln	Phe	Gly	Thr	Ser	Cys	Glu	Ile	Thr	Gly	Phe	Gly	Lys	Glu
				245					250					255	
Asn	Ser	Thr	Asp	Tyr	Leu	Tyr	Pro	Glu	Gln	Leu	Lys	Met	Thr	Val	Val
			260					265					270		
Lys	Leu	Ile	Ser	His	Arg	Glu	Cys	Gln	Gln	Pro	His	Tyr	Tyr	Gly	Ser
		275					280						285		

80

Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

Pro Gly Gly Phe Pro Arg Pro Gly Gly Gly Gly Asn Gly Asp Phe Glu
370 375 380

Glu Ile Pro Glu Glu Tyr Leu *
385 390

(2) INFORMAÇÃO PARA SEQ ID NO: 31:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1197 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

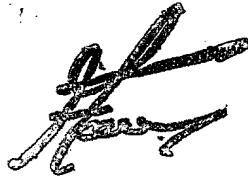
(iv) ANTI-SENTIDO: não

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1197
- (D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada proteoliticamente em E.coli"
/product = "gene sintético para a proteína M15 (pSJ79)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 31:

ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG	48
Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys	
1 5 10 15	
GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT	96
Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala	
20 25 30	
ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG	144
Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln	
35 40 45	
CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT	192
Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg	
50 55 60	



CGT Arg 65	CCG Pro	TGG Trp	TGC Cys	TAC Tyr	GTT Val 70	CAG Gln	GTT Val	GGT Gly	CTG Leu	AAA Lys 75	CCG Pro	CTA Leu	GTT Val	CAG Gln	GAA Glu 80	240
TGC Cys	ATG Met	GTT Val	CAC His	GAC Asp 85	TGC Cys	GCT Ala	GAC Asp	GGT Gly	AAA Lys 90	AAA Lys	CCG Pro	TCT Ser	TCT Ser	CCG Pro 95	CCG Pro	288
GAA Glu	GAG Glu	CTC Leu	AAA Lys 100	TTC Phe	CAG Gln	TGC Cys	GGT Gly	CAA Gln 105	AAA Lys	ACC Thr	CTA Leu	CGT Arg	CCG Pro 110	CGT Arg	TTT Phe	336
AAA Lys	ATC Ile	ATC Ile 115	GGT Gly	GGT Gly	GAG Glu	TTC Phe	ACC Thr 120	ACC Thr	ATC Ile	GAA Glu	AAC Asn	CAG Gln 125	CCG Pro	TGG Trp	TTT Phe	384
GCT Ala	GCT Ala	ATC Ile	TAC Tyr	CGT Arg	CGT Arg	CAC His 135	CGT Arg	GGT Gly	GGT Gly	TCT Ser	GTT Val 140	ACC Thr	TAC Tyr	GTT Val	TGC Cys	432
GGT Gly 145	GGT Gly	TCT Ser	CTG Leu	ATC Ile	TCT Ser 150	CCG Pro	TGC Cys	TGG Trp	GTT Val	ATC Ile 155	TCT Ser	GCT Ala	ACC Thr	CAC His	TGC Cys 160	480
TTC Phe	ATC Ile	GAC Asp	TAC Tyr	CCG Pro 165	AAA Lys	AAA Lys	GAA Glu	GAC Asp	TAC Tyr 170	ATC Ile	GTT Val	TAC Tyr	CTC Leu	GGC Gly 175	CGT Arg	528
TCT Ser	CGT Arg	TTA Leu 180	AAC Asn	TCT Ser	AAC Asn	ACC Thr	CAG Gln 185	GGT Gly	GAA Glu	ATG Met	AAA Lys	TTC Phe 190	GAA Glu	GTT Val	GAA Glu	576
AAC Asn	CTG Leu	ATC Ile 195	CTG Leu	CAC His	AAA Lys	GAC Asp	TAC Tyr 200	TCT Ser	GCT Ala	GAC Asp	ACC Thr	CTG Leu 205	GCT Ala	CAC His	CAC His	624
AAC Asn 210	GAC Asp	ATC Ile	GCT Ala	CTG Leu	CTA Leu	AAA Lys 215	ATC Ile	CGT Arg	TCT Ser	AAA Lys	GAA Glu 220	GGT Gly	CGT Arg	TGC Cys	GCT Ala	672
CAG Gln 225	CCG Pro	TCT Ser	CGT Arg	ACC Thr 230	ATC Ile	CAG Gln	ACC Thr	ATC Ile	TGC Cys	CTG Leu 235	CCG Pro	TCT Ser	ATG Met	TAC Tyr 240	AAC Asn	720
GAC Asp	CCG Pro	CAG Gln	TTC Phe	GGT Gly 245	ACC Thr	TCT Ser	TGC Cys	GAA Glu	ATC Ile 250	ACC Thr	GGT Gly	TTC Phe	GGT Gly	AAA Lys 255	GAA Glu	768
AAC Asn	TCT Ser	ACC Thr	GAC Asp 260	TAC Tyr	CTG Leu	TAC Tyr	CCG Pro	GAA Glu 265	CAG Gln	CTG Leu	AAA Lys	ATG Met 270	ACC Thr	GTT Val	GTT Val	816
AAA Lys	CTG Leu	ATC Ile 275	TCT Ser	CAC His	CGT Arg	GAA Glu	TGC Cys 280	CAG Gln	CAG Gln	CCG Pro	CAC His	TAC Tyr 285	TAC Tyr	GGT Gly	TCT Ser	864
GAA Glu 290	GTT Val	ACC Thr	ACC Thr	AAA Lys	ATG Met	CTG Leu 295	TGC Cys	GCT Ala	GCT Ala	GAC Asp	CCG Pro 300	CAG Gln	TGG Trp	AAA Lys	ACC Thr	912
GAC Asp 305	TCT Ser	TGC Cys	CAA Gln	GGT Gly 310	GAC Asp	TCT Ser	GGT Gly	GGT Gly	CCA Pro	CTA Leu 315	GTT Val	TGC Cys	TCT Ser	CTC Leu	CAG Gln 320	960



GGT Gly	CGT Arg	ATG Met	ACC Thr	CTG Leu	ACC Thr	GGT Gly	ATT Ile	GTT Val	TCT Ser	TGG Trp	GGT Gly	CGT Arg	GGT Gly	TGC Cys	GCT Ala	1008
				325					330					335		
CTG Leu	AAA Lys	GAC Asp	AAA Lys	CCG Pro	GGT Gly	GTT Val	TAC Tyr	ACC Thr	CGT Arg	GTT Val	TCT Ser	CAC His	TTC Phe	CTG Leu	CCG Pro	1056
			340					345					350			
TGG Trp	ATC Ile	CGT Arg	TCT Ser	CAC His	ACC Thr	AAA Lys	GAA Glu	GAA Glu	AAC Asn	GGT Gly	CTG Leu	GCT Ala	CTG Leu	AGC Ser	CCG Pro	1104
		355					360					365				
CCG Pro	AGC Ser	CCG Pro	CCG Pro	AGC Ser	CCG Pro	CCG Pro	GGT Gly	GGT Gly	TTC Phe	GGT Gly	CCG Pro	AGG Arg	CCT Pro	GGT Gly	GGT Gly	1152
	370					375					380					
GGT Gly	GGT Gly	AAC Asn	GGT Gly	GAC Asp	TTC Phe	GAA Glu	GAA Glu	ATC Ile	CCG Pro	GAA Glu	GAG Glu	TAC Tyr	CTG Leu	TGA *		1197
385					390					395						

(2) INFORMAÇÃO PARA SEQ ID NO: 32:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 399 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 32:

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

Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190

Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205

Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220

Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

Pro Ser Pro Pro Ser Pro Pro Gly Gly Phe Gly Pro Arg Pro Gly Gly
370 375 380

Gly Gly Asn Gly Asp Phe Glu Glu Ile Pro Glu Glu Tyr Leu *
385 390 395

(2) INFORMAÇÃO PARA SEQ ID NO: 33:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1185 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não



(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: CDS

(B) LOCALIZAÇÃO: 1..1185

(D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada
proteoliticamente em E.coli"
/product = "gene sintético para a proteína M16 (pSJ81)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 33:

ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG	48
Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys	
1 5 10 15	
GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT	96
Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala	
20 25 30	
ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG	144
Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln	
35 40 45	
CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT	192
Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg	
50 55 60	
CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA	240
Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu	
65 70 75 80	
TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG	288
Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro	
85 90 95	
GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT	336
Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe	
100 105 110	
AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC	384
Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe	
115 120 125	
GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC	432
Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys	
130 135 140	
GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC	480
Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys	
145 150 155 160	
TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT	528
Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg	
165 170 175	
TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA	576
Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu	
180 185 190	
AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC	624
Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His	
195 200 205	

AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala 210 215 220	672
CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn 225 230 235 240	720
GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu 245 250 255	768
AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val 260 265 270	816
AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser 275 280 285	864
GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr 290 295 300	912
GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln 305 310 315 320	960
GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala 325 330 335	1008
CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro 340 345 350	1056
TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro 355 360 365	1104
GTT AAG GCT TGG CCG GGA CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC Val Lys Ala Trp Pro Gly Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp 370 375 380	1152
TTC GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA Phe Glu Glu Ile Pro Glu Glu Tyr Leu Gln * 385 390 395	1185

(2) INFORMAÇÃO PARA SEQ ID NO: 34:

- (i) CARACTERÍSTICAS DA SEQUÊNCIA:
 (A) COMPRIMENTO: 395 aminoácidos
 (B) TIPO: aminoácido
 (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 34:

Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys
 1 5 10 15
 Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala
 20 25 30



Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
35 40 45

Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
50 55 60

Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
65 70 75 80

Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
85 90 95

Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
100 105 110

Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
115 120 125

Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
130 135 140

Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
145 150 155 160

Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
165 170 175

Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190

Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205

Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220

Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

Val Lys Ala Trp Pro Gly Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp
 370 375 380
 Phe Glu Glu Ile Pro Glu Glu Tyr Leu Gln *
 385 390 395

(2) INFORMAÇÃO PARA SEQ ID NO: 35:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1185 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

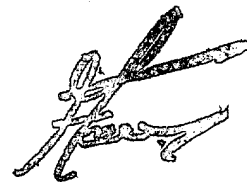
(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1185
- (D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada proteoliticamente em E.coli"
/product = "gene sintético para a proteína M17 (pSJ83)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 35:

ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG	48
Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys	
1 5 10 15	
GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT	96
Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala	
20 25 30	
ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG	144
Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln	
35 40 45	
CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT	192
Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg	
50 55 60	
CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA	240
Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu	
65 70 75 80	
TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG	288
Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro	
85 90 95	
GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT	336
Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe	
100 105 110	
AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC	384
Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe	
115 120 125	

GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys 130 135 140	432
GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys 145 150 155 160	480
TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg 165 170 175	528
TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu 180 185 190	576
AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His 195 200 206	624
AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala 210 215 220	672
CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn 225 230 235 240	720
GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu 245 250 255	768
AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val 260 265 270	816
AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser 275 280 285	864
GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr 290 295 300	912
GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln 305 310 315 320	960
GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala 325 330 335	1008
CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro 340 345 350	1056
TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro 355 360 365	1104
GTT AAG GCT TTC GGA CCG CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC Val Lys Ala Phe Gly Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp 370 375 380	1152



TTC GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
Phe Glu Glu Ile Pro Glu Glu Tyr Leu Gln *
385 390 395

1135

(2) INFORMAÇÃO PARA SEQ ID NO: 36:

- (i) CARACTERÍSTICAS DA SEQUÊNCIA:
 - (A) COMPRIMENTO: 395 aminoácidos
 - (B) TIPO: aminoácido
 - (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 36:

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

Asn	Ser	Thr	Asp	Tyr	Leu	Tyr	Pro	Glu	Gln	Leu	Lys	Met	Thr	Val	Val
			260					265					270		
Lys	Leu	Ile	Ser	His	Arg	Glu	Cys	Gln	Gln	Pro	His	Tyr	Tyr	Gly	Ser
		275					280					285			
Glu	Val	Thr	Thr	Lys	Met	Leu	Cys	Ala	Ala	Asp	Pro	Gln	Trp	Lys	Thr
	290					295					300				
Asp	Ser	Cys	Gln	Gly	Asp	Ser	Gly	Gly	Pro	Leu	Val	Cys	Ser	Leu	Gln
305					310					315					320
Gly	Arg	Met	Thr	Leu	Thr	Gly	Ile	Val	Ser	Trp	Gly	Arg	Gly	Cys	Ala
				325					330					335	
Leu	Lys	Asp	Lys	Pro	Gly	Val	Tyr	Thr	Arg	Val	Ser	His	Phe	Leu	Pro
			340					345					350		
Trp	Ile	Arg	Ser	His	Thr	Lys	Glu	Glu	Asn	Gly	Leu	Ala	Leu	Ser	Pro
		355					360					365			
Val	Lys	Ala	Phe	Gly	Pro	Arg	Pro	Leu	Gly	Gly	Gly	Gly	Asn	Gly	Asp
	370					375					380				
Phe	Glu	Glu	Ile	Pro	Glu	Glu	Tyr	Leu	Gln	*					
385					390					395					

(2) INFORMAÇÃO PARA SEQ ID NO: 37:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1185 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

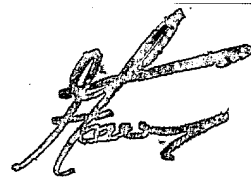
(iv) ANTI-SENTIDO: não

(ix) CARACTERÍSTICA:

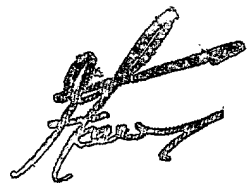
- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1185
- (D) OUTRA INFORMAÇÃO: /function= "metionina N-terminal é separada proteoliticamente em E.coli"
/product= "gene sintético para a proteína M18 (pSJ90)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 37:

ATG	AGC	AAA	ACT	TGC	TAC	GAA	GGT	AAC	GGT	CAC	TTC	TAC	CGT	GGT	AAG	48
Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys	
1				5				10						15		
GCT	TCT	ACC	GAC	ACC	ATG	GGT	CGT	CCG	TGC	CTG	CCG	TGG	AAC	TCT	GCT	96
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala	
			20					25					30			
ACC	GTT	CTG	CAG	CAG	ACC	TAC	CAC	GCT	CAC	CGT	TCT	GAT	GCA	TTG	CAG	144
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln	
			35				40					45				



CTG Leu	GGT Gly	CTG Leu	GGT Gly	AAA Lys	CAC His	AAC Asn	TAC Tyr	TGC Cys	CGT Arg	AAC Asn	CCG Pro	GAC Asp	AAC Asn	CGT Arg	CGT Arg	192
	50					55					60					
CGT Arg	CCG Pro	TGG Trp	TGC Cys	TAC Tyr	GTT Val	CAG Gln	GTT Val	GGT Gly	CTG Leu	AAA Lys	CCG Pro	CTA Leu	GTT Val	CAG Gln	GAA Glu	240
	65				70					75					80	
TGC Cys	ATG Met	GTT Val	CAC His	GAC Asp	TGC Cys	GCT Ala	GAC Asp	GGT Gly	AAA Lys	AAA Lys	CCG Pro	TCT Ser	TCT Ser	CCG Pro	CCG Pro	288
				85					90					95		
GAA Glu	GAG Glu	CTC Leu	AAA Lys	TTC Phe	CAG Gln	TGC Cys	GGT Gly	CAA Gln	AAA Lys	ACC Thr	CTA Leu	CGT Arg	CCG Pro	CGT Arg	TTT Phe	336
			100					105					110			
AAA Lys	ATC Ile	ATC Ile	GGT Gly	GGT Gly	GAG Glu	TTC Phe	ACC Thr	ACC Thr	ATC Ile	GAA Glu	AAC Asn	CAG Gln	CCG Pro	TGG Trp	TTC Phe	384
		115					120					125				
GCT Ala	GCT Ala	ATC Ile	TAC Tyr	CGT Arg	CGT Arg	CAC His	CGT Arg	GGT Gly	GGT Gly	TCT Ser	GTT Val	ACC Thr	TAC Tyr	GTT Val	TGC Cys	432
	130					135					140					
GGT Gly	GGT Gly	TCT Ser	CTG Leu	ATC Ile	TCT Ser	CCG Pro	TGC Cys	TGG Trp	GTT Val	ATC Ile	TCT Ser	GCT Ala	ACC Thr	CAC His	TGC Cys	480
	145				150					155					160	
TTC Phe	ATC Ile	GAC Asp	TAC Tyr	CCG Pro	AAA Lys	AAA Lys	GAA Glu	GAC Asp	TAC Tyr	ATC Ile	GTT Val	TAC Tyr	CTC Leu	GGC Gly	CGT Arg	528
				165					170					175		
TCT Ser	CGT Arg	TTA Leu	AAC Asn	TCT Ser	AAC Asn	ACC Thr	CAG Gln	GGT Gly	GAA Glu	ATG Met	AAA Lys	TTC Phe	GAA Glu	GTT Val	GAA Glu	576
			180					185					190			
AAC Asn	CTG Leu	ATC Ile	CTG Leu	CAC His	AAA Lys	GAC Asp	TAC Tyr	TCT Ser	GCT Ala	GAC Asp	ACC Thr	CTG Leu	GCT Ala	CAC His	CAC His	624
		195					200					205				
AAC Asn	GAC Asp	ATC Ile	GCT Ala	CTG Leu	CTA Leu	AAA Lys	ATC Ile	CGT Arg	TCT Ser	AAA Lys	GAA Glu	GGT Gly	CGT Arg	TGC Cys	GCT Ala	672
	210					215					220					
CAG Gln	CCG Pro	TCT Ser	CGT Arg	ACC Thr	ATC Ile	CAG Gln	ACC Thr	ATC Ile	TGC Cys	CTG Leu	CCG Pro	TCT Ser	ATG Met	TAC Tyr	AAC Asn	720
	225			230						235					240	
GAC Asp	CCG Pro	CAG Gln	TTC Phe	GGT Gly	ACC Thr	TCT Ser	TGC Cys	GAA Glu	ATC Ile	ACC Thr	GGT Gly	TTC Phe	GGT Gly	AAA Lys	GAA Glu	768
				245				250						255		
AAC Asn	TCT Ser	ACC Thr	GAC Asp	TAC Tyr	CTG Leu	TAC Tyr	CCG Pro	GAA Glu	CAG Gln	CTG Leu	AAA Lys	ATG Met	ACC Thr	GTT Val	GTT Val	816
			260					265					270			
AAA Lys	CTG Leu	ATC Ile	TCT Ser	CAC His	CGT Arg	GAA Glu	TGC Cys	CAG Gln	CAG Gln	CCG Pro	CAC His	TAC Tyr	TAC Tyr	GGT Gly	TCT Ser	864
		275					280					285				
GAA Glu	GTT Val	ACC Thr	ACC Thr	AAA Lys	ATG Met	CTG Leu	TGC Cys	GCT Ala	GCT Ala	GAC Asp	CCG Pro	CAG Gln	TGG Trp	AAA Lys	ACC Thr	912
	290					295					300					



GAC Asp 305	TCT Ser	TGC Cys	CAA Gln	GGT Gly	GAC Asp	TCT Ser	GGT Gly	GGT Gly	CCA Pro	CTA Leu	GTT Val	TGC Cys	TCT Ser	CTC Leu	CAG Gln	960
310				310					315						320	
GGT Gly	CGT Arg	ATG Met	ACC Thr	CTG Leu	ACC Thr	GGT Gly	ATT Ile	GTT Val	TCT Ser	TGG Trp	GGT Gly	CGT Arg	GGT Gly	TGC Cys	GCT Ala	1008
				325					330						335	
CTG Leu	AAA Lys	GAC Asp	AAA Lys	CCG Pro	GGT Gly	GTT Val	TAC Tyr	ACC Thr	CGT Arg	GTT Val	TCT Ser	CAC His	TTC Phe	CTG Leu	CCG Pro	1056
			340					345						350		
TGG Trp	ATC Ile	CGT Arg	TCT Ser	CAC His	ACC Thr	AAA Lys	GAA Glu	GAA Glu	AAC Asn	GGT Gly	CTG Leu	GCT Ala	CTG Leu	AGC Ser	CCG Pro	1104
		355					360					365				
GTT Val	CGG Arg	GCT Ala	TTC Phe	GGA Gly	CCG Pro	CGG Arg	CCG Pro	CTG Leu	GGT Gly	GGT Gly	GGT Gly	GGT Gly	AAC Asn	GGT Gly	GAC Asp	1152
		370				375										
TTC Phe	GAA Glu	GAA Glu	ATC Ile	CCG Pro	GAA Glu	GAA Glu	TAC Tyr	CTG Leu	CAA Gln	TAA *						1155
					390					395						

(2) INFORMAÇÃO PARA SEQ ID NO: 38:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 395 aminoácidos

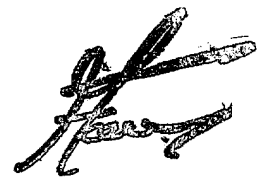
(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 38:

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



Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
165 170 175

Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190

Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205

Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220

Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

Val Arg Ala Phe Gly Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp
370 375 380

Phe Glu Glu Ile Pro Glu Glu Tyr Leu Gln *
385 390 395

(2) INFORMAÇÃO PARA SEQ ID NO: 39:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1185 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não



(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: CDS

(B) LOCALIZAÇÃO: 1..1185

(D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada proteoliticamente em E.coli"

/product = "gene sintético para a proteína M19 (pSJ91)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 39:

ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG	48
Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys	
1 5 10 15	
GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT	96
Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala	
20 25 30	
ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG	144
Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln	
35 40 45	
CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT	192
Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg	
50 55 60	
CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA	240
Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu	
65 70 75 80	
TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG	288
Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro	
85 90 95	
GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT	336
Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe	
100 105 110	
AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC	384
Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe	
115 120 125	
GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC	432
Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys	
130 135 140	
GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC	480
Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys	
145 150 155 160	
TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT	528
Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg	
165 170 175	
TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA	576
Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu	
180 185 190	
AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC	624
Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His	
195 200 205	



AAC Asn	GAC Asp	ATC Ile	GCT Ala	CTG Leu	CTA Leu	AAA Lys	ATC Ile	CGT Arg	TCT Ser	AAA Lys	GAA Glu	GGT Gly	CGT Arg	TGC Cys	GCT Ala	672
210						215					220					
CAG Gln	CCG Pro	TCT Ser	CGT Arg	ACC Thr	ATC Ile	CAG Gln	ACC Thr	ATC Ile	TGC Cys	CTG Leu	CCG Pro	TCT Ser	ATG Met	TAC Tyr	AAC Asn	720
225					230					235					240	
GAC Asp	CCG Pro	CAG Gln	TTC Phe	GGT Gly	ACC Thr	TCT Ser	TGC Cys	GAA Glu	ATC Ile	ACC Thr	GGT Gly	TTC Phe	GGT Gly	AAA Lys	GAA Glu	768
				245					250					255		
AAC Asn	TCT Ser	ACC Thr	GAC Asp	TAC Tyr	CTG Leu	TAC Tyr	CCG Pro	GAA Glu	CAG Gln	CTG Leu	AAA Lys	ATG Met	ACC Thr	GTT Val	GTT Val	816
			260					265					270			
AAA Lys	CTG Leu	ATC Ile	TCT Ser	CAC His	CGT Arg	GAA Glu	TGC Cys	CAG Gln	CAG Gln	CCG Pro	CAC His	TAC Tyr	TAC Tyr	GGT Gly	TCT Ser	864
		275					280					285				
GAA Glu	GTT Val	ACC Thr	ACC Thr	AAA Lys	ATG Met	CTG Leu	TGC Cys	GCT Ala	GCT Ala	GAC Asp	CCG Pro	CAG Gln	TGG Trp	AAA Lys	ACC Thr	912
290						295					300					
GAC Asp	TCT Ser	TGC Cys	CAA Gln	GGT Gly	GAC Asp	TCT Ser	GGT Gly	GGT Gly	CCA Pro	CTA Leu	GTT Val	TGC Cys	TCT Ser	CTC Leu	CAG Gln	960
305					310					315					320	
GGT Gly	CGT Arg	ATG Met	ACC Thr	CTG Leu	ACC Thr	GGT Gly	ATT Ile	GTT Val	TCT Ser	TGG Trp	GGT Gly	CGT Arg	GGT Gly	TGC Cys	GCT Ala	1008
				325				330						335		
CTG Leu	AAA Lys	GAC Asp	AAA Lys	CCG Pro	GGT Gly	GTT Val	TAC Tyr	ACC Thr	CGT Arg	GTT Val	TCT Ser	CAC His	TTC Phe	CTG Leu	CCG Pro	1056
			340				345						350			
TGG Trp	ATC Ile	CGT Arg	TCT Ser	CAC His	ACC Thr	AAA Lys	GAA Glu	GAA Glu	AAC Asn	GGT Gly	CTG Leu	GCT Ala	CTG Leu	AGC Ser	CCG Pro	1104
		355					360				365					
GTT Val	AAG Lys	GCT Ala	TAC Tyr	GGA Gly	CCG Pro	CGG Arg	CCG Pro	CTG Leu	GGT Gly	GGT Gly	GGT Gly	GGT Gly	AAC Asn	GGT Gly	GAC Asp	1152
		370				375					380					
TTC Phe	GAA Glu	GAA Glu	ATC Ile	CCG Pro	GAA Glu	GAA Glu	TAC Tyr	CTG Leu	CAA Gln	TAA *						1185
385					390					395						

(2) INFORMAÇÃO PARA SEQ ID NO: 40:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 395 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 40:

Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys
1				5					10					15	
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala
			20					25					30		



Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
35 40 45

Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
50 55 60

Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
65 70 75 80

Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
85 90 95

Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
100 105 110

Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
115 120 125

Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
130 135 140

Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
145 150 155 160

Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
165 170 175

Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190

Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205

Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220

Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240

Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255

Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285

Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365



Val Lys Ala Tyr Gly Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp
370 375 380

Phe Glu Glu Ile Pro Glu Glu Tyr Leu Gln *
385 390 395

(2) INFORMAÇÃO PARA SEQ ID NO: 41:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1185 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1185
- (D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada proteoliticamente em E.coli"
/product = "gene sintético para a proteína M20 (pSJ92)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 41:

ATG AGC AAA ACT TGC TAC GAA GGT AAC GGT CAC TTC TAC CGT GGT AAG	43
Met Ser Lys Thr Cys Tyr Glu Gly Asn Gly His Phe Tyr Arg Gly Lys	
1 5 10 15	
GCT TCT ACC GAC ACC ATG GGT CGT CCG TGC CTG CCG TGG AAC TCT GCT	96
Ala Ser Thr Asp Thr Met Gly Arg Pro Cys Leu Pro Trp Asn Ser Ala	
20 25 30	
ACC GTT CTG CAG CAG ACC TAC CAC GCT CAC CGT TCT GAT GCA TTG CAG	144
Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln	
35 40 45	
CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT	192
Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg	
50 55 60	
CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA	240
Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu	
65 70 75 80	
TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG	288
Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro	
85 90 95	
GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT	336
Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe	
100 105 110	
AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC	384
Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe	
115 120 125	



GCT Ala	GCT Ala	ATC Ile	TAC Tyr	CGT Arg	CGT Arg	CAC His	CGT Arg	GGT Gly	GGT Gly	TCT Ser	GTT Val	ACC Thr	TAC Tyr	GTT Val	TGC Cys	432
	130					135					140					
GGT Gly	GGT Gly	TCT Ser	CTG Leu	ATC Ile	TCT Ser	CCG Pro	TGC Cys	TGG Trp	GTT Val	ATC Ile	TCT Ser	GCT Ala	ACC Thr	CAC His	TGC Cys	480
145					150					155					160	
TTC Phe	ATC Ile	GAC Asp	TAC Tyr	CCG Pro	AAA Lys	AAA Lys	GAA Glu	GAC Asp	TAC Tyr	ATC Ile	GTT Val	TAC Tyr	CTC Leu	GGC Gly	CGT Arg	528
				165					170					175		
TCT Ser	CGT Arg	TTA Leu	AAC Asn	TCT Ser	AAC Asn	ACC Thr	CAG Gln	GGT Gly	GAA Glu	ATG Met	AAA Lys	TTC Phe	GAA Glu	GTT Val	GAA Glu	576
			180					185					190			
AAC Asn	CTG Leu	ATC Ile	CTG Leu	CAC His	AAA Lys	GAC Asp	TAC Tyr	TCT Ser	GCT Ala	GAC Asp	ACC Thr	CTG Leu	GCT Ala	CAC His	CAC His	624
		195					200					205				
AAC Asn	GAC Asp	ATC Ile	GCT Ala	CTG Leu	CTA Leu	AAA Lys	ATC Ile	CGT Arg	TCT Ser	AAA Lys	GAA Glu	GGT Gly	CGT Arg	TGC Cys	GCT Ala	672
	210					215					220					
CAG Gln	CCG Pro	TCT Ser	CGT Arg	ACC Thr	ATC Ile	CAG Gln	ACC Thr	ATC Ile	TGC Cys	CTG Leu	CCG Pro	TCT Ser	ATG Met	TAC Tyr	AAC Asn	720
225					230					235					240	
GAC Asp	CCG Pro	CAG Gln	TTC Phe	GGT Gly	ACC Thr	TCT Ser	TGC Cys	GAA Glu	ATC Ile	ACC Thr	GGT Gly	TTC Phe	GGT Gly	AAA Lys	GAA Glu	768
				245				250						255		
AAC Asn	TCT Ser	ACC Thr	GAC Asp	TAC Tyr	CTG Leu	TAC Tyr	CCG Pro	GAA Glu	CAG Gln	CTG Leu	AAA Lys	ATG Met	ACC Thr	GTT Val	GTT Val	816
			260					265				270				
AAA Lys	CTG Leu	ATC Ile	TCT Ser	CAC His	CGT Arg	GAA Glu	TGC Cys	CAG Gln	CAG Gln	CCG Pro	CAC His	TAC Tyr	TAC Tyr	GGT Gly	TCT Ser	864
		275				280						285				
GAA Glu	GTT Val	ACC Thr	ACC Thr	AAA Lys	ATG Met	CTG Leu	TGC Cys	GCT Ala	GCT Ala	GAC Asp	CCG Pro	CAG Gln	TGG Trp	AAA Lys	ACC Thr	912
	290					295					300					
GAC Asp	TCT Ser	TGC Cys	CAA Gln	GGT Gly	GAC Asp	TCT Ser	GGT Gly	GGT Gly	CCA Pro	CTA Leu	GTT Val	TGC Cys	TCT Ser	CTC Leu	CAG Gln	960
305					310					315					320	
GGT Gly	CGT Arg	ATG Met	ACC Thr	CTG Leu	ACC Thr	GGT Gly	ATT Ile	GTT Val	TCT Ser	TGG Trp	GGT Gly	CGT Arg	GGT Gly	TGC Cys	GCT Ala	1008
				325					330					335		
CTG Leu	AAA Lys	GAC Asp	AAA Lys	CCG Pro	GGT Gly	GTT Val	TAC Tyr	ACC Thr	CGT Arg	GTT Val	TCT Ser	CAC His	TTC Phe	CTG Leu	CCG Pro	1056
			340				345						350			
TGG Trp	ATC Ile	CGT Arg	TCT Ser	CAC His	ACC Thr	AAA Lys	GAA Glu	GAA Glu	AAC Asn	GGT Gly	CTG Leu	GCT Ala	CTG Leu	AGC Ser	CCG Pro	1104
		355					360				365					
GTT Val	GTT Val	GCT Ala	TTC Phe	GGA Gly	CCG Pro	CGG Arg	CCG Pro	CTG Leu	GGT Gly	GGT Gly	GGT Gly	GGT Gly	AAC Asn	GGT Gly	GAC Asp	1152
	370					375					380					



TTC GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
Phe Glu Glu Ile Pro Glu Glu Tyr Leu Gln *
385 390 395

1135

(2) INFORMAÇÃO PARA SEQ ID NO: 42:

- (i) CARACTERÍSTICAS DA SEQUÊNCIA:
 - (A) COMPRIMENTO: 395 aminoácidos
 - (B) TIPO: aminoácido
 - (D) TOPOLOGIA: linear

- (ii) TIPO DE MOLÉCULA: proteína
- (xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 42:

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

Asn	Ser	Thr	Asp	Tyr	Leu	Tyr	Pro	Glu	Gln	Leu	Lys	Met	Thr	Val	Val	
			260					265					270			
Lys	Leu	Ile	Ser	His	Arg	Glu	Cys	Gln	Gln	Pro	His	Tyr	Tyr	Gly	Ser	
		275					280					285				
Glu	Val	Thr	Thr	Lys	Met	Leu	Cys	Ala	Ala	Asp	Pro	Gln	Trp	Lys	Thr	
	290					295					300					
Asp	Ser	Cys	Gln	Gly	Asp	Ser	Gly	Gly	Pro	Leu	Val	Cys	Ser	Leu	Gln	
305					310					315					320	
Gly	Arg	Met	Thr	Leu	Thr	Gly	Ile	Val	Ser	Trp	Gly	Arg	Gly	Cys	Ala	
				325					330					335		
Leu	Lys	Asp	Lys	Pro	Gly	Val	Tyr	Thr	Arg	Val	Ser	His	Phe	Leu	Pro	
			340					345					350			
Trp	Ile	Arg	Ser	His	Thr	Lys	Glu	Glu	Asn	Gly	Leu	Ala	Leu	Ser	Pro	
		355					360					365				
Val	Val	Ala	Phe	Gly	Pro	Arg	Pro	Leu	Gly	Gly	Gly	Gly	Asn	Gly	Asp	
	370					375					380					
Phe	Glu	Glu	Ile	Pro	Glu	Glu	Tyr	Leu	Gln	*						
385					390					395						

(2) INFORMAÇÃO PARA SEQ ID NO: 43:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1182 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(ix) CARACTERÍSTICA:

- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1182
- (D) OUTRA INFORMAÇÃO: /function= "metionina N-terminal é separada proteoliticamente em E.coli"
/product= "gene sintético para a proteína M21 (pSJ93)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 43:

ATG	AGC	AAA	ACT	TGC	TAC	GAA	GGT	AAC	GGT	CAC	TTC	TAC	CGT	GGT	AAG	48
Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys	
1			5				10						15			
GCT	TCT	ACC	GAC	ACC	ATG	GGT	CGT	CCG	TGC	CTG	CCG	TGG	AAC	TCT	GCT	96
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala	
			20				25						30			
ACC	GTT	CTG	CAG	CAG	ACC	TAC	CAC	GCT	CAC	CGT	TCT	GAT	GCA	TTG	CAG	144
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln	
		35				40						45				



CTG GGT CTG GGT AAA CAC AAC TAC TGC CGT AAC CCG GAC AAC CGT CGT Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg 50 55 60	192
CGT CCG TGG TGC TAC GTT CAG GTT GGT CTG AAA CCG CTA GTT CAG GAA Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu 65 70 75 80	240
TGC ATG GTT CAC GAC TGC GCT GAC GGT AAA AAA CCG TCT TCT CCG CCG Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro 85 90 95	288
GAA GAG CTC AAA TTC CAG TGC GGT CAA AAA ACC CTA CGT CCG CGT TTT Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe 100 105 110	336
AAA ATC ATC GGT GGT GAG TTC ACC ACC ATC GAA AAC CAG CCG TGG TTC Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe 115 120 125	384
GCT GCT ATC TAC CGT CGT CAC CGT GGT GGT TCT GTT ACC TAC GTT TGC Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys 130 135 140	432
GGT GGT TCT CTG ATC TCT CCG TGC TGG GTT ATC TCT GCT ACC CAC TGC Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys 145 150 155 160	480
TTC ATC GAC TAC CCG AAA AAA GAA GAC TAC ATC GTT TAC CTC GGC CGT Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg 165 170 175	528
TCT CGT TTA AAC TCT AAC ACC CAG GGT GAA ATG AAA TTC GAA GTT GAA Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu 180 185 190	576
AAC CTG ATC CTG CAC AAA GAC TAC TCT GCT GAC ACC CTG GCT CAC CAC Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His 195 200 205	624
AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala 210 215 220	672
CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn 225 230 235 240	720
GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu 245 250 255	768
AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val 260 265 270	816
AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser 275 280 285	864
GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr 290 295 300	912



GAC	TCT	TGC	CAA	GGT	GAC	TCT	GGT	GGT	CCA	CTA	GTT	TGC	TCT	CTC	CAG	960
Asp	Ser	Cys	Gln	Gly	Asp	Ser	Gly	Gly	Pro	Leu	Val	Cys	Ser	Leu	Gln	
305					310					315					320	
GGT	CGT	ATG	ACC	CTG	ACC	GGT	ATT	GTT	TCT	TGG	GGT	CGT	GGT	TGC	GCT	1008
Gly	Arg	Met	Thr	Leu	Thr	Gly	Ile	Val	Ser	Trp	Gly	Arg	Gly	Cys	Ala	
				325					330					335		
CTG	AAA	GAC	AAA	CCG	GGT	GTT	TAC	ACC	CGT	GTT	TCT	CAC	TTC	CTG	CCG	1056
Leu	Lys	Asp	Lys	Pro	Gly	Val	Tyr	Thr	Arg	Val	Ser	His	Phe	Leu	Pro	
			340					345					350			
TGG	ATC	CGT	TCT	CAC	ACC	AAA	GAA	GAA	AAC	GGT	CTG	GCT	CTG	AGC	CCG	1104
Trp	Ile	Arg	Ser	His	Thr	Lys	Glu	Glu	Asn	Gly	Leu	Ala	Leu	Ser	Pro	
		355					360					365				
GTT	CGG	GCT	TTC	CCG	CGG	CCG	CTG	GGT	GGT	GGT	GGT	AAC	GGT	GAC	TTC	1152
Val	Arg	Ala	Phe	Pro	Arg	Pro	Leu	Gly	Gly	Gly	Gly	Asn	Gly	Asp	Phe	
	370					375					380					
GAA	GAA	ATC	CCG	GAA	GAA	TAC	CTG	CAA	TAA							1182
Glu	Glu	Ile	Pro	Glu	Glu	Tyr	Leu	Gln	*							
385					390											

(2) INFORMAÇÃO PARA SEQ ID NO: 44:

- (i) CARACTERÍSTICAS DA SEQUÊNCIA:
 - (A) COMPRIMENTO: 394 aminoácidos
 - (B) TIPO: aminoácido
 - (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 44:

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



Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
165 170 175
Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
180 185 190
Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
195 200 205
Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
210 215 220
Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
225 230 235 240
Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
245 250 255
Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270
Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 280 285
Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300
Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320
Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335
Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350
Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365
Val Arg Ala Phe Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
370 375 380
Glu Glu Ile Pro Glu Glu Tyr Leu Gln *
385 390

(2) INFORMAÇÃO PARA SEQ ID NO: 45:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1182 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não



(ix) CARACTERÍSTICA:

(A) NOME/CHAVE: CDS

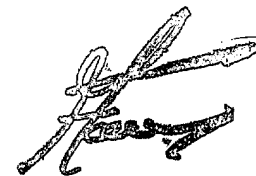
(B) LOCALIZAÇÃO: 1..1182

(D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada
proteoliticamente em E.coli"

/product = "gene sintético para a proteína M22 (pSJ94)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 45:

ATG	AGC	AAA	ACT	TGC	TAC	GAA	GGT	AAC	GGT	CAC	TTC	TAC	CGT	GGT	AAG	48
Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys	
1				5					10					15		
GCT	TCT	ACC	GAC	ACC	ATG	GGT	CGT	CCG	TGC	CTG	CCG	TGG	AAC	TCT	GCT	96
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala	
			20					25					30			
ACC	GTT	CTG	CAG	CAG	ACC	TAC	CAC	GCT	CAC	CGT	TCT	GAT	GCA	TTG	CAG	144
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln	
		35					40					45				
CTG	GGT	CTG	GGT	AAA	CAC	AAC	TAC	TGC	CGT	AAC	CCG	GAC	AAC	CGT	CGT	192
Leu	Gly	Leu	Gly	Lys	His	Asn	Tyr	Cys	Arg	Asn	Pro	Asp	Asn	Arg	Arg	
	50					55					60					
CGT	CCG	TGG	TGC	TAC	GTT	CAG	GTT	GGT	CTG	AAA	CCG	CTA	GTT	CAG	GAA	240
Arg	Pro	Trp	Cys	Tyr	Val	Gln	Val	Gly	Leu	Lys	Pro	Leu	Val	Gln	Glu	
	65				70					75					80	
TGC	ATG	GTT	CAC	GAC	TGC	GCT	GAC	GGT	AAA	AAA	CCG	TCT	TCT	CCG	CCG	298
Cys	Met	Val	His	Asp	Cys	Ala	Asp	Gly	Lys	Lys	Pro	Ser	Ser	Pro	Pro	
				85					90					95		
GAA	GAG	CTC	AAA	TTC	CAG	TGC	GGT	CAA	AAA	ACC	CTA	CGT	CCG	CGT	TTT	336
Glu	Glu	Leu	Lys	Phe	Gln	Cys	Gly	Gln	Lys	Thr	Leu	Arg	Pro	Arg	Phe	
			100					105					110			
AAA	ATC	ATC	GGT	GGT	GAG	TTC	ACC	ACC	ATC	GAA	AAC	CAG	CCG	TGG	TTC	384
Lys	Ile	Ile	Gly	Gly	Glu	Phe	Thr	Thr	Ile	Glu	Asn	Gln	Pro	Trp	Phe	
		115					120					125				
GCT	GCT	ATC	TAC	CGT	CGT	CAC	CGT	GGT	GGT	TCT	GTT	ACC	TAC	GTT	TGC	432
Ala	Ala	Ile	Tyr	Arg	Arg	His	Arg	Gly	Gly	Ser	Val	Thr	Tyr	Val	Cys	
	130					135					140					
GGT	GGT	TCT	CTG	ATC	TCT	CCG	TGC	TGG	GTT	ATC	TCT	GCT	ACC	CAC	TGC	480
Gly	Gly	Ser	Leu	Ile	Ser	Pro	Cys	Trp	Val	Ile	Ser	Ala	Thr	His	Cys	
	145				150					155					160	
TTC	ATC	GAC	TAC	CCG	AAA	AAA	GAA	GAC	TAC	ATC	GTT	TAC	CTC	GGC	CGT	528
Phe	Ile	Asp	Tyr	Pro	Lys	Lys	Glu	Asp	Tyr	Ile	Val	Tyr	Leu	Gly	Arg	
				165					170				175			
TCT	CGT	TTA	AAC	TCT	AAC	ACC	CAG	GGT	GAA	ATG	AAA	TTC	GAA	GTT	GAA	576
Ser	Arg	Leu	Asn	Ser	Asn	Thr	Gln	Gly	Glu	Met	Lys	Phe	Glu	Val	Glu	
			180					185					190			
AAC	CTG	ATC	CTG	CAC	AAA	GAC	TAC	TCT	GCT	GAC	ACC	CTG	GCT	CAC	CAC	624
Asn	Leu	Ile	Leu	His	Lys	Asp	Tyr	Ser	Ala	Asp	Thr	Leu	Ala	His	His	
		195					200					205				



AAC GAC ATC GCT CTG CTA AAA ATC CGT TCT AAA GAA GGT CGT TGC GCT Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala 210 215 220	672
CAG CCG TCT CGT ACC ATC CAG ACC ATC TGC CTG CCG TCT ATG TAC AAC Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn 225 230 235 240	720
GAC CCG CAG TTC GGT ACC TCT TGC GAA ATC ACC GGT TTC GGT AAA GAA Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu 245 250 255	768
AAC TCT ACC GAC TAC CTG TAC CCG GAA CAG CTG AAA ATG ACC GTT GTT Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val 260 265 270	816
AAA CTG ATC TCT CAC CGT GAA TGC CAG CAG CCG CAC TAC TAC GGT TCT Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser 275 280 285	864
GAA GTT ACC ACC AAA ATG CTG TGC GCT GCT GAC CCG CAG TGG AAA ACC Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr 290 295 300	912
GAC TCT TGC CAA GGT GAC TCT GGT GGT CCA CTA GTT TGC TCT CTC CAG Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln 305 310 315 320	960
GGT CGT ATG ACC CTG ACC GGT ATT GTT TCT TGG GGT CGT GGT TGC GCT Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala 325 330 335	1008
CTG AAA GAC AAA CCG GGT GTT TAC ACC CGT GTT TCT CAC TTC CTG CCG Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro 340 345 350	1056
TGG ATC CGT TCT CAC ACC AAA GAA GAA AAC GGT CTG GCT CTG AGC CCG Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro 355 360 365	1104
GTT AAA GCT TAC CCG CGG CCG CTG GGT GGT GGT GGT AAC GGT GAC TTC Val Lys Ala Tyr Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe 370 375 380	1152
GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA Glu Glu Ile Pro Glu Glu Tyr Leu Gln * 385 390	1182

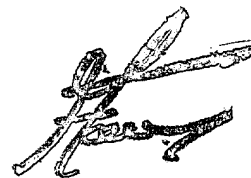
(2) INFORMAÇÃO PARA SEQ ID NO: 46:

- (i) CARACTERÍSTICAS DA SEQUÊNCIA:
 (A) COMPRIMENTO: 394 aminoácidos
 (B) TIPO: aminoácido
 (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 46:

Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys
1				5				10						15	
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala
			20					25						30	



Thr Val Leu Gln Gln Thr Tyr His Ala His Arg Ser Asp Ala Leu Gln
 35 40 45
 Leu Gly Leu Gly Lys His Asn Tyr Cys Arg Asn Pro Asp Asn Arg Arg
 50 55 60
 Arg Pro Trp Cys Tyr Val Gln Val Gly Leu Lys Pro Leu Val Gln Glu
 65 70 75 80
 Cys Met Val His Asp Cys Ala Asp Gly Lys Lys Pro Ser Ser Pro Pro
 85 90 95
 Glu Glu Leu Lys Phe Gln Cys Gly Gln Lys Thr Leu Arg Pro Arg Phe
 100 105 110
 Lys Ile Ile Gly Gly Glu Phe Thr Thr Ile Glu Asn Gln Pro Trp Phe
 115 120 125
 Ala Ala Ile Tyr Arg Arg His Arg Gly Gly Ser Val Thr Tyr Val Cys
 130 135 140
 Gly Gly Ser Leu Ile Ser Pro Cys Trp Val Ile Ser Ala Thr His Cys
 145 150 155 160
 Phe Ile Asp Tyr Pro Lys Lys Glu Asp Tyr Ile Val Tyr Leu Gly Arg
 165 170 175
 Ser Arg Leu Asn Ser Asn Thr Gln Gly Glu Met Lys Phe Glu Val Glu
 180 185 190
 Asn Leu Ile Leu His Lys Asp Tyr Ser Ala Asp Thr Leu Ala His His
 195 200 205
 Asn Asp Ile Ala Leu Leu Lys Ile Arg Ser Lys Glu Gly Arg Cys Ala
 210 215 220
 Gln Pro Ser Arg Thr Ile Gln Thr Ile Cys Leu Pro Ser Met Tyr Asn
 225 230 235 240
 Asp Pro Gln Phe Gly Thr Ser Cys Glu Ile Thr Gly Phe Gly Lys Glu
 245 250 255
 Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
 260 265 270
 Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
 275 280 285
 Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
 290 295 300
 Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
 305 310 315 320
 Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
 325 330 335
 Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
 340 345 350
 Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
 355 360 365



Val Lys Ala Tyr Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
 370 375 380
 Glu Glu Ile Pro Glu Glu Tyr Leu Gln *
 385 390

(2) INFORMAÇÃO PARA SEQ ID NO: 47:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1182 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(iii) HIPOTÉTICO: não

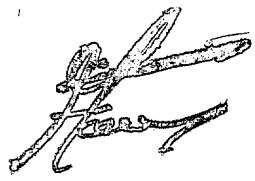
(iv) ANTI-SENTIDO: não

(ix) CARACTERÍSTICA:

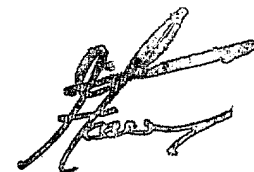
- (A) NOME/CHAVE: CDS
- (B) LOCALIZAÇÃO: 1..1182
- (D) OUTRA INFORMAÇÃO: /function = "metionina N-terminal é separada proteoliticamente em E.coli"
 /product = "gene sintético para a proteína M23 (pSJ95)"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 47:

ATG	AGC	AAA	ACT	TGC	TAC	GAA	GGT	AAC	GGT	CAC	TTC	TAC	CGT	GGT	AAG	43
Met	Ser	Lys	Thr	Cys	Tyr	Glu	Gly	Asn	Gly	His	Phe	Tyr	Arg	Gly	Lys	
1				5				10					15			
GCT	TCT	ACC	GAC	ACC	ATG	GGT	CGT	CCG	TGC	CTG	CCG	TGG	AAC	TCT	GCT	96
Ala	Ser	Thr	Asp	Thr	Met	Gly	Arg	Pro	Cys	Leu	Pro	Trp	Asn	Ser	Ala	
			20				25						30			
ACC	GTT	CTG	CAG	CAG	ACC	TAC	CAC	GCT	CAC	CGT	TCT	GAT	GCA	TTG	CAG	144
Thr	Val	Leu	Gln	Gln	Thr	Tyr	His	Ala	His	Arg	Ser	Asp	Ala	Leu	Gln	
			35				40					45				
CTG	GGT	CTG	GGT	AAA	CAC	AAC	TAC	TGC	CGT	AAC	CCG	GAC	AAC	CGT	CGT	192
Leu	Gly	Leu	Gly	Lys	His	Asn	Tyr	Cys	Arg	Asn	Pro	Asp	Asn	Arg	Arg	
	50					55					60					
CGT	CCG	TGG	TGC	TAC	GTT	CAG	GTT	GGT	CTG	AAA	CCG	CTA	GTT	CAG	GAA	240
Arg	Pro	Trp	Cys	Tyr	Val	Gln	Val	Gly	Leu	Lys	Pro	Leu	Val	Gln	Glu	
65					70					75					80	
TGC	ATG	GTT	CAC	GAC	TGC	GCT	GAC	GGT	AAA	AAA	CCG	TCT	TCT	CCG	CCG	288
Cys	Met	Val	His	Asp	Cys	Ala	Asp	Gly	Lys	Lys	Pro	Ser	Ser	Pro	Pro	
				85					90					95		
GAA	GAG	CTC	AAA	TTC	CAG	TGC	GGT	CAA	AAA	ACC	CTA	CGT	CCG	CGT	TTT	336
Glu	Glu	Leu	Lys	Phe	Gln	Cys	Gly	Gln	Lys	Thr	Leu	Arg	Pro	Arg	Phe	
			100					105					110			
AAA	ATC	ATC	GGT	GGT	GAG	TTC	ACC	ACC	ATC	GAA	AAC	CAG	CCG	TGG	TTC	384
Lys	Ile	Ile	Gly	Gly	Glu	Phe	Thr	Thr	Ile	Glu	Asn	Gln	Pro	Trp	Phe	
			115				120						125			



GCT Ala	GCT Ala	ATC Ile	TAC Tyr	CGT Arg	CGT Arg	CAC His	CGT Arg	GGT Gly	GGT Gly	TCT Ser	GTT Val	ACC Thr	TAC Tyr	GTT Val	TGC Cys	432
	130					135					140					
GGT Gly	GGT Gly	TCT Ser	CTG Leu	ATC Ile	TCT Ser	CCG Pro	TGC Cys	TGG Trp	GTT Val	ATC Ile	TCT Ser	GCT Ala	ACC Thr	CAC His	TGC Cys	480
145					150					155					160	
TTC Phe	ATC Ile	GAC Asp	TAC Tyr	CCG Pro	AAA Lys	AAA Lys	GAA Glu	GAC Asp	TAC Tyr	ATC Ile	GTT Val	TAC Tyr	CTC Leu	GGC Gly	CGT Arg	528
				165					170					175		
TCT Ser	CGT Arg	TTA Leu	AAC Asn	TCT Ser	AAC Asn	ACC Thr	CAG Gln	GGT Gly	GAA Glu	ATG Met	AAA Lys	TTC Phe	GAA Glu	GTT Val	GAA Glu	576
			180					185					190			
AAC Asn	CTG Leu	ATC Ile	CTG Leu	CAC His	AAA Lys	GAC Asp	TAC Tyr	TCT Ser	GCT Ala	GAC Asp	ACC Thr	CTG Leu	GCT Ala	CAC His	CAC His	624
	195						200					205				
AAC Asn	GAC Asp	ATC Ile	GCT Ala	CTG Leu	CTA Leu	AAA Lys	ATC Ile	CGT Arg	TCT Ser	AAA Lys	GAA Glu	GGT Gly	CGT Arg	TGC Cys	GCT Ala	672
	210					215					220					
CAG Gln	CCG Pro	TCT Ser	CGT Arg	ACC Thr	ATC Ile	CAG Gln	ACC Thr	ATC Ile	TGC Cys	CTG Leu	CCG Pro	TCT Ser	ATG Met	TAC Tyr	AAC Asn	720
225				230						235					240	
GAC Asp	CCG Pro	CAG Gln	TTC Phe	GGT Gly	ACC Thr	TCT Ser	TGC Cys	GAA Glu	ATC Ile	ACC Thr	GGT Gly	TTC Phe	GGT Gly	AAA Lys	GAA Glu	768
				245				250						255		
AAC Asn	TCT Ser	ACC Thr	GAC Asp	TAC Tyr	CTG Leu	TAC Tyr	CCG Pro	GAA Glu	CAG Gln	CTG Leu	AAA Lys	ATG Met	ACC Thr	GTT Val	GTT Val	816
			260					265				270				
AAA Lys	CTG Leu	ATC Ile	TCT Ser	CAC His	CGT Arg	GAA Glu	TGC Cys	CAG Gln	CAG Gln	CCG Pro	CAC His	TAC Tyr	TAC Tyr	GGT Gly	TCT Ser	864
	275						280					285				
GAA Glu	GTT Val	ACC Thr	ACC Thr	AAA Lys	ATG Met	CTG Leu	TGC Cys	GCT Ala	GCT Ala	GAC Asp	CCG Pro	CAG Gln	TGG Trp	AAA Lys	ACC Thr	912
	290					295					300					
GAC Asp	TCT Ser	TGC Cys	CAA Gln	GGT Gly	GAC Asp	TCT Ser	GGT Gly	GGT Gly	CCA Pro	CTA Leu	GTT Val	TGC Cys	TCT Ser	CTC Leu	CAG Gln	960
305				310						315				320		
GGT Gly	CGT Arg	ATG Met	ACC Thr	CTG Leu	ACC Thr	GGT Gly	ATT Ile	GTT Val	TCT Ser	TGG Trp	GGT Gly	CGT Arg	GGT Gly	TGC Cys	GCT Ala	1008
				325					330					335		
CTG Leu	AAA Lys	GAC Asp	AAA Lys	CCG Pro	GGT Gly	GTT Val	TAC Tyr	ACC Thr	CGT Arg	GTT Val	TCT Ser	CAC His	TTC Phe	CTG Leu	CCG Pro	1056
		340					345						350			
TGG Trp	ATC Ile	CGT Arg	TCT Ser	CAC His	ACC Thr	AAA Lys	GAA Glu	GAA Glu	AAC Asn	GGT Gly	CTG Leu	GCT Ala	CTG Leu	AGC Ser	CCG Pro	1104
		355					360					365				
GTT Val	GTT Val	GCT Ala	TTC Phe	CCG Pro	CGG Arg	CCG Pro	CTG Leu	GGT Gly	GGT Gly	GGT Gly	GGT Gly	AAC Asn	GGT Gly	GAC Asp	TTC Phe	1152
	370					375						380				



GAA GAA ATC CCG GAA GAA TAC CTG CAA TAA
Glu Glu Ile Pro Glu Glu Tyr Leu Gln *
385 390

1182

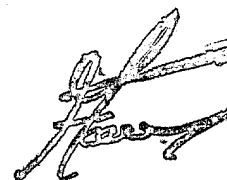
(2) INFORMAÇÃO PARA SEQ ID NO: 48:

- (i) CARACTERÍSTICAS DA SEQUÊNCIA:
 - (A) COMPRIMENTO: 394 aminoácidos
 - (B) TIPO: aminoácido
 - (D) TOPOLOGIA: linear

- (ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 48:

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



Asn Ser Thr Asp Tyr Leu Tyr Pro Glu Gln Leu Lys Met Thr Val Val
260 265 270

Lys Leu Ile Ser His Arg Glu Cys Gln Gln Pro His Tyr Tyr Gly Ser
275 290 295

Glu Val Thr Thr Lys Met Leu Cys Ala Ala Asp Pro Gln Trp Lys Thr
290 295 300

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Leu Gln
305 310 315 320

Gly Arg Met Thr Leu Thr Gly Ile Val Ser Trp Gly Arg Gly Cys Ala
325 330 335

Leu Lys Asp Lys Pro Gly Val Tyr Thr Arg Val Ser His Phe Leu Pro
340 345 350

Trp Ile Arg Ser His Thr Lys Glu Glu Asn Gly Leu Ala Leu Ser Pro
355 360 365

Val Val Ala Phe Pro Arg Pro Leu Gly Gly Gly Gly Asn Gly Asp Phe
370 375 380

Glu Glu Ile Pro Glu Glu Tyr Leu Gln *
385 390

(2) INFORMAÇÃO PARA SEQ ID NO: 49:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 63 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0105"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 49

TATGAGCAAACTTGCTACG AAGGTAACGG TCACTTCTAC CGTGGTAAGG CTTCTACCGA 60
CAC 63

(2) INFORMAÇÃO PARA SEQ ID NO: 50:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 65 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0106"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) Descrição da sequência: SEQ ID NO: 50:

CATGGTGTCTG GTAGAAGCCT TACCACGGTA GAAGTGACCG TTACCTTCGT AGCAAGTTTT	60
GCTCA	65

(2) INFORMAÇÃO PARA SEQ ID NO: 51:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 83 pares de bases

(B) TIPO: nucleótido

(C) TIPO DE CADEIA: cadeia simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0220"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) Descrição da sequência: SEQ ID NO: 51:

CGGTAAAGGC TTCCCGAGG CCTGGTGGTG GTGGTAACGG TGACTTCGAA GAAATCCCGG	60
AAGAGTACCT GTGATAGGAT CAA	83

(2) INFORMAÇÃO PARA SEQ ID NO: 52:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 91 pares de bases

(B) TIPO: nucleótido

(C) TIPO DE CADEIA: cadeia simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0221"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) Descrição da sequência: SEQ ID NO: 52:

CTAGTTGATC CTATCACAGG TACTCTTCCG GGATTTCTTC GAAGTCACCG TTACCACCAC	60
CACCAGGCCT CGGAAAGCC TTAACCGGC T	91

(2) INFORMAÇÃO PARA SEQ ID NO: 53:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 101 pares de bases

- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0222"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) Descrição da sequência: SEQ ID NO: 53:

CGCCGAGCCC	GCCGAGCCCC	CCGGGTGGTT	TCCCGAGGCC	TGGTGGTGGT	GGTAACGGTG	60
ACTTCGAAGA	AATCCCGGAA	GAGTACCTGT	GATAGGATCA	A		101

(2) INFORMAÇÃO PARA SEQ ID NO: 54:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 109 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0223"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) Descrição da sequência: SEQ ID NO: 54:

CTAGTTGATC	CTATCACAGG	TACTCTTCCG	GGATTTCTTC	GAAGTCACCG	TTACCACCAC	60
CACCAGGCC	CGGAAACCA	CCCGCGGGC	TCGCGGGCT	CGGCGGGCT		109

(2) INFORMAÇÃO PARA SEQ ID NO: 55:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 83 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0224"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) Descrição da sequência: SEQ ID NO: 55:



CGCCGGGTGG TTTCCCGAGG CCTGGTGGTG GTGGTAACGG TGACTTCGAA GAAATCCCGG 60
AAGAGTACCT GTGATAGGAT CAA 83

(2) INFORMAÇÃO PARA SEQ ID NO: 56:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 91 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0225"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) Descrição da sequência: SEQ ID NO: 56:

CTAGTTGATC CTATCACAGG TACTCTTCCG GGATTTCTTC GAAGTCACCG TTACCACCAC 60
CACCAGGCCT CGGGAAACCA CCCGGCGGGC T 91

(2) INFORMAÇÃO PARA SEQ ID NO: 57:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 104 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0226"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não



(xi) Descrição da sequência: SEQ ID NO: 57:

CGCCGAGCCC	GCCGAGCCCG	CCGGGTGGTT	TCGGTCCGAG	GCCTGGTGGT	GGTGGTAACG	60
GTGACTTCGA	AGAAATCCCG	GAAGAGTACC	TGTGATAGGA	TCAA		104

(2) INFORMAÇÃO PARA SEQ ID NO: 58:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 112 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0227"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) Descrição da sequência: SEQ ID NO: 58:

CTAGTTGATC	CTATCACAGG	TACTCTCCG	GGATTTCTTC	GAAGTCACCG	TTACCACCAC	60
CACCAGGCTT	CGGACCGAAA	CCACCCGGCG	GGCTCGGCGG	GCTCGGCGGG	CT	112

(2) INFORMAÇÃO PARA SEQ ID NO: 59:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

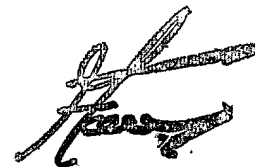
- (A) COMPRIMENTO: 58 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0265"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não



(xi) Descrição da sequência: SEQ ID NO: 59:

CACCCGGCGG AGACGGCGGG CTCAGAGCCA GACCGTTTTTTC TTCTTTGGTG TGAGAACG 58

(2) INFORMAÇÃO PARA SEQ ID NO: 60:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 58 pares de bases

(B) TIPO: nucleótido

(C) TIPO DE CADEIA: cadeia simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0281"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) Descrição da sequência: SEQ ID NO: 60:

CGTCCGGGTG GTGGTGGTAA CGGTGACTTC GAAGAAATCC CGGAAGAATA CCTGTAAG 58

(2) INFORMAÇÃO PARA SEQ ID NO: 61:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 70 pares de bases

(B) TIPO: nucleótido

(C) TIPO DE CADEIA: cadeia simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0282"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) Descrição da sequência: SEQ ID NO: 61:

GATCCGTTCT CACACCAAAG AAGAAAACGG TCTGGCTCTG AGCCCGCCGT CTCCGCCGGG 60
TGGTTTCCCG 70

(2) INFORMAÇÃO PARA SEQ ID NO: 62:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 70 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0283"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) Descrição da sequência: SEQ ID NO: 62:

```
CTAGCTTACA GGTATTCTTC CGGGATTTCT TCGAAGTCAC CGTTACCACC ACCACCCGGA 60
CGCGGGAAAC 70
```

(2) INFORMAÇÃO PARA SEQ ID NO: 63:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 30 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0329"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) Descrição da sequência: SEQ ID NO: 63:

```
AAGAAATCCC GGAAGAATAC CTGCAATAAG 30
```

(2) INFORMAÇÃO PARA SEQ ID NO: 64:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 54 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0330"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) Descrição da sequência: SEQ ID NO: 64:

CGGTTAAGGC TTGGGGACCG CGGCCGCTGG GTGGTGGTGG TAACGGTGAC TTCG

54

(2) INFORMAÇÃO PARA SEQ ID NO: 65:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 42 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0331"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

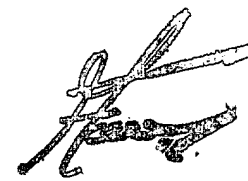
(xi) Descrição da sequência: SEQ ID NO: 65:

ACCACCACC AGCGGCCGCG GTCCCCAAGC CTTAACCGGG CT

42

(2) INFORMAÇÃO PARA SEQ ID NO: 66:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:



- (A) COMPRIMENTO: 50 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0332"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) Descrição da sequência: SEQ ID NO: 66:

CTAGCTTATT GCAGGTATTC TTCCGGGATT TCTTCGAAGT CACCGTTACC

50

(2) INFORMAÇÃO PARA SEQ ID NO: 67:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 21 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0333"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) Descrição da sequência: SEQ ID NO: 67:

CGGTTAAGGC TTTCGGACCG C

21

(2) INFORMAÇÃO PARA SEQ ID NO: 68:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 29 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0334"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) Descrição da sequência: SEQ ID NO: 68:

GGCCGCGGTC CGAAAGCCTT AACCGGGCT

29

(2) INFORMAÇÃO PARA SEQ ID NO: 69:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 21 pares de bases

(B) TIPO: nucleótido

(C) TIPO DE CADEIA: cadeia simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0335"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) Descrição da sequência: SEQ ID NO: 69:

CGGTTTCGGGC TTTCGGTCCG C

21

(2) INFORMAÇÃO PARA SEQ ID NO: 70:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 29 pares de bases

(B) TIPO: nucleótido

(C) TIPO DE CADEIA: cadeia simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0336"

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 72:

GGCCGCGGTC CGTAAGCCTT AACCGGGCT

29

(2) INFORMAÇÃO PARA SEQ ID NO: 73:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 21 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0339"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 73:

CGGTTGTTGC TTTCGGTCCG C

21

(2) INFORMAÇÃO PARA SEQ ID NO: 74:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 29 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0340"

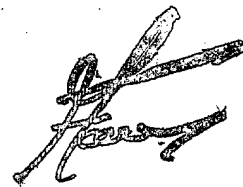
(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 74:

GGCCGCGGAC CGAAAGCAAC AACCGGGCT

29



(2) INFORMAÇÃO PARA SEQ ID NO: 75:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0341"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 75:

CGGTTCGGGC TTTCCCGC

18

(2) INFORMAÇÃO PARA SEQ ID NO: 76:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 26 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0342"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 76:

GGCCGCGGGA AAGCCCGAAC CGGGCT

26

(2) INFORMAÇÃO PARA SEQ ID NO: 77:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0343"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 77:

CGGTTAAGGC TTACCCGC

18

(2) INFORMAÇÃO PARA SEQ ID NO: 78:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 26 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0344"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 78:

GGCCGCGGGT AAGCCTTAAC CGGGCT

26

(2) INFORMAÇÃO PARA SEQ ID NO: 79:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0347"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 79:

CGGTTGTTGC TTTCCCGC

18

(2) INFORMAÇÃO PARA SEQ ID NO: 80:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 26 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0348"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 80:

GGCCGCGGGA AAGCAACAAC CGGGCT

26

(2) INFORMAÇÃO PARA SEQ ID NO: 81:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0381"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 81:

CGGTTAAGGC TTGGCCGC

18

(2) INFORMAÇÃO PARA SEQ ID NO: 82:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 26 pares de bases

(B) TIPO: nucleótido

(C) TIPO DE CADEIA: cadeia simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0383"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 82:

GGCCGCGGCC AAGCCTTAAC CGGGCT

26

(2) INFORMAÇÃO PARA SEQ ID NO: 83:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 18 pares de bases

(B) TIPO: nucleótido

(C) TIPO DE CADEIA: cadeia simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0384"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 83:

CGGTTAAGGC TTTCCCGC

18

(2) INFORMAÇÃO PARA SEQ ID NO: 84:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 26 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0385"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 84:

GGCCGCGGGA AAGCCTTAAC CGGGCT

26

(2) INFORMAÇÃO PARA SEQ ID NO: 85:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0386"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 85:

CGGTTGTAGT TTTCCCGC

18

(2) INFORMAÇÃO PARA SEQ ID NO: 86:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0387"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 86:

CGGTTGAAGT TTTCCCGC

18

(2) INFORMAÇÃO PARA SEQ ID NO: 87:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 26 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0388"

(iii) HIPOTÉTICO: não

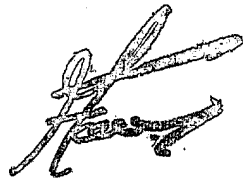
(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 87:

GGCCGCACTA CAACTACAAC CGGGCT

26

(2) INFORMAÇÃO PARA SEQ ID NO: 88:



(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0389"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 88:

CGGTTGTAGT TGTAGTGC

18

(2) INFORMAÇÃO PARA SEQ ID NO: 89:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 26 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0390"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 89:

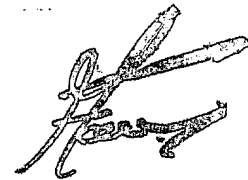
GGCCGCGGGA AAACCTCAAC CGGGCT

26

(2) INFORMAÇÃO PARA SEQ ID NO: 90:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 26 pares de bases
- (B) TIPO: nucleótido



(C) TIPO DE CADEIA: cadeia simples
(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0391"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 90:

GGCCGCGGGA AACTACAAC CGGGCT

26

(2) INFORMAÇÃO PARA SEQ ID NO: 91:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 72 pares de bases

(B) TIPO: nucleótido

(C) TIPO DE CADEIA: cadeia simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0392"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 91:

CTAGCTTATT CGTTTTTTC TTCGTCTTCC CAGAACGGTT CGTATTTGTC GTTCGGGTTT
CGCAGCAGGA AC

60

72

(2) INFORMAÇÃO PARA SEQ ID NO: 92:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 72 pares de bases

(B) TIPO: nucleótido

(C) TIPO DE CADEIA: cadeia simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0393"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 92:

GGCCGTTTCCT GCTGCGGAAC CCGAACGACA AATACGAACC GTTCTGGGAA GACGAAGAAA 60
AAAACGAATA AG 72

(2) INFORMAÇÃO PARA SEQ ID NO: 93:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0453"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 93:

TGGTTAAAGC TTTCCCGC 18

(2) INFORMAÇÃO PARA SEQ ID NO: 94:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 26 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

(A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0454"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 94:

GGCCGCGGGA AAGCTTTAAC CAGGCT

26

(2) INFORMAÇÃO PARA SEQ ID NO: 95:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 18 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0455"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 95:

TGGTTGTTGC TTTCCCGC

18

(2) INFORMAÇÃO PARA SEQ ID NO: 96:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 26 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0456"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 96:

GGCCGCGGGA AAGCAACAAC CAGGCT

26

(2) INFORMAÇÃO PARA SEQ ID NO: 97:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 47 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0465"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 97:

GGCCGCGGGA ACAGAGCCAG ACCGTTTTCT TCTTTGGTGT GAGAACG

47

(2) INFORMAÇÃO PARA SEQ ID NO: 98:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

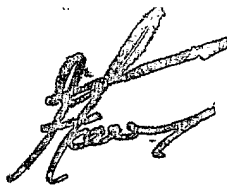
- (A) COMPRIMENTO: 47 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0466"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não



(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 98:

GATCCGTTCT CACACCAAAG AAGAAAACGG TCTGGCTCTG TTCCCGC

47

(2) INFORMAÇÃO PARA SEQ ID NO: 99:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 38 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0467"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 99:

CGGTTAAGGC TTTCCCGCGG CCGTTCCTGC TGC GGAAC

38

(2) INFORMAÇÃO PARA SEQ ID NO: 100:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 54 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0468"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 100:

TTTGTCGTTT GGGTTCCGCA GCAGGAACGG CCGCGGGAAA GCCTTAACCG GGCT

54

(2) INFORMAÇÃO PARA SEQ ID NO: 101:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 44 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0469"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 101:

CTAGCTTATT CGTTTTTTTC TTCGTCTTCC CAGAACGGTT CGTA

44

(2) INFORMAÇÃO PARA SEQ ID NO: 102:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 52 pares de bases
- (B) TIPO: nucleótido
- (C) TIPO DE CADEIA: cadeia simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: outro ácido nucleico

- (A) Descrição: /desc = "ADN sintético; sequência nucleotídica para o oligo 0470"

(iii) HIPOTÉTICO: não

(iv) ANTI-SENTIDO: não

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID NO: 102:


CCGAACGACA AATACGAACC GTTCTGGGAA GACGAAGAAA AAAACGAATA AG

52

Lisboa, 26. ABR. 2000

Por GRÜNENTHAL GmbH

- O AGENTE OFICIAL -


O ADJUNTO

ENG.º ANTÓNIO JOÃO DA CUNHA FERREIRA Ag. Of. Pr. Ind. Rua das Flores, 74 - 4.º 1200 LISBOA
--

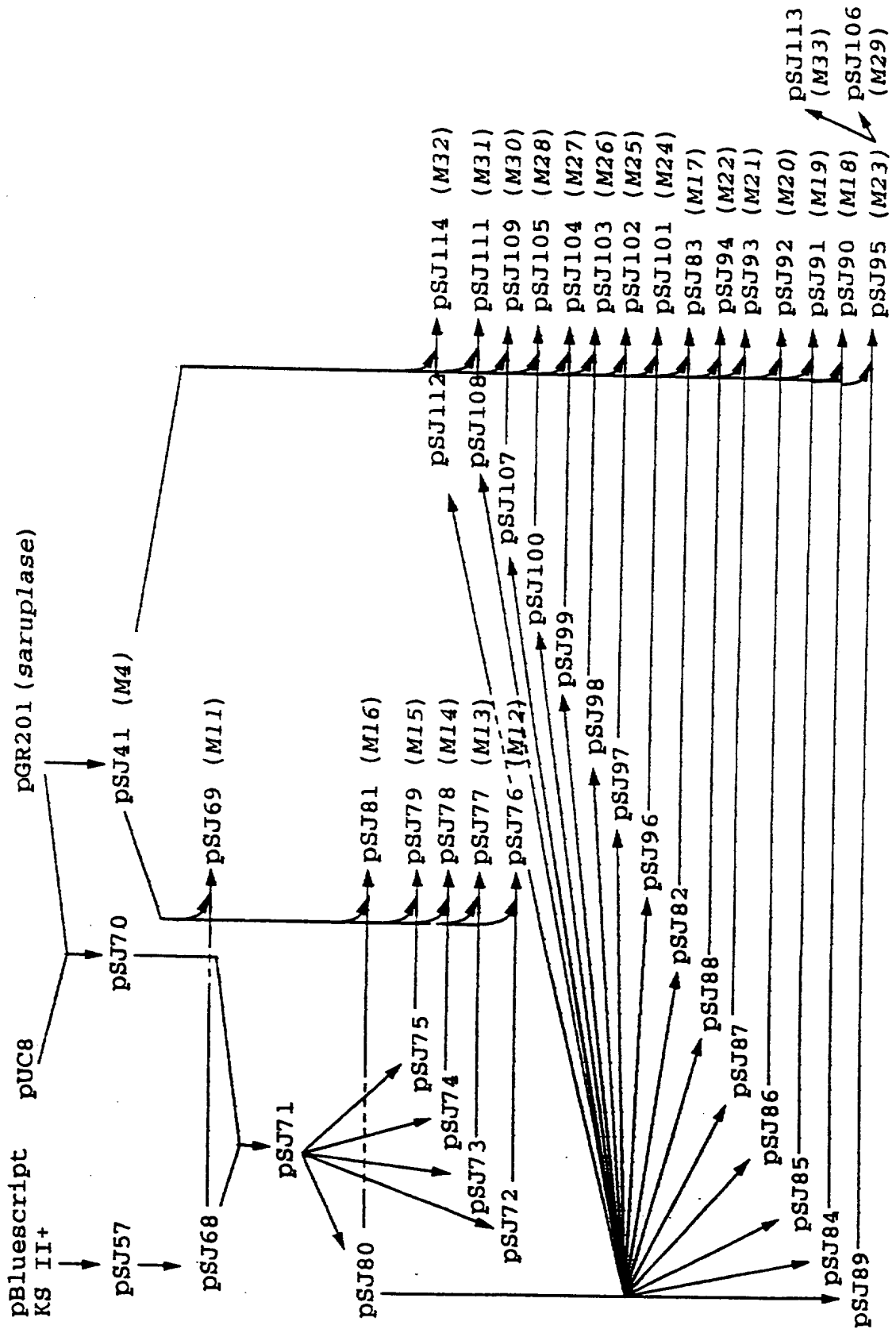


Figura 2

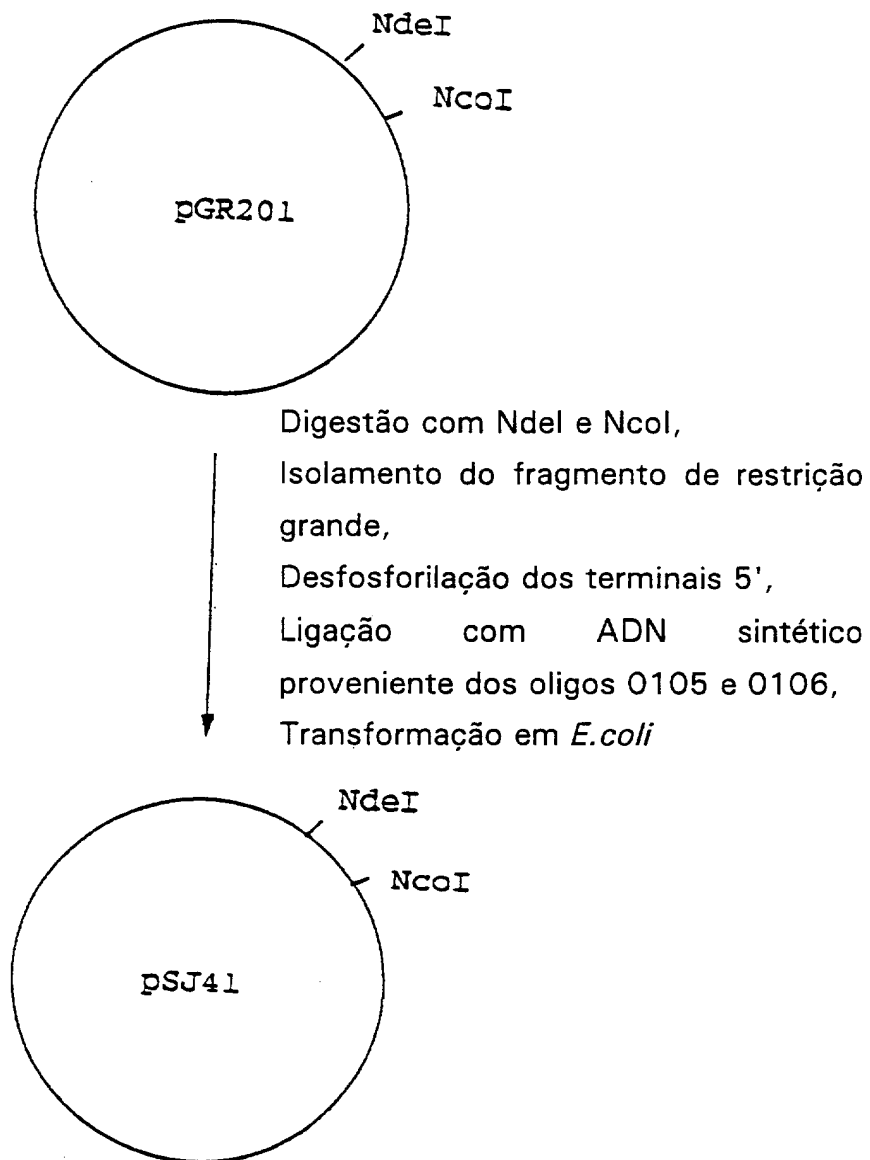


Figura 2a

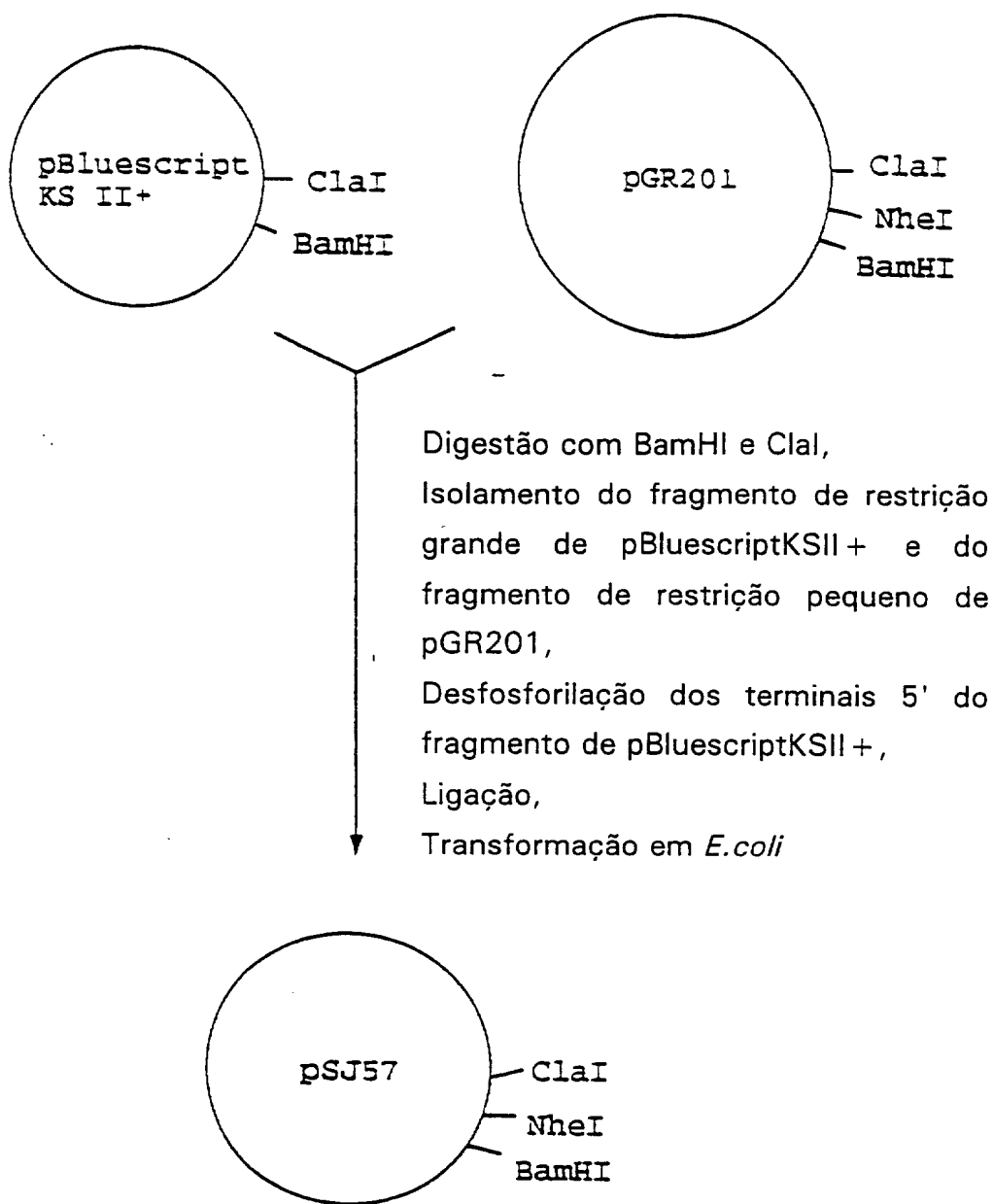


Figura 2b

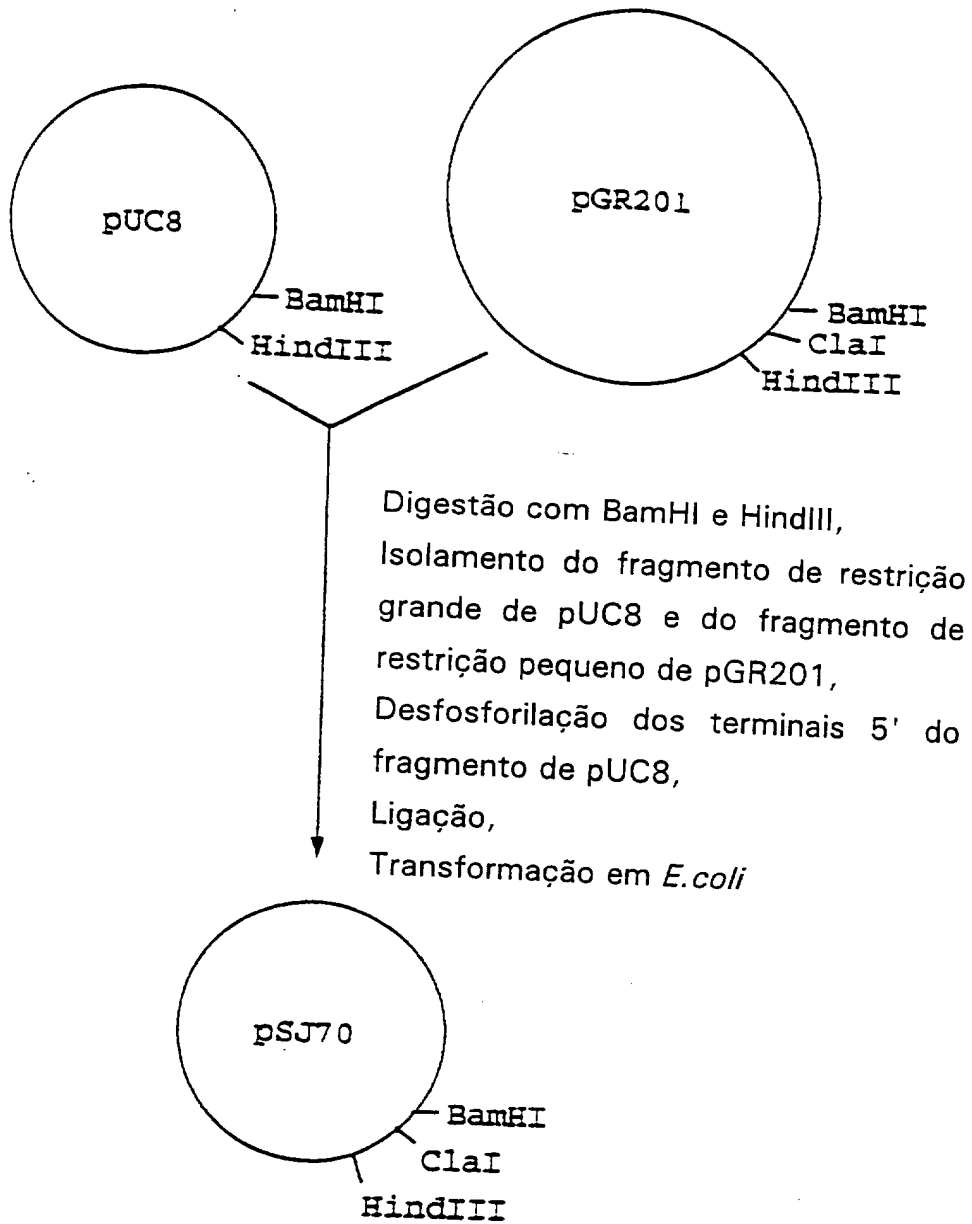
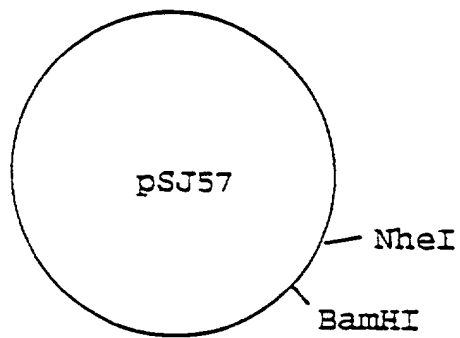


Figura 2c



Digestão com NheI e BamHI,
Isolamento do fragmento de restrição^o grande,
grande, desfosforilação dos terminais 5', te dos
Ligação com ADN sintético proveniente
dos oligos 0265, 0281, 0282, 0283,
Transformação em *E.coli*

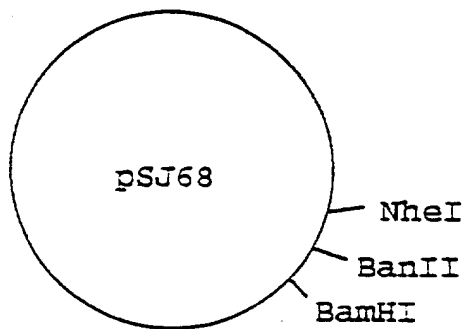


Figura 2d:

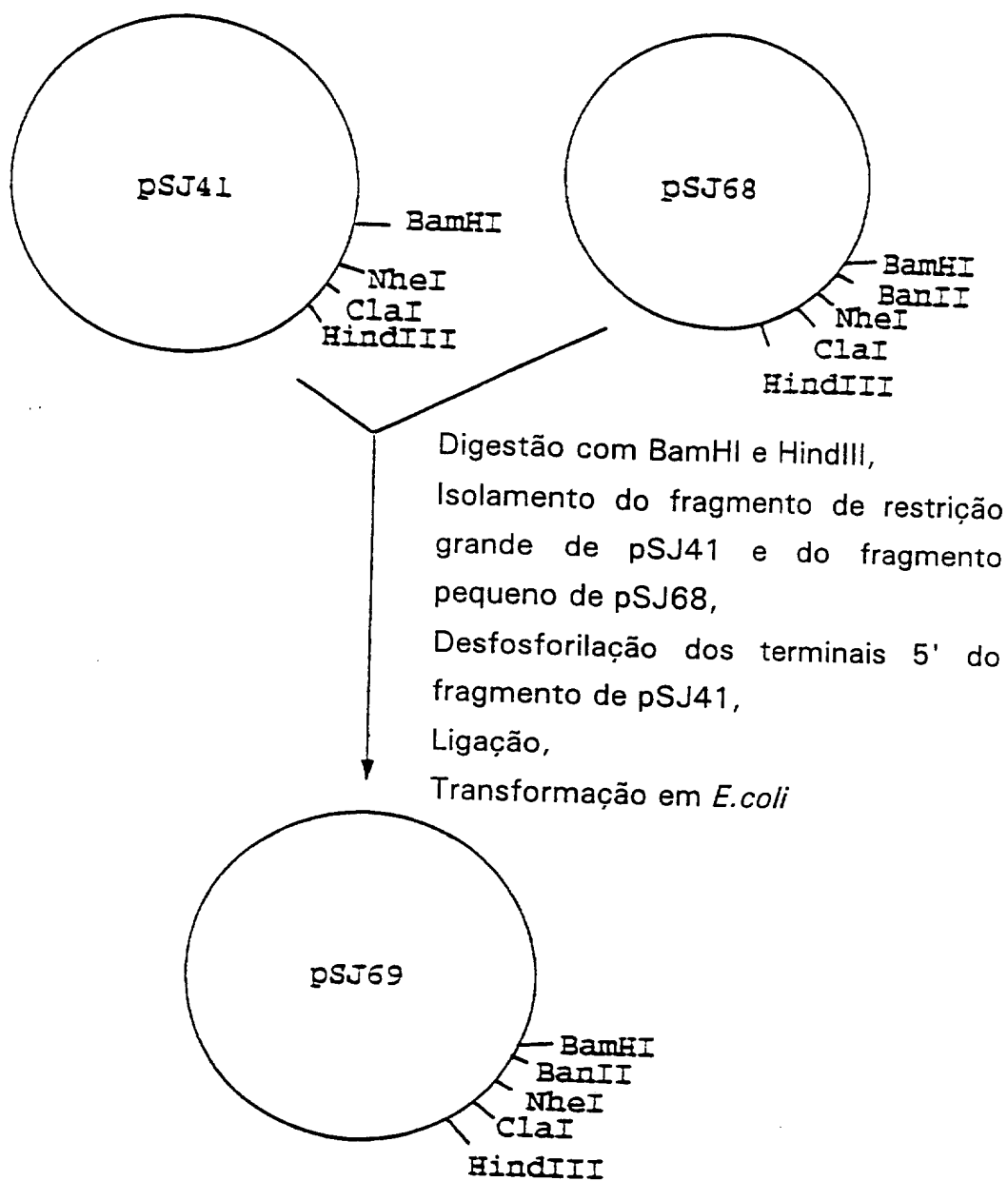


Figura 2e:

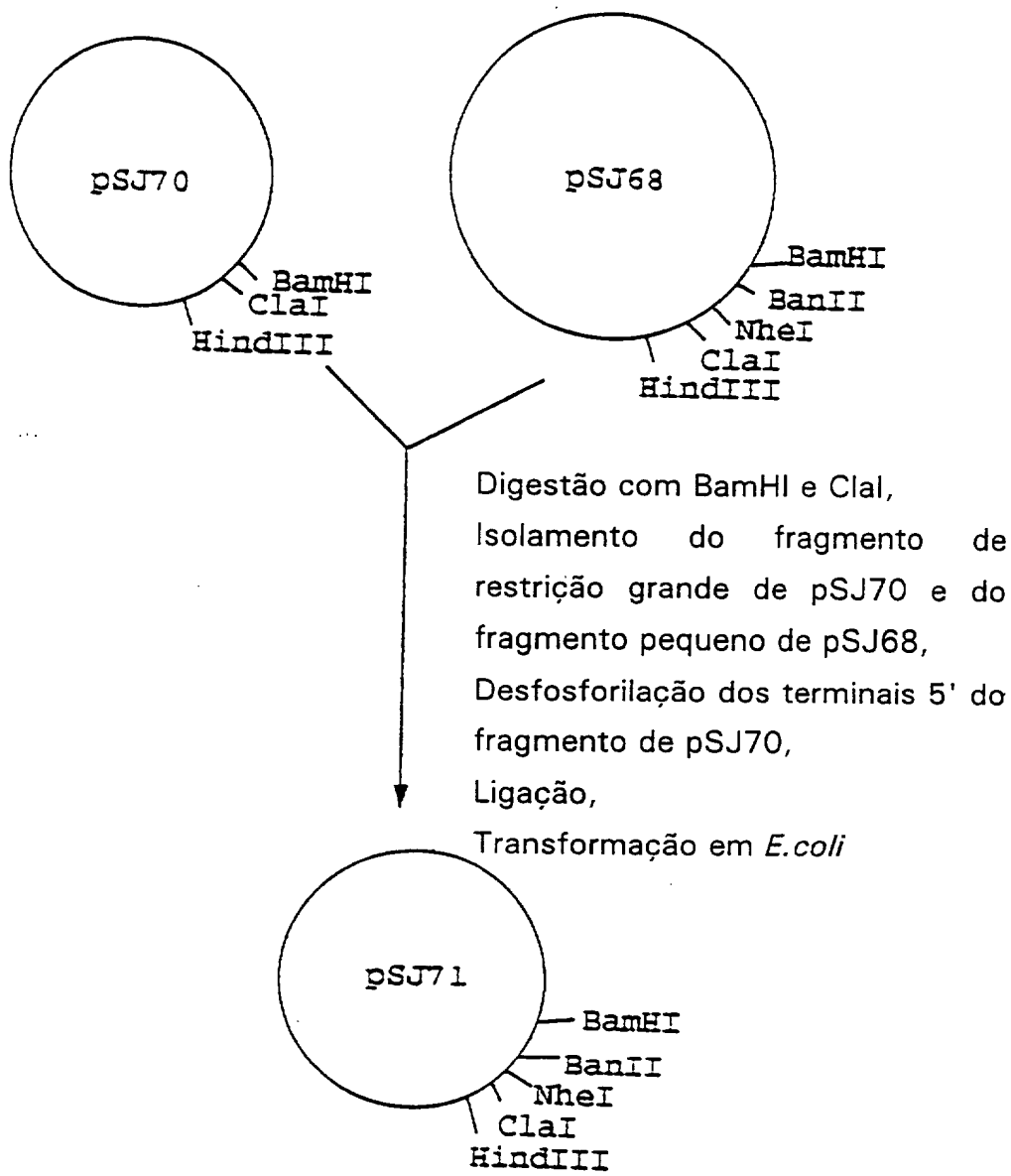


Figura 2f:

9/19

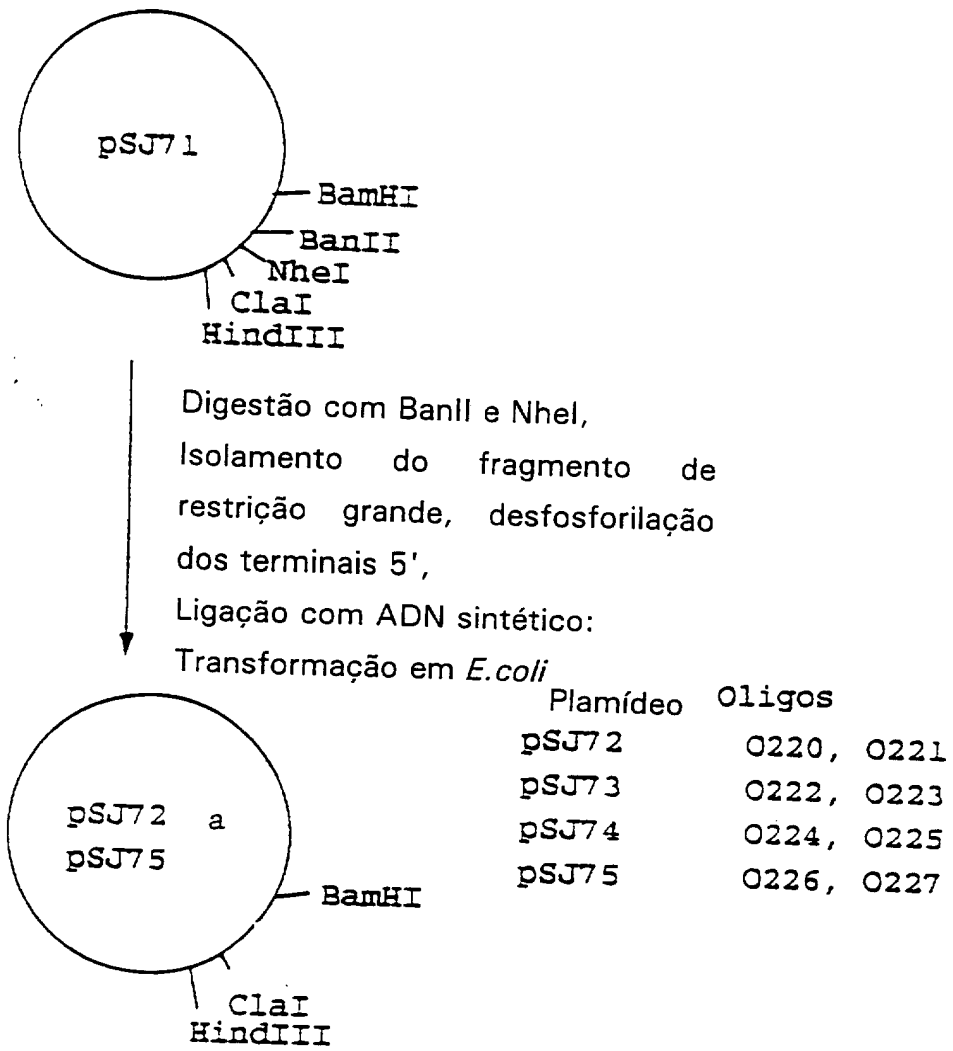


Figura 2g:

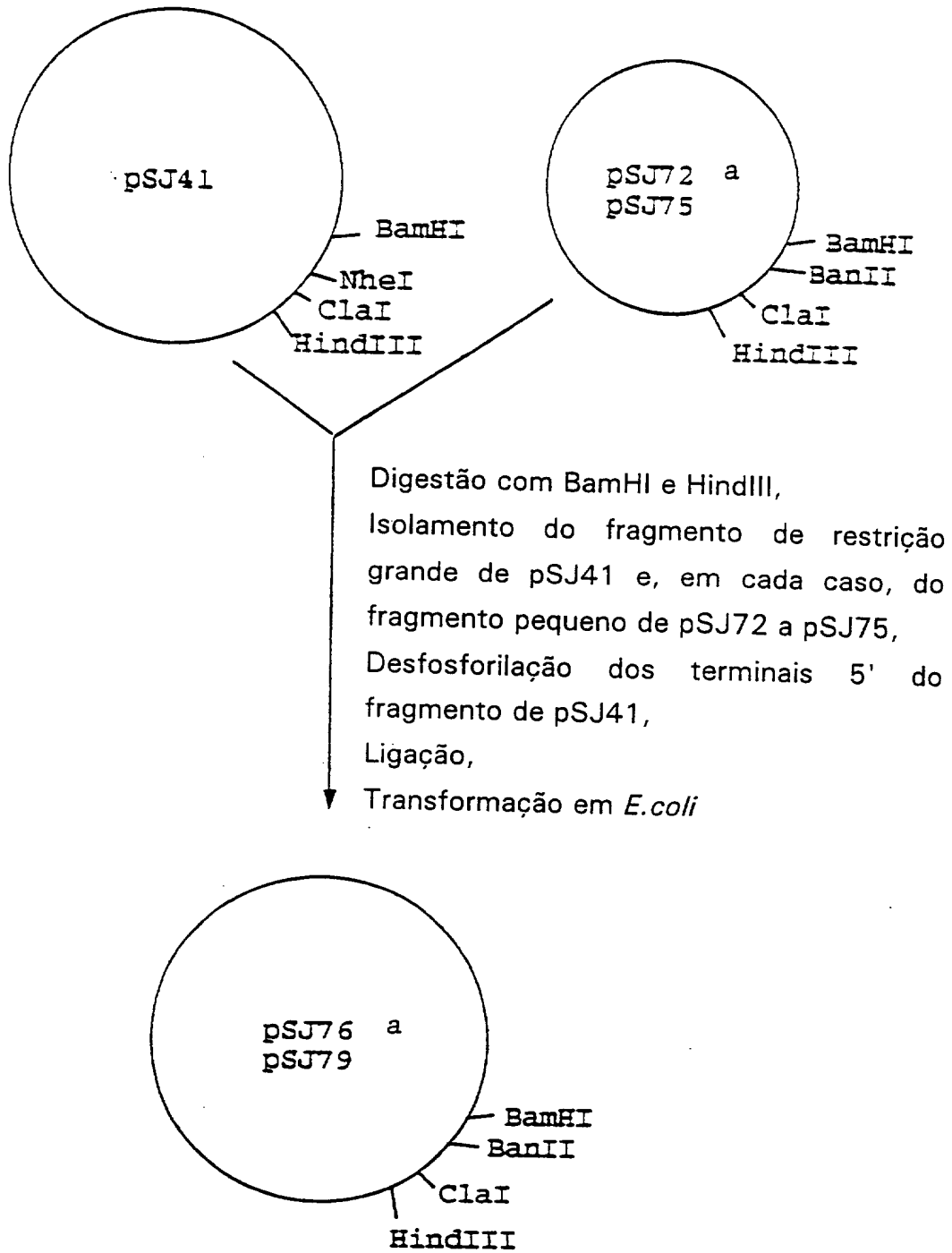


Figura 2h:

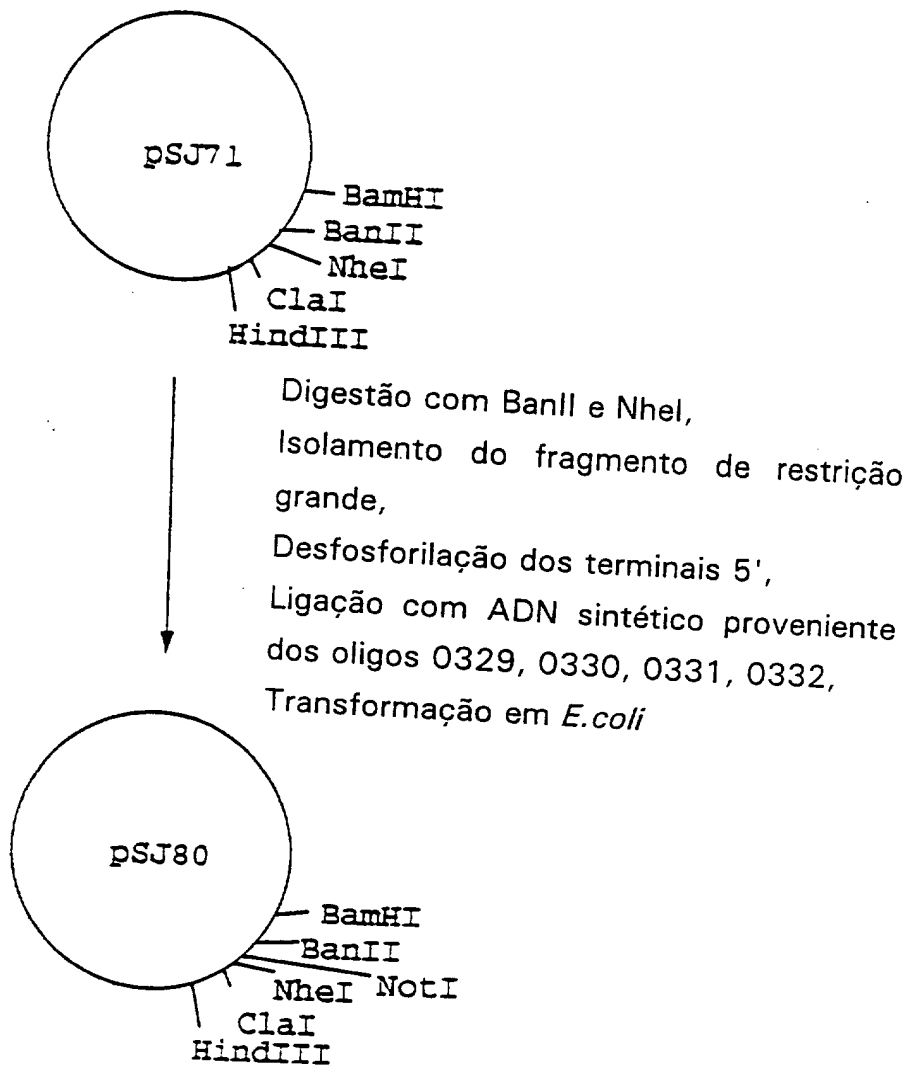


Figura 2i:

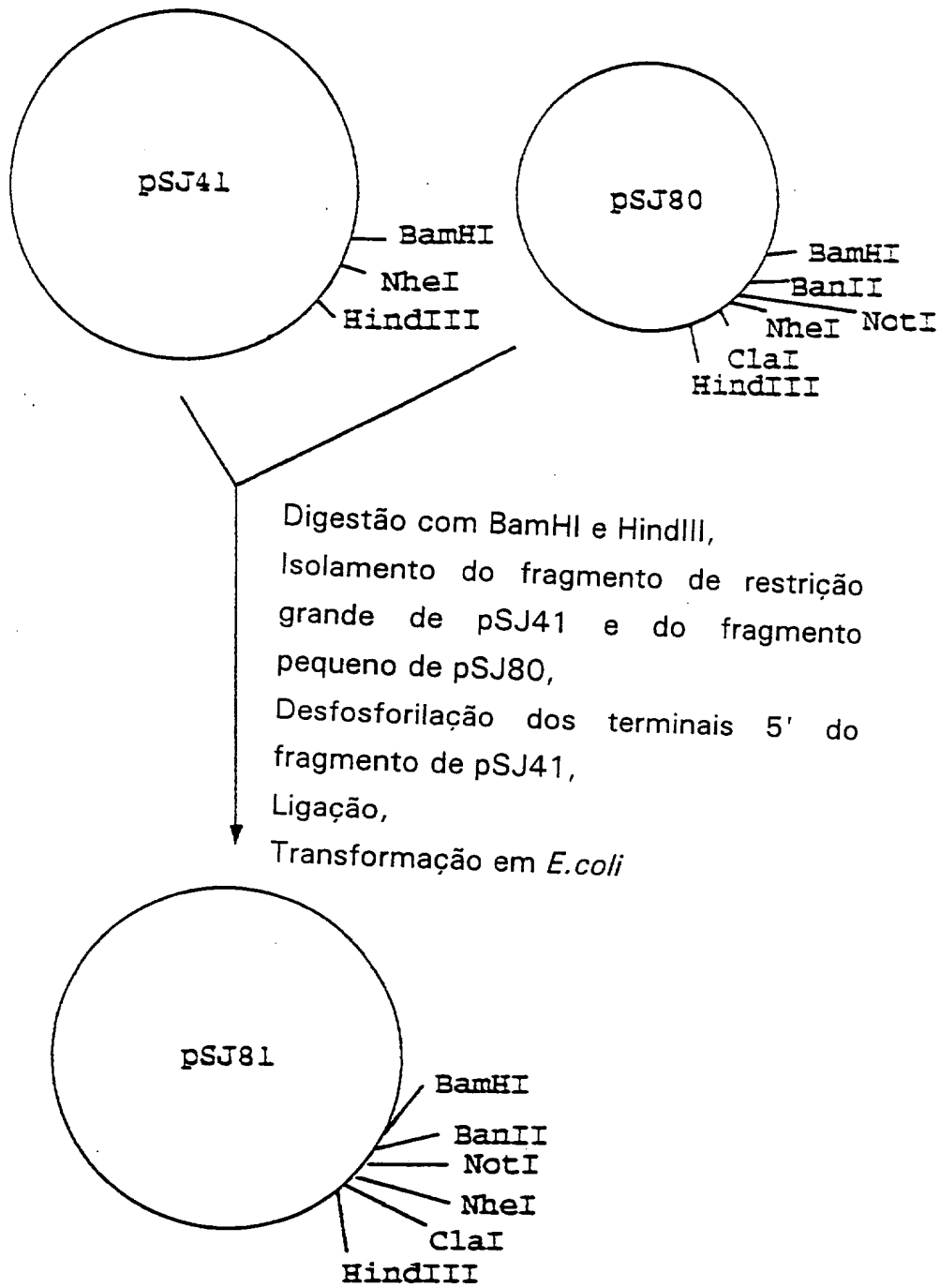


Figura 2j:

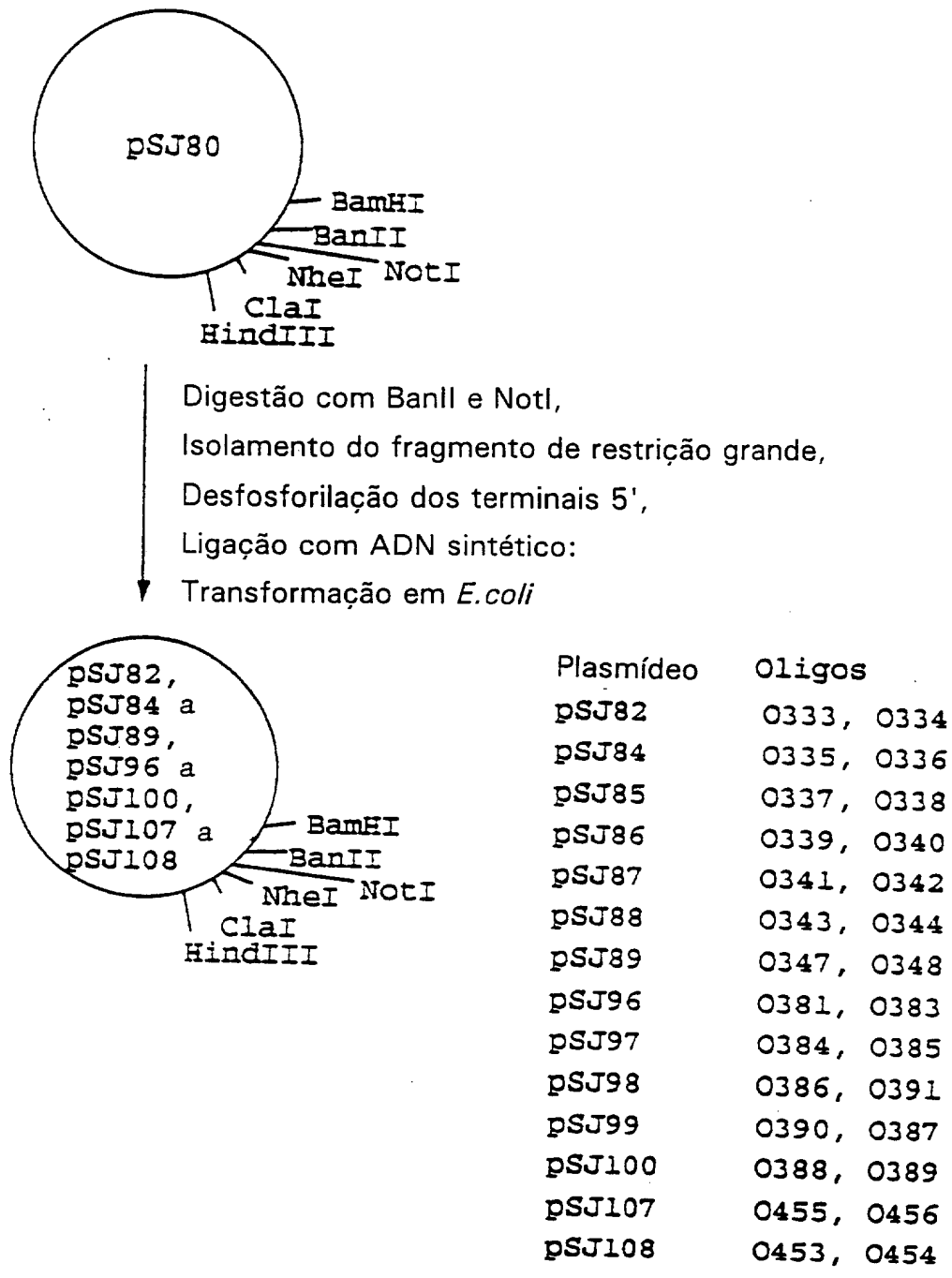


Figura 2k:

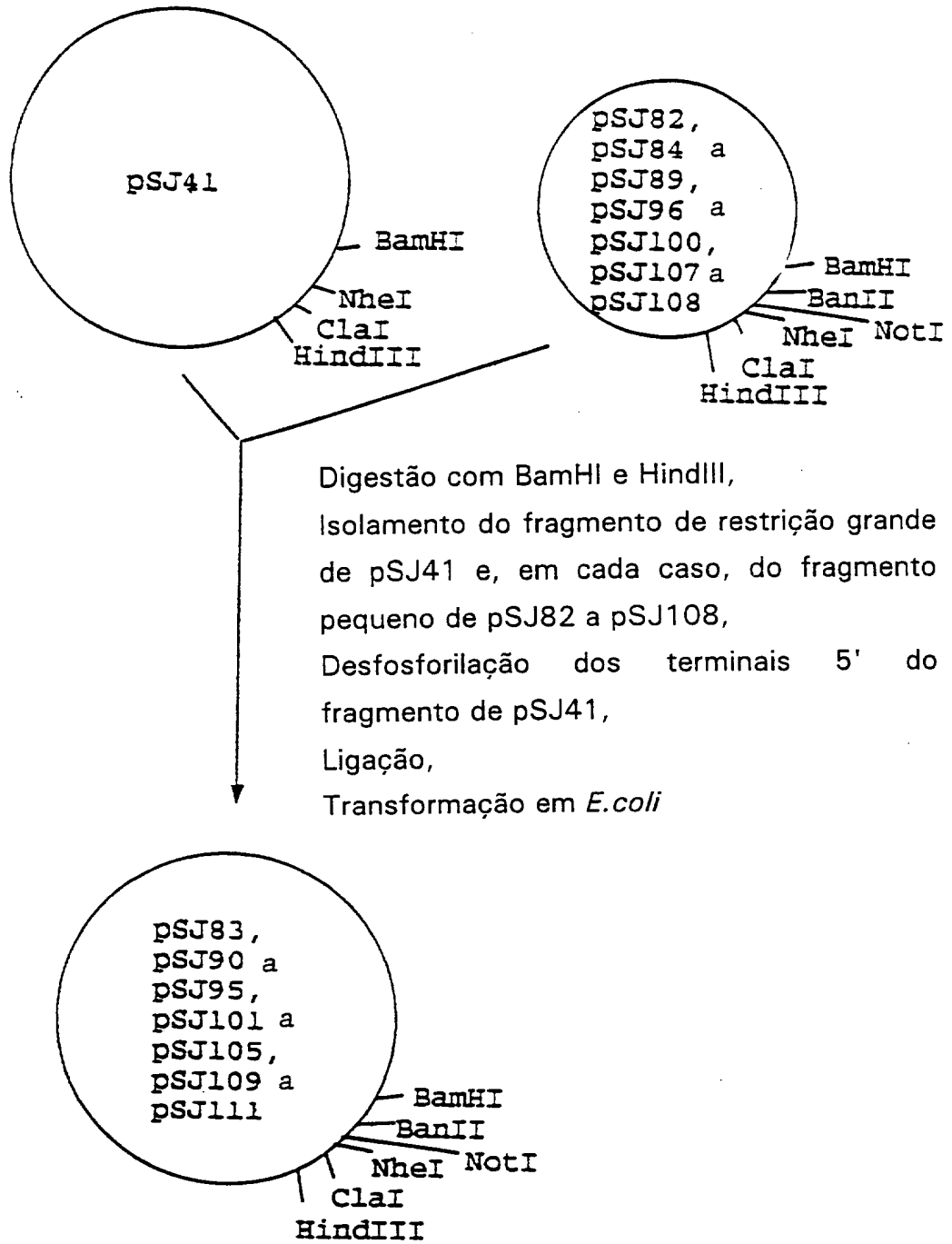


Figura 2I:

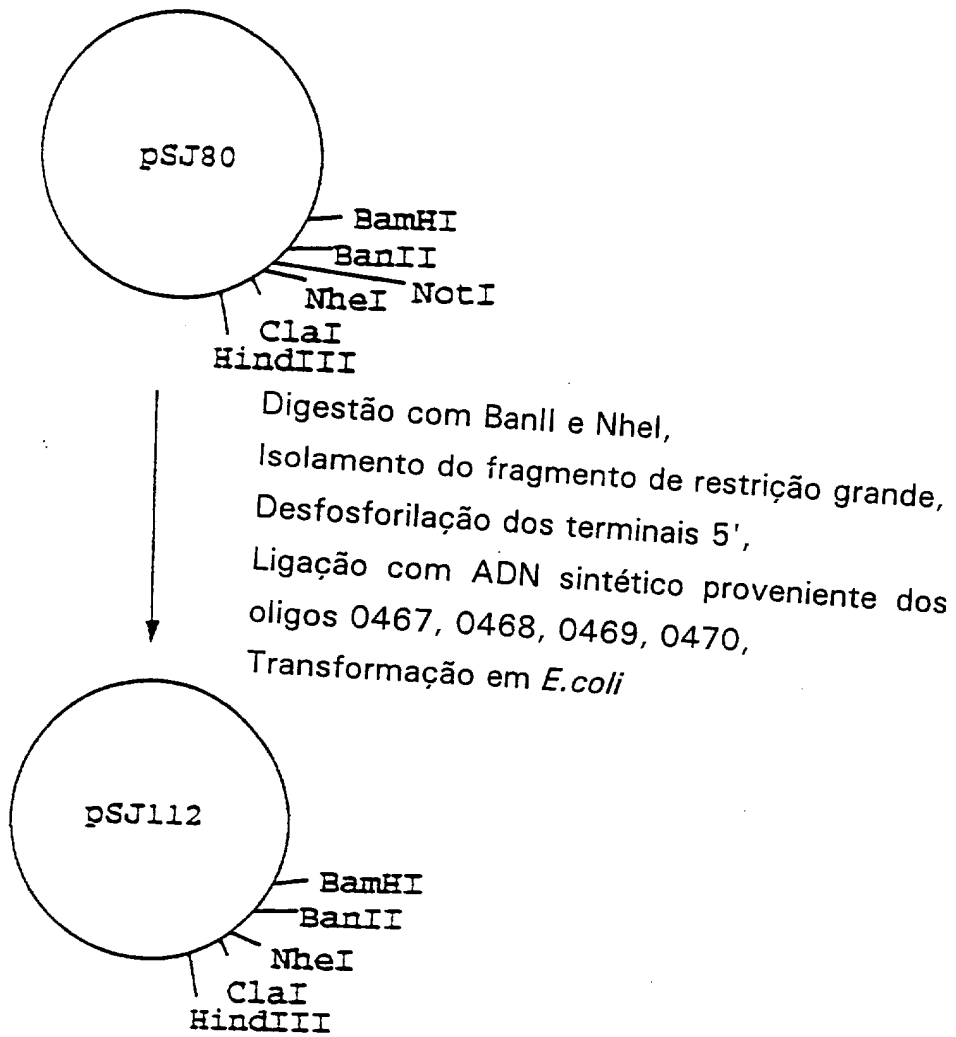


Figura 2m:

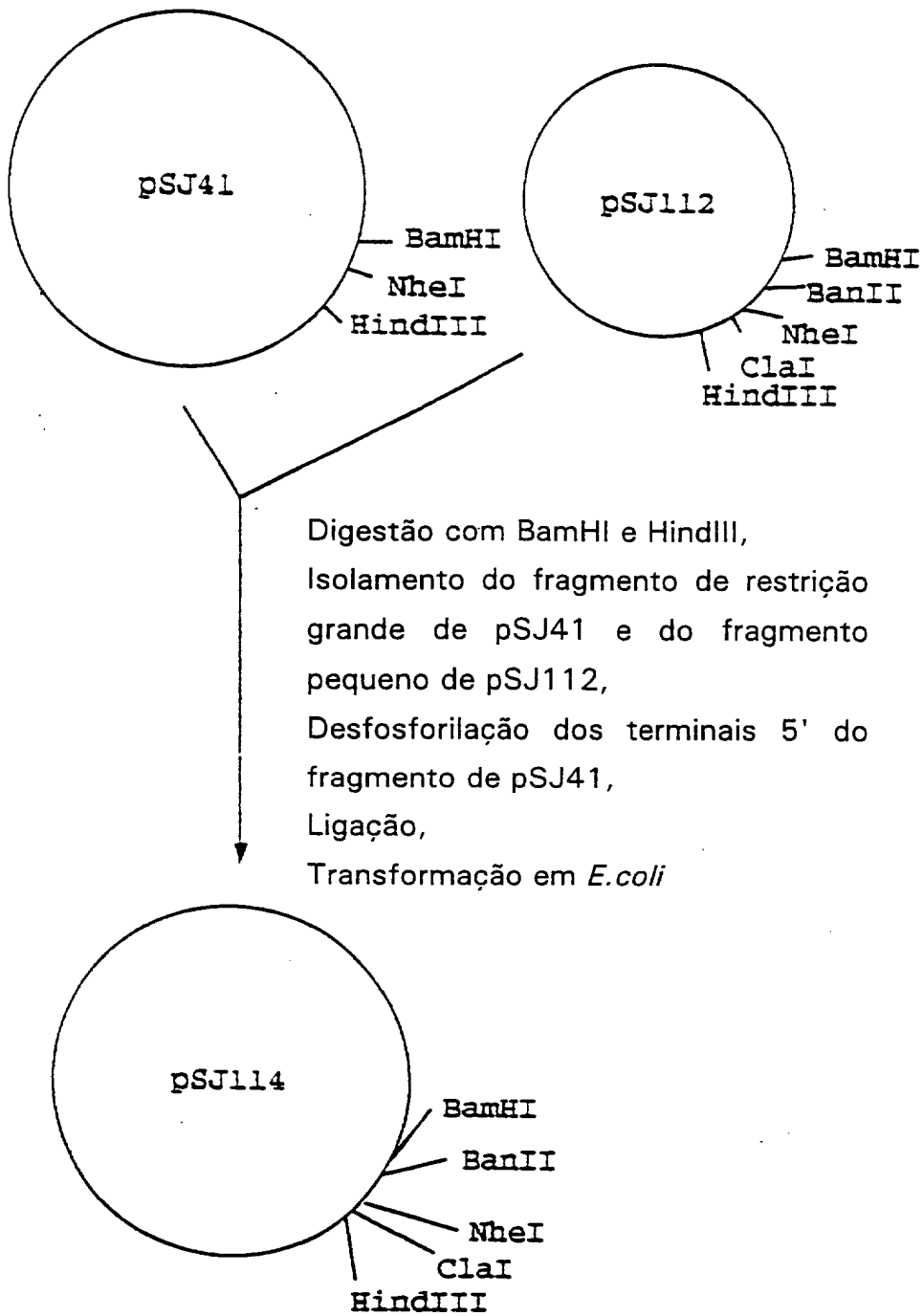


Figura 2n:

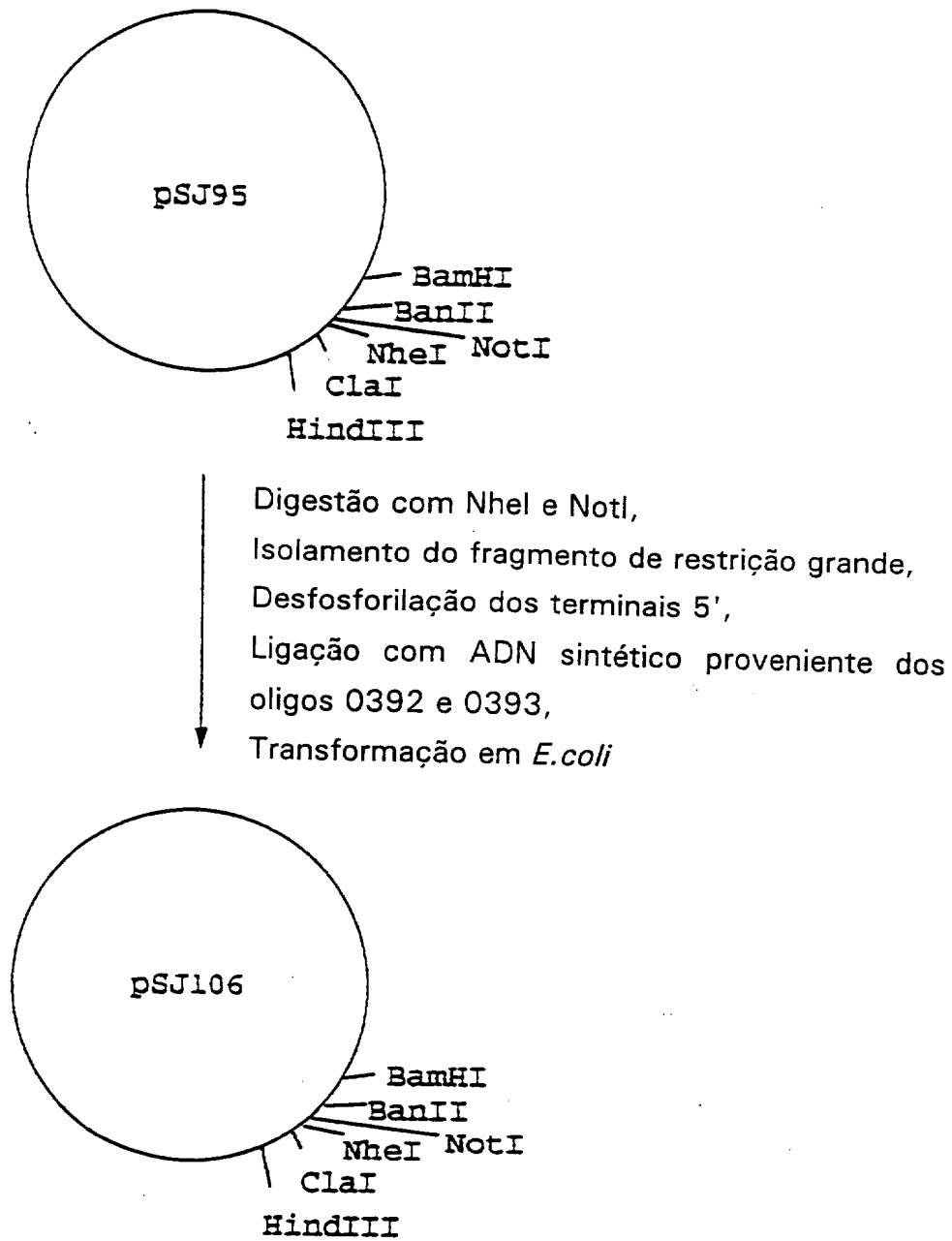


Figura 2o:

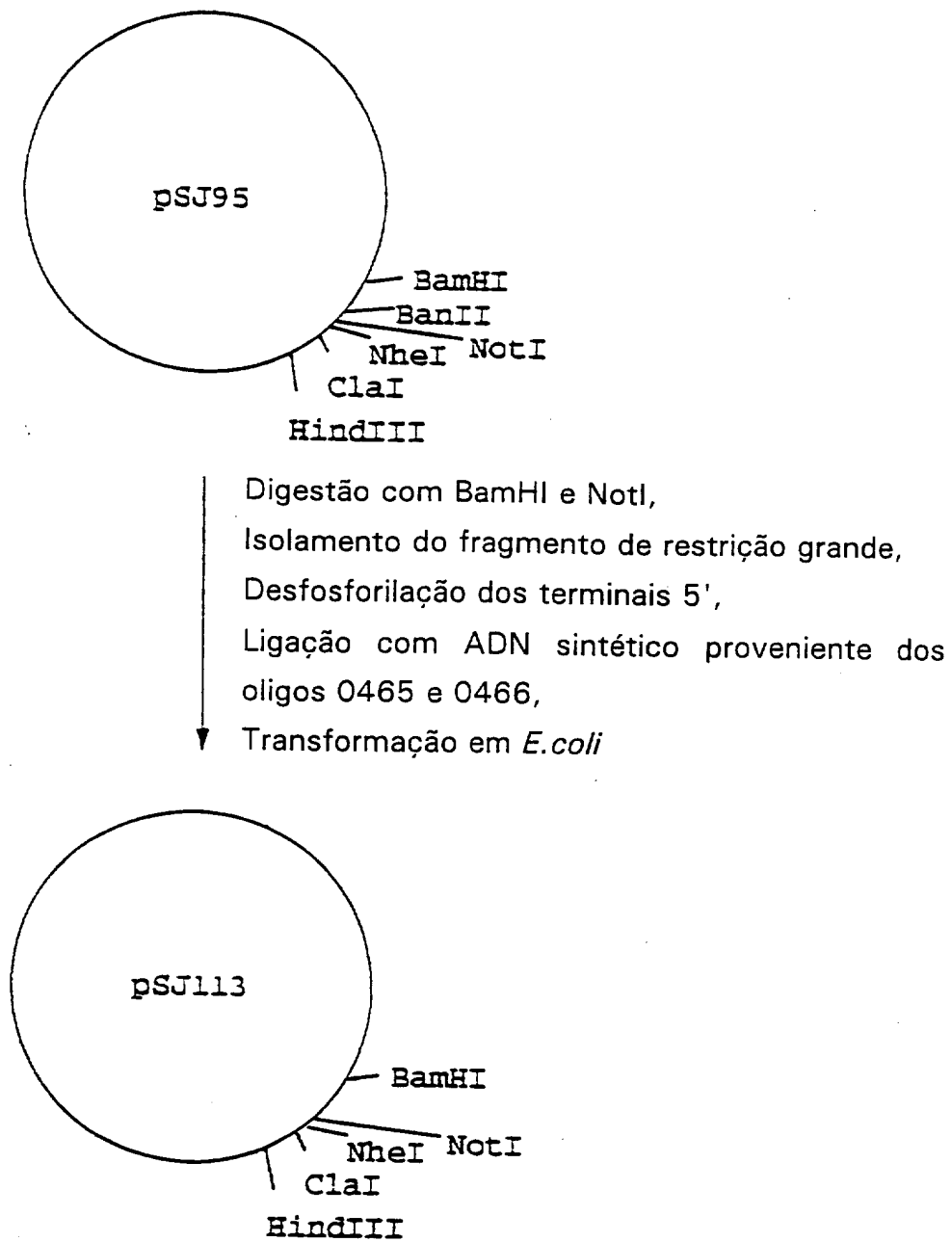
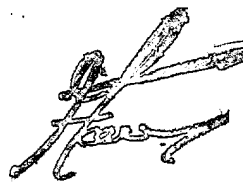


Figura 2p:

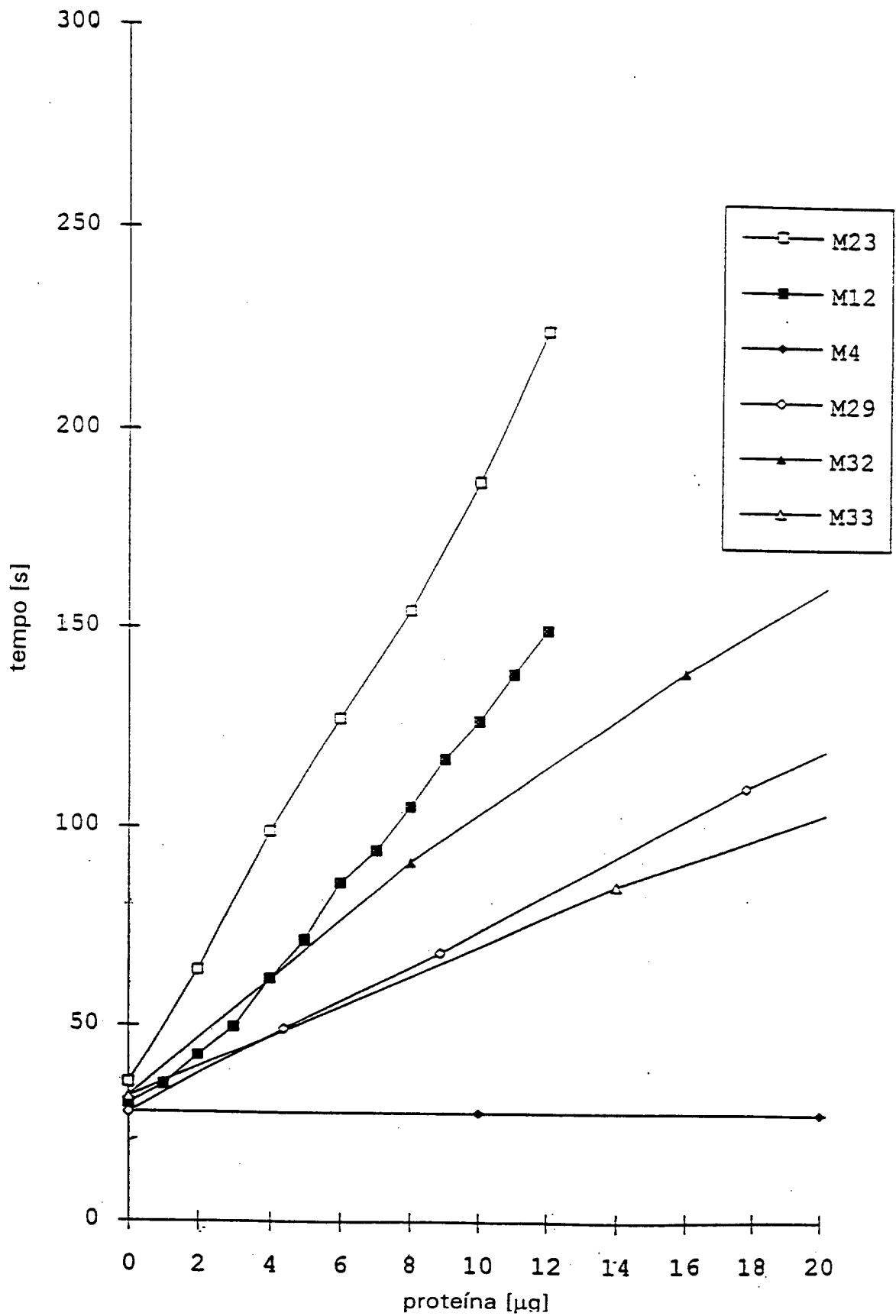
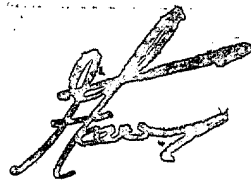
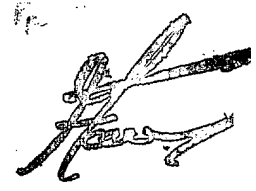
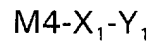


Figura 3: Aumento do tempo de trombina em função da concentração



REIVINDICAÇÕES

1. Variante de uroquinase bifuncional, de fórmula geral I



em que

M4 significa a sequência de aminoácidos ⁴⁷Ser a ⁴¹¹Leu da pro-uroquinase não glicosilada, de acordo com a figura 1 (SEQ ID NO: 21 e 22),

X₁ é uma ligação directa entre M4 e Y₁ ou representa um péptido de sequência

Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe

ou

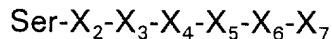
Ser-Pro-Pro-Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe

ou

Ser-Pro-Pro-Ser-Pro-Pro-Ser-Pro-Pro-Gly-Gly-Phe-Gly

ou

uma sequência peptídica de fórmula geral II



em que X₂ significa Pro ou Leu, X₃ significa Val ou Pro, X₄ significa Lys, Val, Arg, Gly ou Glu, X₅ significa Ala, Val, Gly, Leu ou Ile, X₆ significa Phe, Trp, Tyr ou Val, e X₇ significa Gly ou uma ligação directa entre X₆ e Y₁,

e

Y₁ representa um péptido de sequência

Y₂-Arg-Pro-Y₃-Gly-Gly-Gly-Gly-Asn-Gly-Asp-Phe-Glu-Glu-Ile-Pro-Glu-Glu-Tyr-Leu-Y₄

ou

Y_2 -Arg-Pro-Phe-Leu-Leu-Arg-Asn-Pro-Asn-Asp-Lys-Tyr-Glu-Pro-Phe-Trp-Glu-Asp-Glu-Glu-Lys-Asn-Glu

ou

Y_2 -Arg-Pro-Ser-Ser-Glu-Phe-Glu-Glu-Phe-Glu-Ile-Asp-Glu-Glu-Glu-Lys,

onde Y_2 significa Pro ou Val, Y_3 significa Leu ou uma ligação directa entre Pro e Gly, e Y_4 significa Gln ou um grupo hidroxilo.

2. Variante de uroquinase de acordo com a reivindicação 1, caracterizada por Y_1 significar um péptido de sequência

Y_2 -Arg-Pro- Y_3 -Gly-Gly-Gly-Gly-Asn-Gly-Asp-Phe-Glu-Glu-Ile-Pro-Glu-Glu-Tyr-Leu- Y_4 .

3. Variante de uroquinase de acordo com a reivindicação 1, caracterizada por Y_1 significar um péptido da sequência

Y_2 -Arg-Pro-Phe-Leu-Leu-Arg-Asn-Pro-Asn-Asp-Lys-Tyr-Glu-Pro-Phe-Trp-Glu-Asp-Glu-Glu-Lys-Asn-Glu.

4. Variante de uroquinase de acordo com a reivindicação 1 e/ou a reivindicação 2, caracterizada por X_1 representar uma sequência peptídica de fórmula geral II em que X_2 significa Pro ou Leu, X_3 significa Val, X_4 significa Lys, Val ou Arg, X_5 significa Ala, Val ou Gly, X_6 significa Phe, Trp, Tyr ou Val, e X_7 significa Gly ou uma ligação directa entre X_6 e Y_1 .

5. Variante de uroquinase de acordo com a reivindicação 4, caracterizada por X_4 significar Lys ou Val, X_5 significar Ala ou Val, X_6 significar Phe, Trp ou Tyr, e X_7 significar Gly ou uma ligação directa entre X_6 e Y_1 .

6. Variante de uroquinase de acordo com a reivindicação 4 e/ou a reivindicação 5, caracterizada por X_7 ser uma ligação directa entre X_6 e Y_1 .

7. Variante de uroquinase de acordo com a reivindicação 1 e/ou a reivindicação 3, caracterizada por X_1 representar uma sequência peptídica de



fórmula geral II em que X_2 significa Pro ou Leu, X_3 significa Val, X_4 significa Lys ou Val, X_5 significa Ala ou Val, X_6 significa Phe ou Trp, e X_7 significa uma ligação directa entre X_6 e Y_1 .

8. Plasmídeo para a utilização na obtenção de uma variante de uroquinase bifuncional de acordo com as reivindicações 1 a 7, caracterizado por o operão apresentar um promotor regulável, uma sequência de Shine-Dalgarno que actua como local de ligação de ribossomas, um códon de iniciação, um gene estrutural sintético para uma variante de uroquinase bifuncional de fórmula geral I de acordo com as reivindicações 1 a 7 e, a jusante do gene estrutural, 1 ou 2 terminadores, e por o plasmídeo ser apropriado para a expressão da variante de uroquinase bifuncional em estirpes de *Escherichia coli*.

9. Plasmídeo de acordo com a reivindicação 8, caracterizado por a distância entre a sequência de Shine-Dalgarno e o codão de iniciação ser de 6-12, de preferência 8-10 nucleótidos.

10. Plasmídeo de acordo com a reivindicação 8 e/ou a reivindicação 9, seleccionado de entre o grupo que consiste em pSJ 69, pSJ 76, pSJ 77, pSJ 78, pSJ 79, pSJ 81, pSJ 83, pSJ 90, pSJ 91, pSJ 92, pSJ 93, pSJ 94, pSJ 95, pSJ 101, pSJ 102, pSJ 103, pSJ 104, pSJ 105, pSJ 106, pSJ 109, pSJ 111, pSJ 114 e pSJ 113.

11. Plasmídeo de acordo com a reivindicação 10, seleccionado de entre o grupo que consiste em pSJ 76, pSJ 81, pSJ 83, pSJ 90, pSJ 91, pSJ 92, pSJ 93, pSJ 94, pSJ 95, pSJ 101, pSJ 102, pSJ 103, pSJ 105, pSJ 106, pSJ 109, pSJ 111 e pSJ 114.

12. Plasmídeo de acordo com a reivindicação 10 e/ou a reivindicação 11, seleccionado de entre o grupo que consiste em pSJ 76, pSJ 81, pSJ 83, pSJ 91, pSJ 92, pSJ 94, pSJ 95, pSJ 101, pSJ 102, pSJ 103, pSJ 106, pSJ 109, pSJ 111 e pSJ 114.

13. Plasmídeo de acordo com uma ou várias das reivindicações 10 a 12, seleccionado de entre o grupo que consiste em pSJ 76, pSJ 94, pSJ 95, pSJ 101, pSJ 102, pSJ 103, pSJ 106, pSJ 109, pSJ 111 e pSJ 114.

14. Processo para a produção de plasmídeos de acordo com as reivindicações 8 a 13, caracterizado por se obter os plasmídeos pBluescriptKSII+, pUC 8 e pGR 201, de acordo com as figuras 2 e 2a a 2p.

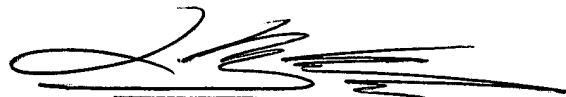
15. Utilização de um plasmídeo de acordo com as reivindicações 8 a 13 na obtenção de uma variante de uroquinase bifuncional de fórmula geral I de acordo com as reivindicações 1 a 7, caracterizada por se transformar com um plasmídeo, de maneira em princípio usual, uma estirpe de *Escherichia coli*, por se induzir a expressão do gene estrutural, por se separar a proteína precursora formada, da variante de uroquinase bifuncional de fórmula geral I, do meio e das células bacterianas lisadas, por se solubilizar a proteína precursora e por se redobrar esta a seguir, por acção de um sistema redox, no polipéptido de fórmula geral I.

16. Agente trombolítico contendo como substância activa uma variante de uroquinase bifuncional de fórmula geral I de acordo com as reivindicações 1 a 7.

17. Agente trombolítico trombolítico de acordo com a reivindicação 16, caracterizado por ser apropriado para aplicação em *bolus*.

Lisboa, 26. ABR. 2000

Por GRÜNENTHAL GmbH
- O AGENTE OFICIAL -



O ADJUNTO

ENG.º ANTÓNIO JOÃO DA CUNHA FERREIRA Ag. Of. Pr. Ind. Rua das Flores, 74 - 4.º 1200 LISBOA
--