

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

(22) Data de pedido: 2007.05.24	(73) Titular(es): CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
(30) Prioridade(s): 2006.05.24 EP 06290843	3, RUE MICHEL-ANGE 75794 PARIS FR
(43) Data de publicação do pedido: 2009.02.04	UNIVERSITÉ DE PROVENCE (AIX-MARSEILLE 1) FR
(45) Data e BPI da concessão: 2011.07.13 170/2011	(72) Inventor(es): CATHERINE RONIN FR GAËLLE GUIRAUDIE-CAPRAZ FR
	(74) Mandatário: ANTÓNIO JOÃO COIMBRA DA CUNHA FERREIRA RUA DAS FLORES, Nº 74, 4º AND 1249-235 LISBOA PT

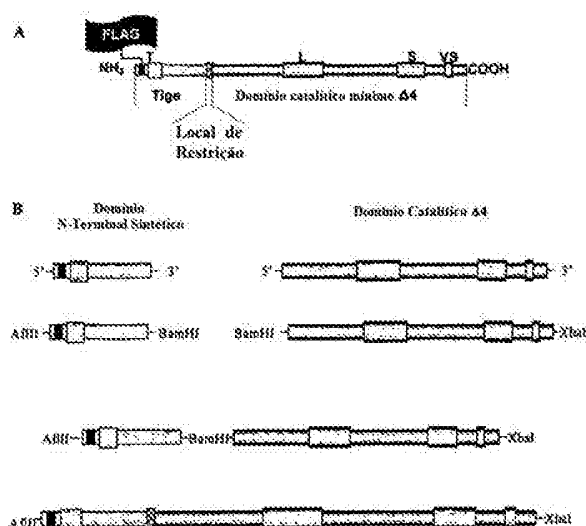
(54) Epígrafe: **PREPARAÇÃO E UTILIZAÇÕES DE SEQUÊNCIAS GÉNICAS CODIFICANDO GLICOSILTRANSFERASES QUIMÉRICAS COM ACTIVIDADE DE GLICOSILAÇÃO OPTIMIZADA**

(57) Resumo:

O PRESENTE INVENTO REFERE-SE À PRODUÇÃO DE SEQUÊNCIAS GÉNICAS CODIFICANDO GLICOSILTRANSFERASES MEMBRANARES QUIMÉRICAS APRESENTANDO UMA ACTIVIDADE DE GLICOSILAÇÃO OPTIMIZADA EM CÉLULAS TRANSFORMADAS COM AS REFERIDAS SEQUÊNCIAS, CORRESPONDENDO AS REFERIDAS SEQUÊNCIAS À FUSÃO DE UM PRIMEIRO ÁCIDO NUCLEICO CODIFICANDO PARA UM FRAGMENTO MÍNIMO C-TERMINAL DO DOMÍNIO CATALÍTICO (CD) DA GLICOSILTRANSFERASE INTEIRA NATIVA COM UM SEGUNDO ÁCIDO NUCLEICO CODIFICANDO PARA UM PÉPTIDO TRANSMEMBRANAR COMPREENDENDO NA SUA REGIÃO N-TERMINAL UMA REGIÃO DE CAUDA CITOPLASMÁTICA (CT) LOCALIZADA A MONTANTE DE UM DOMÍNIO TRANSMEMBRANAR (TMD), ELE PRÓPRIO LOCALIZADO A MONTANTE DE UMA REGIÃO CAULE (SR), DESDE QUE PELO MENOS UM DESTES PÉPTIDOS CT, TMD, SR SEJA DIFERENTE DA ESTRUTURA PRIMÁRIA DOS PÉPTIDOS DE OCORRÊNCIA NATURAL CORRESPONDENTES PRESENTES NA GLICOSILTRANSFERASE NATIVA DA QUAL É DERIVADO O FRAGMENTO CD COM ACTIVIDADE DE GLICOSILTRANSFERASE ÓPTIMA TAL COMO DEFINIDO ACIMA. O INVENTO REFERE-SE TAMBÉM À UTILIZAÇÃO DAS REFERIDAS SEQUÊNCIAS GÉNICAS NO QUADRO DA PREPARAÇÃO DE PROTEÍNAS RECOMBINANTES DE INTERESSE ATRAVÉS DE CÉLULAS TRANSFORMADAS COM AS REFERIDAS SEQUÊNCIAS E SEQUÊNCIAS CODIFICANDO AS REFERIDAS PROTEÍNAS RECOMBINANTES.

RESUMO

"Preparação e utilizações de sequências génicas codificando glicosiltransferases quiméricas com actividade de glicosilação otimizada"



O presente invento refere-se à produção de sequências génicas codificando glicosiltransferases membranares quiméricas apresentando uma actividade de glicosilação otimizada em células transformadas com as referidas sequências, correspondendo as referidas sequências à fusão de um primeiro ácido nucleico codificando para um fragmento mínimo C-terminal do domínio catalítico (CD) da glicosiltransferase inteira nativa com um segundo ácido nucleico codificando para um péptido transmembranar compreendendo na sua região N-terminal uma região de cauda citoplasmática (CT) localizada a montante de um domínio transmembranar (TMD), ele próprio localizado a montante de uma região caule (SR), desde que pelo menos um destes péptidos CT, TMD, SR seja diferente da estrutura primária dos péptidos de ocorrência natural correspondentes presentes na glicosiltransferase nativa da qual é derivado o fragmento CD com actividade de glicosiltransferase óptima tal como definido acima. O invento refere-se também à utilização das referidas sequências génicas no quadro da preparação de proteínas recombinantes de interesse através de células transformadas com as referidas sequências e sequências codificando as referidas proteínas recombinantes.

DESCRIÇÃO

"Preparação e utilizações de sequências génicas codificando glicosiltransferases quiméricas com actividade de glicosilação otimizada"

O presente invento refere-se à produção de sequências génicas codificando glicosiltransferases quiméricas apresentando actividade de glicosilação otimizada, e às suas utilizações no quadro da preparação de proteínas recombinantes de interesse através de células transformadas com as referidas sequências e sequências codificando as referidas proteínas recombinantes.

Glicosilação de proteínas e lípidos

As proteínas e os lípidos são componentes principais das membranas celulares. Os hidratos de carbono associados à membrana estão exclusivamente na forma de oligossacáridos covalentemente ligados a proteínas formando glicoproteínas e a lípidos formando glicolípidos. Os glicoconjugados (incluindo glicolípidos e glicoproteínas) são muito frequentemente moléculas chave da superfície celular que se considera estarem envolvidas em interacções célula-célula e na adesão celular (Feizi, 1993; Crocker & Feizi, 1996).

Os monossacáridos predominantes verificados em glicoproteínas eucarióticas são a glicose (Glc), galactose (Gal), manose (Man), fucose (Fuc), N-Acetilgalactosamina (GalNAc), N-acetilglucosamina (GlcNAc) e ácido siálico (Sia) muito frequentemente como ácido neuramínico (NeuAc) que pode ser N-acetilado ou N-glicolilado em mamíferos mas apenas N-acetilado em humanos. As cadeias de hidratos de carbono também designadas como glicanos são ligadas à estrutura polipeptídica através de ligações O-glicosídicas ou N-glicosídicas. A ligação N-glicosídica é através do grupo amida da asparagina. A ligação O-glicosídica é ao hidroxilo da serina, treonina ou hidroxilisina. Em glicoproteínas ligadas em O, do tipo ser ou thr, o monossacárido ligado directamente à proteína é frequentemente GalNAc enquanto em glicoproteínas ligadas em N, apenas se verifica GlcNAc.

A ligação N-glicosídica é conservada ao longo do reino eucariótico incluindo leveduras, plantas, insectos, mamíferos e humanos. O local da glicosilação ligada em N ocorre dentro de uma sequência de consenso de aminoácidos, Asn-X-Ser/Thr (N-X-S(T)), onde X é qualquer aminoácido excepto prolina. Quando é efectuada uma análise proteica nas bases de dados públicas, pode-se mostrar que aproximadamente 65% de todas as proteínas contêm pelo menos uma tal sequência de consenso. As glicoproteínas ligadas em N contêm todas um pentassacárido comum (invariante) ligado ao polipéptido. Este núcleo consiste em três resíduos Man e dois GlcNAc. Antenas de sequência variável completam o glicano e permitem a classificação em três subclasses principais ligadas em N:

1. Os glicanos ricos em manose contêm apenas manose como açúcar terminal.
2. Glicanos híbridos contêm pelo menos uma antena de GlcNAc-Gal.
3. Os glicanos complexos contêm 2 a 5 das antenas terminadas em GlcNAc, Gal ou ácido siálico.

Em todas as células eucarióticas, as glicoproteínas ligadas em N são sintetizadas em co-tradução a partir de polirribossomas ligados ao retículo endoplasmático (RE). O processamento dos glicanos ligados em N ocorre inicialmente no lúmen do RE e continua no aparelho de Golgi e na rede transGolgi onde as glicoproteínas alcançam o seu polimorfismo final e funcional. A ligação dos glicanos ligados em O ocorre após a tradução no aparelho de Golgi onde também ocorre a glicosilação dos lípidos. Os açúcares utilizados para a N- e O-glicosilação são activados através de ligação a nucleótidos específicos no citoplasma e são importados para dentro do lúmen dos organelos através de transportadores específicos.

A glicosilação é o mais sofisticado conjunto de modificações pós-tradução que podem ocorrer nas proteínas. Elas têm como objectivo: i) controlar da actividade biológica das proteínas, ii) proteínas sinal para ligação de receptores semelhantes a lectina e/ou sistemas de degradação, iii) dirigir proteínas para a superfície celular (secreção), iv) direccionar proteínas para compartimentos celulares, v) definir uma identidade imunológica (grupos sanguíneos). Como resultado, as proteínas glicosiladas existem como uma mistura

de glicoformas cujas propriedades físico-químicas e biológicas diferem do produto directamente codificado pelo gene relevante. É amplamente admitido que as modificações pós-tradução das proteínas permitem uma melhor adaptação da proteína à sua função biológica (Helenius & Aeby, 2001).

Muitas glicoproteínas humanas são de elevada relevância clínica. Por exemplo, em superfícies celulares, são importantes para comunicação entre células, para manutenção da estrutura celular e para auto-reconhecimento pelo sistema imunitário.

As glicoproteínas são muito abundantes em formas solúveis em fluidos biológicos tais como leite e sangue. Neste caso, os glicanos protegem as proteínas contra proteases, aumentam a sua solubilidade, governam a sua eliminação plasmática e dirigem-nas para órgãos alvo.

Enzimas de Glicosilação

As reacções de glicosilação são de grande importância biológica tanto para procariotas como eucariotas e requerem a acção coordenada de um grande número de enzimas designadas como Glicosiltransferases (GT) (Breton & Imberty, 1999).

As glicosiltransferases (GT) constituem uma grande família de enzimas envolvidas na biossíntese de polissacáridos e glicoconjugados nos reinos procariótico e eucariótico, respectivamente. As sequências das enzimas procarióticas são não homólogas às glicosiltransferases de mamífero enquanto as enzimas de humanos, mamíferos e drosófila partilham frequentemente uma homologia significativa. Até agora sabe-se que mais de mil GT medeiam uma ampla gama de funções biológicas. Desenvolvimentos na biologia molecular destas enzimas revelaram uma diversidade inesperada sugerindo que o processo de glicosilação provavelmente requer o envolvimento de cerca de 250 genes numa única célula viva humana (Breton *et al.*, 2001). Até à data foram clonadas cerca de 500 glicosiltransferases incluindo 160 enzimas humanas. O número de enzimas clonadas está a crescer em humanos e pode atingir 200 em breve (Narimatsu, 2004). As GT são classificadas de acordo com a

estereoquímica dos substratos e produtos da reacção como enzimas de retenção ou de inversão (Sinnott, 1990). Foi estabelecida uma classificação das glicosiltransferases utilizando difosfo-açúcar nucleotídico, monofosfo-açúcares nucleotídicos e fosfatos de açúcar e proteínas aparentadas em famílias distintas baseadas nas sequências. Mostra que as GT humanas se distribuem até agora em 42 famílias estruturais (<http://afmb.cnrs-mrs.fr/CAZY/Homo-sapiens.html>).

As glicosiltransferases catalisam a transferência de resíduos de açúcar de um açúcar de um nucleótido (substrato dador activado) para um aceitador (lípidio, proteína ou cadeia de hidrato de carbono em crescimento). Além da classificação oficial das enzimas, estas podem ser agrupadas em famílias funcionais com base nas semelhanças das suas sequências, que podem reflectir características enzimáticas tais como especificidade do dador, especificidade do aceitador, e especificidade da ligação entre dador e aceitador. Com base no açúcar que transferem, as GT são nomeadas de acordo com o açúcar que transferem tal como N-Acetilglucosaminiltransferase (GlcNAcT), galactosiltransferases (GalT) e fucosiltransferases (FucT) ou sialiltransferases (SiaT ou ST). Devido à acumulação de dados de genes de glicosiltransferases, pensa-se que as glicosiltransferases têm elevada especificidade para o tipo de ligação nos substratos dadores e aceptadores bem como para a natureza do glicoconjugado aceitador (glicoproteínas ou lípidos ligados em N/O).

Estrutura das Glicosiltransferases:

Todas as GT clonadas até agora em vertebrados apresentam a mesma topologia. São proteínas membranares de tipo II (domínio citoplasmático N-terminal) compostas por quatro domínios principais: uma cauda citoplasmática curta (**CT**) na extremidade N-terminal, uma região âncora membranares (**TMD**) de 10 a 20 aminoácidos, uma região de caule (**SR**) e um grande domínio catalítico (**CD**) C-terminal (**Figura 1**) (Paulson et al., 1987). A região âncora actua como péptido sinal não clivável e também como domínio transmembranares (Wickner & Lodisch, 1985), orientando o domínio catalítico das GT luminal do aparelho de Golgi. A SR é amplamente considerada

uma região flexível permitindo também a orientação do CD luminal do aparelho de Golgi. De um modo geral, o CD é razoavelmente conservado entre as GT de várias espécies enquanto a SR constitui uma porção hipervariável da transferase. Algumas destas enzimas podem ser clivadas na sua SR através de uma protease endógena, ou proteases, e segregada para fora da célula (Paulson & Colley, 1989) para produzir enzimas do leite ou do sangue. Foi bem documentado que a clivagem proteolítica e a secreção das glicosiltransferases são afectadas por várias condições patológicas tais como transformação maligna e inflamação mas os mecanismos moleculares subjacentes à clivagem e secreção não foram ainda clarificados.

As GT e proteínas/lípidos neoglicosilados foram localizados dentro do RE e nos subcompartimentos do Golgi utilizando tanto fraccionamento subcelular de membranas celulares como microscopia imunoeléctronica (Roth, 1987). Estudos iniciais sugerem que a compartimentação das GT podem reflectir a sequência da modificação da cadeia do oligossacárido (Kornfeld & Kornfeld, 1985). Pensou-se que esta localização estrita assegurava uma biossíntese óptima das cadeias de glicanos proporcionando uma vizinhança eficiente entre a enzima, o substrato e o dador do açúcar nucleotídico. Outros estudos mostraram que muitas enzimas de glicosilação se sobrepõem na localização e demonstraram subcompartimentação no Golgi específica do tipo celular (Colley, 1997). As GT estão espalhadas nas cisternas do Golgi e isto pode diferir entre os tipos celulares para uma dada proteína (Roth, 1991).

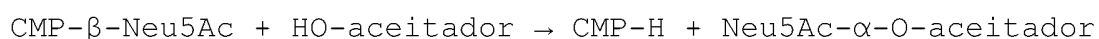
Em geral, foi demonstrado que o domínio transmembranar (TMD) das proteínas é um determinante chave para conferir a localização no Golgi essencial para pelo menos GalNAcT, GlcNAcT, GalT, FucT e SiaT encontrarem os seus aceitadores. Foi demonstrado que a região flanqueadora de TMD e/ou luminal também contribuem para localizar (Munro, 1998; Yang *et al.*, 1996). Além disso, a CT e a SR das GT especifica a sua sublocalização funcional *in vivo* e estabilidade no aparelho de Golgi. A substituição de um dos três domínios (CT, TMD e SR) não muda a actividade catalítica da enzima mas contribui para alterar a sua distribuição nos compartimentos do Golgi

(Grabenhorst & Conradt, 1999). Não foram verificadas claras sequências de direcçãoamento específicas na última década e apenas foram identificadas regiões críticas das GT para a sua compartimentação (Opat *et al.*, 2001). Mais recentemente, os inventores verificaram que o domínio catalítico solúvel de ST6Gal seguia uma via subcelular diferente da da enzima inteira (Ronin, Biochimie 2003).

Entre o grande número de GT, as famílias que são de maior interesse são as que estão encarregues da glicosilação terminal porque estes açúcares têm um papel importante em fenómenos de reconhecimento e sinalização em humanos. Esses incluem essencialmente sialiltransferases (SiaT), fucosiltransferases (FucT) e galactosiltransferases (GalT). A galactose e fucose estão envolvidas no reconhecimento de antígenos do grupo sanguíneo (ABH/Lewis). Os oligossacáridos sialilados das glicoproteínas e dos glicolípidos estão implicados em muitos processos biológicos tais como adesão celular e reconhecimento de receptores em inflamação e cancro (antígeno sialilLewis) bem como crescimento neuronal (antígeno Polissialico) (Paulson, 1989). A diversidade estrutural e a expressão regulada dos sialilglicoconjugados parecem estar bem correlacionadas com as suas funções (Sasaki, 1996).

Sialiltransferases:

A família do ácido siálico é composta por mais de uma centena de derivados entre os quais o ácido neuramínico é o mais frequente em mamíferos e humanos. As sialiltransferases (ST) catalisam a transferência de um resíduo de ácido siálico da sua forma activada, ácido citidil-monofosfo-N-acetil-neuramínico (CMP-Neu5Ac), para uma posição terminal não redutora num glicano aceitador em glicoproteínas ou glicolípidos. A reacção catalisada é como se segue:



Cada ST é classificada de acordo com o tipo de ligação estabelecida entre o resíduo de ácido siálico e a especificidade do substrato aceitador (que pode ser uma proteína ou um lípido). Assim, três grupos são distinguidos:

α 2,3-sialiltransferases (α 2,3-ST), α 2,6-sialiltransferases (α 2,6-ST), e α 2,8-sialiltransferases (α 2,8-ST). α 2,6-ST transferem um resíduo de ácido siálico numa posição alfa2,6 (ST6Gal) para um resíduo de galactose ou para uma N-Acetil-D-galactosamina (ST6GalNAc) ou para uma N-acetil-D-glucosamina (ST6GlcNAc). No entanto, a enzima envolvida na formação deste último tipo de ligação é ainda desconhecido. As α 2,3-ST transferem um ácido siálico para o carbono 3 da galactose (ST3Gal), e as α 2,8-ST do carbono 8 de um outro resíduo de ácido siálico (ST8Sia).

ST foram também identificadas em todos os animais (de aves a humanos), em células bacterianas e em vírus (Sujino et al., 2000). A família de ST humanas é constituída por 20 genes diferentes clonados tal como mostrado na Tabela 1 (Sasaki, 1996, Ronin 2003). As suas especificidades de substrato são como se segue.

1) α 2,6-ST

ST6Gal:

ST6Gal I transfere um resíduo de ácido siálico no dissacárido Gal β 1-4GlcNAc de quase N-glicanos (van den Eijnden et al., 1980; Weinstein et al., 1982). É expressa de um modo ubíquo em tecidos humanos excepto nos testículos e no cérebro onde é expressa a menores níveis (Tabela 2) (Kitagawa & Paulson, 1994a).

ST6Gal II, identificada recentemente (Krzewinski-Recchi et al., 2003; Takashima et al., 2002), reconhece o dissacárido Gal β 1-4GlcNAc como substrato aceitador particularmente quando está num oligossacárido livre (proteína ou lípido desconhecido). O padrão de expressão de ST6Gal II está restrito ao cérebro e mostra um baixo nível de expressão em testículos, tiróide, gânglios linfático e alguns tecidos fetais (Tabela 2).

ST6GalNAc:

A segunda subfamília de α 2,6-ST é representada pelo grupo de ST6GalNAc, que transfere um resíduo de ácido siálico

num resíduo de N-acetilgalactosamina. Foram identificados seis membros em humanos (Tabela 2).

ST6GalNAc I e ST6GalNAc II possuem a especificidade de substrato mais ampla uma vez que são capazes de transferir um ácido siálico nas seguintes estruturas de O-glicanos: Gal β 1-3GalNAc, GalNAc α -O-Ser/Thr, Sia α 2-3 Gal β 1-3GalNAc α -O-Ser/Thr (Ikehara *et al.*, 1999; Kono *et al.*, 2000; Kurosawa *et al.*, 1994; Kurosawa *et al.*, 1996; Harduin-Lepers *et al.*, 2001). Os seu padrão de expressão é diferente, ST6GalNAc I é expressa nas glândulas submaxilares e mamárias, no baço e cólon enquanto ST6GalNAc II é expressa em muito tecidos tais como testículos e glândulas mamárias lactantes (Kurosawa *et al.*, 1996; Kurosawa *et al.*, 2000).

ST6GalNAc III, IV, V e VI possuem uma reduzida especificidade de substrato: reconhecem apenas o trissacárido Sia α 2,3Gal β 1-3GalNAc. No entanto, está agora estabelecido que ST6GalNAc III, V e VI catalisam preferencialmente a formação do glicolípido GM1b (Sjoberg *et al.*, 1996; Lee *et al.*, 1999) enquanto ST6GalNAc IV catalisa a transferência do ácido siálico dos O-glicanos (Harduin-Lepers *et al.*, 2000, Ikehara *et al.*, 1999; Lee *et al.*, 1999, Okajima *et al.*, 2000). ST6GalNAc IV mostra uma especificidade de substrato restrita utilizando apenas a sequência do trissacárido Sia α 2,3Gal β 1-3GalNAc e não discrimina entre GalNAc ligado em α e β . Verificaram-se duas isoformas diferentes. A análise *Northern blot* detectou um transcrito de 2,2 kb em vários tecidos adultos e níveis menores de expressão de um transcrito adicional no cérebro, coração e músculo-esquelético (Harduin-Lepers *et al.*, 2000).

2) α 2,3-ST

Elas são todas ST3Gal transferindo o ácido siálico de um resíduo de galactose na proteína ou no lípido aceitador.

ST3Gal I foi clonada a partir de uma biblioteca de ADNc de glândulas submaxilares e a sua expressão ubíqua (Tabela 2) foi confirmada através de *Northern blot* (Chang *et al.*, 1995). Sintetiza Sia α 2,3Gal β 1-3-GalNAc, uma estrutura comum a muitos oligossacáridos ligados em O. Difere de outras ST na sua

capacidade para utilizar substratos aceitadores de glicolípidos *in vitro* (Kitagawa & Paulson, 1994b).

ST3Gal II foi clonada por Kim *et al.* (1996) a partir de uma biblioteca de ADNc de fígado e a análise de *Northern blot* revelou elevados níveis de expressão no coração, fígado e músculo-esquelético, níveis intermédios no timo, nódulo linfático, apêndice e baço, e níveis inferiores em pulmão e linfócitos do sangue periférico (Tabela 2). ST3Gal II pode transferir resíduos de ácido siálico em Gal β 1-3-GalNAc ou nos gangliósidos (lípidos) GM1 ou asialo-GM1 como substratos aceitadores (Giordanengo *et al.*, 1997).

ST3GalIII foi clonada através de pesquisa de uma biblioteca de ADNc de placenta humana com uma sonda baseada no motivo sialil, o ADNc codificando ST3Gal III foi isolado (Kitagawa & Paulson, 1993). O transcrito é abundantemente expresso em músculo-esquelético e tecido fetal e menos expresso em placenta. Esta enzima catalisa a transferência de ácido siálico para substratos contendo galactose (Tabela 2).

ST3Gal IV está envolvida em sialilação de Gal β 1-3GalNAc ligado em O (Tabela 2) (Tetteroo *et al.*, 1987). Existem 5 ARNm diferentes expressos, e codificam para sequências proteicas idênticas excepto nas extremidades 5'. Estes transcritos são produzidos através de uma combinação de processamento alternativo. A análise de *Northern blot* mostrou que um deles é especificamente expresso em placenta, testículos e ovário, indicando que a sua expressão é regulada independentemente dos outros (Kitagawa *et al.*, 1996).

ST3Gal V foi isolada a partir de uma biblioteca de ADNc de células humanas e designada GM3-sintetase. Um transcrito principal de 2,4 kb é expresso em muitos tecidos particularmente em cérebro, músculo-esquelético, placenta e testículos. Está amplamente distribuído no cérebro humano com uma expressão ligeiramente elevada no córtex cerebral, lobo temporal e putámen. A especificidade de substrato desta enzima é altamente restrita a lacossilceramida como aceitador (Tabela 2) (Ishii *et al.*, 1998).

ST3Gal VI foi clonada a partir de uma biblioteca de ADNc de melanoma humano. Esta ST exibe especificidade de substrato restrita: está envolvida na síntese de sialil-paraglobósido, um precursor do determinante sialil-Lewis X (Okajima *et al.*, 1999). Existem 2 formas de ARNm de ST3Gal VI (designadas tipo 1 e 2), diferindo apenas na região 5' não traduzida. Esta enzima é expressa a níveis semelhantes na maioria dos tecidos (Tabela 2) (Taniguchi *et al.*, 2001).

3) α 2,8-ST

ST8Sia I é também designada gangliósido GD3-sintetase e catalisa a transferência de uma molécula de ácido siálico para o ácido siálico terminal de GM3 através de uma ligação α 2,8 (Sasaki *et al.*, 1994). Mostrou-se que esta enzima também pode utilizar GM1b, GD1a e GT1b como substrato aceitador (Tabela 2) (Nakayama *et al.*, 1996; Nara *et al.*, 1996; Watanabe *et al.*, 1996).

ST8Sia II é também designada STX. Scheidegger *et al.* (1995) utilizaram sequências de STX de roedor para clonar o ADNc humano a partir de uma biblioteca de coração fetal. A STX é primariamente expressa em tecidos embrionários e modestamente no coração, cérebro e timo de adulto (Angata *et al.*, 1997). Esta enzima regula a ligação entre a molécula de adesão celular neural (NCAM) e o ácido polissialico (PSA) e modula deste modo as propriedades adesivas de NCAM. STX catalisa a transferência do primeiro ácido siálico através de uma ligação α 2,8 noutro ácido siálico ligado em α 2,3 ou α 2,6 (Tabela 2) (Angata *et al.*, 1997) num N-Glicano, depois está envolvida no processo de alongamento fazendo ligações α 2,8 sucessivas ao longo dos resíduos anteriores.

ST8Sia III transfere um ácido siálico em estruturas Sia α 2-3Gal β 1-4GlcNAc dentro de N-glicanos e glicolípidos tais como GM3 (Tabela 2) (Lee *et al.*, 1998; Yoshida *et al.*, 1995a; Yoshida *et al.*, 1995b).

ST8Sia IV, também designada PST (Polissialiltransferase), foi clonada por Nakayama *et al.* (1995). A análise *Northern blot* revelou que PST é expressa em muitos tecidos fetais e no coração, baço, timo de adulto e a

níveis menores noutros órgãos e tecidos (Tabela 2). Esta enzima também regula a ligação entre a molécula de adesão celular neural (NCAM) e o ácido polissialílico (PSA). Além disso, mostra-se que catalisa a mesma reacção que STX (Angata *et al.*, 2000; Nakayam *et al.*, 1995).

ST8Sia V apresenta uma actividade de transferência para os gangliósidos GM1b, GD1a, GT1b e GD3 (Tabela 2) (Kono *et al.*, 1996).

ST8Sia VI foi clonada muito recentemente em humano e pouco se sabe sobre ela. A ST8Sia VI de ratinho possui uma actividade de transferência de ácido siálico na estrutura NeuAc α 2,3(6)Gal β verificada na extremidade não reduzida dos O-glicanos, N-glicanos e oligossacáridos livres tais como a sialil-lactose (Tabela 2) (Takashima *et al.*, 2002).

4) Organização estrutural necessária para promoção da actividade de sialiltransferase

Todos os membros da família das ST possuem três regiões conservadas no seu CD designadas motivos Sialil: motivo L para *Large* (grande), S para *Small* (pequeno) e VS para *Very Small* (muito pequeno) (Datta & Paulson, 1995; Geremia *et al.*, 1997). Com base na comparação das suas sequências primárias foi recentemente publicada uma pesquisa dos genomas animais (Harduin-Lepers, 2005) para encontrar sequências desconhecidas possuindo os motivos sialil e assim potencialmente novas ST.

O motivo sialil L consistia em 44 ou 45 aminoácidos e continha entre 5 resíduos invariantes entre todas as enzimas humanas tal como mostrado na Figura 2. Esta região é o centro do CD. O segundo motivo sialil S, na porção COOH-terminal, consistia em 23 resíduos de aminoácidos, sendo dois dos quais idênticos entre todas as ST (Drickamer, 1993). O terceiro motivo sialil VS, na parte terminal da enzima, consistia em 13 aminoácidos com dois resíduos conservados (histidina e glutamato).

Foi demonstrado que o motivo sialil L está envolvido na ligação do substrato dador CMP-ácido siálico (Datta &

paulson, 1995). Alguns aminoácidos deste motivo podem também participar na actividade catalítica das enzimas (Sasaki, 1996). Foi proposto que o motivo sialil S participa na ligação tanto do dador como do aceitador (Datta *et al.*, 1998). O papel exacto de VS é ainda incerto mas estudos recentes sobre STX e PST examinaram o papel funcional da His conservada neste motivo (Kitazume-Kawaguchi *et al.*, 2001). A mudança da Histidina por um resíduo de Lisina afecta a sua actividade catalítica mostrando que o motivo VS está envolvido em catálise. O motivo sialil VS é necessário para uma eficiência catalítica óptima e faz parte do local activo, principalmente no local aceitador ou na vizinhança dos substratos de açúcar tanto dadores como aceitadores (Jeanneau *et al.*, 2004). De trabalhos anteriores, é agora claro que a parte C-terminal do CD (parte do motivo sialil S, motivo 3 e motivo sialil VS) está primariamente dedicada ao reconhecimento de substratos aceitadores enquanto a parte N-terminal (motivos sialil L e parte do S) está principalmente envolvida na ligação do açúcar nucleotídico (Datta & Paulson, 1995; Datta *et al.*, 1998; Laroy *et al.*, 2001; Kitazume-Kawaguchi *et al.*, 2001; Jeanneau *et al.*, 2004). No entanto, não se pode excluir que o motivo Sialil L também pode estar envolvido no reconhecimento do aceitador (Legaigneur *et al.*, 2001).

Tal como mencionado acima, as sequências de aminoácidos deduzidas a partir do ADNc da sialiltransferase humana clonado mostram a mesma organização em quatro domínios partilhados por muitas GT: i) uma curta cauda citoplasmática N-terminal (CT; cerca de 10 aminoácidos), ii) um domínio transmembranar (TMD; cerca de 20 aminoácidos), iii) uma região de caule (SR), de comprimento altamente variável e iv) um domínio catalítico (CD; cerca de 300 aminoácidos).

A SR é definida como a região peptídica entre o TMD e o CD que pode ser removida sem alterar a actividade enzimática (Paulson 1989, Ronin, 2001, Vallejo-Ruiz *et al.*, 2001; Jeanneau *et al.*, 2004). No entanto, os inventores mostraram que esta região hipervariável pode ser essencial para definir um reconhecimento fino do glicano aceitador (tri>bi>tetra-anténario) e propuseram que possa estar envolvida numa mudança de conformação para abrir o local catalítico (Ronin,

2001). Como resultado, a eficiência catalítica aumenta num factor de 35 e o reconhecimento do glicano aceitador é alargado.

O CD é crucial para que as ST apresentem actividade enzimática. Contém pelo menos três sequências altamente conservadas mantendo posições de aminoácidos idênticas verificadas em todas as ST de mamífero clonadas. Esses domínios são os motivos sialil L, S e VS (Livingston & Paulson, 1993; Geremia *et al.*, 1997). Um domínio adicional (motivo 3) foi também recentemente isolado entre os motivos sialil S e VS, contém quatro resíduos altamente conservados, com a seguinte sequência de consenso: (H/y)Y(Y/F/W/h)(E/D/q/g). (As letras maiúsculas e as letras minúsculas indicam uma forte ou uma baixa ocorrência do aminoácido, respectivamente) (Jeanneau *et al.*, 2004). Muitos estudos descritos na literatura procuram conhecimento sobre relações de estrutura/função na grande família das ST e para assim definir particularmente um domínio catalítico mínimo dentro. O CD mínimo foi assim experimentalmente definido através de mutagénese dirigida ao local ou alternativamente através de alinhamento de sequências e comparação para algumas ST (Vallejo-Ruiz *et al.*, 2001; Chen & Colley, 2000, Ronin 2003).

Apesar da ampla definição do CD e da actividade catalítica, ainda não é fácil delinear um CD mínimo de ST. ST6Gal I é a enzima mais estudada entre as ST a maior parte da investigação tem sido realizada nesta para definir o domínio catalítico mínimo e a actividade catalítica. Isto é devido à ausência de ST6Gal em todas as células utilizadas para produzir proteínas recombinantes humanas. Esta falta é um estrangulamento tecnológico para sistemas heterólogo utilizando células não animais (leveduras, insectos e plantas) uma vez que não contém qualquer ácido siálico. Alternativamente, em células CHO apenas uma actividade de ST3 é expressa.

O delineamento entre a SR e o CD nunca foi bem definido excepto experimentalmente. Com base em bioinformática, os inventores desenharam uma estratégia de identificação da extremidade da SR para as 3 famílias de ST (Ronin., 2003).

O CD coincide geralmente com a extremidade da região hipervariável na metade N-terminal das enzimas (Ronin., 2003). Assume-se portanto que o domínio catalítico começa cerca dos resíduos 70-90 a montante do motivo sialil L e cerca dos resíduos 40-45 para ST6GalNAc III a VI. O tamanho médio do CD é estimado cerca de 300 ($\pm$ 20) aminoácidos, incluindo os motivos sialil (Jeanneau *et al.*, 2004) (Tabela 3). O CD mínimo foi definido para algumas ST - através de experiências de truncagem ou comparação de sequências - (Legaigreur *et al.*, 2001; Vallejo-Ruiz *et al.*, 2001; Chen & Colley, 2000), tal como para hST3Gal I (CD mínimo consistindo nos aminoácidos 57-340; Vallejo-Ruiz *et al.*, 2001) e hST8Sia IV (CD mínimo consistindo nos aminoácidos 62-359; Angata *et al.*, 2004).

Quando transfectado para células CHO, verificou-se que o CD solúvel de hST6Gal I (aminoácidos 90-406) mostra uma especificidade alargada para aceitadores endógenos uma vez que segue a via de secreção intracelular dentro do aparelho de Golgi. Foi então possível delinear um CD mínimo para hST6Gal I contendo uma sequência crítica capaz de deslocar o reconhecimento do aceitador de aceitadores intracelulares residentes para glicoconjugados da superfície celular (Donadio *et al.*, 2003). Além disso, o CD segregado solúvel de hST6Gal I mostrou maior eficiência de transferência, independentemente do padrão de ramificação do aceitador glicano (Legaigreur *et al.*, 2001).

Glicosilação de proteínas (fármacos) produzidas em sistemas de expressão heteróloga

Proteínas de interesse terapêutico foram primeiro extraídas de fontes naturais tais como sangue, placenta, tecidos humanos ou animais. No entanto, esta abordagem está limitada pela fonte, quantidade e disponibilidade de tecidos humanos e pode conter contaminantes perigosos para a vida (priões, oncogenes, vírus...) bem como potenciais alergénios gerados por proteínas de animais. Com a subida das proteínas aprovadas como fármacos e das necessidades clínicas, foram desenvolvidas novas abordagens para produzir proteínas utilizando diferentes sistemas de expressão.

Na maioria dos casos, trabalho intensivo está actualmente a tentar a humanização do padrão de glicosilação das proteínas recombinantes para o aproximar do padrão verificado nas glicoproteínas naturais o mais possível para melhorar a farmacocinética e baixar a imunogenicidade do produto.

A maquinaria necessária para a síntese, a activação e a introdução de resíduos sialilo é fracamente expressa nos vários sistemas de expressão de proteínas recombinantes existentes. As proteínas humanas recombinantes produzidas são por isso frequentemente subsialiladas ou mesmo não sialiladas em comparação com os seus equivalentes nativos.

Um dos sistemas mais utilizados é a bactéria *Escherichia coli* (*E. coli*) (Swartz, 2001; Baneyx, 1999), mas o seu maior inconveniente é que as modificações pós-tradução humanas, particularmente as glicosilações de proteínas não são realizadas por este procariota porque tais glicosiltransferases não são expressas em *E. coli*. Isto pode conduzir à rejeição das proteínas terapêuticas de interesse pelo sistema imunitário, à redução da sua semi-vida em circulação e/ou da sua actividade biológica. A proteína produzida pode eventualmente ser mal dobrada e agregar-se como corpos de inclusão.

As leveduras e fungos filamentosos são também sistemas de expressão eucariotas bem conhecidos e possuem maquinaria celular semelhante à das células humanas. As leveduras produzem proteínas complexas e são capazes de realizar várias modificações pós-tradução tais como glicosilações simples. No entanto, embora o processo de N-glicosilação realizado pelas leveduras e pelos fungos seja igual ao processo de mamífero nos seus passos iniciais no retículo endoplasmático, não se verificou nenhum oligossacárido complexo contendo ácido siálico, galactose, fucose e N-acetilgalactosamina dentro das glicoproteínas produzidas por estes organismos (Blanchard, 2004); tanto as leveduras como os fungos produzem, tipicamente, glicanos ricos em manose através da adição de 100 resíduos de manose (no caso das leveduras) no núcleo pentassacarídico (Tanner & Lehele, 1987; Herscovics & Orlean,

1993) no aparelho de Golgi. Essa hipermanosilação cria uma resposta imunitária em humanos.

A modificação de glicosilação em leveduras (Hamilton *et al.*, 2003; Roy *et al.*, 2000) permitiu pela primeira vez a redução do número de resíduos de manose adicionado. Depois, foram adicionadas a estes sistemas enzimas que são necessárias para N-acetilglucosaminilação periférica e galactosilação (Maras *et al.*, 1999; Bretthauer, 2003; Vervecken *et al.*, 2004). No entanto, a adição do ácido siálico terminal é ainda difícil de alcançar, devido ao grande número de enzimas envolvidas neste passo final.

As proteínas expressas em células de insecto são dobradas correctamente, podem sofrer modificações pós-tradução e ser segregadas. A glicosilação ligada em N inicial realizada por estas células é semelhante à realizada por células de mamífero. No entanto, as estruturas de glicano obtidas neste caso são do tipo paucimanosídico, i.e. truncadas devido à presença de uma actividade de N-acetilhexosaminidase que degrada as neoglicoproteínas expressas durante expressão por Baculovírus (Blanchard, 2004). Além disso, nalgumas linhas celulares de insecto, podem ser verificados resíduos α 1,3/fucose e estes resíduos desencadeiam geralmente uma resposta imunitária indesejável em humano. Assim, a utilização deste sistema está restrita à produção de antigénios de vacinação.

Em células de insecto, as GT catalisando a transferência de açúcares imunogénicos foram suprimidas e a sialilação pode ser alcançada através da adição de 3 genes codificando para a N-acetilglucosamina-2-epimerase, N-acetilneuraminil-liase e CMP-Neu5Ac-sintetase (Jarvis *et al.*, 1998; Aumiller *et al.*, 2003). Foi criada uma nova linha celular de insecto (SfSWT-3) desenhada para sintetizar o seu próprio CMP-ácido siálico. As células resultantes expressam todos os 7 genes de mamífero, podem produzir CMP-ácido siálico e sialilam uma proteína heteróloga quando cultivadas num meio de crescimento sem soro (Aumiller *et al.*, 2003).

Como células eucarióticas, as plantas exibem uma maquinaria celular complexa e sofisticada que pode ser

utilizada para produzir proteínas terapêuticas. As proteínas recombinantes possuem uma qualidade farmacológica muito boa porque as plantas têm muitas das enzimas necessárias para a maturação das proteínas. No entanto, a via de glicosilação necessita de maiores ajustamentos para não produzir proteínas alergénicas. De facto, as N-glicosilações em plantas (Lerouge *et al.*, 2000) são semelhantes às realizadas em humanos no que se refere à glicosilação central. No entanto, os glicanos não possuem ainda antenas sialiladas e contêm β 1,2-xilose e α 1,3-fucose internas. Ambos os resíduos são altamente imunogénicos para humanos e comprometem actualmente a aprovação de plantas transgénicas como sistemas de expressão para produtos terapêuticos.

Em plantas, a primeira estratégia utilizada tentou evitar a adição de açúcares alergénicos evitando que as proteínas saíssem do retículo endoplasmático. Como resultado, os glicanos ligados em N não podem amadurecer até ao tipo complexo. Outra estratégia baseou-se na inibição de várias GT dentro do aparelho de Golgi. Esta inibição não pode ser completa e/ou pode entrar em competição com a maquinaria endógena pela maturação.

O sistema de expressão em células de mamífero, designadamente células CHO, é actualmente o único sistema aprovado para fármacos para produzir produtos terapêuticos recombinantes. Estas células mostram uma importante vantagem porque são capazes de sintetizar glicoproteínas ligadas em N complexas de elevado peso molecular e/ou multiméricas. As células de mamífero expressam naturalmente, não só as enzimas envolvidas na síntese e no transporte dos açúcares nucleotídicos, como também as glicosiltransferases necessárias para alcançar glicosilação complexa das proteínas recombinantes com um teor satisfatório de ácido siálico ligado em 3. No entanto, existe a falta de algumas enzimas tais como as α 1,3/4-fucosiltransferases e α 2,6-sialiltransferases, que realizam glicosilação terminal ligada em O e ligada em N. Além disso, em células de mamífero, a sialilação ocorre através do ácido N-glicolineuramínico (NeuGc) que também difere significativamente do derivado N-Acetilado (NeuAc) verificado em células humanas.

No caso das células de mamífero, têm sido realizados trabalhos através da sobre-expressão de uma $\alpha 2,3$ -ST e uma $\beta 1,4$ GalT (Weikert et al., 1999); ambas as enzimas estão presentes no genoma mas as suas actividades são altamente variáveis em condições de cultura celular. Isto conduziu a uma grande variabilidade em relação à presença do Gal terminal e do ácido siálico e a uma extensa micro-heterogeneidade nas proteínas glicosiladas. O trabalho foi também orientado para a optimização da galactosilação e sialilação (Granbenhorst et al., 1999) através da introdução de um $\alpha 2,6$ -ST (Bragonzi et al., 2000) em células de mamífero. Foi estabelecida uma linha celular CHO expressando estavelmente $\alpha 2,6$ -ST nativa de rato. As glicoproteínas produzidas por estas células CHO mostram ambos os resíduos de ácido siálico terminais ligados em $\alpha 2,6$ e $\alpha 2,3$, semelhantes ao humano; a razão observada entre os resíduos de ácido siálico terminais ligados em $\alpha 2,6$ e $\alpha 2,3$ contidos foi de 40,4% de resíduos de ácido siálico em $\alpha 2,6$ e 59,6% em $\alpha 2,3$, o que melhorou a farmacocinética em estudos de eliminação (Bragonzi et al., 2000). Apesar destas melhorias na humanização de células, a razão entre resíduos de ácido siálico terminais ligados em $\alpha 2,6$ e $\alpha 2,3$ não pode ser controlada e nunca se verificou sequer favorável à actividade de 6. Parece assim que, primeiro, a sialilação é um passo crítico para controlar as estruturas dos glicanos e, segundo, a dificuldade reside também na expressão das ST heterólogas num compartimento específico da célula (direccionamento), na optimização da sua actividade desde que o substrato dador para esta enzima esteja presente (transportador).

Deve-se portanto observar que a sialilação terminal de fármacos glicoproteicos aprovados é o passo mais difícil de obter em todos os sistemas de expressão até agora disponíveis.

Objectivo do invento

O objectivo do presente invento é proporcionar um processo para a criação de um painel de novas sequências génicas codificando enzimas membranares quiméricas com actividade de glicosiltransferase, e em particular para desenhar sequências génicas codificando sialiltransferases

quiméricas inovadoras para equipar células com uma actividade de sialilação necessária e melhorar a qualidade e o rendimento de proteínas glicosiladas recombinantes.

O invento refere-se a um processo para a produção de sequências génicas codificando glicosiltransferases membranares quiméricas apresentando uma actividade de glicosilação optimizada em células transformadas com as referidas sequências, quando comparada com a actividade de glicosilação das correspondentes glicosiltransferases nativas, i.e. uma glicosilação menos selectiva e/ou mais eficiente (até pelo menos 30 vezes superior à actividade inicial das glicosiltransferases inteiras nativas) para o substrato aceitador, compreendendo o referido processo a fusão:

- de uma primeira sequência de ácido nucleico codificando para um fragmento mínimo C-terminal do domínio catalítico (CD) da glicosiltransferase inteira nativa, apresentando o referido fragmento de CD mínimo C-terminal uma actividade de transferase (que pode ser aumentada até pelo menos 30 vezes mais que a actividade inicial do CD nativo), sendo a referida primeira sequência nucleica obtida através da remoção dos nucleótidos codificando para um ou vários aminoácidos contíguos estendendo-se desde o primeiro aminoácido da extremidade N-terminal do referido CD, e selecção do ácido nucleico codificando o referido fragmento de CD mínimo C-terminal que é tal que se n representa o número de aminoácidos contíguos tal como definido acima que foram suprimidos, então o fragmento obtido quando se suprime pelo menos $n+1$ aminoácidos contíguos tal como definido acima, não tem actividade de transferase substancial,

- com um segundo ácido nucleico de sequência variável codificando para uma cadeia peptídica transmembranar especificando o ancoramento das glicosiltransferases em compartimentos intracelulares, e compreendendo (ou consistindo em) a sua região N-terminal uma região de cauda citoplasmática (CT) localizada a montante de um domínio transmembranar (TMD), ele próprio localizado a montante de uma região de caule (SR) ou de um fragmento de pelo menos 3 aminoácidos contíguos da SR, estando a referida SR ou parte desta ligada ao referido domínio catalítico, opcionalmente através de um ligador ou péptido de conexão de pelo menos 2

aminoácidos codificado por um local de restrição que não existe na sequência nucleotídica codificando para o CD nativo mencionado acima,

desde que pelo menos um destes péptidos CT, TMD, SR seja diferente da estrutura primária dos péptidos nativos correspondentes na glicosiltransferase nativa da qual é derivado o fragmento de CD com actividade de glicosiltransferase óptima tal como definido acima,

sendo esta fusão realizada de tal modo que o primeiro ácido nucleico fique localizado a jusante do segundo ácido nucleico e proporcione um produto proteico no qual o CD está na metade C-terminal.

Por glicosiltransferases membranares quiméricas apresentando uma actividade de glicosilação optimizada em células transformadas com as referidas sequências, quando comparada com a actividade de glicosilação das correspondentes glicosiltransferases nativas, deve entender-se que as referidas glicosiltransferases membranares quiméricas:

- têm uma actividade de transferência de açúcar menos selectiva para substratos aceitadores que as correspondentes glicosiltransferases nativas quando testadas *in vitro* com glicoproteínas aceitadoras exógenas/comercialmente disponíveis possuindo glicanos bi, tri ou tetra-antenários conhecidos, i.e. possuindo uma actividade de glicosilação *in vivo* em células para todos os aceitadores glicoproteicos celulares, sendo a referida actividade de glicosilação mensurável de preferência em compartimentos intracelulares e da superfície celular de acordo com o seguinte procedimento geral utilizando ligação de lectina SNA,

- e/ou têm uma actividade de transferência de açúcar mais eficiente para os seus substratos aceitadores que as correspondentes glicosiltransferases nativas, i.e. mostram uma actividade de glicosilação até pelo menos 30 vezes maior que a actividade inicial das glicosiltransferases inteiras nativas, sendo a referida actividade de glicosilação mensurável de acordo com o procedimento geral *in vitro* com glicoproteínas aceitadoras exógenas/comercialmente disponíveis possuindo glicanos bi, tri ou tetra-antenários conhecidos.

A expressão "glicosiltransferases quiméricas" utilizada acima corresponde a uma glicosiltransferase cuja sequência inteira não é o produto directo de um gene nativo ou um transcrito mas foi desenhada exclusivamente utilizando sequências de outras glicosiltransferases.

A expressão "glicosiltransferases nativas" utilizada acima corresponde à sequência da enzima inteira tal como representada por uma sequência de codificação de ocorrência natural.

A expressão "actividade de glicosilação" utilizada acima corresponde à reacção catalítica de transferência de um açúcar de um nucleótido dador para um substrato aceitador.

A expressão "domínio catalítico (CD) mínimo" utilizada acima corresponde ao domínio peptídico C-terminal de uma sequência de glicosiltransferase que não pode ser mais suprimido sem perda de actividade de transferência.

A expressão "domínio transmembranar (TMD)" utilizada acima corresponde a uma porção peptídica composta por um trecho de 17-24 aminoácidos essencialmente hidrófobos.

A expressão "cauda citoplasmática (CT)" utilizada acima corresponde ao péptido N-terminal de uma glicosiltransferase que pode englobar pelo menos mais de 3 aminoácidos a montante do TMD.

A expressão "região de caule (SR)" utilizada acima corresponde a um trecho de não mais de 246 aminoácidos a jusante do TMD/a montante do CD.

O invento refere-se mais particularmente a um processo tal como definido acima, caracterizado por o primeiro e o segundo ácidos nucleicos serem derivados de sequências nucleotídicas codificando domínios CD, ou CT, TMD, e regiões SR, respectivamente, em glicosiltransferases de origem eucariótica, de preferência mamíferos e humanos, estando a referida glicosiltransferases envolvida em:

- O-glicosilação das proteínas em células, tais como as N-acetilgalactosaminil-, N-acetilglucosaminil-, glucosil-, fucosil-, galactosil- ou sialiltransferases;
- N-glicosilação das proteínas em células tais como as N-acetilglucosaminil-, galactosil-, fucosil- ou sialiltransferases,
- glicosilação de lípidos tais como as N-acetilgalactosaminil-, N-acetilglucosaminil-, fucosil-, galactosil- ou sialiltransferases.

A expressão "O-glicosilação" utilizada acima corresponde à via biossintética elaborando monossacáridos ou oligossacáridos covalentemente ligados a aminoácidos que não sejam resíduos de asparagina mas que possam ser de preferência hidroxiaminoácidos tais como resíduos de serina, treonina ou hidroxilisina.

A expressão "N-glicosilação" utilizada acima corresponde à via biossintética elaborando oligossacáridos ligados a resíduos de asparagina dentro do tripéptido de consenso Asn-X-Ser/Thr (X não sendo Prolina).

A expressão "galactosiltransferases" utilizada acima corresponde a uma glicosiltransferase que transfere um resíduo de galactose de UDP-Gal para um aceitador proteína ou glicolípido ligado em N/O.

A expressão "sialiltransferases" utilizada acima corresponde a uma glicosiltransferase que transfere um resíduo de ácido siálico, de preferência um derivado de ácido neuramínico de CMP-NeuAc para um aceitador proteína ou glicolípido ligado em N/O.

O invento refere-se mais particularmente a um processo tal como definido acima, caracterizado por o primeiro e segundo ácidos nucleicos serem derivados de sequências nucleotídicas codificando domínios CD, ou CT, TMD, e regiões SR, respectivamente, em N-acetilgalactosaminiltransferases, N-acetilglucosaminiltransferases, glucosiltransferases, galactosiltransferases, fucosiltransferases ou sialiltransferases.

O invento refere-se mais particularmente a um processo tal como definido acima, caracterizado por o primeiro e segundo ácidos nucleicos serem derivados de sequências nucleotídicas codificando domínios CD, ou CT, TMD, e regiões SR, respectivamente, em alfa-6-fucosiltransferases (glicosilação central), beta-2/4/6-N-acetilglucosaminiltransferases (ramificação), beta-4-galactosiltransferases, e alfa-3/6/8-sialiltransferases (glicosilação terminal).

Vantajosamente, o primeiro e segundo ácidos nucleicos são derivados de sequências nucleotídicas codificando domínios CD, ou CT, TMD, e regiões SR, de enzimas implicadas na via de N-glicosilação.

O invento refere-se mais particularmente a um processo tal como definido acima, caracterizado por o primeiro e segundo ácidos nucleicos serem derivados de sequências nucleotídicas codificando domínios CD, ou CT, TMD, e regiões SR, respectivamente, em sialiltransferases.

O invento refere-se mais particularmente a um processo tal como definido acima, caracterizado por o primeiro e segundo ácidos nucleicos serem derivados de sequências nucleotídicas codificando domínios CD, ou CT, TMD, e regiões SR, respectivamente, em α 2,6-sialiltransferases, α 2,3-sialiltransferases ou α 2,8-sialiltransferases.

Vantajosamente, as sequências nucleotídicas codificando CT, TMD e/ou SR ou fragmento de SR mencionadas acima são de preferência de origem sintética.

A expressão " α 2,6-sialiltransferases" utilizada acima corresponde a uma glicosiltransferase capaz de transferir um resíduo de ácido siálico, de preferência um derivado de ácido neuramínico de CMP-NeuAc para a posição 6 de um aceitador hidrato de carbono do tipo proteína ou glicolípido ligado em N/O.

A expressão " α 2,3-sialiltransferases" utilizada acima corresponde a uma glicosiltransferase capaz de transferir um resíduo de ácido siálico, de preferência um derivado de ácido

neuramínico de CMP-NeuAc para a posição 3 de um aceitador hidrato de carbono do tipo proteína ou glicolípido ligado em N/O.

A expressão " α 2,8-sialiltransferases" utilizada acima corresponde a uma glicosiltransferase capaz de transferir um resíduo de ácido siálico, de preferência um derivado de ácido neuramínico de CMP-NeuAc para a posição 8 de um aceitador sialilado do tipo proteína ou glicolípido ligado em N/O.

O invento refere-se mais particularmente a um processo tal como definido acima, caracterizado por o primeiro e segundo ácidos nucleicos serem derivados de sequências nucleotídicas codificando domínios CD, ou CT, TMD, e regiões SR, respectivamente, em:

- α 2,6-sialiltransferases escolhidas de entre:

- * as β 1,4-galactósido- α 2,6-sialiltransferases I e II humanas (hST6Gal I e hST6Gal II) representadas por SEQ ID NO:2, e SEQ ID NO:4, respectivamente, codificadas pelas sequências nucleotídicas SEQ ID NO:1, e SEQ ID NO:3, respectivamente, ou

- * as N-acetilgalactosaminida- α 2,6-sialiltransferases I a VI humanas (hST6GalNAc I a VI) representadas por SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, e SEQ ID NO:16, respectivamente, codificadas pelas sequências nucleotídicas SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, e SEQ ID NO:15, respectivamente,

- α 2,3-sialiltransferases escolhidas de entre as galactósido- α 2,3-sialiltransferases I a VI humanas (hST3Gal I a VI) representadas por SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, e SEQ ID NO:28, respectivamente, codificadas pelas sequências nucleotídicas SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, e SEQ ID NO:27, respectivamente, ou as galactósido- α 2,3-sialiltransferases I a VI de rato (rST3Gal I a VI), tal como a rST3Gal III representada por SEQ ID NO:30 codificada pela sequência nucleotídica SEQ ID NO:29, ou qualquer ST de outra origem animal que partilhe pelo menos 85% de homologia com a enzima humana,

- α 2,8-sialiltransferases escolhidas de entre as ácido siálico- α 2,8-sialiltransferases I a VI humanas (hST8Sia I a

VI) representadas por SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, e SEQ ID NO:42, respectivamente, codificadas pelas sequências nucleotídicas SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, e SEQ ID NO:41, respectivamente.

A expressão "galactósido- α 2,6-sialiltransferases" utilizada acima corresponde a uma glicosiltransferase que transfere um resíduo de ácido siálico, de preferência um derivado de ácido neuramínico de CMP-NeuAc para a posição 6 de um aceitador galactosilado do tipo proteína ou lípido ligado em N/O.

A expressão "N-acetilgalactosaminida- α 2,6-sialiltransferases" utilizada acima corresponde a uma glicosiltransferase que transfere um resíduo de ácido siálico, de preferência um derivado de ácido neuramínico de CMP-NeuAc para a posição 6 de um resíduo N-acetilgalactosaminilo de um aceitador proteína ou glicolípido ligado em N/O.

A expressão "galactósido- α 2,3-sialiltransferases" utilizada acima corresponde a uma glicosiltransferase que transfere um resíduo de ácido siálico, de preferência um derivado de ácido neuramínico de CMP-NeuAc para a posição 3 de um aceitador proteína ou lípido ligado em N/O galactosilado.

A expressão "ácido siálico- α 2,8-sialiltransferases" utilizada acima corresponde a uma glicosiltransferase que transfere um resíduo de ácido siálico, de preferência um derivado de ácido neuramínico de CMP-NeuAc para um aceitador proteína sialilada ou glicolípido ligado em N/O.

O invento refere-se mais particularmente a um processo tal como definido acima, caracterizado por a sequência nucleotídica codificando o domínio CD compreendido no primeiro ácido nucleico ser escolhida de entre as sequências constituídas por, ou compreendendo:

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 268 a 330 e na sua extremidade 3' pelo nucleótido localizado na posição 1218 de

SEQ ID NO:1, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST6Gal I delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 90 a 110 e na sua extremidade C-terminal pelo aminoácido localizado na posição 406 de SEQ ID NO:2,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 307 a 369 e na sua extremidade 3' pelo nucleótido localizado na posição 1587 de SEQ ID NO:3, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST6Gal II delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 103 a 123 e na sua extremidade C-terminal pelo aminoácido localizado na posição 529 de SEQ ID NO:4,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 814 a 876 e na sua extremidade 3' pelo nucleótido localizado na posição 1800 de SEQ ID NO:5, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST6GalNac I delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 272 a 292 e na sua extremidade C-terminal pelo aminoácido localizado na posição 600 de SEQ ID NO:6,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 172 a 234 e na sua extremidade 3' pelo nucleótido localizado na posição 1122 de SEQ ID NO:7, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST6GalNac II delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 58 a 78 e na sua extremidade C-terminal pelo aminoácido localizado na posição 374 de SEQ ID NO:8,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 73 a 135 e na sua extremidade 3' pelo nucleótido localizado na posição 915 de SEQ ID NO:9, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST6GalNac III delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 25 a 45 e na sua extremidade C-terminal pelo aminoácido localizado na posição 305 de SEQ ID NO:10,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 61 a 123 e na sua extremidade 3' pelo nucleótido localizado na posição 906 de SEQ ID NO:11, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST6GalNac IV delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 21 a 41 e na sua extremidade C-terminal pelo aminoácido localizado na posição 302 de SEQ ID NO:12,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 121 a 183 e na sua extremidade 3' pelo nucleótido localizado na posição 1008 de SEQ ID NO:13, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST6GalNac V delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 41 a 61 e na sua extremidade C-terminal pelo aminoácido localizado na posição 336 de SEQ ID NO:14,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 154 a 216 e na sua extremidade 3' pelo nucleótido localizado na posição 999 de SEQ ID NO:15, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST6GalNac VI delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 52 a 72 e na sua extremidade C-terminal pelo aminoácido localizado na posição 333 de SEQ ID NO:16,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 145 a 207 e na sua extremidade 3' pelo nucleótido localizado na posição 1020 de SEQ ID NO:17, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST3Gal I delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 49 a 69 e na sua extremidade C-terminal pelo aminoácido localizado na posição 340 de SEQ ID NO:18,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 175 a 237 e na sua extremidade 3' pelo nucleótido localizado na posição 1050 de SEQ ID NO:19, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST3Gal II delimitada na sua extremidade N-terminal pelo

aminoácido localizado na posição 59 a 79 e na sua extremidade C-terminal pelo aminoácido localizado na posição 350 de SEQ ID NO:20,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 199 a 261 e na sua extremidade 3' pelo nucleótido localizado na posição 1332 de SEQ ID NO:21, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST3Gal III delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 67 a 87 e na sua extremidade C-terminal pelo aminoácido localizado na posição 444 de SEQ ID NO:22,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 79 a 141 e na sua extremidade 3' pelo nucleótido localizado na posição 987 de SEQ ID NO:23, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST3Gal IV delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 27 a 47 e na sua extremidade C-terminal pelo aminoácido localizado na posição 329 de SEQ ID NO:24,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 136 a 198 e na sua extremidade 3' pelo nucleótido localizado na posição 1086 de SEQ ID NO:25, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST3Gal V delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 46 a 66 e na sua extremidade C-terminal pelo aminoácido localizado na posição 362 de SEQ ID NO:26,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 73 a 135 e na sua extremidade 3' pelo nucleótido localizado na posição 993 de SEQ ID NO:27, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST3Gal VI delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 25 a 45 e na sua extremidade C-terminal pelo aminoácido localizado na posição 331 de SEQ ID NO:28,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 103 a 165 e na sua extremidade 3' pelo nucleótido localizado na posição 1122 de

SEQ ID NO:29, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de RatST3Gal III delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 35 a 55 e na sua extremidade C-terminal pelo aminoácido localizado na posição 374 de SEQ ID NO:30,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 133 a 195 e na sua extremidade 3' pelo nucleótido localizado na posição 1068 de SEQ ID NO:31, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST8Sia I delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 45 a 65 e na sua extremidade C-terminal pelo aminoácido localizado na posição 356 de SEQ ID NO:32,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 190 a 252 e na sua extremidade 3' pelo nucleótido localizado na posição 1125 de SEQ ID NO:33, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST8Sia II delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 64 a 84 e na sua extremidade C-terminal pelo aminoácido localizado na posição 375 de SEQ ID NO:34,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 205 a 267 e na sua extremidade 3' pelo nucleótido localizado na posição 1140 de SEQ ID NO:35, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST8Sia III delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 69 a 89 e na sua extremidade C-terminal pelo aminoácido localizado na posição 380 de SEQ ID NO:36,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 145 a 207 e na sua extremidade 3' pelo nucleótido localizado na posição 1077 de SEQ ID NO:37, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST8Sia IV delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 49 a 69 e na sua extremidade C-terminal pelo aminoácido localizado na posição 359 de SEQ ID NO:38,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 211 a 273 e na sua extremidade 3' pelo nucleótido localizado na posição 1128 de SEQ ID NO:39, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST8Sia V delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 71 a 91 e na sua extremidade C-terminal pelo aminoácido localizado na posição 376 de SEQ ID NO:40,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 277 a 339 e na sua extremidade 3' pelo nucleótido localizado na posição 1194 de SEQ ID NO:41, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST8Sia VI delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 93 a 113 e na sua extremidade C-terminal pelo aminoácido localizado na posição 398 de SEQ ID NO:42.

O invento refere-se mais particularmente a um processo tal como definido acima, caracterizado por o primeiro ácido nucleico ser a sequência nucleotídica SEQ ID NO:43 correspondendo à sequência delimitada pelos nucleótidos localizados na posição 268 e 1218 de SEQ ID NO:1, a referida sequência nucleotídica SEQ ID NO:43 codificando a sequência polipeptídica SEQ ID NO:44 correspondendo ao fragmento mínimo C-terminal do domínio CD de hST6Gal I delimitada pelos aminoácidos localizados nas posições 90 a 406 de SEQ ID NO:2.

O invento refere-se mais particularmente a um processo tal como definido acima, caracterizado por:

- a sequência nucleotídica codificando a região CT compreendida no segundo ácido nucleico ser escolhida de entre:

* a sequência nucleotídica SEQ ID NO:45 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 27 de SEQ ID NO:1, a referida sequência nucleotídica SEQ ID NO:45 codificando a sequência polipeptídica SEQ ID NO:46 correspondendo à região CT de hST6Gal I delimitada pelos aminoácidos localizados nas posições 1 a 9 de SEQ ID NO:2,

* a sequência nucleotídica SEQ ID NO:47 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 30 de SEQ ID NO:3, a referida sequência nucleotídica SEQ ID NO:47 codificando a sequência polipeptídica SEQ ID NO:48 correspondendo à região CT de hST6Gal II delimitada pelos aminoácidos localizados nas posições 1 a 10 de SEQ ID NO:4,

* a sequência nucleotídica SEQ ID NO:49 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 42 de SEQ ID NO:5, a referida sequência nucleotídica SEQ ID NO:49 codificando a sequência polipeptídica SEQ ID NO:50 correspondendo à região CT de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 1 a 14 de SEQ ID NO:6,

* a sequência nucleotídica SEQ ID NO:51 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 21 de SEQ ID NO:7, a referida sequência nucleotídica SEQ ID NO:51 codificando a sequência polipeptídica SEQ ID NO:52 correspondendo à região CT de hST6GalNAc II delimitada pelos aminoácidos localizados nas posições 1 a 7 de SEQ ID NO:8,

* a sequência nucleotídica SEQ ID NO:53 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 24 de SEQ ID NO:9, a referida sequência nucleotídica SEQ ID NO:53 codificando a sequência polipeptídica SEQ ID NO:54 correspondendo à região CT de hST6GalNAc III delimitada pelos aminoácidos localizados nas posições 1 a 8 de SEQ ID NO:10,

* a sequência nucleotídica SEQ ID NO:55 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 18 de SEQ ID NO:11, a referida sequência nucleotídica SEQ ID NO:55 codificando a sequência polipeptídica SEQ ID NO:56 correspondendo à região CT de hST6GalNAc IV delimitada pelos aminoácidos localizados nas posições 1 a 6 de SEQ ID NO:12,

* a sequência nucleotídica SEQ ID NO:57 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 24 de SEQ ID NO:13, a referida sequência nucleotídica SEQ ID NO:57 codificando a sequência polipeptídica SEQ ID NO:58 correspondendo à região CT de hST6GalNAc V delimitada pelos aminoácidos localizados nas posições 1 a 8 de SEQ ID NO:14,

* a sequência nucleotídica SEQ ID NO:59 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 129 de SEQ ID NO:15, a referida sequência nucleotídica SEQ ID NO:59 codificando a sequência polipeptídica SEQ ID NO:60 correspondendo à região CT de hST6GalNAc VI delimitada pelos aminoácidos localizados nas posições 1 a 43 de SEQ ID NO:16,

* a sequência nucleotídica SEQ ID NO:61 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 39 de SEQ ID NO:17, a referida sequência nucleotídica SEQ ID NO:61 codificando a sequência polipeptídica SEQ ID NO:62 correspondendo à região CT de hST3Gal I delimitada pelos aminoácidos localizados nas posições 1 a 13 de SEQ ID NO:18,

* a sequência nucleotídica SEQ ID NO:63 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 18 de SEQ ID NO:19, a referida sequência nucleotídica SEQ ID NO:63 codificando a sequência polipeptídica SEQ ID NO:64 correspondendo à região CT de hST3Gal II delimitada pelos aminoácidos localizados nas posições 1 a 6 de SEQ ID NO:20,

* a sequência nucleotídica SEQ ID NO:65 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 24 de SEQ ID NO:21, a referida sequência nucleotídica SEQ ID NO:65 codificando a sequência polipeptídica SEQ ID NO:66 correspondendo à região CT de hST3Gal III delimitada pelos aminoácidos localizados nas posições 1 a 8 de SEQ ID NO:22,

* a sequência nucleotídica SEQ ID NO:67 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 24 de SEQ ID NO:23, a referida sequência nucleotídica SEQ ID NO:67 codificando a sequência polipeptídica SEQ ID NO:68 correspondendo à região CT de hST3Gal IV delimitada pelos aminoácidos localizados nas posições 1 a 8 de SEQ ID NO:24,

* a sequência nucleotídica SEQ ID NO:69 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 15 de SEQ ID NO:25, a referida sequência nucleotídica SEQ ID NO:69 codificando a sequência polipeptídica SEQ ID NO:70 correspondendo à região CT de hST3Gal V delimitada pelos aminoácidos localizados nas posições 1 a 5 de SEQ ID NO:26,

* a sequência nucleotídica SEQ ID NO:71 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 12 de SEQ ID NO:27, a referida sequência nucleotídica SEQ ID NO:71 codificando a sequência polipeptídica SEQ ID NO:72 correspondendo à região CT de hST3Gal VI delimitada pelos aminoácidos localizados nas posições 1 a 4 de SEQ ID NO:28,

* a sequência nucleotídica SEQ ID NO:73 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 24 de SEQ ID NO:29, a referida sequência nucleotídica SEQ ID NO:73 codificando a sequência polipeptídica SEQ ID NO:74 correspondendo à região CT de ratST3Gal III delimitada pelos aminoácidos localizados nas posições 1 a 8 de SEQ ID NO:30,

* a sequência nucleotídica SEQ ID NO:75 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 87 de SEQ ID NO:31, a referida sequência nucleotídica SEQ ID NO:75 codificando a sequência polipeptídica SEQ ID NO:76 correspondendo à região CT de hST8Sia I delimitada pelos aminoácidos localizados nas posições 1 a 29 de SEQ ID NO:32,

* a sequência nucleotídica SEQ ID NO:77 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 18 de SEQ ID NO:33, a referida sequência nucleotídica SEQ ID NO:77 codificando a sequência polipeptídica SEQ ID NO:78 correspondendo à região CT de hST8Sia II delimitada pelos aminoácidos localizados nas posições 1 a 6 de SEQ ID NO:34,

* a sequência nucleotídica SEQ ID NO:79 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 27 de SEQ ID NO:35, a referida sequência nucleotídica SEQ ID NO:79 codificando a sequência polipeptídica SEQ ID NO:80 correspondendo à região CT de hST8Sia III delimitada pelos aminoácidos localizados nas posições 1 a 9 de SEQ ID NO:36,

* a sequência nucleotídica SEQ ID NO:81 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 21 de SEQ ID NO:37, a referida sequência nucleotídica SEQ ID NO:81 codificando a sequência polipeptídica SEQ ID NO:82 correspondendo à região CT de hST8Sia IV delimitada pelos aminoácidos localizados nas posições 1 a 7 de SEQ ID NO:38,

* a sequência nucleotídica SEQ ID NO:83 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 51 de SEQ ID NO:39, a referida sequência nucleotídica SEQ ID NO:83 codificando a sequência polipeptídica SEQ ID NO:84 correspondendo à região CT de hST8Sia V delimitada pelos aminoácidos localizados nas posições 1 a 17 de SEQ ID NO:40,

* a sequência nucleotídica SEQ ID NO:85 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 9 de SEQ ID NO:41, a referida sequência nucleotídica SEQ ID NO:85 codificando a sequência polipeptídica SEQ ID NO:86 correspondendo à região CT de hST8Sia VI delimitada pelos aminoácidos localizados nas posições 1 a 3 de SEQ ID NO:42,

- a sequência nucleotídica codificando a região TMD compreendida no segundo ácido nucleico ser escolhida de entre:

* a sequência nucleotídica SEQ ID NO:87 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 28 e 78 de SEQ ID NO:1, a referida sequência nucleotídica SEQ ID NO:87 codificando a sequência polipeptídica SEQ ID NO:88 correspondendo à região TMD de hST6Gal I delimitada pelos aminoácidos localizados nas posições 10 a 26 de SEQ ID NO:2,

* a sequência nucleotídica SEQ ID NO:89 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 31 e 90 de SEQ ID NO:3, a referida sequência nucleotídica SEQ ID NO:89 codificando a sequência polipeptídica SEQ ID NO:90 correspondendo à região TMD de hST6Gal II delimitada pelos aminoácidos localizados nas posições 11 a 30 de SEQ ID NO:4,

* a sequência nucleotídica SEQ ID NO:91 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 43 e 105 de SEQ ID NO:5, a referida sequência nucleotídica SEQ ID NO:91 codificando a sequência polipeptídica SEQ ID NO:92 correspondendo à região TMD de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 15 a 35 de SEQ ID NO:6,

* a sequência nucleotídica SEQ ID NO:93 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 22 e 84 de SEQ ID NO:7, a referida sequência nucleotídica SEQ ID NO:93 codificando a

sequência polipeptídica SEQ ID NO:94 correspondendo à região TMD de hST6GalNAc II delimitada pelos aminoácidos localizados nas posições 8 a 28 de SEQ ID NO:8,

* a sequência nucleotídica SEQ ID NO:95 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 25 e 84 de SEQ ID NO:9, a referida sequência nucleotídica SEQ ID NO:95 codificando a sequência polipeptídica SEQ ID NO:96 correspondendo à região TMD de hST6GalNAc III delimitada pelos aminoácidos localizados nas posições 9 a 28 de SEQ ID NO:10,

* a sequência nucleotídica SEQ ID NO:97 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 19 e 81 de SEQ ID NO:11, a referida sequência nucleotídica SEQ ID NO:97 codificando a sequência polipeptídica SEQ ID NO:98 correspondendo à região TMD de hST6GalNAc IV delimitada pelos aminoácidos localizados nas posições 7 a 27 de SEQ ID NO:12,

* a sequência nucleotídica SEQ ID NO:99 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 25 e 87 de SEQ ID NO:13, a referida sequência nucleotídica SEQ ID NO:99 codificando a sequência polipeptídica SEQ ID NO:100 correspondendo à região TMD de hST6GalNAc V delimitada pelos aminoácidos localizados nas posições 9 a 29 de SEQ ID NO:14,

* a sequência nucleotídica SEQ ID NO:101 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 130 e 177 de SEQ ID NO:15, a referida sequência nucleotídica SEQ ID NO:101 codificando a sequência polipeptídica SEQ ID NO:102 correspondendo à região TMD de hST6GalNAc VI delimitada pelos aminoácidos localizados nas posições 44 a 59 de SEQ ID NO:16,

* a sequência nucleotídica SEQ ID NO:103 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 40 e 102 de SEQ ID NO:17, a referida sequência nucleotídica SEQ ID NO:103 codificando a sequência polipeptídica SEQ ID NO:104 correspondendo à região TMD de hST3Gal I delimitada pelos aminoácidos localizados nas posições 14 a 34 de SEQ ID NO:18,

* a sequência nucleotídica SEQ ID NO:105 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 19 e 81 de SEQ ID NO:19, a referida sequência nucleotídica SEQ ID NO:105 codificando a

sequência polipeptídica SEQ ID NO:106 correspondendo à região TMD de hST3Gal II delimitada pelos aminoácidos localizados nas posições 7 a 27 de SEQ ID NO:20,

* a sequência nucleotídica SEQ ID NO:107 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 25 e 84 de SEQ ID NO:21, a referida sequência nucleotídica SEQ ID NO:107 codificando a sequência polipeptídica SEQ ID NO:108 correspondendo à região TMD de hST3Gal III delimitada pelos aminoácidos localizados nas posições 9 a 28 de SEQ ID NO:22,

* a sequência nucleotídica SEQ ID NO:109 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 25 e 78 de SEQ ID NO:23, a referida sequência nucleotídica SEQ ID NO:109 codificando a sequência polipeptídica SEQ ID NO:110 correspondendo à região TMD de hST3Gal IV delimitada pelos aminoácidos localizados nas posições 9 a 26 de SEQ ID NO:24,

* a sequência nucleotídica SEQ ID NO:111 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 16 e 78 de SEQ ID NO:25, a referida sequência nucleotídica SEQ ID NO:111 codificando a sequência polipeptídica SEQ ID NO:112 correspondendo à região TMD de hST3Gal V delimitada pelos aminoácidos localizados nas posições 6 a 26 de SEQ ID NO:26,

* a sequência nucleotídica SEQ ID NO:113 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 13 e 75 de SEQ ID NO:27, a referida sequência nucleotídica SEQ ID NO:113 codificando a sequência polipeptídica SEQ ID NO:114 correspondendo à região TMD de hST3Gal VI delimitada pelos aminoácidos localizados nas posições 5 a 25 de SEQ ID NO:28,

* a sequência nucleotídica SEQ ID NO:115 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 25 e 84 de SEQ ID NO:29, a referida sequência nucleotídica SEQ ID NO:115 codificando a sequência polipeptídica SEQ ID NO:116 correspondendo à região TMD de RatST3Gal III delimitada pelos aminoácidos localizados nas posições 9 a 28 de SEQ ID NO:30,

* a sequência nucleotídica SEQ ID NO:117 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 88 e 144 de SEQ ID NO:31, a referida sequência nucleotídica SEQ ID NO:117

codificando a sequência polipeptídica SEQ ID NO:118 correspondendo à região TMD de hST8Sia I delimitada pelos aminoácidos localizados nas posições 30 a 48 de SEQ ID NO:32,

* a sequência nucleotídica SEQ ID NO:119 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 19 e 69 de SEQ ID NO:33, a referida sequência nucleotídica SEQ ID NO:119 codificando a sequência polipeptídica SEQ ID NO:120 correspondendo à região TMD de hST8Sia II delimitada pelos aminoácidos localizados nas posições 7 a 23 de SEQ ID NO:34,

* a sequência nucleotídica SEQ ID NO:121 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 28 e 99 de SEQ ID NO:35, a referida sequência nucleotídica SEQ ID NO:121 codificando a sequência polipeptídica SEQ ID NO:122 correspondendo à região TMD de hST8Sia III delimitada pelos aminoácidos localizados nas posições 10 a 33 de SEQ ID NO:36,

* a sequência nucleotídica SEQ ID NO:123 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 22 e 60 de SEQ ID NO:37, a referida sequência nucleotídica SEQ ID NO:123 codificando a sequência polipeptídica SEQ ID NO:124 correspondendo à região TMD de hST8Sia IV delimitada pelos aminoácidos localizados nas posições 8 a 20 de SEQ ID NO:38,

* a sequência nucleotídica SEQ ID NO:125 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 52 e 114 de SEQ ID NO:39, a referida sequência nucleotídica SEQ ID NO:125 codificando a sequência polipeptídica SEQ ID NO:126 correspondendo à região TMD de hST8Sia V delimitada pelos aminoácidos localizados nas posições 18 a 38 de SEQ ID NO:40,

* a sequência nucleotídica SEQ ID NO:127 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 10 e 72 de SEQ ID NO:41, a referida sequência nucleotídica SEQ ID NO:127 codificando a sequência polipeptídica SEQ ID NO:128 correspondendo à região TMD de hST8Sia VI delimitada pelos aminoácidos localizados nas posições 4 a 24 de SEQ ID NO:42,

- a sequência nucleotídica codificando a região SR compreendida no segundo ácido nucleico ou codificando um fragmento de pelo menos 2 aminoácidos desta, ser escolhida de entre sequências constituídas por, ou compreendendo:

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 79 e na sua extremidade 3' pelo nucleótido localizado na posição 267 a 327 de SEQ ID NO:1, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST6Gal I delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 27 e na sua extremidade C-terminal pelo aminoácido localizado na posição 89 a 109 de SEQ ID NO:2,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 91 e na sua extremidade 3' pelo nucleótido localizado na posição 306 a 336 de SEQ ID NO:3, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST6Gal II delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 31 e na sua extremidade C-terminal pelo aminoácido localizado na posição 102 a 112 de SEQ ID NO:4,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 106 e na sua extremidade 3' pelo nucleótido localizado na posição 813 a 873 de SEQ ID NO:5, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST6GalNAc I delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 36 e na sua extremidade C-terminal pelo aminoácido localizado na posição 271 a 291 de SEQ ID NO:6,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 85 e na sua extremidade 3' pelo nucleótido localizado na posição 171 a 231 de SEQ ID NO:7, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST6GalNAc II delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 29 e na sua extremidade C-terminal pelo aminoácido localizado na posição 57 a 77 de SEQ ID NO:8,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 85 e na sua extremidade 3' pelo nucleótido localizado na posição 102 a 132 de SEQ ID NO:9, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST6GalNAc III delimitada na sua extremidade N-terminal

pelo aminoácido localizado na posição 29 e na sua extremidade C-terminal pelo aminoácido localizado na posição 34 a 44 de SEQ ID NO:10,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 82 e na sua extremidade 3' pelo nucleótido localizado na posição 90 a 120 de SEQ ID NO:11, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST6GalNAc IV delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 28 e na sua extremidade C-terminal pelo aminoácido localizado na posição 30 a 40 de SEQ ID NO:12,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 88 e na sua extremidade 3' pelo nucleótido localizado na posição 120 a 180 de SEQ ID NO:13, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST6GalNAc V delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 30 e na sua extremidade C-terminal pelo aminoácido localizado na posição 40 a 60 de SEQ ID NO:14,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 178 e na sua extremidade 3' pelo nucleótido localizado na posição 183 de SEQ ID NO:15, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST6GalNAc VI delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 60 e na sua extremidade C-terminal pelo aminoácido localizado na posição 61 de SEQ ID NO:16,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 103 e na sua extremidade 3' pelo nucleótido localizado na posição 144 a 204 de SEQ ID NO:17, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST3Gal I delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 35 e na sua extremidade C-terminal pelo aminoácido localizado na posição 48 a 68 de SEQ ID NO:18,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 82 e na sua extremidade 3' pelo nucleótido localizado na posição 174

a 234 de SEQ ID NO:19, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST3Gal II delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 28 e na sua extremidade C-terminal pelo aminoácido localizado na posição 58 a 78 de SEQ ID NO:20,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 85 e na sua extremidade 3' pelo nucleótido localizado na posição 198 a 258 de SEQ ID NO:21, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST3Gal III delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 29 e na sua extremidade C-terminal pelo aminoácido localizado na posição 66 a 86 de SEQ ID NO:22,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 79 e na sua extremidade 3' pelo nucleótido localizado na posição 108 a 138 de SEQ ID NO:23, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST3Gal IV delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 27 e na sua extremidade C-terminal pelo aminoácido localizado na posição 36 a 46 de SEQ ID NO:24,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 79 e na sua extremidade 3' pelo nucleótido localizado na posição 135 a 195 de SEQ ID NO:25, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST3Gal V delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 27 e na sua extremidade C-terminal pelo aminoácido localizado na posição 45 a 65 de SEQ ID NO:26,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 76 e na sua extremidade 3' pelo nucleótido localizado na posição 102 a 132 de SEQ ID NO:27, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST3Gal VI delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 26 e na sua extremidade C-terminal pelo aminoácido localizado na posição 34 a 44 de SEQ ID NO:28,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 85 e na sua extremidade 3' pelo nucleótido localizado na posição 105 a 165 de SEQ ID NO:29, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de rST3Gal III delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 29 e na sua extremidade C-terminal pelo aminoácido localizado na posição 35 a 55 de SEQ ID NO:30,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 145 e na sua extremidade 3' pelo nucleótido localizado na posição 162 a 192 de SEQ ID NO:31, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST8Sia I delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 49 e na sua extremidade C-terminal pelo aminoácido localizado na posição 54 a 64 de SEQ ID NO:32,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 70 e na sua extremidade 3' pelo nucleótido localizado na posição 189 a 249 de SEQ ID NO:33, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST8Sia II delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 24 e na sua extremidade C-terminal pelo aminoácido localizado na posição 63 a 83 de SEQ ID NO:34,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 100 e na sua extremidade 3' pelo nucleótido localizado na posição 204 a 264 de SEQ ID NO:35, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST8Sia III delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 34 e na sua extremidade C-terminal pelo aminoácido localizado na posição 68 a 88 de SEQ ID NO:36,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 61 e na sua extremidade 3' pelo nucleótido localizado na posição 144 a 204 de SEQ ID NO:37, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST8Sia IV delimitada na sua extremidade N-terminal

pelo aminoácido localizado na posição 21 e na sua extremidade C-terminal pelo aminoácido localizado na posição 48 a 68 de SEQ ID NO:38,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 115 e na sua extremidade 3' pelo nucleótido localizado na posição 210 a 270 de SEQ ID NO:39, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST8Sia V delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 39 e na sua extremidade C-terminal pelo aminoácido localizado na posição 70 a 90 de SEQ ID NO:40,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 73 e na sua extremidade 3' pelo nucleótido localizado na posição 276 a 336 de SEQ ID NO:41, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST8Sia VI delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 25 e na sua extremidade C-terminal pelo aminoácido localizado na posição 92 a 112 de SEQ ID NO:42,

* ou qualquer fragmento de pelo menos 6 nucleótidos das sequências nucleotídicas codificando sequências polipeptídicas correspondendo às regiões SR mencionadas acima, e codificando pelo menos 2 aminoácidos contíguos das referidas regiões SR, tais como:

** o fragmento de SEQ ID NO:129 delimitado pelos nucleótidos localizados nas posições 106 a 222 de SEQ ID NO:5, codificando a sequência polipeptídica SEQ ID NO:130 correspondendo ao fragmento da região SR de hST6GalNAc I delimitado pelos aminoácidos localizados nas posições 36 a 74 de SEQ ID NO:6,

** o fragmento de SEQ ID NO:131 delimitado pelos nucleótidos localizados nas posições 109 a 222 de SEQ ID NO:5, codificando a sequência polipeptídica SEQ ID NO:132 correspondendo ao fragmento da região SR de hST6GalNAc I delimitado pelos aminoácidos localizados nas posições 37 a 74 de SEQ ID NO:6,

** o fragmento de SEQ ID NO:133 delimitado pelos nucleótidos localizados nas posições 106 a 420 de SEQ ID NO:5, codificando a sequência polipeptídica SEQ ID NO:134 correspondendo ao fragmento da região SR de hST6GalNAc I

delimitado pelos aminoácidos localizados nas posições 36 a 140 de SEQ ID NO:6,

** o fragmento de SEQ ID NO:135 delimitado pelos nucleótidos localizados nas posições 106 a 774 de SEQ ID NO:5, codificando a sequência polipeptídica SEQ ID NO:136 correspondendo ao fragmento da região SR de hST6GalNAc I delimitado pelos aminoácidos localizados nas posições 36 a 258 de SEQ ID NO:6,

** o fragmento de SEQ ID NO:137 delimitado pelos nucleótidos localizados nas posições 85 a 138 de SEQ ID NO:21, codificando a sequência polipeptídica SEQ ID NO:138 correspondendo ao fragmento da região SR de hST3Gal III delimitado pelos aminoácidos localizados nas posições 29 a 46 de SEQ ID NO:22,

** o fragmento de SEQ ID NO:139 delimitado pelos nucleótidos localizados nas posições 85 a 138 de SEQ ID NO:29, codificando a sequência polipeptídica SEQ ID NO:140 correspondendo ao fragmento da região SR de ratST3Gal III delimitado pelos aminoácidos localizados nas posições 29 a 46 de SEQ ID NO:30,

* o fragmento de SEQ ID NO:141 delimitado pelos nucleótidos localizados nas posições 70 a 237 de SEQ ID NO:33, codificando a sequência polipeptídica SEQ ID NO:142 correspondendo ao fragmento da região SR de hST8Sia II delimitado pelos aminoácidos localizados nas posições 24 a 79 de SEQ ID NO:34,

* o fragmento de SEQ ID NO:143 delimitado pelos nucleótidos localizados nas posições 61 a 201 de SEQ ID NO:37, codificando a sequência polipeptídica SEQ ID NO:144 correspondendo ao fragmento da região SR de hST8Sia IV delimitado pelos aminoácidos localizados nas posições 21 a 67 de SEQ ID NO:38.

O invento refere-se mais particularmente a um processo tal como definido acima, caracterizado por:

- a sequência nucleotídica codificando a região CT compreendida no segundo ácido nucleico ser escolhida de entre:

* a sequência nucleotídica SEQ ID NO:49 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 42 de SEQ ID NO:5, a referida sequência nucleotídica SEQ ID NO:49 codificando a

sequência polipeptídica SEQ ID NO:50 correspondendo à região CT de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 1 a 14 de SEQ ID NO:6,

* a sequência nucleotídica SEQ ID NO:65 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 24 de SEQ ID NO:21, a referida sequência nucleotídica SEQ ID NO:65 codificando a sequência polipeptídica SEQ ID NO:66 correspondendo à região CT de hST3Gal III delimitada pelos aminoácidos localizados nas posições 1 a 8 de SEQ ID NO:22,

* a sequência nucleotídica SEQ ID NO:73 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 24 de SEQ ID NO:29, a referida sequência nucleotídica SEQ ID NO:73 codificando a sequência polipeptídica SEQ ID NO:74 correspondendo à região CT de ratST3Gal III delimitada pelos aminoácidos localizados nas posições 1 a 8 de SEQ ID NO:30,

* a sequência nucleotídica SEQ ID NO:77 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 18 de SEQ ID NO:33, a referida sequência nucleotídica SEQ ID NO:77 codificando a sequência polipeptídica SEQ ID NO:78 correspondendo à região CT de hST8Sia II delimitada pelos aminoácidos localizados nas posições 1 a 6 de SEQ ID NO:34,

* a sequência nucleotídica SEQ ID NO:81 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 21 de SEQ ID NO:37, a referida sequência nucleotídica SEQ ID NO:81 codificando a sequência polipeptídica SEQ ID NO:82 correspondendo à região CT de hST8Sia IV delimitada pelos aminoácidos localizados nas posições 1 a 7 de SEQ ID NO:38,

- a sequência nucleotídica codificando a região TMD compreendida no segundo ácido nucleico ser escolhida de entre:

* a sequência nucleotídica SEQ ID NO:91 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 43 e 105 de SEQ ID NO:5, a referida sequência nucleotídica SEQ ID NO:91 codificando a sequência polipeptídica SEQ ID NO:92 correspondendo à região TMD de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 15 a 35 de SEQ ID NO:6,

* a sequência nucleotídica SEQ ID NO:107 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 25 e 84 de SEQ ID NO:21, a referida sequência nucleotídica SEQ ID NO:107 codificando a sequência polipeptídica SEQ ID NO:108 correspondendo à região TMD de hST3Gal III delimitada pelos aminoácidos localizados nas posições 9 a 28 de SEQ ID NO:22,

* a sequência nucleotídica SEQ ID NO:115 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 25 e 84 de SEQ ID NO:29, a referida sequência nucleotídica SEQ ID NO:115 codificando a sequência polipeptídica SEQ ID NO:116 correspondendo à região TMD de rST3Gal III delimitada pelos aminoácidos localizados nas posições 9 a 28 de SEQ ID NO:30,

* a sequência nucleotídica SEQ ID NO:119 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 19 e 69 de SEQ ID NO:33, a referida sequência nucleotídica SEQ ID NO:119 codificando a sequência polipeptídica SEQ ID NO:120 correspondendo à região TMD de hST8Sia II delimitada pelos aminoácidos localizados nas posições 7 a 23 de SEQ ID NO:34,

* a sequência nucleotídica SEQ ID NO:123 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 22 e 60 de SEQ ID NO:37, a referida sequência nucleotídica SEQ ID NO:123 codificando a sequência polipeptídica SEQ ID NO:124 correspondendo à região TMD de hST8Sia IV delimitada pelos aminoácidos localizados nas posições 8 a 20 de SEQ ID NO:38,

- a sequência nucleotídica codificando a região SR ou um fragmento desta compreendida no segundo ácido nucleico, ser escolhida de entre:

* a sequência SEQ ID NO:129 delimitada pelos nucleótidos localizados nas posições 106 a 222 de SEQ ID NO:5, codificando a sequência polipeptídica SEQ ID NO:130 correspondendo ao fragmento da região SR de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 36 a 74 de SEQ ID NO:6,

* a sequência SEQ ID NO:131 delimitada pelos nucleótidos localizados nas posições 109 a 222 de SEQ ID NO:5, codificando a sequência polipeptídica SEQ ID NO:132 correspondendo ao fragmento da região SR de hST6GalNAc I

delimitada pelos aminoácidos localizados nas posições 37 a 74 de SEQ ID NO:6,

* a sequência SEQ ID NO:133 delimitada pelos nucleótidos localizados nas posições 106 a 420 de SEQ ID NO:5, codificando a sequência polipeptídica SEQ ID NO:134 correspondendo ao fragmento da região SR de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 36 a 140 de SEQ ID NO:6,

* a sequência SEQ ID NO:135 delimitada pelos nucleótidos localizados nas posições 106 a 774 de SEQ ID NO:5, codificando a sequência polipeptídica SEQ ID NO:136 correspondendo ao fragmento da região SR de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 36 a 258 de SEQ ID NO:6,

* a sequência SEQ ID NO:137 delimitada pelos nucleótidos localizados nas posições 85 a 138 de SEQ ID NO:21, codificando a sequência polipeptídica SEQ ID NO:138 correspondendo ao fragmento da região SR de hST3Gal III delimitada pelos aminoácidos localizados nas posições 29 a 46 de SEQ ID NO:22,

* a sequência SEQ ID NO:139 delimitada pelos nucleótidos localizados nas posições 85 a 138 de SEQ ID NO:29, codificando a sequência polipeptídica SEQ ID NO:140 correspondendo ao fragmento da região SR de ratST3Gal III delimitada pelos aminoácidos localizados nas posições 29 a 46 de SEQ ID NO:30,

* a sequência SEQ ID NO:141 delimitada pelos nucleótidos localizados nas posições 70 a 237 de SEQ ID NO:33, codificando a sequência polipeptídica SEQ ID NO:142 correspondendo ao fragmento da região SR de hST8Sia II delimitado pelos aminoácidos localizados nas posições 24 a 79 de SEQ ID NO:34,

* a sequência SEQ ID NO:143 delimitada pelos nucleótidos localizados nas posições 61 a 201 de SEQ ID NO:37, codificando a sequência polipeptídica SEQ ID NO:144 correspondendo ao fragmento da região SR de hST8Sia IV delimitado pelos aminoácidos localizados nas posições 21 a 67 de SEQ ID NO:38.

O invento refere-se mais particularmente a um processo tal como definido acima, caracterizado por os péptidos CT, TMD, SR ou fragmento de SR compreendidos no segundo ácido

nucleico, serem sequências homólogas derivadas da mesma glicosiltransferase nativa, sendo esta última diferente dos péptidos na glicosiltransferase nativa da qual é derivado o fragmento de CD com actividade de glicosiltransferase óptima tal como definido acima.

O invento refere-se mais particularmente a um processo tal como definido acima, caracterizado por o segundo ácido nucleico ser escolhido de entre as seguintes sequências:

- a sequência SEQ ID NO:145 delimitada pelos nucleótidos localizados nas posições 1 a 222 de SEQ ID NO:5, contendo SEQ ID NO:49, 91 e 129, e codificando a sequência polipeptídica SEQ ID NO:146 correspondendo ao fragmento de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 1 a 74 de SEQ ID NO:6, e contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 130, respectivamente,

- a sequência SEQ ID NO:147 delimitada pelos nucleótidos localizados nas posições 1 a 420 de SEQ ID NO:5, contendo SEQ ID NO:49, 91 e 133, e codificando a sequência polipeptídica SEQ ID NO:148 correspondendo ao fragmento da região SR de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 1 a 140 de SEQ ID NO:6, e contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 134, respectivamente,

- sequência SEQ ID NO:149 delimitada pelos nucleótidos localizados nas posições 1 a 774 de SEQ ID NO:5, contendo SEQ ID NO:49, 91 e 135, e codificando a sequência polipeptídica SEQ ID NO:150 correspondendo ao fragmento da região SR de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 1 a 258 de SEQ ID NO:6, e contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 136, respectivamente,

- a sequência SEQ ID NO:151 delimitada pelos nucleótidos localizados nas posições 1 a 138 de SEQ ID NO:21, contendo SEQ ID NO:65, 107 e 137, e codificando a sequência polipeptídica SEQ ID NO:152 correspondendo ao fragmento da região SR de hST3Gal III delimitada pelos aminoácidos localizados nas posições 1 a 46 de SEQ ID NO:22, e contendo as regiões CT, TMD e fragmento de SR de hST3Gal III correspondendo a SEQ ID NO:66, 108 e 138, respectivamente,

- a sequência SEQ ID NO:153 delimitada pelos nucleótidos localizados nas posições 1 a 138 de SEQ ID NO:29, contendo SEQ ID NO:73, 115 e 139, e codificando a sequência polipeptídica SEQ ID NO:154 correspondendo ao fragmento da região SR de rST3Gal III delimitada pelos aminoácidos localizados nas posições 1 a 46 de SEQ ID NO:30, e contendo as regiões CT, TMD e fragmento de SR de hST3Gal III correspondendo a SEQ ID NO:74, 116 e 140, respectivamente,

- a sequência SEQ ID NO:155 delimitada pelos nucleótidos localizados nas posições 1 a 237 de SEQ ID NO:33, contendo SEQ ID NO:77, 119 e 141, e codificando a sequência polipeptídica SEQ ID NO:156 correspondendo ao fragmento da região SR de hST8Sia II delimitada pelos aminoácidos localizados nas posições 1 a 79 de SEQ ID NO:34, e contendo as regiões CT, TMD e SR de hST8Sia II correspondendo a SEQ ID NO:78, 120 e 142, respectivamente,

- a sequência SEQ ID NO:157 delimitada pelos nucleótidos localizados nas posições 1 a 201 de SEQ ID NO:37, contendo SEQ ID NO:81, 123 e 143, e codificando a sequência polipeptídica SEQ ID NO:158 correspondendo ao fragmento da região SR de hST8Sia IV delimitada pelos aminoácidos localizados nas posições 1 a 67 de SEQ ID NO:38, e contendo as regiões CT, TMD e SR de hST8Sia IV correspondendo a SEQ ID NO:82, 124 e 144, respectivamente.

O invento refere-se mais particularmente a um processo tal como definido acima, caracterizado por os péptidos CT, TMD, SR ou fragmento de SR compreendidos no segundo ácido nucleico, serem sequências heterólogas derivando de um gene ou transcrito de uma glicosiltransferase natural diferente.

O invento refere-se mais particularmente a um processo tal como definido acima, caracterizado por o segundo ácido nucleico ser a sequência SEQ ID NO:159 correspondendo à fusão da sequência nucleotídica SEQ ID NO:177 contendo SEQ ID NO:65 e 107 codificando as regiões CT e TMD de hST3Gal III correspondendo a SEQ ID NO:66 e 108 respectivamente, com a sequência nucleotídica SEQ ID NO:131 codificando a sequência polipeptídica SEQ ID NO:132 correspondendo ao fragmento da região SR de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 37 a 74 de SEQ ID NO:6, a referida sequência SEQ ID NO:159 codificando o polipéptido de fusão

SEQ ID NO:160 entre as regiões CT e TMD de hST3Gal III, por um lado e o fragmento 37-74 da região SR de hST6GalNAc I, por outro.

O invento refere-se mais particularmente a um processo tal como definido acima, caracterizado por o segundo ácido nucleico ser a sequência SEQ ID NO:179 correspondendo à fusão da sequência nucleotídica SEQ ID NO:65 codificando o CT de hST3Gal III correspondendo a SEQ ID NO:66, com a sequência nucleotídica SEQ ID NO:119 codificando a região TMD de hST8Sia II correspondendo a SEQ ID NO:120, e com a sequência nucleotídica SEQ ID NO:129 codificando a sequência polipeptídica SEQ ID NO:130 correspondendo ao fragmento da região SR de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 36 a 74 de SEQ ID NO:6, a referida sequência SEQ ID NO:179 codificando o polipéptido de fusão SEQ ID NO:180 entre a região CT de hST3Gal III, a região TMD de hST8Sia II e o fragmento 36-74 da região SR de hST6GalNAc I.

O invento refere-se mais particularmente a um processo tal como definido acima, caracterizado por compreender a fusão da sequência SEQ ID NO:43 como primeiro ácido nucleico, com o segundo ácido nucleico escolhido de entre:

- a sequência SEQ ID NO:145, conduzindo à sequência nucleotídica SEQ ID NO:161 codificando a proteína de fusão SEQ ID NO:162 contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 130, respectivamente, ligadas através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I,

- a sequência SEQ ID NO:147, conduzindo à sequência nucleotídica SEQ ID NO:163 codificando a proteína de fusão SEQ ID NO:164 contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 134, respectivamente, ligadas através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I,

- a sequência SEQ ID NO:149, conduzindo à sequência nucleotídica SEQ ID NO:165 codificando a proteína de fusão SEQ ID NO:166 contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 136,

respectivamente, ligadas através de um ligador SR ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I,

- a sequência SEQ ID NO:151, conduzindo à sequência nucleotídica SEQ ID NO:167 codificando a proteína de fusão SEQ ID NO:168 contendo as regiões CT, TMD e fragmento de SR de hST3Gal III correspondendo a SEQ ID NO:66, 108 e 138, respectivamente, ligadas através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I,

- a sequência SEQ ID NO:153, conduzindo à sequência nucleotídica SEQ ID NO:169 codificando a proteína de fusão SEQ ID NO:170 contendo as regiões CT, TMD e fragmento de SR de ratST3Gal III correspondendo a SEQ ID NO:74, 116 e 140, respectivamente, ligadas através de um ligador SR ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I,

- a sequência SEQ ID NO:155, conduzindo à sequência nucleotídica SEQ ID NO:171 codificando a proteína de fusão SEQ ID NO:172 contendo as regiões CT, TMD e SR de hST8Sia II correspondendo a SEQ ID NO:78, 120 e 142, respectivamente, ligadas através de um ligador KL ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I,

- a sequência SEQ ID NO:157, conduzindo à sequência nucleotídica SEQ ID NO:173 codificando a proteína de fusão SEQ ID NO:174 contendo as regiões CT, TMD e SR de hST8Sia IV correspondendo a SEQ ID NO:82, 124 e 144, respectivamente, ligadas através de um ligador KL ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I,

- a sequência SEQ ID NO:159, conduzindo à sequência nucleotídica SEQ ID NO:175 codificando a proteína de fusão SEQ ID NO:176 contendo as regiões CT e TMD de hST3Gal III correspondendo a SEQ ID NO:66 e 108, respectivamente, e ao fragmento 37-74 da região SR de hST6GalNac I correspondendo a SEQ ID NO:132, ligadas através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I,

- a sequência SEQ ID NO:179, conduzindo à sequência nucleotídica SEQ ID NO:181 codificando a proteína de fusão SEQ ID NO:182 contendo a região CT de hST3Gal III, a região TMD de hST8Sia II e o fragmento 36-74 da região SR de hST6GalNac I, correspondendo a SEQ ID NO:66, 120 e 130 respectivamente, ligadas através de um ligador GS ao

fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I.

O invento refere-se também a sequências génicas codificando glicosiltransferases quiméricas tais como obtidas de acordo com o processo tal como definido acima.

O invento refere-se mais particularmente a sequências génicas tal como definidas acima, escolhidas de entre:

- a sequência SEQ ID NO:161 codificando a proteína de fusão SEQ ID NO:162 contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 130, respectivamente, ligadas através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I, a referida sequência SEQ ID NO:161 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:145,

- a sequência SEQ ID NO:163 codificando a proteína de fusão SEQ ID NO:164 contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 134, respectivamente, ligadas através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I, a referida sequência SEQ ID NO:163 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:147,

- a sequência SEQ ID NO:165 codificando a proteína de fusão SEQ ID NO:166 contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 136, respectivamente, ligadas através de um ligador SR ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I, a referida sequência SEQ ID NO:165 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:149,

- a sequência SEQ ID NO:167 codificando a proteína de fusão SEQ ID NO:168 contendo as regiões CT, TMD e fragmento de SR de hST3Gal III correspondendo a SEQ ID NO:66, 108 e 138, respectivamente, ligadas através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I, a referida sequência SEQ ID NO:167 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:151,

- a sequência SEQ ID NO:169 codificando a proteína de fusão SEQ ID NO:170 contendo as regiões CT, TMD e fragmento de SR de ratST3Gal III correspondendo a SEQ ID NO:74, 116 e 140, respectivamente, ligadas através de um ligador SR ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal

I, a referida sequência SEQ ID NO:169 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:153,

- a sequência SEQ ID NO:171 codificando a proteína de fusão SEQ ID NO:172 contendo as regiões CT, TMD e SR de hST8Sia II correspondendo a SEQ ID NO:78, 120 e 142, respectivamente, ligadas através de um ligador KL ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I, a referida sequência SEQ ID NO:171 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:155,

- a sequência SEQ ID NO:173 codificando a proteína de fusão SEQ ID NO:174 contendo as regiões CT, TMD e SR de hST8Sia IV correspondendo a SEQ ID NO:82, 124 e 144, respectivamente, ligadas através de um ligador KL ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I, a referida sequência SEQ ID NO:173 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:157,

- a sequência SEQ ID NO:175 codificando a proteína de fusão SEQ ID NO:176 contendo as regiões CT e TMD de hST3Gal III correspondendo a SEQ ID NO:66 e 108, respectivamente, e o fragmento 37-74 da região SR de hST6GalNAc I correspondendo a SEQ ID NO:132, ligadas através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I, a referida sequência SEQ ID NO:175 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:159,

- a sequência SEQ ID NO:181 codificando a proteína de fusão SEQ ID NO:182 contendo a região CT de hST3Gal III, a região TMD de hST8Sia II e o fragmento 36-74 da região SR de hST6GalNAc I, correspondendo a SEQ ID NO:66, 120 e 130, respectivamente, ligadas através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I, a referida sequência SEQ ID NO:181 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:179.

O invento refere-se também a vectores, tais como plasmídeos, construções virais ou bacterianas, compreendendo pelo menos uma sequência génica tal como definida acima.

O invento refere-se também a células hospedeiras eucarióticas de origem em levedura, fungos, insecto, plantas, mamífero ou humano, transformadas com pelo menos uma sequência génica tal como definida acima, utilizando pelo menos um vector tal como mencionado acima.

O invento refere-se também à utilização de pelo menos uma sequência génica tal como definida acima ou de um vector tal como mencionado acima, para a transformação de células tal como definido acima ou à utilização de animais transgénicos obtidos a partir de tais células transformadas, no quadro da produção de proteínas recombinantes de interesse.

O invento refere-se também a um método para a preparação de proteínas recombinantes de interesse compreendendo a transformação de células tal como definido acima, com um vector contendo pelo menos uma sequência nucleotídica codificando as referidas proteínas recombinantes de interesse.

As proteínas recombinantes de interesse preferidas que podem ser preparadas de acordo com um método tal como mencionado acima de acordo com o invento são escolhidas de entre hormonas, enzimas, factores de coagulação, antigénios hidratos de carbono/biomarcadores séricos, citocinas, factores de crescimento, anticorpos ou receptores.

As células hospedeiras preferidas para a preparação de proteínas recombinantes de interesse tal como mencionado acima, são escolhidas de entre células ou organismos aprovados para fármacos, de preferência células de roedor, mamífero ou humanas.

DESCRIÇÃO DAS FIGURAS

A **Figura 1** representa a topologia membranar das glicosiltransferases. AS glicosiltransferases são proteínas membranares de tipo II incluindo uma cauda citoplasmática N-terminal (**CT**), um sinal de âncora transmembranar (**TMD**) seguido de uma região de caule (**SR**) e um grande domínio catalítico (**CD**) C-terminal.

Figura 2 Sequência de aminoácidos do motivo sialil L, S e VS em 20 sialiltransferases humanas. Os resíduos de aminoácidos de consenso em todas as sialiltransferases são mostrados a negrito.

A **Figura 3** representa a distribuição esquemática de CT, TMD, SR e CD da ST6Gal I de rato mostrando resíduos chave que são tyr123 e 7 cys conservados. CT: cauda citoplasmática (9 aminoácidos); TMD: domínio transmembranar (17 aa); SR: região de caule (70 aa); CD: domínio catalítico (307 aa); L: motivo sialil L; S: motivo sialil L; e VS: motivo sialil VS.

A **Figura 4** representa construções de ADN para produzir quimeras e avaliar o papel da cauda citoplasmática, o domínio transmembranar em duas isoformas enzimáticas (rST6Gal I Tyr ou Cys 123), respectivamente (Fenteany & Colley, 2005).

A **Figura 5** representa alinhamento de sequências da ST6Gal I de rato e da humana. Pode observar-se que a sequência N-terminal compreendendo o CT(1-9), TMD (10-26) é completamente conservada enquanto que a porção (27-89) justamembranar de SR é mais variável embora o péptido justamembranar e especialmente positivamente os resíduos de lisina e cisteína são bem conservados.

A **Figura 6** representa a via do ácido siálico em células humanas no contexto da glicosilação celular global. As enzimas envolvidas neste processo sequencial são: (a e b) UDP-N-acetilglucosamina/2-epimerase/N-acetilmanosamina-cinase (UDP-GlcNAc 2-epimerase/ManNAc 6-cinase) - Reacções: (a) epimerase e (b) cinase, (c) ácido N-acetilneuramínico-fosfato-sintetase; (NANS SAS; ácido N-acetilneuramínico-fosfato-sintetase) - Reacção e N-acetilneuraminato-piruvato-liase de *Homo sapiens* (NPL; N-acetilneuraminato-piruvato-liase) - Reacção, (d) NeuAc 9-fosfatase, (e) Citidina-5'-monofosfato-ácido N-acetilneuramínico-sintetase, (CMP-NeuAc-sintetase), (f) transportador de citidina-monofosfato-ácido siálico (transportador de CMP-NeuAc no Golgi) e (g) sialiltransferases.

De nota, o derivado de ácido siálico mostrado neste diagrama, designadamente "Neu5Ac," é amplamente considerado como a forma "humana" do ácido siálico. Em todos os outros animais, com excepção das galinhas, existe um passo adicional na via (mostrado neste diagrama) onde CMP-Neu5Ac é ainda hidroxilado e convertido em CMP-Neu5Gc através da acção enzimática de CMP-Ácido N-Acetilneuramínico-Hidroxilase.

A **Figura 7** representa uma perspectiva esquemática das várias construções quiméricas sintéticas geradas através do invento.

A **Figura 7A** representa a construção geral de uma quimera sintética incluindo o domínio sintético N-terminal marcado com o epitopo FLAG fundido com o domínio catalítico (CD) mínimo através da adição de um local de restrição.

A **Figura 7B** mostra a ligação entre o domínio sintético N-terminal e o catalítico utilizando um local de restrição *BamHI*. Observe-se que as construções são introduzidas no vector utilizando locais de restrição, designadamente *AflIII* e *XbaI* distintos do local de ligação.

A **Figura 8** mostra a topologia e característica do vector pcDNA3.1 (+) (Invitrogen).

A **Figura 9** representa as construções do CD de FLAG-hST6Gal I e das formas quiméricas deste CD fundidas com vários fragmentos N-terminais de outras sialiltransferases de comprimento variável.

A Figura 9A corresponde a FLAG-CD.

A Figura 9B corresponde a hST8SiaII-79/CD.

A Figura 9C corresponde a hST8SiaIV-67/CD.

A Figura 9D corresponde a hST6GalNAc 1-258/CD.

A Figura 9E corresponde a r/hST3Gal III -46/CD.

A **Figura 10** representa o domínio catalítico mínimo digerido de hST6Gal I e o seu produto de PCR amplificado a utilizar em cada construção da quimera sintética. As amostras foram carregadas num gel de agarose a 1,5% com o marcador nucleico SmartLadder (SL).

A Figura 10A mostra o domínio catalítico de hST6Gal I digerido a partir do vector CMV, proveniente da clonagem em laboratório.

A Figura 10B mostra o produto de PCR do domínio catalítico mínimo a 966 pb.

A Figura 10C mostra o produto de PCR concentrado do domínio catalítico mínimo de hST6Gal I com 5 µl carregados no gel.

A **Figura 11** mostra um gel de agarose (2%) mostrando a banda de ADN da região N-terminal sintética reconstituída de hST3Gal III. A pista 1 corresponde a um produto de PCR de 174 pb; a pista 2 corresponde ao SmartLadder SF; a pista 3 corresponde a 5 µl do produto de PCR concentrado de 174 pb de tamanho.

A **Figura 12** representa o gene total da enzima reconstruído de hST3Gal III/CD.

Na Figura 12A, um gel de agarose (1,5%) mostra a banda de ADN do hST3Gal III/CD sintético reconstituído amplificado por PCR.

A Figura 12B corresponde a 5 µl dos produtos de PCR concentrados de cerca de 1200 pb de tamanho.

A Figura 12C representa o produto da digestão do vector recombinante pelas enzimas de restrição *AflIII* e *XbaI*, mostrando a inserção do gene da quimera (tamanho esperado de 1200 pb).

A **Figura 13** mostra um gel de agarose (2%) mostrando a banda de ADN da região N-terminal sintética reconstituída de hST6GalNAc 1-74. A pista 1 corresponde a um produto de PCR de 270 pb; a pista 2 corresponde ao SmartLadder SF; a pista 3 corresponde a 5 µl dos produtos de PCR concentrados de 270 pb de tamanho.

A **Figura 14** representa o gene sintético reconstruído completo de hST6GalNAc I-74/CD.

A Figura 14A mostra um gel de agarose (1,5%) com a banda de ADN do hST6GalNAc I-74/CD sintético reconstituído amplificado por PCR.

A Figura 14B mostra 5 µl dos produtos de PCR concentrados da quimera de cerca de 1200 pb de tamanho.

A Figura 14C mostra o produto da digestão do vector recombinante pelas enzimas de restrição *AflIII* e *XbaI*, mostrando a inserção do gene da quimera (tamanho esperado de 1225 pb).

A **Figura 15** mostra um gel de agarose (1,5%) mostrando a banda de ADN de da região N-terminal sintética reconstituída de hST6GalNAc 1-140. Pista 1: SmartLadder SF; pista 2: produto de PCR de 468 pb.

A **Figura 16** mostra um gel de agarose (1,5%) mostrando a banda de ADN do gene híbrido sintético reconstituído composto por hST3Gal III-29/37-hST6GalNAc 1-74. A pista 1 corresponde ao SmartLadder SF; a pista 2 corresponde a um produto de PCR de 249 pb.

A **Figura 17** mostra um gel de agarose (1,5%) mostrando a banda de ADN do gene híbrido sintético reconstituído composto por hST3Gal III-29/37-hST6GalNAc 1-74 ligado ao CD. A pista 1 corresponde ao SmartLadder SF; a pista 2 corresponde a um produto de PCR de 1197 pb.

A **Figura 18** representa a expressão do domínio catalítico (CD) mínimo (Painel A) de hST6Gal I em células CHO, com base numa dupla marcação com ligação de FITC-SNA a ácido siálico ligado em 6 e com um mAb anti-FLAG e utilizando microscopia confocal. A expressão transiente do CD foi analisada através de marcação anti-FLAG (Painel B) e para a ligação de SNA-FITC (Painel C). O painel D mostra a sobreposição do sinal dos painéis B e C (Donadio *et al.*, 2003).

A **Figura 19** mostra a expressão de duas enzimas quiméricas, hST8SiaIV-67/CD, hST8SiaII-79/CD em células CHO. As células transfectadas foram duplamente marcadas com um mAb anti-FLAG (A e D) e com FITC-SNA (B e E). As camadas são representadas em C and F.

A **Figura 20** mostra a expressão de ST3GalIII/CD e ST3GalIII-hST6GalNAc 1-74/CD em células CHO. As células transfectadas foram duplamente marcadas com um mAb anti-FLAG (A e D) e com FITC-SNA (B e E). As camadas são representadas em C e F.

A **Figura 21** mostra a expressão de hST6GalNAc I-74/CD e 258/CD em células CHO. As células transfectadas foram duplamente marcadas com um mAb anti-FLAG (A e D) e com FITC-SNA (B e E). As camadas são representadas em C e F.

A **Figura 22** mostra um gel de agarose (2%) mostrando a banda de ADN da região N-terminal sintética reconstituída de hST3Gal III/hST8Sia II/hST6GalNAc 1. A pista 1 corresponde a

um produto de PCR de 240 pb; a pista 2 corresponde ao SmartLadder SF; a pista 3 corresponde a 3 µl do produto de PCR concentrado de 240 pb.

A **Figura 23** representa o produto da digestão do vector recombinante pelas enzimas de restrição *XbaI* e *BamHI*, mostrando a inserção do gene do domínio catalítico (tamanho esperado de 960 pb). A amostra foi carregada num gel de agarose a 1,5% com o marcador nucleico SL.

A **Figura 24** representa o gene reconstruído da enzima hST3Gal III/hST8Sia II/hST6GalNAc I/CD.

Na Figura 24a, um gel de agarose (1,5%) mostra o produto de digestão do vector recombinante pelas enzimas de restrição *XbaI* e *AflIII*, mostrando a inserção do gene da quimera (tamanho esperado de 1200 pb).

A Figura 24b representa o produto de PCR obtido após a amplificação de uma amostra de minipreparação, de acordo com a inserção do gene da quimera.

A **Figura 25** mostra a expressão da enzima quimérica hST3Gal III/hST8Sia II/hST6GalNAc I/CD em células CHO. As células transfectadas foram duplamente marcadas com um mAb anti-FLAG (A) e com FITC-SNA (B). A camada está representada em C.

LEGENDAS das TABELAS 1 a 5

Tabela 1: Sequências e números de entrada das 20 sialiltransferases conhecidas em humanos.

Tabela 2: Substratos aceitadores e locais de expressão das sialiltransferases humanas tal como descrito em Harduin-Lepers *et al.*, 2001 e Jeanneau *et al.*, 2003.

Tabela 3: Distribuição dos 4 domínios peptídicos partilhados pelas ST humanas. Uma curta cauda citoplasmática N-terminal (**CT**; cerca de 10 aminoácidos), um domínio transmembranar (**TMD**; cerca de 20 aminoácidos), uma região de caule (**SR**), altamente variável em comprimento e um domínio catalítico (**CD**; cerca de 300 aminoácidos), incluindo os motivos sialil L, S e VS.

Tabela 4: Péptidos conservados propostos de cerca de 31-85 aminoácidos identificados dentro da região SR hipervariável. Foram encontrados 5 motivos: motivo A comum a ST6Gal e ST6GalNAc I, motivo B comum a ST6GalNAc I e ST6GalNAc II, motivos C e D comuns a todos os ST8Sia e motivo E presente em ST3Gal excepto ST3Gal I e II (Donadio *et al.*, 2003).

Tabela 5: Desenho dos quatro pares de iniciadores utilizados para amplificação por PCR.

Tabela 1

SIALILTRANSFERASES HUMANAS	Nº de ENTRADA em GenBank	Nº de ENTRADA em Swiss Prot
ST6Gal I	A17362	P15907
ST6Gal II	NM_032528	Q8IUG7
ST6GalNAc I	NM_018414	Q9NSC7
ST6GalNAc II	NM_006456	Q9UJ37
ST6GalNAc III	NM_152996	Q8NDV1
ST6GalNAc IV	NM_014403	Q9H4F1
ST6GalNAc V	NM_030965	Q9BVH7
ST6GalNAc VI	AB035173	Q969X2
ST3Gal I	L29555	Q11201
ST3Gal II	U63090	Q16842
ST3Gal III	L23768	Q11203
ST3Gal IV	L23767	Q11206
ST3Gal V	AF105026	Q9UNP4
ST3Gal VI	AF119391	Q9Y274
ST8Sia I	L32867	Q92185
ST8Sia II	U33551	Q92186
ST8Sia III	AF004668	043173
ST8Sia IV	L41680	Q92187
ST8Sia V	U91641	015466
ST8Sia VI	AJ621583	P61647

Tabela 2

SIALILTRANSFERASES HUMANAS	Aceitador	Expressão
ST6Gal I ST6Gal II	Gal β 1-4GlcNAc Gal β 1-4GlcNAc	Ubíqua Cérebro, testículos, tiróide, tecido fetal, gânglios linfáticos
ST6GalNAc I	Gal β 1-3GalNAc	Glândulas submaxilares e mamárias, baço, cólon
ST6GalNAc II	GalNAc α -O-Ser/Thr Sia α 2-3 Gal β 1- 3GalNAc α -O-Ser/Thr Gal β 1- 3GalNAc α -O- Ser/Thr	Glândulas mamárias lactantes, testículos
ST6GalNAc III	Sia α 2-3Gal β 1- 3GalNAc α -O-Ser/Thr	
ST6GalNAc IV	Sia α 2-3Gal β 1-3GalNAc	Ubíqua
ST6GalNAc V	Sia α 2-3Gal β 1- 3GalNAc, GM1b	
ST6GalNAc VI	Sia α 2-3Gal β 1- 3GalNAc, GM1b GM1b, GT1b	
ST3Gal I ST3Gal II	Gal β 1-3GalNAc Gal β 1-3GalNAc, GM1, asialo-GM1	Ubíqua Coração, fígado, músculo- esquelético, timo, nódulo linfático, apêndice, glândulas salivares, baço
ST3Gal III	Gal β 1-3GlcNAc	Músculo- esquelético, tecido fetal
ST3Gal IV	Gal β 1-4GlcNAc Gal β 1-4GlcNAc-O Gal β 1-3GalNAc-O	Placenta, testículos, ovário

ST3Gal V ST3Gal VI	Gal β 1-4Glc β -Cer Gal β 1-4GlcNAc	Ubíqua Coração, placenta, fígado e a maioria dos outros tecidos
ST8Sia I ST8Sia II ST8Sia III ST8Sia IV ST8Sia V ST8Sia VI	Sia α 2-3Gal β 1-4Glc β 1- O-Cer, GM3 SiaGal β 1-4GalNAc Sia2-3Gal β 1-9GlcNAc SiaGal β 1-4GalNAc GM1b, GD1a, GT1b, GD3 NeuAc α 2,3(6)Gal β	Tecidos embrionários, cérebro Cérebro, tecidos fetais, coração de adulto, baço, timo

Tabela 3

ESTIMATIVA DO COMPRIMENTO DOS DOMÍNIOS								
Sialiltrans- ferases Humanas	Total	Cauda Citoplasmá- tica (CT)	Domínio Transmem- brar (TMD)	Região de Caule (SR; $\pm$ 10 AA)	Domínio Catalí- tico (CD; $\pm$ 10 AA)	Motivo Sialil L (47 AA)	Motivo Sialil S (24 AA)	Motivo Sialil VS (12 AA)
ST6Gal I	406	1-9 (9 AA)	10-26 (17 AA)	27-100 (74 AA)	101-406 (306 AA)	181-227	320-343	366-378
ST6Gal II	529	1-10 (10 AA)	11-30 (20 AA)	31-112 (82 AA)	113-529 (417 AA)	293-339	433-456	479-491
ST6GalNAc I	600	1-14 (14 AA)	15-35 (41 AA)	36-281 (246 AA)	282-600 (319 AA)	362-408	518-541	563-575
ST6GalNAc II	374	1-7 (7 AA)	8-28 (21 AA)	29-67 (39 AA)	68-374 (307 AA)	148-194	302-325	347-359
ST6GalNAc III	305	1-8 (8 AA)	9-28 (20 AA)	29-34 (6 AA)	35-305 (271 AA)	77-123	214-237	273-285
ST6GalNAc IV	302	1-6 (6 AA)	7-27 (21 AA)	28-30 (3 AA)	31-302 (272 AA)	73-119	210-233	270-282
ST6GalNAc V	336	1-8 (8 AA)	9-29 (21 AA)	30-50 (21 AA)	51-336 (286 AA)	93-139	230-253	291-303
ST6GalNAc VI	333	1-43 (43 AA)	44-59 (16 AA)	60-61 (2 AA)	62-333 (272 AA)	105-151	241-264	303-315
ST3Gal I	340	1-13 (13 AA)	14-34 (21 AA)	35-58 (24 AA)	59-340 (282 AA)	139-185	266-289	312-324
ST3Gal II	350	1-6 (6 AA)	7-27 (21 AA)	28-68 (41 AA)	69-350 (282 AA)	149-195	276-299	322-334
ST3Gal III	444	1-8 (8 AA)	9-28 (20 AA)	29-76 (48 AA)	17-444 (368 AA)	157-203	299-322	346-358
ST3Gal IV	329	1-8	9-26	27-36	37-329	117-163	258-281	307-319

		(8 AA)	(18 AA)	(10 AA)	(293 AA)			
ST3Gal V	362	1-5 (5 AA)	6-26 (21 AA)	27-55 (29 AA)	56-362 (307 AA)	136-182	282-305	329-341
ST3Gal VI	331	1-4 (4 AA)	5-25 (21 AA)	26-34 (9 AA)	35-331 (297 AA)	115-161	256-279	303-315
ST8Sia I	356	1-29 (29 AA)	30-48 (19 AA)	49-54 (6 AA)	55-356 (302 AA)	135-182	272-295	318-330
ST8Sia II	375	1-6 (6 AA)	7-23 (17 AA)	24- 73 (50 AA)	74-375 (302 AA)	157-200	292-315	342-354
ST8Sia III	380	1-9 (9 AA)	10-33 (24 AA)	34-78 (45 AA)	79-380 (302 AA)	159-205	298-321	350-362
ST8Sia IV	359	1-7 (7 AA)	8-20 (13 AA)	21-58 (38 AA)	59-359 (301 AA)	139-185	277-300	327-339
ST8Sia V	376	1-17 (17 AA)	18-38 (21 AA)	39-80 (42 AA)	81-376 (296 AA)	93-208	298-321	344-350
ST8Sia VI	398	1-3 (3 AA)	4-24 (21 AA)	25-102 (78 AA)	103-398 (296 AA)	183-230	320-343	360-378

Tabela 4

ENZIMAS	região conservada	Motivo A	Motivo B	Motivo C	Motivo D	Motivo E
hST6Gal I	93-165	159-165	X	X	X	X
hST6GalNAc I	259-331	310-315	324-331	X	X	X
hST8Sia IV	71-133	X	X	71-90	119-133	X
hST3Gal IV	75-106	X	X	X	X	84-91

Tabela 5

Fragmentos amplificados	Conjunto de iniciadores	Tamanho do produto de PCR (Pb)
hST3Gal III	5' <i>Afl</i> III-ST3 3' ST3- <i>Bam</i> HI	186
hST6GalNAc I-74	5' <i>Afl</i> III-ST6 3 'ST6- <i>Bam</i> HI	270
hST6GalNAc I-140	5' <i>Afl</i> III-ST6 3' ST6- <i>Bam</i> HI ^o 2	468
hST3Gal III-29/37- hST6GalNAc-74	5' <i>Afl</i> III-ST3 3' ST6- <i>Bam</i> HI	246

Descrição do invento

O invento será melhor descrito na descrição detalhada que se segue.

I - Introdução

Bases do invento:

Os inventores centraram-se no estudo de hST6Gal I porque esta actividade enzimática está ausente em todas as células hospedeiras utilizadas até agora para a produção de proteínas heterólogas, especialmente em sistemas celulares aprovados para fármacos. A enzima humana foi clonada pelos inventores em Ronin, 2001.

Em contraste com a literatura, os inventores foram inicialmente capazes de mostrar que a enzima inteira pode sialilar diferencialmente glicoproteínas aceitadoras com estrutura de glicano bi, tri e/ou tetra-anténaria (Ronin 2001) e nesse momento, os inventores colocaram a hipótese de que dentro da estrutura da transferase, deveria ser exercido um controlo espacial localizado na região SR hipervariável no CD para regular e naturalmente constranger a especificidade enzimática. Com o objectivo de produzir proteínas sialiladas de interesse biomédico, este controlo regulador deve ser abolido e, como resultado, a especificidade pode ser alargada. Dentro da região conservada (93-165) de hST6Gal I, foi observada uma sequência hidrófoba: 93-QVWxKDS-100. A importância desta sequência foi estudada através da deleção progressiva de hST6Gal I da sua parte N-terminal dentro da região conservada recentemente identificada. Os mutantes suprimidos $\Delta 35$, $\Delta 48$, $\Delta 60$ e $\Delta 82$ mostram uma eficiência de transferência crescente. As deleções realizadas antes e depois do domínio conservado (93-165 de ST6GalI) mostraram que o mutante $\Delta 100$ perde a sua actividade catalítica, enquanto $\Delta 89$ (contendo a sequência hidrófoba QVWxKDS) possui uma actividade catalítica óptima. Os resultados indicam claramente que esta curta sequência é crucial para promover a actividade. Esta sequência hidrófoba pode contribuir com alterações conformacionais locais essenciais para que o local activo promova a transferência do ácido siálico.

Um segundo estudo foi realizado pelos inventores com o objectivo de elucidar a base molecular e celular que governa a preferência das ST pelos aceitadores (Ronin 2003). Uma vez que foi difícil delinear o CD da região hipervariável da SR, eles colocaram a hipótese de a SR das ST dever conter características estruturais relacionadas com uma preferência pelo aceitador. Foram analisadas 53 enzimas animais e humanas de especificidade conhecida através de bioinformática para determinar se as ST podem partilhar uma porção peptídica para restringir o reconhecimento do aceitador a uma subfamília de enzimas. Uma região altamente conservada de cerca de 31-85 aminoácidos foi identificada e dentro desta região foram verificados 5 motivos: o motivo A comum a ST6Gal e ST6GalNAc I, o motivo B comum a ST6GalNAc I e ST6GalNAc II, os motivos C e D comuns a todas as ST8Sia e o motivo E presente em ST3Gal excepto a ST3Gal I a II (Tabela 4).

Esses 5 motivos são típicos de um subgrupo de ST partilhando uma actividade catalítica semelhante e assim envolvidos do mesmo modo na transferência do ácido siálico. Estão localizados no final da SR próximo do motivo sialil L e podem ser considerados como parte do CD como característica chave para o reconhecimento do aceitador.

Também de especial interesse, foi também a descoberta de que o mutante $\Delta 89$ é extremamente eficiente em células CHO e que segue a via intracelular desde o Golgi inicial até ao *trans*-Golgi/rede *trans*-Golgi. $\Delta 89$ é expresso juntamente com a via intracelular das glicoproteínas e glicolípidos nas cisternas inteiras do aparelho de Golgi (Ronin, 2003). Este trabalho tornou possível delinear o CD mínimo para hST6Gal I contendo a sequência hidrófoba crucial (QVWxKDS) e deslocar o reconhecimento do aceitador de aceitadores residentes intracelulares para glicoconjugados da superfície celular. Estas descobertas deram a base para o desenho de novas quimeras ancoradas à membrana que podem apresentar um tráfego intracelular semelhante para encontrar glicoproteínas neossintetizadas dentro da via secretora uma vez que são empacotadas em células hospedeiras modificadas quando produzidas como fármacos.

O CD mínimo de hST6Gal I Δ 89 foi utilizado no invento como um "CD otimizado" de especificidade alargada e eficiência de transferência aumentada e será designado CD.

A distribuição das sialiltransferases no aparelho de Golgi

A localização no Golgi das GT não está estritamente organizada, as enzimas não estão distribuídas e isoladas num único compartimento do aparelho de Golgi, a maioria delas sobrepõe-se e estão co-compartimentadas (Colley, 1997; Berger *et al.*, 1998; Berger, 2002). Não foram ainda identificados sinais claros de retenção mas as enzimas formam gradientes sobreponíveis através das cisternas (Opat *et al.*, 2001) e apenas foram identificadas as regiões cruciais.

Existem duas hipóteses em relação aos mecanismos de retenção no Golgi: i) o comprimento do TMD dirige a retenção no Golgi (modelo da espessura da bicamada; Bretscher & Munro, 1993; Colley, 1997; Munro, 1998) e ii) a oligomerização das proteínas conduz a retenção no Golgi (hipótese de Oligomerização/reconhecimento da mesma família; Machamer, 1991; Colley, 1997), na qual os autores postulam que as enzimas interagem na cisterna do Golgi para formar estruturas demasiado grandes para entrar em vesículas de transporte (Munro, 1998).

O comprimento do TMD parece também crucial como sinal de retenção. O TMD representa um factor chave no processo de retenção, na maioria dos casos é suficiente para promover uma localização no Golgi, i.e. através dos compartimentos *cis*, médio e *trans*. Aumentando o comprimento do TMD das ST resulta em retenção reduzida e o TMD sintético (criado através de mutagénesis) de 17 leucinas dá retenção no Golgi enquanto um de 23 leucinas não (Munro, 1991, 1995, 1998). Em relação aos sinais de retenção das ST, várias regiões estão envolvidas de um modo independente para manter as enzimas no aparelho de Golgi. O TMD com as suas regiões flangeadoras são suficientes para retenção no Golgi (Colley, 1997). Outros trabalhos sugerem que não são necessárias sequências específicas no TMD para retenção no Golgi, especialmente na presença de sequências flangeadoras citoplasmáticas e

luminais espaçadas apropriadas e/ou SR (Colley, 1997). Verificou-se que a presença de sequências de aminoácidos carregados negativamente em ST, particularmente próximo da âncora de membrana, localiza erradamente as proteínas no aparelho de Golgi. A presença de epitopo FLAG ou MYC no CT perturba a localização no Golgi de ST6Gal I, enquanto a utilização de uma sequência espaçadora adicional, para espaçar cargas negativas do TMD, revela uma forte vantagem para a retenção no Golgi (Yang *et al.*, 1996). As sequências de SR sozinhas parecem funcionar como um outro tipo de retenção no Golgi (Colley, 1997).

Há provavelmente mais de um mecanismo de retenção, explicando a co-localização das ST dentro do *trans*-Golgi e da rede *trans*-Golgi. Além disso a localização das enzimas também depende do tipo celular onde são expressas (Colley, 1997). O TMD das ST sozinho pode ser suficiente para a retenção no Golgi em células MDCK mas o mesmo TMD e as suas sequências flanqueadoras são claramente necessárias para retenção no Golgi em células CHO (Colley, 1997). De interesse, os dados disponíveis revelam que a alteração do CT, do TMD e da SR não perturba a actividade catalítica das ST, mas permite deslocalizá-los dentro do aparelho de Golgi (Grabenhorst e Conradt, 1999).

A enzima mais estudada quanto ao sinal de retenção no Golgi, é a ratST6Gal I (rST6Gal I), que foi localizada no *trans*-Golgi e na rede *trans*-Golgi dos hepatócitos (Roth *et al.*, 1985; Opat *et al.*, 2001) e em compartimentos pós-Golgi noutras células (Colley, 1997). Existem duas isoformas naturais de rST6Gal I, diferem apenas num aminoácido na posição 123 no CD. Isto é devido a uma única mudança de nucleótidos de A para G, resultando na mudança de aminoácidos de Tyr para Cys. Não se verificaram diferenças na actividade catalítica entre as duas isoformas da enzima (Chen *et al.*, 2003). A forma STcys é verificada no Golgi em células expressando moderadamente e nunca nas superfícies celulares ou clivada ou segregada para o meio de células COS-1 ou HeLa, enquanto a isoforma STtyr é verificada no Golgi, a níveis baixos à superfície celular e é clivada e segregada a partir de células COS-1 e HeLa (Ma *et al.*, 1997). Para explicar as duas localizações diferentes das proteínas, os autores

propuseram duas hipóteses sobre o mecanismo de retenção: espessura da bicamada e oligomerização. Primeiro, a análise do aumento do comprimento do TMD de STcys e tyr sugere que a espessura da bicamada não é um processo predominante para a retenção no Golgi para ST6Gal I. Segundo, a análise de oligómeros revela que 100% e 13% das STcys e STtyr, respectivamente, são insolúveis a pH 6,3 (pH do Golgi tardio), mas a pH 8,0 ambas as isoformas são solúveis. Os resultados sugerem que as mudanças conformacionais e a formação de ligações dissulfureto no CD representam a base para uma maior capacidade para STcys formar oligómeros e a sua localização estável no aparelho de Golgi. A natureza do aminoácido 123 influencia as formações de oligómeros (Chen *et al.*, 2000). Além disso, o processo de oligomerização e a actividade catalítica dependem também de outros sete resíduos de cisteína conservados (C24, 139, 181, 332, 359, 361 e 403). O Cys24, dentro do TMD, é necessário para dimerização, enquanto os resíduos de cisteína presentes no CD são necessários para o tráfego e a actividade catalítica. As enzimas Cys181 e Cys332 (no motivo sialil L e S, respectivamente) são retidas no retículo endoplasmático e são minimamente activas ou inactivas, respectivamente. Cys359 e 361 (entre os motivos sialil S e VS) são enzimas inactivas sem comprometer a sua localização e o seu tráfego. As enzimas mutadas Cys139 ou Cys403 não produzem alteração da actividade catalítica nem da localização no Golgi. A substituição destes dois resíduos de cisteína diminui a clivagem e secreção sugerindo que são necessários para a clivagem do sinal (Qian *et al.*, 2001) (Figura 3).

Foram caracterizados mutantes das enzimas STtyr, partilhando a deleção na SR: STtyr Δ 4 e STtyr Δ 5 (deleção dos aminoácidos 32 a 104 e 86 a 104, respectivamente) não são activas e/ou segregadas, STtyr Δ 1, Δ 2 e Δ 3 (deleção entre os aminoácidos 32 a 86) representam formas não cliváveis e mostram uma crescente expressão na superfície celular (Kitazume *et al.*, 1999).

Um trabalho recente (Fenteany & Colley, 2005) mostra que são, de facto, necessários múltiplos sinais para a oligomerização de rST6Gal I e para a localização no Golgi. Os autores tentaram reavaliar o papel do CT, do TMD das duas

isoformas da enzima (Tyr ou Cys 123). Eles produziram várias construções de ADN para produzir a quimera tal como representada na Figura 4. O aumento do comprimento do domínio transmembranar (com mais de 17 aminoácidos) não aumenta a saída do Golgi, não altera o tráfego nem a localização no Golgi. Em relação ao papel do CT e do TMD na isoforma STcys, existe um sinal composto por 3 resíduos de Lisina no CT que é reconhecido como sinal de exportação para fora do RE. Os autores mostram que a alteração de CT e TMD perturba o caminho da enzima. A única diferença entre as isoformas de rST6Gal I é a sua capacidade para formar oligómeros de acordo com o pH. Para STcys, a formação de oligómeros, de acordo com o pH, está comprometida apenas quando CT e TMD são alterados. Para STtyr, aumenta ligeiramente a velocidade de saída do Golgi (Fenteany & Colley, 2005).

Todas estas descobertas recentes em relação à rST6Gal I podem ser consideradas aplicáveis à ST6Gal I humana uma vez que ambas as enzimas partilham 85% de homologia e que ambos os CD são virtualmente idênticos. De facto, o alinhamento das duas sequências (Figura 5) mostra claramente que as sete cisteínas anteriormente descritas são conservadas na hST6Gal I e a Tirosina 123 também está presente.

Apesar de esforços consideráveis dos cientistas, não se conseguiu demonstrar qualquer sinal na sequência de ST ou FuT. Crê-se que um processo multifactorial pode atrasar as enzimas no seu caminho para a superfície para as direccionar para os compartimentos apropriados que são a Rede *Trans*-Golgi para ST6Gal. Subsequentemente, os inventores colocaram a hipótese de que CT, TMD e SR possam ser alternativamente mudados para ancorar um CD definido em células hospedeiras que não têm o gene relevante e não conseguem expressar a actividade de transferase relevante. Assim, é portanto possível gerar uma ampla variedade de combinações uma vez que estas 3 porções podem ser seleccionadas a partir de 55 genes animais conhecidos. Cada um deles pode ser potencialmente fundido com 20 domínios catalíticos distintos conhecidos para apresentar a capacidade de transferir ácido siálico. Isto é de interesse decisivo para a humanização de leveduras (*Pichia pastoris* ou *Shyosaccharomyces piombae*), plantas e insectos que não expressem estas actividade e contudo sejam utilizados

para produzir fármacos recombinantes. Embora não testados, esta consideração também pode aplicar-se a FuT, GalT, GalNAcT e GlcNAc uma vez que governam a biossíntese de glicotopos de grande relevância clínica.

Importância de vários resíduos

Tal como anteriormente descrito, um quarto motivo (motivo 3) foi identificado na família das ST, entre os motivos S e VS. É composto por quatro aminoácidos altamente conservados com a seguinte sequência de consenso: **(H/y)Y(Y/F/W/h)(E/D/q/g)**. A mutagénese dirigida ao local na hST3Gal I mostrou a importância funcional de dois aminoácidos neste motivo: His299 e Tyr300. Os resultados sugerem a importância dos resíduos aromáticos, o seu possível envolvimento no reconhecimento do aceitador e a sua contribuição para uma eficiência catalítica óptima. Particularmente o Tyr300 invariante tem um papel conformacional principal. A análise mutacional mostra que os mutantes H299A e Y300A não apresentam actividades catalíticas, enquanto o mutante Y300F restaura parcialmente a actividade (Jeanneau *et al.*, 2004). Este motivo está também presente no domínio catalítico de hST6Gal I. Além disso vários resíduos de cisteína são da maior importância na dimerização e actividade catalítica tal como anteriormente invocado (Qian *et al.*, 2001). Pelo menos uma ligação dissulfureto foi colocada em evidência entre dois resíduos de cisteína conservados dentro dos motivos sialil L e S. Esta ligação é essencial para manter a conformação da proteína. Estas mesmas observações revelaram a existência desta ligação na PST (Angata *et al.*, 2001), além disso, existe uma segunda ligação dissulfureto entre motivos sialil e a região C-terminal. Foi demonstrada a dimerização da ST6Gal I, através de ligação dissulfureto. Esta enzima dentro do aparelho de Golgi está cerca de 20 a 30% na forma dimérica que tem uma actividade inferior porque a sua afinidade é reduzida para o CMP-NeuAc (Jeanneau, 2003).

"Autoglicosilações"

As GT são frequentemente elas próprias glicosiladas. Estão disponíveis poucos dados em relação ao estado da

glicosilação das ST e a sua importância na função biológica das enzimas. Parece que a N-glicosilação na ST6Gal I não é necessária para a actividade biológica da enzima *in vivo*. Por outro lado, quando testada *in vitro* no mutante, a actividade enzimática é apenas observada para a proteína mutada no primeiro local de glicosilação. Estes resultados sugerem a existência de dois N-glicanos (Asn 146 e Asn 158) que podem estabilizar e/ou proteger a proteína contra degradação. A mutação no segundo local de glicosilação pode conduzir a agregação ou degradação da proteína (Chen & Colley, 2000; Chen *et al.*, 2000). No entanto, a N-glicosilação na ST8Sia I pode afectar a sua actividade e a sua localização subcelular (Martina *et al.*, 1998). A eliminação dos N-glicanos reduziu a sua actividade *in vivo* com menos de 10% da actividade inicial. Foi recentemente mostrado que ST8SiaII e ST8SiaIV possuem, para além da sua actividade clássica, uma actividade de autopoli-sialilação (Muhlenhoff *et al.*, 2001). Esta autopoli-sialilação ocorre nos N-glicanos na posição Asn 74 na PST e nas posições Asn 69 e 219 na STX. Mutagénese dirigida ao local nestes locais inactiva a enzima tanto *in vivo* como *in vitro*. Como resultado, é necessário cuidado para manter o local de glicosilação de Asn 74 e Asn 69 quando se utiliza a SR de ST8SiaIV e ST8SiaII, respectivamente, para construir genes quiméricos.

II - Utilização das sequências do invento para introdução de nova actividade de transferase em sistemas de expressão

IIa - O papel dos glicanos e sua importância no sistema imunitário

Os glicanos são sinais de reconhecimento na maioria dos organismos vivos mas são também determinantes estruturais chave para a biopotência das proteínas.

Eles têm um papel principal na oligomerização da dobragem das proteínas, no controlo de qualidade, na separação e no transporte (Helenius & Aebi, 2001). Os glicanos representam epitopos oligosídicos que podem ser considerados antígenos de hidratos de carbono. A natureza dos glicanos modifica a imunorreactividade. Existem várias

famílias de antigénios oligosídicos: os determinantes dos grupos sanguíneos ABH, grupos de tecidos de Lewis, membros de antigénios da família T e antigénios específicos da adesão celular polissialilada ou sulfatada (Legaigneur *et al.*, 1999). Os glicanos têm múltiplos papéis: estão também envolvidos em mecanismos de interacções entre células e entre células e matriz. Algumas proteínas celulares designadas Lectinas reconhecem especificamente estruturas glicânicas e actuam como receptores específicos (Gabiús *et al.*, 2004). Na maioria dos casos, elas são componentes dos glicoconjugados da superfície celular.

As porções hidrato de carbono actuam como sinais de reconhecimento no sistema imunitário e influenciam o reconhecimento imunitário de pelo menos duas maneiras: i) a conformação da proteína é alterada e modula a sua função biológica; ii) os oligossacáridos servem como determinantes de reconhecimento. Particularmente, os ácidos siálicos contribuem grandemente para ambos estes efeitos porque a sua natureza altamente electronegativa e hidrófila pode influenciar a conformação de macromoléculas sialiladas e isto é particularmente relevante para as glicoproteínas. Os ácidos siálicos possuem a capacidade de actuar como máscaras biológicas para impedir ou reduzir a acessibilidade e têm um papel significativo na superfície celular no processo de reconhecimento de discriminação de próprio/não próprio (Pilatte *et al.*, 1993, Glicobiology 2006 a adicionar).

Uma vez que os N-glicanos estão bem representados nas glicoproteínas séricas e muitas outras proteínas teciduais do corpo humano, é altamente desejável que não sejam antigénicos quando presentes em fármacos recombinantes glicosilados. No entanto desde que foi desenvolvida como 1º fármaco aprovado a HuEPO produzida em células CHO, parece que vários açúcares terminais podem ser antigénicos em humanos: poliLacNAc, alfa1,3-Gal e Ácido N-glicolilneuramínico. De acordo com os corpos reguladores mais recentes, estes devem deixar de estar presentes em produtos terapêuticos recombinantes.

I Ib - A eliminação

Para além da sua influência nas propriedades físico-químicas e nas funções biológicas das proteínas, os glicanos possuem também um papel predominante na duração das glicoproteínas no sangue. Este fenómeno é designado de eliminação metabólica e representa a taxa a que um composto natural ou um fármaco é removido do corpo através do fígado ou do rim. É definida como o volume plasmático purificado de acordo com o tempo (ml/min). Este mecanismo permite ao organismo eliminar os fármacos.

A maioria das proteínas plasmáticas em circulação pertencem à família das glicoproteínas partilhando oligossacáridos ligados em N e todas elas terminam em ácido siálico, particularmente em humanos. A remoção do ácido siálico através de uma neuraminidase diminui drasticamente a duração da vida plasmática das glicoproteínas séricas para minutos e promove a sua tomada pelo fígado (van den Hamer *et al.*, 1970; Morell *et al.*, 1971). Por exemplo, quando a hormona da gravidez (mais recentemente HuEPO recombinante) não tem ácido siálico, a sua semi-vida é reduzida para 2 minutos, enquanto, habitualmente, a sua duração de vida é de cerca de 48 horas. É amplamente conhecido actualmente que o ácido siálico é necessário para manter as glicoproteínas no sangue porque protege as proteínas em circulação da eliminação através de um receptor de sialoglicoproteína (ASGPR ou Lectina hepática) composto por duas subunidades e incluído nas membranas de hepatócitos (Hudgin *et al.*, 1974; Kawasaki & Ashwell, 1976; Bianucci & Chiellini, 2000). Este receptor liga-se a galactose ou resíduos de N-acetilglucosamina dos N-glicanos desialilados (Meier *et al.*, 2000). As glicoproteínas reconhecidas são então internalizadas em vesículas endocíticas cobertas de clatrina e redireccionadas para lisossomas onde são degradadas (Ashwell & Hardford, 1982).

Este passo pareceu crucial para as propriedades farmacocinéticas das glicoproteínas terapêuticas recombinantes uma vez que os organismos inferiores não são capazes de sialilação de proteínas e em células CHO a adição do ácido siálico é extremamente sensível ao estado energético

das células em cultura (NeuAc é sintetizado através de condensação do piruvato e ácido láctico) e é quase completa. Todas as proteínas mais famosas sialiladas (EPO, IFN, GM-CSF, FSH, Ab...) devem ser purificadas para que o fabricante apresente um elevado teor de ácido siálico e serem controlados para consistência de lote para lote.

IIc - Controlo da glicosilação & Alergenicidade

Os resíduos de ácido siálico na posição terminal dos N-glicanos são da maior importância para as proteínas terapêuticas uma vez que a sialilação das proteínas confere propriedades importantes às glicoproteínas. A maquinaria necessária para a síntese, a activação e a introdução dos resíduos sialilados está fracamente representada nos diferentes sistemas de expressão de proteínas recombinantes. Depois, as proteínas humanas recombinantes produzidas estão muito frequentemente subsialiladas ou mesmo não sialiladas em comparação com os seus equivalentes nativos. Além disso, se forem expressas em células de mamífero, a sialilação pode ocorrer através do ácido N-glicolilneuramínico (NeuGc) que difere significativamente do derivado N-Acetilado (NeuAc) uma vez que está presente em células de mamífero mas não em células humanas.

Como existe uma questão importante com o ácido siálico no mercado de fármacos recombinantes, muitas equipas de investigação estão a trabalhar na modificação do processo de N-glicosilação nos diferentes sistemas de expressão existentes. Trabalho intenso tenta a humanização do padrão de glicosilação das proteínas recombinantes para o aproximar tanto quanto possível do padrão verificado nas glicoproteínas naturais e melhorar a farmacocinética bem como a segurança do produto. Entretanto, o procedimento tenderá a reduzir a presença de determinantes imunogénicos. A este respeito, um estudo muito recente confirmou que o ácido siálico ligado em 6 e o ácido siálico ligado em 3 impedem as glicoproteínas de se ligarem aos receptores do sistema imunitário e gerar activação (Glicobiology 2006). A adição do ácido siálico ligado em 6 a proteínas glicosiladas recombinantes beneficiará portanto grandemente a segurança do fármaco e seria um importante avanço no campo.

As proteínas de interesse terapêutico foram primeiro extraídas a partir de fontes naturais tais como sangue, placenta, tecidos humanos ou animais. No entanto, esta abordagem está limitada à quantidade de tecidos humanos disponíveis e pode acarretar riscos de contaminação (vírus, príons, oncogenes) e/ou gerar reacções alérgicas devido a vestígios de proteínas animais ou toxinas. Com o surgimento da biologia molecular, novas abordagens foram desenvolvidas para produzir proteínas utilizando sistemas de expressão bastante diferentes. Uma grande variedade de sistemas de expressão heterólogos está disponível (Andersen & Krummen, 2002) e cada um deles possui vantagens e desvantagens com particular atenção dada ao padrão de N-glicosilação das proteínas recombinantes produzidas nestes sistemas.

IIId- Biossíntese de N-glicanos em sistemas de expressão conhecidos: necessidade de humanização da via de glicosilação

Bactérias

Um dos sistemas mais utilizados é a bactéria *Escherichia coli* (*E. coli*) (Swartz, 2001; Baneyx, 1999), mas o seu principal inconveniente é que as modificações pós-tradução humanas, particularmente as glicosilações, não são realizadas por este procariota porque nenhuma dessas glicosiltransferases é expressa em *E. coli*. Isto pode conduzir a dobragem errada e subsequente rejeição das proteínas terapêuticas de interesse pelo sistema imunitário, redução na sua duração de vida e actividade biológica.

Nenhuma N-glicosilação do tipo humano foi verificada até agora neste microrganismo.

Leveduras e fungos filamentosos

As leveduras e fungos filamentosos são também sistemas de expressão em eucariotas bem estabelecidos e possuem uma maquinaria celular aproximando-se das células humanas. As leveduras produzem proteínas complexas e realizam várias modificações pós-tradução incluindo glicosilações. Tanto as

leveduras como os cogumelos produzem tipicamente glicanos ricos em manose através da adição de até 100 resíduos de manose (em relação às leveduras) sobre o núcleo pentassacarídico (Tanner & Lehele, 1987; Herscovics & Orlean, 1993). Essa hipermanosilação recebe definitivamente uma resposta imunitária em humano. Além disso, até agora, não foi observado nenhum oligossacárido complexo contendo ácido siálico, galactose, fucose e N-acetilgalactosamina dentro das glicoproteínas produzidas por estes organismos (Blanchard, 2004).

O processo de N-glicosilação realizado pelas leveduras e cogumelos é semelhante ao processo em mamífero em relação aos passos iniciais no RE mas a presença de polímeros adicionados no aparelho de Golgi impede a obtenção de aprovação pelos corpos reguladores.

Células de insecto

As proteínas expressas em células de insecto são correctamente dobradas, segregadas e podem receber modificações pós-tradução. As modificações pós-tradução realizadas por estas células são semelhantes às realizadas por células de mamífero, em particular em relação à N-glicosilação da proteína. As estruturas dos glicanos obtidos neste caso, são no entanto incompletas e designadas como paucimanose devido à presença de uma actividade de N-acetilglucosaminidase indesejável que degrada as neoglicoproteínas expressas durante expressão em Baculovírus (Blanchard, 2004).

A falta de ácido neuramínico em insectos é ainda activamente debatida (Marchal *et al.*, 2001, Lerouge *et al* 2005). Nalgumas linhas celulares de insecto, são verificados resíduos de fucose ligados em $\alpha 1,3$ e isto pode desencadear uma resposta imunitária em humano. Actualmente, a utilização deste sistema está portanto restringida à produção de antígenos de vacinação.

Plantas transgénicas

Como outras células eucarióticas, as plantas exibem uma maquinaria celular complexa e sofisticada que pode ser utilizada para produzir proteínas terapêuticas. As proteínas recombinantes possuem uma qualidade farmacológica muito boa porque as plantas expressam as enzimas necessárias para a maturação das proteínas.

No entanto, o processo de glicosilação necessita de vários ajustes para não produzir proteínas alergénicas. De facto, as N-glicosilações em plantas (Lerouge *et al.*, 2000) são semelhantes às realizadas em humanos na medida em que se refere à glicosilação central. As glicosilações não possuem ainda antenas sialiladas e existe uma adição de β 1,2-xilose e α 1,3-fucose. Ambos os resíduos são altamente imunogénicos para humanos e actualmente comprometem consideravelmente a aprovação de plantas transgénicas como sistemas de expressão para produtos terapêuticos.

Células de mamífero

O sistema de expressão em células de mamífero, designadamente células CHO, é actualmente o único sistema aprovado para fármacos para produzir produtos terapêuticos recombinantes. Tais células mostram uma importante vantagem porque são capazes de sintetizar proteínas complexas, alcançar N-glicosilação de tipo complexo de elevada massa molecular. As células de mamífero utilizam naturalmente, não só as enzimas envolvidas na síntese e o transporte dos nucleótidos-açúcares, mas também as glicosiltransferases necessárias para garantir uma glicosilação complexa de proteínas recombinantes heterólogas com um elevado nível de α 2,3-sialilação. No entanto, existe falta de outras enzimas tais como as α 1,3/4-fucosiltransferases e α 2,6-sialiltransferases, que transfiram motivos glicosídicos específicos dos N-glicanos de tecidos humanos. Em roedores, o ácido N-glicolilneuramínico é substancialmente preferido em N-glicanos, a O-acetilação do ácido N-Acetilneuramínico na posição 4, 7, 8 e muito frequentemente 9 ocorre também frequentemente e cada um destes derivados pode ser potencialmente imunogénico em humano. Isto representa uma

limitação à utilização de células de ratinho ou de ratinhos/coelhos lactantes para expressar proteínas recombinantes (Blanchard, 2004).

IIE - "Glicoengenharia"

Tal como discutido acima, a sialilação terminal de glicoproteínas aprovadas é o passo mais difícil de obter em todos os sistemas de expressão disponíveis até agora. Como exemplo, não se conseguiu adicionar nenhum ácido siálico ligado em $\alpha 2,6$ embora seja a ligação prevalente no sangue humano. Em todos os organismos, a maquinaria enzimática necessária simplesmente não existe (Figura 6).

O objectivo final da investigação neste campo é melhorar o processo de glicosilação, o rendimento da produção e a qualidade das proteínas recombinantes. A utilização dos sistemas de expressão geneticamente modificados para glicosilação permitiria produzir glicoproteínas com uma homogeneidade excepcional nas suas estruturas glicânicas. Tais sistemas podiam então ser utilizados para desenvolver um elevado nível de produção para proteínas de interesse biomédico de estruturas bem definidas. A engenharia de glicosilação em levedura (Hamilton *et al.*, 2003; Roy *et al.*, 2000) preveniu primeiro a adição de cadeias polimanosídicas. Depois, foram adicionadas aos sistemas as enzimas que são necessárias à galactosilação e N-acetilglucosaminilação (Maras *et al.*, 1999; Bretthauer, 2003; Vervecken *et al.*, 2004). No entanto, a adição do ácido siálico terminal é ainda difícil de realizar, devido ao número e à localização das enzimas envolvidas neste processo mas está actualmente a ser feita no Japão (Figura 6).

Em células de insecto, as GT catalisando a transferência de açúcares imunogénicos foram removidas e a sialilação foi realizada através da adição de 3 genes codificando para a N-acetilglucosamina-2-epimerase, N-acetilneuraminil-liase e CMP-Neu5Ac-sintetase (Jarvis *et al.*, 1998; Aumiller *et al.*, 2003). Os autores criaram novas linhas celulares de insecto (SfSWT-3) desenhadas para sintetizar o seu próprio CMP-ácido siálico. As células resultantes expressam todos os 7 genes de mamífero necessários para formar um N-glicano, produzir CMP-

ácido siálico e sialilaram uma proteína recombinante quando cultivadas num meio de crescimento sem soro (Aumiller *et al.*, 2003). Este trabalho foi patenteado.

Em plantas, a primeira estratégia utilizada tentou prevenir a adição de açúcares alergénicos através de proteínas de aprovisionamento no RE. Mas, neste caso, os glicanos não podem ser do tipo complexo. Uma outra estratégia baseou-se na inibição de várias GT dentro do aparelho de Golgi. Esta inibição pode ser completada e/ou pode entrar em competição com a maquinaria endógena para maturação. A adição da maquinaria sialil está também em processo.

No caso de células de mamífero, o trabalho foi realizado através da sobre-expressão de uma $\alpha 2,3$ -ST e uma $\beta 1,4$ Gal (Weikert *et al.*, 1999), estas enzimas estão ambas presentes no genoma mas as suas actividades variam com as condições de cultura. Isto conduz a uma ampla variabilidade em relação à presença da Gal terminal e do ácido siálico e também a extensa micro-heterogeneidade nas estruturas de glicano das proteínas de secreção.

As investigações foram principalmente orientadas para a optimização da galactosilação e sialilação (Granbenhorst *et al.*, 1999) através da introdução de uma $\alpha 2,6$ -ST (Bragonzi *et al.*, 2000). Uma linha celular CHO expressando estavelmente $\alpha 2,6$ -ST inteira foi estabelecida há muito tempo e foi também decepcionante. A razão observada entre os resíduos de ácido siálico terminais ligados em $\alpha 2,6$ e $\alpha 2,3$ contidos foi de 40,4% de resíduos de ácido siálico em $\alpha 2,6$ e de 59,6% em $\alpha 2,3$, melhorando a farmacocinética em estudos de eliminação (Bragonzi *et al.*, 2000). Apesar da melhoria na humanização de células, a razão entre resíduos de ácido siálico terminais ligados em $\alpha 2,6$ e em $\alpha 2,3$ não pode ser controlada e nem é favorável à actividade em 6.

Em resumo, a sialilação é um passo crítico para controlar as estruturas de glicano em proteínas produzidas através de engenharia genética. A principal dificuldade reside não tanto na expressão da actividade em falta mas na obtenção do sistema de produção heterólogo para funcionar com sucesso. De facto, 3 objectivos principais devem ser

verificados para humanizar a glicosilação dos actuais sistemas de expressão: 1) obtenção do substrato dador e do transportador relevante necessário para que esta enzima funcione dentro do Golgi. 2) obtenção da enzima correctamente compartimentada para eventualmente competir com a actividade de sialiltransferase endógena. 3) obtenção da enzima capaz de catalisar a transferência de siálico para qualquer substrato aceitador relevante.

IIIf - Antecedentes do invento

A estratégia dos inventores baseia-se na consideração de que a enzima ST6 inteira foi até agora incapaz de verificar os objectivos acima, especialmente a competição com a actividade ST3 endógena existente na maioria das células de mamífero, provavelmente porque o produto génico não é inserido em compartimentos intracelulares envolvidos na via secretora. Os inventores desenvolveram assim um procedimento que pode conferir a melhor oportunidade para que um CD de transferase alcance os compartimentos do Golgi onde as neoglicoproteínas são separadas para vesículas secretoras e/ou onde encontram a ST modificada antes/em vez da ST3 existente.

III - Desenho de ST membranares sintéticas

O presente invento consiste num procedimento que distribui um painel de ST quiméricas de actividade catalítica conhecida, possuindo uma âncora membranares para as direccionar para o aparelho de Golgi de células eucarióticas. Nenhuma delas existe em células vivas e são consideradas como "ST sintéticas" porque a sua construção não necessita de nenhuma das sequências de ADN codificando para CT, TMD ou SR. Os oligonucleótidos relevantes estão dentro de 200 pb (60 aminoácidos) de comprimento e podem ser desenhados através de informática e obtidos comercialmente. Utilizando métodos de PCR, é construída uma âncora sintética marcada para codificar para a metade N-terminal da proteína ST e fundida com o domínio catalítico optimizado da hST6GalI.

O invento descreve vários métodos moleculares para construir 3 tipos de ST membranares quiméricas:

- *Construção homóloga:* a âncora da ST fundida com o CD Δ89 de ST6Gal. Sintética ou eventualmente copiada através de PCR. Neste caso, as porções CT+TMD+SR são da mesma ST: ST3Gal/ST6GalNAc/ST8Sia. O exemplo descreve uma sequência de 200 pb.
- *Construção híbrida:* CT+TMD são de uma ST e a SR de outra ST. Aqui também, o tamanho mínimo é de 200 pb.
- *Construção heteróloga:* os fragmentos de CT, TMD e SR são de ST diferentes. Aqui também, o tamanho mínimo é de 200 pb.
- *Construção longa:* uma construção de CT+TMD+SR1 de 200 pb é preparada e fundida com uma segunda construção de SR2 no máximo de 200 pb de um modo tal que SR2 fique a jusante de SR1. A construção duplicada pode ainda ser ligada na extremidade 3' a qualquer outra construção de SR repetidamente conforme necessário.

O tamanho de 200 pb baseou-se na observação de que a porção N-terminal mais curta de todas as ST humanas é ST6GalNAcIII com CT+TMD+SR=200 pb. Esta enzima não tem virtualmente caule.

Actualmente, os inventores validaram um método para construir um fragmento de ADN sintético e fundi-lo com um domínio catalítico mínimo para gerar uma nova enzima transferase de actividade desejada. Várias construções representativas estão disponíveis para mais expressão. As quimeras sintéticas são activas e estão a ser estudadas em células CHO utilizando expressão transiente e microscopia confocal.

EXEMPLOS

Exemplo 1: Produção do domínio catalítico (CD) mínimo de hST6Gal I (SEQ ID NO:43 codificando para SEQ ID NO:44)

A ST6Gal I humana (hST6Gal I) clonada por Legaigneur *et al.*, 2001 foi utilizada para amplificar o CD mínimo clonado no vector pFLAG-CMV (Sigma) (Donadio *et al.*, 2003) que foi utilizado em todas as construções de quimeras.

O vector recombinante foi primeiro digerido para verificar a presença do CD de hST6Gal I. O gel de agarose corrido revela a presença de uma banda nucleotídica a cerca de 1000 pb, correspondendo ao tamanho esperado para o domínio catalítico mínimo (Figura 10A-C).

O CD correspondendo aos aminoácidos 90 a 406 de hST6Gal I foi obtido através de PCR utilizando os seguintes iniciadores:

5'BamHI-Δ4: 5'-GAGCCCGGATCCGAGGCCTCCTTC-3' e

3'Δ4-XbaI: 5'TAACCTCTAGATTAGCAGTGAATGGTCCGGAAGC-3'.

Estes iniciadores contêm os locais de restrição *Bam*H1 (5'BamHI-Δ4) e *Xba*I (3'Δ4-XbaI) e o codão de terminação natural da sequência. O local de restrição *Bam*H1 foi utilizado para ligar o CD de hST6Gal 90-406 à parte N-terminal sintética de outras ST para formar enzimas quiméricas. O local de restrição *Xba*I na extremidade 3' foi utilizado para introduzir a quimera dentro do vector pcDNA3.1 (Invitrogen) (**Figura 11**).

A PCR foi efectuada num aparelho I-Cycler (Biorad), utilizando 2,5 unidades de ADN-polimerase ProofStart (Qiagen), 300 μM de cada dNTP (Sigma Aldrich), 1 μM de cada iniciador, tampão do fabricante ProofStart 1× (Qiagen), solução Q do fabricante 1× (Qiagen) e 1,5 mM de Mg²⁺ num volume final de 50 μl com 200 ng de pFlag-CMV-hST6Gal I-90-406. A reacção foi efectuada como se segue: 95°C durante 5 min, seguido de 40 ciclos de 30 s a 94°C, 30 s a 55°C e 1 min a 72°C. O produto de PCR amplificado foi analisado através de electroforese num gel de agarose a 1,5% (amplificação de um produto de PCR no tamanho esperado de 1000 pb; Figura 17B) e purificado utilizando o estojo de Extração em Gel Qiagen de acordo com as instruções do fabricante.

Vários produtos de amplificação por PCR foram reunidos e concentrados para estimar a quantidade de ADN num gel de agarose (**Figura 10C**). Os produtos de PCR purificados foram mantidos a -20°C até à utilização.

Exemplo 2: Montagem dos domínios N-terminal sintéticos

Para criar uma grande gama de SR tal como escolhidas através de análise bioinformática, o CT, o TMD e a SR de ST distintas foram sintetizados e montados.

Todas as partes N-terminais seleccionadas da quimera contêm as 3 regiões típicas da família de ST: o CT, o TMD e a SR. De nota, a SR pode ser de comprimento variável e pode ser originária da duplicação do gene sintético. Todos estes fragmentos, i.e. CT+TMD+SR, foram inteiramente reconstituídos utilizando passos complementares de hibridação e fosforilação tal como descrito abaixo.

a - Construção do domínio não catalítico de hST3Gal III (SEQ ID NO:151 codificando SEQ ID NO:152)

Seis oligonucleótidos com sentido e cinco oligonucleótidos anti-sentido correspondendo a parte da sequência da região N-terminal (1 a 138 pb) de hST3Gal III foram desenhados e sintetizados (Eurogentec). Este exemplo mostra o método de síntese de uma sequência de interesse de ocorrência natural. Um epitopo FLAG (a negrito e sublinhado) foi adicionado entre o codão de iniciação ATG e o segundo codão (GGA) da sequência hST3Gal III:

ATG **GACTACAAAGACGATGACGACAAG** GGA.

Um micrograma de cada nucleótido interno foi fosforilado utilizando 1 unidade de polinucleótido-cinase (Eurogentec) num tampão cinase contendo 2 mM de ATP (Sigma). A reacção foi efectuada durante 60 min a 37°C e a incubação finalmente inactivada durante 10 min a 65°C. Todos os oligonucleótidos foram separadamente desnaturados durante 10 min a 80°C e depois misturados para correspondência. A correspondência foi efectuada de um dia para o outro com um gradiente de temperatura decrescente de 80°C para 20°C. O fragmento resultante foi então sujeito a PCR.

b - Construção do domínio não catalítico de hST6GalNAc I-74 (SEQ ID NO:145 codificando SEQ ID NO: 146)

Um dos caules mais curtos dentro da família de ST é o caule da hST6GalNAc III, que é composto por cerca de 6 aminoácidos. Um possível domínio N-terminal desta ST é de 74 aminoácidos (CT, TMD e SR) que corresponde a 222 pb. Portanto, para delinear uma construção com a SR mais curta, foi reconstituído um domínio N-terminal sintético de 222 pb correspondendo aos primeiros 74 aminoácidos da hST6GalNAc I.

Nove oligonucleótidos com sentido e oito oligonucleótidos anti-sentido correspondendo à sequência da região N-terminal completa (1 a 222 pb) de hST6GalNAc I foram desenhados e sintetizados (Eurogentec). Uma sequência FLAG (a negrito e sublinhado) foi adicionada entre o codão de iniciação ATG e o segundo codão (GGA) da sequência de hST6GalNAc I:

ATG **GACTACAAAGACGATGACGACAAG** GGA.

A fosforilação e as reacções de correspondência foram efectuadas tal como anteriormente descrito.

c - Duplicação do oligonucleótido sintético para montar o domínio não catalítico hST6GalNAc II-140 (SEQ ID NO:148)

hST6GalNAc I foi seleccionada como a sialiltransferase que exibia a maior SR (246 aa). Neste exemplo, o comprimento da porção CT+TMD+SR sinteticamente reconstituída (resíduos de aminoácidos 1-35) foi unida a outro trecho de SR (resíduos de aminoácidos 36-140) para duplicar o comprimento e ser capaz de adicionar uma âncora membrana de 140 resíduos de aminoácidos tal como relatado em Donadio et al., 2003.

d - Construção de um domínio sintético híbrido: hST3Gal III 1-28-hST6GalNAc I 37-74 (SEQ ID NO:160)

Foi construída uma região N-terminal híbrida contendo primeiro o CT e o TMD da hST3Gal III (resíduos de aminoácidos 1 a 28) e segundo o início da SR de hST6GalNAc I (aminoácidos 37 a 74).

O comprimento total deste híbrido sintético é de 66 aminoácidos correspondendo a 198 nucleótidos (SEQ ID NO:159). Nove oligonucleótidos com sentido e oito oligonucleótidos anti-sentido correspondendo à região N-terminal híbrida descrita acima (1 a 225 pb), utilizando as sequências de hST3Gal III/hST6GalNAc I, foram desenhados e sintetizados (Eurogentec). Um epitopo FLAG (a negrito) foi adicionado entre o codão de iniciação ATG e o segundo codão (GGA) da sequência de hST3Gal III:

5'- ATG **GACTACAAAGACGATGACGACAAG** GGA -3'.

A fosforilação e a reacção de correspondência foram efectuadas tal como anteriormente descrito.

e - Amplificação por PCR das construções sintéticas

Após o passo de reconstituição sintética, o produto foi amplificado através da técnica de PCR utilizando iniciadores específicos que trazem os locais de restrição desejados em cada extremidade do fragmento de ADN reconstituído para os ligar primeiro ao CD seleccionado e depois no vector pcDNA3.1.

Foram desenhados os seguintes iniciadores:

5'AflIII-ST3: 5'-GAGCCCCTTAAGATGGACTACAAAGACGACGATGACG-3'
e 3'ST3-BamHI: 5'-TAAGGGGGATCCGCTAGAGTGACTATACTTACTGGA-3.

5'AflIII-ST6: 5'-GAGCCCCTTAAGATGGACTACAAAGACGACGATGACG-3'
e 3'ST6-BamHI: 5'-TAAGGGGGATCCGTTGTGGAGGAACGGGA-3';

3'ST6-BamHI n°2: 5'-TAAGGGGGATCCTCTGGGTGACAGTGTGTTCAC-3'.

Estes iniciadores contêm os locais de restrição AflIII (5'AflIII-ST3 e 5'AflIII-ST6) e BamHI (3' ST3-BamHI, 3' ST6-BamHI e 3' ST6-BamHI n°2) para assegurar as ligações posteriores.

As PCR foram realizadas com o conjunto de pares de iniciadores descrito na Tabela 5. Estas foram efectuadas num aparelho de PCR (I-Cycler, Biorad) com 5 µl de cada fragmento sintético numa solução contendo 2,5 unidades de ADN-polimerase ProofStart (Qiagen), 300 µM de cada dNTP (Sigma Aldrich), 1 µM de cada iniciador, tampão do fabricante

ProofStart (Qiagen) 1×, solução Q do fabricante (Qiagen) 1× e 1,5 mM de Mg^{2+} num volume final de 50 μ l.

A reacção foi realizada como se segue: 95°C durante 5 min, seguido de 40 ciclos de 30 s a 94°C, 30 s a 55°C, 53°C, 53°C e 55°C, respectivamente, para hST3Gal III (SEQ ID NO:151), hST6GaINAc I-74 (SEQ ID NO:145), hST6GalNAC 1-140 (SEQ ID NO:147) e hST3Gal III-28/37-hST6GaINAc I-74 (SEQ ID NO:159) e 1 min a 72°C. O produto de PCR amplificado foi analisado através de electroforese num gel de agarose a 2%, de acordo com o tamanho do produto de PCR (Tabela 5):

- hST3 Gal III: foi amplificado apenas um produto de PCR a um tamanho estimado de 170 pb, que corresponde ao tamanho esperado de 174 pb (**Figura 11**).

- hST6 Gal NAc I: a **Figura 13** mostra um produto de PCR amplificado com um tamanho estimado de 270 pb, que corresponde ao tamanho esperado de 270 pb.

- hST6GalNAC I-140: foi amplificada uma sequência nucleotídica de 468 pb (**Figura 15**).

- hST3Gal III-28-hST6GaINAc I37-74: foi amplificada uma sequência nucleotídica de 249 pb (**Figura 17**).

Os produtos de PCR foram purificados utilizando o estojo de Extracção em Gel de Qiagen de acordo com as instruções do fabricante.

- O produto de PCR de hST6GaINAc I-140 foi submetido a sequenciação directa de ambas as cadeias de ADN por Genome Express (Meylan, França).

- Os produtos de PCR de hST3Gal III e hST6GalNAC 1-74 foram ligados ao CD (SEQ ID NO:43) antes de serem introduzidos no vector pcDNA3.1 (Invitrogen) para serem também sequenciados por Genome Express (Meylan, França).

- O fragmento sintético de hST3Gal III-28-hST6GalNAC 137-74 foi ligado com o CD (SEQ ID NO:43) e sequenciado directamente em ambas as cadeias antes da sua introdução no vector pcDNA3.1 (Invitrogen).

f - Quantificação dos produtos de PCR

Todos os produtos de PCR foram reunidos e concentrados de acordo com cada construção sintética. Dois volumes de etanol e 0,1 volume de acetato de sódio 3 M, pH 5,2 foram

adicionados a amostras dos produtos de PCR. A mistura foi incubada durante 30 min a -20°C e centrifugada 20 min a 10000 g a 4°C . Os sedimentos foram lavados com 100 μl de etanol a 70% e a centrifugação foi corrida durante 10 min a 10000 g a 4°C . Depois os sobrenadantes foram removidos e os sedimentos secos e ressuspensos em 50 μl de água purificada. As amostras foram conservadas a -20°C até à utilização. Cinco microlitros dos quatro fragmentos sintéticos concentrados foram carregados num gel de agarose a 2% para estimar a sua quantidade (**Figuras 11, 12, 13 e 14**).

Exemplo 3: Montagem dos domínios membranares N-terminais não catalíticos com uma actividade de sialiltransferase

a - Digestões e purificação

Todos os produtos de PCR (as construções N-terminais sintéticas ou o domínio catalítico) foram digeridos pela enzima de restrição *Bam*HI.

200 nanogramas das construções N-terminais sintéticas foram digeridos numa solução contendo 3 unidades da enzima de restrição *Bam*HI, 0,1 $\mu\text{g}.\mu\text{l}^{-1}$ de Albumina de Soro Bovino e tampão E do fabricante apropriado 1 $\times$ (Promega) num volume final de 20 μl .

500 nanogramas do CD foram digeridos com 5 unidades de *Bam*HI, 0,1 $\mu\text{g}.\mu\text{l}^{-1}$ de Albumina de Soro Bovino e tampão E 1 $\times$ (Promega) num volume final de 20 μl .

As digestões foram realizadas durante 90 min a 37°C e a incubação foi finalmente inactivada durante 15 min a 65°C . O ADN digerido foi purificado utilizando o estojo de purificação de PCR (Qiagen).

b - Ligação

Cada fragmento N-terminal foi ligado ao fragmento de CD numa solução contendo 1,5 unidades de ADN-ligase T4, tampão de ligação 1 $\times$, ambos comercializados (Promega), 62,5 ng, 78,2 ng e 75 ng, respectivamente, para hST3Gal III, hST6GalNAc I-74 e o fragmento híbrido hST3Gal III-28-hST6GalNAc I37-74 e

100 ng do CD num volume final de 20 µl. A mistura foi incubada a 15°C de um dia para o outro e seguida de um passo de inactivação durante 10 min a 70°C (**Figura 17**).

c - Amplificação da inserção sintética marcada

Para verificar a ligação correcta, os produtos de ligação foram directamente submetidos a PCR utilizando os seguintes pares de iniciadores: 5'AflIII-ST3/3'Δ4-XbaI, 5'AflIII-ST6/3'Δ4-XbaI e 5'AflIII-ST3/3'Δ4-XbaI, respectivamente, para o ST3/CD, o ST6GalNAc-74/CD e o híbrido/CD.

As reacções foram realizadas utilizando 2,5 unidades de ADN-polimerase ProofStart (Qiagen), 300 µM de cada dNTP (Sigma Aldrich), 1 µM de cada iniciador, tampão do fabricante ProofStart 1× (Qiagen), solução Q do fabricante 1× (Qiagen) e 3 mM de Mg²⁺ num volume final de 50 µl. A reacção foi realizada como se segue: 95°C durante 5 min, seguido de 40 ciclos de 1 min a 94°C, 1 min a 57°C e 1 min 30 s a 72°C.

Os produtos de PCR amplificados foram analisados através de electroforese num gel de agarose a 1,5%:

- inserção sintética hST3Gal III/CD (SEQ ID NO:167): foi amplificado um produto de PCR de 1200 pb (**Figura 12**), que corresponde ao tamanho esperado do fragmento de ADN ligado: ST3Gal III mais CD.

- inserção hST6GalNAc 1-74/CD (SEQ ID NO:161): foi amplificado um produto de PCR de 1200 pb (**Figura 14**), que corresponde ao tamanho esperado do fragmento de ADN ligado: ST6GalNAc I mais CD que tem exactamente 1225 pb.

- inserção hST3GalIII-28/37hST6GalNAc-74/CD (SEQ ID NO:175): foi amplificado um produto de PCR de 1200 pb (**Figura 17**) que corresponde ao tamanho esperado do fragmento de ADN ligado hST3GalIII-29/hST6GalNAc37-74 mais CD.

Os produtos de PCR amplificados foram purificados utilizando o estojo de Extração em Gel de Qiagen de acordo com as instruções do fabricante.

Os produtos de PCR amplificados da mesma ST sintética foram reunidos, concentrados e quantificados num gel de

agarose a 1,5% (Figuras **12**, **13 e 16**), tal como anteriormente descrito acima, para obter uma quantidade suficiente de ADN para ligar a quimera no vector pcDNA3.1.

d - Clonagem da inserção de ST sintética num vector de expressão

1 - Digestão e purificação

As inserções sintéticas tais como a inserção hST3Gal III/CD ou hST6GalNAc I-74/CD foram ligadas no vector pcDNA3.1 nos locais de restrição *Afl*III e *Xba*I.

A inserção sintética e o vector pcDNA3.1 (**Figura 12**) (Invitrogen) foram primeiro digeridos pela enzima de restrição *Xba*I:

- para cada construção, 500 ng do vector foram digeridos utilizando 2,5 unidades de *Xba*I (Promega), 0,1 $\mu\text{g} \cdot \mu\text{l}^{-1}$ de BSA e tampão D apropriado 1 $\times$, fornecido pelo fabricante (Promega), num volume final de 20 μl , seguindo as instruções do fabricante;

- a inserção foi digerida com 2 a 2,5 unidades de *Xba*I (Promega), dependendo da quantidade de inserção disponível após PCR (p.ex. 200 ng de hST3Gal III/CD, hST3Gal III-28/hST6GalNAc I 37-74/CD e 500 ng de hST6GalNAc I-74/CD foram, respectivamente, digeridos com 2 e 2,5 unidades de *Xba*I), 0,1 $\mu\text{g} \cdot \text{pL}^{-1}$ de BSA e Tampão D 1 $\times$ (Promega), num volume final de 20 μl .

As digestões foram realizadas a 37°C durante 60 min e inactivadas a 65°C durante 15 min.

Os produtos digeridos foram então submetidos a uma digestão através da enzima de restrição *Afl*III seguindo as instruções do fabricante (Biolabs): utilizando o número apropriado de unidades de enzima, dependendo da quantidade de ADN (2, 2, 2,5 e 2,5 unidades, respectivamente, para hST3Gal III/CD, hST3Gal III-28/hST6GalNAc I 37-74/CD, hST6GalNAc I-74/CD e pcDNA3.1), 0,1 $\mu\text{g} \cdot \mu\text{l}^{-1}$ de BSA e tampão 2 1 $\times$ (Biolabs) num volume final de 50 μl . As digestões foram realizadas a 37°C durante 60 min e inactivadas a 65°C durante 20 min. As

digestões foram purificadas com o estojo de purificação de PCR (Qiagen).

2 - Ligação

As condições de ligação foram calculadas de acordo com a seguinte fórmula:

$$\text{X ng inserção} = \frac{(\text{Y pb de inserção}) (\text{x ng de vector pcDNA3.1})}{\text{tamanho em pb do vector pcDNA3.1} = 5428} \times \text{razão molar} \frac{\text{Inserção}}{\text{Vector}}$$

onde o tamanho da inserção é de cerca de 1200 pb para a construção sintética (p.ex. hST3Gal III/CD, hST3Gal III-28/hST6GalNAc I 37-74/CD ou hST6GalNAc I-74/CD), o tamanho do vector pcDNA3.1 é de 5428 pb, a quantidade de vector utilizada é de 50 ng ou 100 ng e finalmente a razão molar de Inserção/Vector é de 3/1.

- A construção do vector recombinante pcDNA3.1/ST3/CD foi realizada utilizando 33 ng da quimera digerida, hST3Gal III/CD (SEQ ID NO:167), 50 ng do vector digerido pcDNA3.1, 3 unidades da enzima ADN-ligase T4 (Promega) e tampão de ligação 1× num volume final de 15 µl. A mistura foi incubada de um dia para o outro a 4°C e a reacção foi inactivada durante 10 min a 70°C.

- A segunda construção do vector recombinante pcDNA3.1/ST6/CD foi realizada utilizando 66,5 ng da quimera digerida hST6GalNAc I-74/CD (SEQ ID NO:161), 100 ng do vector digerido pcDNA3.1, 3 unidades de ADN-ligase T4 (Promega) e tampão de ligação 1× num volume final de 15 µl. A mistura foi incubada de um dia para o outro a 15°C e a reacção foi inactivada durante 10 min a 70°C.

- A terceira construção do vector recombinante pcDNA3.1/HYB/CD foi realizada utilizando 66,5 ng da quimera digerida hST3Gal III-28/hST6GalNAc I 37-74/CD (SEQ ID NO:175), 100 ng do vector pcDNA3.1, 3 unidades de enzima ADN-ligase T4 (Promega) e tampão de ligação 1× num volume final de 15 µl. A mistura foi incubada de um dia para o outro a 15°C e a reacção foi inactivada durante 10 min a 70°C.

3 - Clonagem

Transformação química de células competentes

E. coli quimicamente competente Shot® TOP10 (Invitrogen) foi transformada com o vector recombinante seguindo as instruções do fabricante. Dois microlitros de cada reacção de ligação foram adicionados a 25 µl de células quimicamente competentes e misturadas batendo suavemente. Os frascos foram incubados em gelo durante 30 min. As transformações químicas foram efectuadas durante exactamente 40 s num banho-maria a 42°C. Os frascos foram removidos do banho a 42°C e colocados em gelo durante 2 min. Duzentos e cinquenta microlitros do meio S.O.C. previamente aquecido, fornecidos por Invitrogen, foram adicionados a cada frasco sob condições estéreis. Depois, as misturas foram agitadas a 37°C durante exactamente 1 h a 225 rpm numa incubadora com agitação. Cada uma das transformações foi espalhada em placas de LB-agar (10 g de bacto-triptona, 5 g de extracto de bacto-levedura, 10 g de NaCl e 15 g de agar por litro de solução de GibcoBRL) contendo 50 µg/mL de ampicilina sob condições estéreis. As placas foram invertidas, incubadas a 37°C de um dia para o outro e mantidas a 4°C no dia seguinte até à selecção dos transformantes.

Amplificação

No final do dia, colónias individuais foram isoladas e inoculadas em 3 ml de LB (10 g de bacto-triptona, 5 g de extracto de bacto-levedura, 10 g de NaCl de Difco) contendo 50 µg/ml de ampicilina. O crescimento foi assegurado a 37°C numa incubadora com agitação de um dia para o outro. No dia seguinte, reservas de glicerol das culturas foram preparadas através da mistura de 0,85 ml das culturas com 0,15 ml de glicerol estéril e transferidas para um frasco de criopreservação. As reservas de glicerol foram armazenadas a -80°C até à verificação da sequência e outras utilizações.

Minipreparação dos plasmídeos

Os vectores recombinantes foram purificados através do procedimento de Minipreparação (Maniatis), 1,5 ml das

culturas foram isolados e centrifugados durante 5 min a 10000 g para sedimentar as bactérias. Solução I (glicose 5 mM, Tris-HCl 25 mM pH 8,0, EDTA 10 mM pH 8,0) fria foi adicionada (100 µl) para ressuspender o sedimento, 200 µl de Solução II (NaOH 0,2 N e SDS a 1%) e 150 µl de Solução III (3 M de potássio e 5 M de acetato) foram também adicionados. As misturas foram suavemente agitadas através de vórtice, colocadas em gelo 5 min e centrifugadas a 4°C, 5 min a 15000 g. Os sobrenadantes foram tomados e foi adicionado 1 volume de isopropanol. As minipreparações foram incubadas 5 min à temperatura ambiente e centrifugadas a 4°C, 5 min a 15000 g. Depois, os sobrenadantes foram removidos. O sedimento de ADN foi seco e ressuspenso em 50 µl de água. A quantidade de ADN foi medida utilizando um aparelho espectrofotómetro DO (Biophotometer, Eppendorf). São efectuadas pelo menos 30 minipreparações para pesquisar a presença da inserção de interesse para cada construção.

Um micrograma de cada uma das 30 minipreparações foi duplamente digerido utilizando a enzima de restrição *Afl*III e *Xba*I para verificar a presença da inserção. Primeiro, a solução de digestão continha 10 unidades de *Xba*I (Promega), 0,1 µg.µl⁻¹ de BSA e o tampão D apropriado do fabricante (Promega) 1× numa solução final de 20 µl. As digestões foram efectuadas num banho-maria a 37°C durante 1 h e inactivadas a 65°C durante 15 min. As amostras foram seguidamente digeridas utilizando 10 unidades da enzima de restrição *Afl*III (Biolabs), BSA 0,1× e Tampão 2 (Biolabs) 1× num volume final de 50 µl. As reacções foram realizadas num banho-maria a 37°C durante 1 h e inactivadas a 65°C durante 20 min. As amostras digeridas foram todas purificadas utilizando o estojo de purificação de PCR (Qiagen). Os produtos de digestão foram carregados num gel de agarose a 1,5% para detectar a presença da inserção quimérica. A electroforese foi corrida a 100 V durante cerca de 35 min.

No caso da pcDNA3.1/ST3/CD e pcDNA3.1/ST6/CD, foi detectada uma inserção de 1200 pb (**Figura 19C ou 23C**). O tamanho desta inserção corresponde exactamente ao tamanho esperado da quimera reconstituída ST3Gal III/CD ou ST6GalNAc 1-74/CD.

Os clones positivos foram completamente sequenciados para avaliar a sequência de ADN inserida esperada.

Sequenciação

Os passos de clonagem foram verificados através de sequenciação de ADN de ambas as cadeias por Genome Express (Meylan, França) utilizando iniciadores universais presentes dentro do vector: Promotor T7 e BGH Reverso.

As sequências finais esperadas das três quimeras ST3Gal III/CD (SEQ ID NO:167), hST3Gal III-28/hST6GalNAc I 37-74/CD (SEQ ID NO:175) e ST6GalNAc I-74/CD (SEQ ID NO:161). A sequência nucleotídica obtida foi alinhada com a sequência esperada e o alinhamento revelou uma identidade de 100% entre estas três sequências. A sequência de aminoácidos deduzida foi também alinhada com a sequência esperada da quimera (<http://www.infobiogen.fr/services/analyse/cgi-bin/alignp.in.pl>) e mostra uma identidade de 100% entre duas destas sequências de aminoácidos.

A sequência nucleotídica obtida para hST6GalNAc I-140 mostrou uma identidade de 84% com a sequência teórica. As partes em falta da sequência são os primeiros 36 e os últimos 39 nucleótidos, elas correspondem ambas aos iniciadores desenhados para sequenciar o fragmento de ADN em ambas as cadeias. Além disso o alinhamento de aminoácidos entre as sequências esperadas e as sequências resultantes mostrou uma identidade de 76,9%. Os primeiros 12 e os últimos 24 aminoácidos não se verificam devido à sua correspondência com os iniciadores específicos utilizados para a sequenciação.

Minipreparação

Após a verificação da sequência da inserção de ADN, os plasmídeos recombinantes foram amplificados em culturas de 100 ml para obter uma elevada quantidade de cada construção (pcDNA3.1/ST3/CD e pcDNA3.1/ST6/CD e pcDNA3.1/HYB/CD). Uma única colónia de cada plasmídeo de uma placa selectiva riscada de fresco foi apanhada e inoculada numa cultura iniciadora de 3 ml de meio LB contendo o antibiótico selectivo (ampicilina a 50 $\mu\text{g}.\mu\text{l}^{-1}$). As culturas foram

incubadas de um dia para o outro a 37°C com agitação vigorosa. As culturas iniciadas foram diluídas em 100 ml de meio LB selectivo e incubadas uma vez mais de um dia para o outro a 37°C com agitação vigorosa. As células bacterianas foram colhidas e os vectores recombinantes foram purificados utilizando o Estojó QIAGEN® Plasmid Midi (Qiagen) seguindo as instruções do fabricante. O rendimento foi determinado com uma espectrofotometria de UV para quantificar a concentração de ADN (biophotometer, Eppendorf). Cada construção foi armazenada a -20°C até à utilização.

Exemplo 4: Expressão funcional das construções sintéticas em células CHO

As inserções expressas são representadas na **Figura 15**.

A - Expressão transiente de ST6 sintéticas em células CHO

A linha celular CHO-K1 foi utilizada para expressar construções tal como descrito por Donadio *et al.* (2003). Resumidamente, as células são criadas em meio de Ham suplementado com 10% de Soro Fetal de Vitelo (FCS), fungizona (2,5 µg/mL) e gentamicina (50 µg/ml) a 37°C numa incubadora de CO₂ a 5%. As transfecções com as construções do vector FLAG-CMV foram realizadas utilizando o reagente LipoFectamine seguindo o procedimento recomendado pelo fabricante, com 3 µg de ADN plasmídico recombinante. As experiências de imunofluorescência foram corridas após 36-48 h de transfecção.

A dupla marcação com FITC-SNA e mAb anti-FLAG foi realizada tal como descrito por Donadio *et al.* (2003). As células foram fixadas numa solução de paraformaldeído (PFA) a 1%, saturada com BSA a 1%, lavada com solução salina tamponada com fosfato (PBS) e incubada com FITC-SNA. As células expressando uma actividade de ST6 são marcadas a verde. As amostras foram lavadas duas vezes com PBS e fixadas em PFA a 1%, mais lavadas com PBS, incubadas em NH₄Cl 0,05 M e lavadas com PBS. As células foram permeabilizadas com Triton a 0,1% em PBS e ainda incubadas em PBS contendo soro de cabra a 10%. A marcação anti-FLAG foi realizada utilizando

o mAb M2 anti-FLAG (1:600) em PBS contendo soro de cabra ou de cavalo inactivado a 5% seguido de incubação com um anticorpo secundário anti-IgG de ratinho marcado com Alexa Fluor 568 (1:200) no mesmo tampão. As células expressando a sialiltransferase marcada com FLAG são marcadas a vermelho. Após lavagens, as células foram montadas em Mowiol e mantidas a -20°C no escuro. Foi realizada microscopia confocal com um instrumento Olympus ou Leica. As imagens de confocal foram processadas com um Sistema Metamorph Imaging. Os volumes foram originalmente seguidos como Imagens TrueColor de 24-bit e transferidos para Adobe Photoshop como ficheiros RGB TIFF ou JPG. Sob tais condições, as células CHO-K1 não transfectadas não são marcadas com nenhum dos reagentes acima.

B - Actividade enzimática de ST6 sintéticas em células CHO

B1 Actividade do domínio catalítico optimizado de hST6Gal I

O CD de hST6Gal I foi clonado no vector pFLAG-citomegalovírus (pFLAG-CMV1) de Sigma, utilizando um péptido sinal de preprotipsinogénio. Esta construção e estes plasmídeos foram anteriormente caracterizados *in vitro* por Legaigneur *et al.* (2001) e mostrou-se darem uma proteína enzimática do tamanho esperado libertada para o meio celular. Este CD solúvel verificou-se igualmente activo em aceitadores glicoproteicos exógenos conhecidos com vários graus de ramificação.

Para gerar a forma solúvel do CD mínimo de hST6Gal I, foi gerado um fragmento de PCR codificado entre o 5' (5'-TAATAAAGCTTGAGGCCTCCTTCCAG-3') introduzindo um local de restrição *HindIII* e o 3' (5'-CTATTGGATCCTTAGCAGTGAATGGT-3') codificando um local *BamHI*. O fragmento foi então digerido com *HindIII/BamHI* e inserido no vector de expressão de mamífero pFLAG-CMV1 e subclonado no vector pBluescript II KS (Stratagene) para mais construções.

O CD mínimo de ST6Gal I (SEQ ID NO:44) marcado com o epitopo FLAG (**Figura 18**) é expresso a um elevado nível em células CHO (Figura 18B, D). A sua actividade e localização foram caracterizadas utilizando dupla marcação com Anticorpo monoclonal anti-FLAG e FITC-SNA. Interessantemente, o FLAG-CD mutante foi altamente eficiente na sialilação de aceptadores da superfície celular tal como mostrado nas **Figuras 18C e D**.

B2 - Expressão funcional de ST8/CD

As células CHO-K1 foram transientemente transfectadas com formas quiméricas de hST8SiaIV (1-67)/CD (SEQ ID NO:173), hST8SiaII (1-79)/CD (SEQ ID NO:171). A sua actividade e localização foram caracterizadas utilizando dupla marcação com Anticorpo monoclonal anti-FLAG e FITC-SNA, tal como mostrado na **Figura 19**.

Verificou-se que ambas as quimeras hST8SiaIV (1-67)/CD (SEQ ID NO:174) e hST8SiaII (1-79)/CD (SEQ ID NO:172) quimeras eram activas em aceptadores celulares endógenos tal como a SNA-lectina revela numa intensa marcação da superfície celular (Figura 19B, C, E e F), indicando que o CD é activo independentemente da origem da parte N-Terminal fundida com o domínio catalítico. A actividade de α 2,6-sialiltransferase é mantida para ambas as quimeras, indicando que na enzima quimérica, a informação contida no CD mínimo é necessária e suficiente para reconhecimento do aceptador e eficiência da transferência.

B3 - Expressão funcional de ST3/CD e ST3-ST6GalNAc híbrida/CD

Células CHO-K1 foram transientemente transfectadas com construções quiméricas de hST3Gal III/CD (SEQ ID NO:167) e ST3 1-28-ST6GalNAc37-74/CD (SEQ ID NO:175). A sua actividade e localização foram caracterizadas utilizando dupla marcação com o Anticorpo monoclonal anti-FLAG e FITC-SNA, tal como mostrado na **Figura 19**.

Novamente, verificou-se que ambas as quimeras eram activas em aceptadores endógenos tal como a ligação de SNA revela uma intensa marcação da superfície celular (Figura

19B, C, E e F). Isto indica que o CD é activo independentemente da origem do CT+TMD+SR adicionado. Uma vez que a actividade de $\alpha 2,6$ -sialiltransferase é mantida para ambas as quimeras, também se pode concluir que a informação contida no CD mínimo é necessária e suficiente para total eficiência de transferência.

Portanto, a fusão de sequências ST3/8 a montante do CD, não altera a expressão de uma actividade catalítica de ST6. Uma vez que a ST3 GalIII de rato e humana e mais geralmente a ST3Gal lIII de roedor e de mamífero são idênticas na sua porção N-terminal (CT+TMD+SR justamembranar), parece também que a permuta de qualquer uma destas porções peptídicas por uma heteróloga não-humana não tem efeito na actividade catalítica de ST6. Além disso, a substituição de CT, TMD e SR com porções heterólogas seleccionadas de entre enzimas de O-glicosilação tais como ST GalNAcI não altera o reconhecimento de aceitadores intracelulares nem a actividade enzimática.

B4 - Expressão funcional de hST6GalNAc I -74/CD e hST6GalNAc 258/CD

Uma abordagem semelhante foi utilizada para estimar a expressão e actividade das enzimas hST6GalNAc I (1-74)/CD (SEQ ID NO:162) e hST6GalNAc I-258/CD (SEQ ID NO:166). Ambas as transferases quiméricas são activas: a ligação de SNA revelou uma intensa sialilação semelhante da superfície celular e a marcação de FLAG mostra uma localização idêntica dentro do Golgi (**Figura 21G, H e I**).

Pode-se concluir assim que o comprimento da região SR pode ser amplamente variável e prolongado até aos 258 resíduos sem afectar a expressão funcional e a localização intracelular de uma actividade catalítica ST6Gal quimérica. Isto confirma também que o comprimento da âncora membranar N-terminal heteróloga não afecta a conformação biologicamente activa da sialiltransferase desenhada de novo.

Com base nos exemplos acima, pode-se afirmar que o processo descrito no invento pode abolir o controlo regulador da região SR sobre a actividade de CD. Como resultado, qualquer combinação possível de CT+TMD+SR entre

glicosiltransferases (sialiltransferases) da via de N- ou O-glicosilação pode ser seleccionada para equipar as células hospedeiras com uma actividade de transferase necessária através de um CD relevante. Além disso, o processo descrito no invento é também capaz de direccionar uma actividade necessária para compartimentos intracelulares nos quais as glicoproteínas neo-sintetizadas migram para a superfície celular. Esta descoberta é de crucial importância para separar e segregar correctamente proteínas sialiladas de interesse terapêutico em células (de levedura para humano) desenhadas tal como descrito no invento.

Exemplo 5: Construção de um domínio sintético híbrido: hST3Gal III/hST8Sia II/hST6GalNAc I e fusão com o CD de hST6Gal I

Foi construída uma região N-terminal híbrida (SEQ ID NO:180) contendo primeiro o CT de hST3Gal III (resíduos de aminoácidos 1 a 8; SEQ ID NO:66), depois o TMD de hST8Sia II (resíduos de aminoácidos 7 a 23; SEQ ID NO:120) e finalmente a SR de hST6GalNAc I (aminoácidos 36 a 74; SEQ ID NO:130). Um epitopo FLAG (a negrito) foi adicionado entre o codão de iniciação ATG e o segundo codão (GGA) da sequência hST3Gal III:

5'- ATG GACTACAAAGACGATGACGACAAG GGA -3'

O comprimento total deste híbrido é de 72 aminoácidos (SEQ ID NO:180) correspondendo a 216 nucleótidos (SEQ ID NO:179). Oito oligonucleótidos com sentido e sete oligonucleótidos anti-sentido correspondendo à região N-terminal híbrida descrita acima (nucleótidos 1 a 216), utilizando as sequências de hST3Gal III/hST8Sia II/hST6GalNAc I, foram desenhados e sintetizados (Eurogentec).

A fosforilação e a reacção de correspondência foram efectuadas tal como anteriormente descrito.

Amplificação por PCR da construção sintética

Após o passo de reconstituição sintética, o produto foi amplificado através de uma PCR de alta fidelidade utilizando

iniciadores específicos que trazem os locais de restrição desejados para cada extremidade do fragmento de ADN reconstituído para o ligar no vector pcDNA3.1. Os locais de restrição *Afl*III e *Bam*HI foram respectivamente introduzidos nas extremidades 5' e 3'. Foram designados os seguintes iniciadores:

5' *Afl*III-ST3: 5'-GAGCCCCTTAAGATGGACTACAAAGACGATGACG-3' e
3' ST6-*Bam*HI[°]3: 5'-TAAGGGGGATCCGGTTGTCCTCCTTGCCCT-3'

A PCR foi realizada num aparelho de PCR (I-cycler, Biorad) com 5 µl do fragmento sintético numa solução contendo tampão do fabricante ProofStart 1× e solução Q do fabricante 1×, 1,5 mM de Mg²⁺, 300 µM de cada dNTP (Sigma Aldrich), iniciador directo *Afl*III-ST3 1 µM, iniciador inverso ST6-*Bam*HI[°]3 1 µM, 2,5 unidades de ADN-polimerase ProofStart (Qiagen) num volume final de 50 µl. O perfil de ciclo térmico utilizado foi: 95°C durante 5 minutos, seguido de 40 ciclos de 30 segundos a 94°C, 30 segundos a 55°C e 1 minuto a 72°C.

Os produtos de PCR amplificados foram analisados através de electroforese num gel de agarose a 2%. Foi amplificada uma sequência nucleotídica de 240 pb que corresponde ao tamanho esperado do fragmento (**Figura 22** pista 1). Os produtos de PCR foram purificados utilizando o Estojo de Extração em Gel (QIAGEN) de acordo com as instruções do fabricante.

Quantificação do produto de PCR

Todos os produtos de PCR foram reunidos e concentrados. Foram adicionados aos produtos de PCR dois volumes de etanol e 0,1 volume de acetato de sódio 3 M, pH 5,2. A mistura foi incubada durante 30 min a -20°C e centrifugada 20 min a 10000 g a 4°C. O sedimento foi lavado com 100 µl de etanol a 70% e centrifugado novamente 10 min a 10000 g a 4°C. Depois, o sobrenadante foi removido e o sedimento seco e ressuspenso em 30 µl de água bidestilada. A amostra de ADN foi conservada a -20°C até à utilização. Três microlitros do fragmento concentrado foram carregados num gel de agarose a 2% para estimar a sua quantidade (**Figura 22** pista 3).

Ligação do domínio catalítico num vector de expressão, pcDNA3.1(+)

1 - Digestões e purificações

A sequência nucleotídica do domínio catalítico CD de hST6Gal I (SEQ ID NO:43) foi ligada no vector pcDNA3.1 nos locais de restrição *Bam*HI e *Xba*I.

Tanto o vector pcDNA3.1 (+) como o CD foram primeiro digeridos pela enzima de restrição *Bam*HI: 500 ng do CD foram digeridos com 5 unidades de *Bam*HI (Promega), 0,1 $\mu\text{g} \cdot \mu\text{L}^{-1}$ de Albumina de Soro Bovino e tampão E (Promega) 1× num volume final de 20 μL . 1 μg do vector pcDNA3.1 foi digerido com 10 unidades de *Bam*HI (Promega), 0,1 $\mu\text{g} \cdot \mu\text{L}^{-1}$ de Albumina de Soro Bovino e tampão E (Promega) 1× num volume final de 20 μL . As digestões foram realizadas durante 60 min a 37°C e as incubações foram finalmente inactivadas durante 15 min a 65°C. Os produtos digeridos foram purificados utilizando o Estojo de purificação de PCR (QIAGEN).

Os produtos digeridos foram então submetidos a uma digestão pela enzima de restrição *Xba*I (Promega), seguindo as instruções do fabricante: utilizando o número de unidades da enzima apropriado, dependendo da quantidade de ADN (5 e 6, respectivamente para CD e pcDNA3.1), 0,1 $\mu\text{g} \cdot \mu\text{L}^{-1}$ de BSA e tampão D (Promega) 1× num volume final de 50 μL . As digestões foram realizadas a 37°C durante 60 min. As digestões foram purificadas utilizando o Estojo de purificação de PCR (QIAGEN).

2 - Ligação

As condições de ligação foram calculadas de acordo com a fórmula anteriormente mencionada. O tamanho da inserção CD é de cerca de 960 pb, o tamanho do vector pcDNA3.1 é de 5428 pb, a quantidade de vector utilizada é de 100 ng e finalmente a razão molar de Inserção/Vector é de 3/1.

A construção do vector recombinante pcDNA3.1/CD foi realizada utilizando 53 ng do CD duplamente digerido, 100 ng de vector pcDNA3.1 duplamente digerido, 6 unidades de enzima

ADN-ligase T4 (Promega) e tampão de ligação 1× num volume final de 20 µl. A mistura foi incubada de um dia para o outro a 15°C e a reacção foi inactivada durante 10 min a 70°C.

3 - Clonagem

Transformação química de células competentes

E. coli Quimicamente Competente Shot® TOP10 (Invitrogen, Life Technologies) foi transformada com o vector recombinante seguindo as instruções do fabricante. Cinco microlitros do produto de ligação foram adicionados a uma alíquota de 50 µl de células TOP 10 e misturadas batendo suavemente. A mistura foi incubada em gelo durante 30 minutos, depois sujeita a choque térmico durante exactamente 30 segundos num banho-maria a 42°C e colocada em gelo durante 2 min. Duzentos e cinquenta microlitros de meio S.O.C. previamente aquecido (Invitrogen) foram adicionados à reacção de transformação e incubados a 37°C durante uma hora com agitação a 250 rpm. A reacção de transformação foi espalhada em placas de LB-agar (composição anteriormente mencionada) contendo 50 µg.ml⁻¹ de ampicilina. As placas foram incubadas de um dia para o outro a 37°C.

Amplificação

No final do dia seguinte, as colónias individuais foram apanhadas e incubadas, com 3 ml de caldo LB (composição anteriormente mencionada) contendo 50 µg.ml⁻¹ de ampicilina, de um dia para o outro a 37°C com agitação a 250 rpm.

Minipreparação de plasmídeos

Os vectores recombinantes foram purificados a partir de alíquotas de 1,5 ml das culturas anteriores utilizando o Estojo QIAprep Spin Miniprep proporcionado por QIAGEN e seguindo as instruções do fabricante. Foram efectuadas vinte e quatro minipreparações para pesquisar a presença da inserção.

Cinco microlitros de cada uma das 24 minipreparações foram duplamente digeridos utilizando as enzimas de restrição

XbaI e *BamHI* para verificar a presença da inserção. Primeiro, a solução de digestão continha 12 unidades de *XbaI* (Promega), 0,1 $\mu\text{g} \cdot \mu\text{l}^{-1}$ de BSA e tampão D (Promega) 1× num volume final de 10 μl . As digestões foram efectuadas durante 2 horas a 37°C. As amostras foram depois digeridas utilizando 20 unidades de *BamHI* (Promega) e tampão E (Promega) 1× num volume final de 20 μl . As reacções foram realizadas de um dia para o outro a 37°C. Os produtos de digestão foram carregados num gel de agarose a 1,5% para detectar a presença do CD. Foi detectada uma inserção de cerca de 1000 pb (**Figura 23**) que corresponde ao tamanho esperado do CD (960 pb).

Foi seleccionado um clone positivo para os passos seguintes da construção.

Ligação da região N-terminal no vector recombinante pcDNA3.1/CD

1 - Digestões e purificações

A sequência nucleotídica da região N-terminal (SEQ ID NO:179) foi ligada no vector recombinante pcDNA3.1/CD nos locais de restrição *BamHI* e *AflIII*. Tanto o fragmento N-terminal como o vector recombinante pcDNA3.1/CD foram primeiro digeridos pela enzima de restrição *BamHI*: 300 ng do fragmento N-terminal foram digeridos com 5 unidades de *BamHI* (Promega), 0,1 $\mu\text{g} \cdot \mu\text{l}^{-1}$ de BSA e tampão E (Promega) 1× num volume final de 20 μl . 2 μg do vector recombinante pcDNA3.1/CD foram digeridos com 20 unidades de *BamHI* (Promega), 0,1 $\mu\text{g} \cdot \mu\text{l}^{-1}$ de BSA e tampão E (Promega) 1× num volume final de 20 μl .

As digestões foram efectuadas durante 100 min a 37°C e as incubações foram finalmente inactivadas durante 15 min a 65°C. Os produtos digeridos foram purificados utilizando o Estojó de purificação de PCR (QIAGEN).

Os produtos digeridos foram então submetidos a uma digestão pela enzima de restrição *AflIII* (New England BIOLABS), seguindo as instruções do fabricante: utilizando o número apropriado de unidades da enzima, dependendo da quantidade de ADN (10 e 20, respectivamente para o fragmento

N-terminal e vector recombinante pcDNA3.1/CD), BSA 1× e tampão 2 (New England BIOLABS) 1× num volume final de 50 µl. As digestões foram realizadas a 37°C durante 60 min e as incubações foram finalmente inactivadas durante 20 min a 65°C. As digestões foram purificadas utilizando o Estojó de purificação de PCR (QIAGEN).

2 - Ligação

As condições de ligação foram calculadas de acordo com a fórmula anteriormente mencionada. O tamanho do fragmento N-terminal da inserção é de 240 pb, o tamanho do vector recombinante pADNc3.1/CD é de 6388 pb, a quantidade de vector utilizada é de 100 ng e finalmente a razão molar de Inserção/Vector é de 1/1.

A construção do vector recombinante pADNc3.1/fragmento N-terminal/CD foi realizada utilizando 5 ng do fragmento N-terminal digerido, 100 ng do vector recombinante digerido pADNc3.1/CD, 6 unidades da enzima ADN-ligase T4 (Promega) e tampão de ligação 1× num volume final de 20 µl. A mistura foi incubada de um dia para o outro a 15°C e a reacção foi inactivada durante 10 min a 70°C.

3 - Clonagem

Transformação química de células competentes

E. coli Quimicamente Competente Shot® TOP10 (Invitrogen) foi transformada com o produto de ligação seguindo as instruções do fabricante e tal com anteriormente descrito.

Amplificação

No final do dia seguinte, colónias individuais foram apanhadas e incubadas com 3 ml de caldo LB (composição anteriormente mencionada) contendo 50 µg.ml⁻¹ de ampicilina, de um dia para o outro a 37°C com agitação a 250 rpm.

Minipreparação de plasmídeos

Os vectores recombinantes foram purificados a partir de alíquotas de 1,5 ml das culturas anteriores utilizando o Estojo QIAprep Spin Miniprep proporcionado por QIAGEN seguindo as instruções do fabricante. Foram efectuadas vinte e quatro minipreparações para pesquisar a presença da inserção.

Cinco microlitros de cada uma das 24 minipreparações foram duplamente digeridos utilizando as enzimas de restrição *XbaI* e *AflIII* para verificar a presença da inserção (fragmento N-terminal + CD; SEQ ID NO:181).

A solução de digestão continha 20 unidades de *XbaI* (BIOLABS), 20 unidades de *AflIII* (BIOLABS), BSA (BIOLABS) 1× e tampão 2 (BIOLABS) 1× num volume final de 20 µl. As digestões foram efectuadas durante 2 horas e 30 min a 37°C. Os produtos de digestão foram carregados num gel de agarose a 1,5% para detectar a presença tanto da região N-terminal como de CD. Foi detectada uma inserção de cerca de 1200 pb (**Figura 24a**) que corresponde ao tamanho esperado da inserção.

Foi finalmente efectuada uma PCR em cinco amostras de minipreparação para verificar a presença do ADN da inserção (fragmento N-terminal seguido do CD: SEQ ID NO:181). A PCR foi realizada num aparelho de PCR (I-cycler, Biorad) com 1 µl de cada uma das amostras da minipreparação numa solução contendo tampão do fabricante ProofStart 1× e Solução Q do fabricante 1×, 3 mM de Mg²⁺, 300 µM de cada dNTP (Sigma Aldrich), 1 µM de iniciador directo *AflIII*-ST3, iniciador reverso Δ4-*XbaI* 1 µM, 2,5 unidades de ADN-polimerase ProofStart (Qiagen) num volume final de 50 µl. Os oligonucleótidos incluídos na reacção foram como se segue:

5'*AflIII*-ST3: 5'-GAGCCCCTTAAGATGGACTACAAAGACGATGACG-3' e
3' Δ4-*XbaI*: 5'-TAACCCTCTAGATTAGCAGTGAATGGTCCGGAAGC-3'

O perfil de ciclo térmico utilizado foi: 95°C durante 5 minutos, seguido de 40 ciclos de 1 min a 94°C, 1 min a 57°C e 1 minuto 30 s a 72°C. Os produtos de PCR amplificados foram analisados através de electroforese num gel de agarose a 1,5%: uma sequência nucleotídica de 1200 pb foi amplificada

para cada amostra de minipreparação que corresponde ao tamanho esperado da inserção total (**Figura 24b**).

Sequenciação

Os passos de clonagem foram verificados através de sequenciação de ambas as cadeias de ADN através de GENOME express (Meylan, França) utilizando iniciadores universais presentes dentro do vector: promotor T7 e BGH inverso.

A sequência nucleotídica obtida foi alinhada com a sequência esperada e o alinhamento revelou uma identidade de 100%. A sequência de aminoácidos deduzida foi também alinhada com a sequência esperada SEQ ID NO:181 da quimera e mostra uma identidade de 100%.

Midipreparação

Após a verificação da sequência da inserção de ADN, o plasmídeo recombinante foi amplificado em 100 ml de cultura para obter uma elevada quantidade da construção. Quatrocentos microlitros da correspondente cultura em caldo LB de 3 ml mencionada anteriormente foram inoculados em 100 ml de caldo LB contendo 50 $\mu\text{g}.\text{ml}^{-1}$ de ampicilina e incubados de um dia para o outro a 37°C com agitação a 250 rpm. As células bacterianas foram colhidas e os vectores recombinantes foram purificados utilizando o Estojo QIAGEN Plasmid Midi (Qiagen) seguindo as instruções do fabricante. O rendimento foi determinado com um espectrofotómetro de UV para quantificar a concentração do ADN. A construção foi armazenada a -20°C.

Cultura celular: expressão da enzima quimérica hST3Gal III/hST8Sia II/hST6GalNAc I/CD (SEQ ID NO:182) em células CHO

Células CHO-K1 foram rotineiramente cultivadas a 37°C com CO₂ a 5% em DMEM suplementado com FBS a 10%, 2,5 $\mu\text{g}/\text{ml}$ de fungizona e 50 $\mu\text{g}/\text{ml}$ de gentamicina. No dia anterior à transfecção, as células foram semeadas a uma densidade de 100000 células/ml numa placa de 6 poços contendo 3 lamelas de vidro por poço. As células CHO-K1 foram transientemente transfectadas com o reagente de transfecção FuGENE 6 (Roche) utilizando uma razão de 3:1.

Após um dia de cultura, as células foram fixadas com paraformaldeído a 1% durante 15 minutos à temperatura ambiente (TA). As células foram lavadas três vezes com PBS, bloqueadas durante 30 minutos à TA com BSA a 1% e depois incubadas com 10 µg/ml de SNA-FITC (vector laboratories) durante uma hora a 4°C. As células foram sucessivamente lavadas três vezes com PBS, fixadas com paraformaldeído a 3% durante 20 minutos à TA, mais lavadas com PBS e incubadas com NH₄Cl 0,05 M durante 10 minutos à TA. Após lavagens com PBS, as células foram permeabilizadas com Triton X-100 a 0,25% durante 5 minutos à TA e bloqueadas com soro de cabra inativado a 10% e BSA a 1% durante 30 minutos à TA. As células foram incubadas durante uma hora à TA com 1 µg/ml de anticorpo monoclonal anti-FLAG M2 (Sigma) em PBS contendo soro de cabra inativado a 5% e BSA a 0,5%. As células foram lavadas três vezes com PBS e incubadas durante 45 minutos com 5,7 µg/ml de anticorpo secundário de cabra anti-ratinho conjugado com Alexa Fluor®594 (Molecular Probes) em PBS contendo soro de cabra inativado a 5% e BSA a 0,5%. Após lavagens com PBS, as células foram montadas em Mowiol e analisadas utilizando um microscópio Olympus (**Figura 25**).

Referências

Andersen, D.C. & Krummen, L. (2002) "Recombinant protein expression for therapeutic applications". *Curr. Opin. Biotechnol.*, 13(2): 117-23.

Angata, K., Nakayama, J., Fredette, B., Chong, K., Ranscht, B. & Fukuda, M. (1997) "Human STX polysialyltransferase forms the embryonic form of the neural cell adhesion molecule. Tissue-specific expression, neurite outgrowth, and chromosomal localization in comparison with another polysialyltransferase, PST". *J. Biol. Chem.*, 272(11): 7182-90.

Angata, K., Suzuki, M., McAuliffe, J., Ding, Y., Hindsgaul, O. & Fukuda, M. (2000) "Differential biosynthesis of polysialic acid on neural cell adhesion molecule (NCAM) and oligosaccharide acceptors by three distinct alpha 2,8-

sialyltransferases, ST8Sia IV (PST), ST8Sia II (STX), and ST8Sia III". *J. Biol. Chem.*, 275(24): 18594-601.

Angata, K., Yen, T.Y., El-Battari, A., Macher, B.A. & Fukuda, M. (2001) "Unique disulfide bond structures found in ST8Sia IV polysialyltransferase are required for its activity". *J. Biol. Chem.*, 276(18): 15369-77.

Ashwell, G. & Harford, J. (1982) "Carbohydrate-specific receptors of the liver". *Annu. Rev. Biochem.*, 51: 531-54.

Aumiller, J.J., Hollister, J.R., & Jarvis, D.L. (2003) "A transgenic insect cell line engineered to produce CMP-sialic acid and sialylated glycoproteins". *Glycobiology*, 13(6): 497-507.

Baneyx, F. (1999) "Recombinant protein expression in *Escherichia coli*". *Curr. Opin. Biotechnol.*, 10(5): 411-21.

Berger, E.G. (2002) "Ectopic localizations of Golgi glycosyltransferases". *Glycobiology*, 12(2): 29R-36R.

Berger, E.G., Burger, P., Borsig, L., Malissard, M., Felner, K.M., Zeng, S., Dinter, A. (1998) "Immunodetection of glycosyltransferases: prospects and pitfalls". *Adv. Exp. Med. Biol.*, 435: 119-32.

Blanchard, S. (2004) "Ingenierie de glicoside hydrolases pour la glycosylation des protéines recombinantes". Thèse de l'université Joseph Fourier, pag. 272.

Bragonzi, A., Distefano, G., Buckberry, L.D., Acerbis, G., Foglieni, C., Lamotte, D., Campi, G., Marc, A., Soria, M.R., Jenkins, N., Monaco, L. (2000) "A new Chinese hamster ovary cell line expressing alpha2,6-sialyltransferase used as universal host for the production of human-like sialylated recombinant glycoproteins". *Biochim. Biophys. Acta*, 1474(3): 273-82.

Breton, C. & Imberty, A. (1999) "Structure/function studies of glycosyltransferases". *Curr. Opin. Struct. Biol.*, 9(5): 563-71.

Breton, C., Mucha, J. & Jeanneau, C. (2001) "Structural and functional features of glycosyltransferases". *Biochimie.*, 83(8): 713-8.

Bretscher, M.S. & Munro, S. (1993) "Cholesterol and the Golgi apparatus". *Science*, 261(5126): 1280-1.

Bretthauer, R.K. (2003) "Genetic engineering of *Pichia pastoris* to humanize N-glycosylation of proteins". *Trends Biotechnol.*, 21(11): 459-62.

Bianucci, A.M. & Chiellini, F. (2000) "A 3D model for the human hepatic asialoglycoprotein receptor (ASGP-R)". *J. Biomol. Struct. Dyn.*, 18(3): 435-51.

Chang, M.L., Eddy, R.L., Shows, T.B., Lau, J.T. (1995) "Three genes that encode human beta-galactoside alpha2,3-sialyltransferases. Structural analysis and chromosomal mapping studies". *Glycobiology*, 5(3): 319-25.

Chen, C. & Colley, K.J. (2000) "Minimal structural and glycosylation requirements for ST6Gal I activity and trafficking". *Glycobiology*, 10(5): 531-83.

Chen, C., Ma, J., Lazic, A., Backovic, M., Colley, K.J. (2000) "Formation of insoluble oligomers correlates with ST6Gal I stable localization in the golgi". *J. Biol. Chem.*, 275(18): 13819-26.

Chen, T.L., Chen, C., Bergeron, N.Q., Close, B.E., Bohrer, T.J., Vertel, B.M. & Colley, K.J. (2003) "The two rat alpha2,6-sialyltransferase (ST6Gal I) isoforms: evaluation of catalytic activity and intra-Golgi localization". *Glycobiology*, 13(2): 109-17.

Colley, K.J. (1997) "Golgi localization of glycosyltransferases: more questions than answers". *Glycobiology*, 7(1): 1-13.

Crocker, P.R. & Feizi, T. (1996) "Carbohydrate recognition systems: functional triads in cell-cell interactions". *Curr. Opin. Struct. Biol.*, 6(5): 679-91.

Datta, A.K., Sinha, A. & Paulson, J.C. (1998) "Mutation of the sialyltransferase S-sialylmotif alters the kinetics of the donor and acceptor substrates". *J. Biol. Chem.*, 273(16): 9608-14.

Datta, A.K. & Paulson, J.C. (1995) "Sialylmotifs of sialyltransferases". *Indian J. Biochem. Biophys.*, 34(1-2): 157-65.

Donadio, S., Dubois, C., Fichant, G., Roybon, L., Guillemot, J.C., Breton, C. & Ronin, C. (2003) "Recognition of cell surface acceptors by two human alpha-2,6-sialyltransferases produced in CHO cells". *Biochimie*, 85(3-4): 311-21.

Drickamer, K. (1993) "A conserved disulphide bond in sialyltransferases". *Glycobiology*, 3(1): 2-3.

Feizi, T. (1993) "Carbohydrate-protein interactions in capillary morphogenesis" *Glycobiology*, 3(6): 547-8.

Fenteany, F.H. & Colley, K.J. (2005) "Multiple signals are required for alpha2,6-sialyltransferase (ST6Gal I) oligomerization and Golgi localization". *J. Biol. Chem.*, (7): 5423-9.

Gabius, H.J., Siebert, H.C., Andre, S., Jimenez-Barbero, J. & Rudiger, H. (2004) "Chemical biology of the sugar code". *Chembiochem.*, 5(6): 740-64.

Geremia, R.A., Harduin-Lepers, A. & Delannoy, P. (1997) "Identification of two novel conserved amino acid residues in eukaryotic sialyltransferases: implications for their mechanism of action". *Glycobiology*, 7(2): v-vii.

Giordanengo, V., Bannwarth, S., Laffont, C., Van Miegem, V., Harduin-Lepers, A., Delannoy, P. & Lefebvre, J.C. (1997) "Cloning and expression of cDNA for a human Gal(beta1-

3)GalNAc alpha2,3-sialyltransferase from the CEM T-cell line". *Eur. J. Biochem.*, 247(2): 558-66.

Grabenhorst, E. & Conradt, H.S. (1999) "The cytoplasmic, transmembrane, and stem regions of glycosyltransferases specify their *in vivo* functional sublocalization and stability in the Golgi". *J. Biol. Chem.*, 274(51): 36107-16.

Grabenhorst, E., Schlenke, P., Pohl, S., Nimtz, M. & Conradt, H.S. (1999) "Genetic engineering of recombinant glycoproteins and the glycosylation pathway in mammalian host cells". *Glycoconj. J.*, 16(2): 81-97.

Hamilton, S.R., Bobrowicz, P., Bobrowicz, B., Davidson, R.C., Li, H., Mitchell, T., Nett, J.H., Rausch, S., Stadheim, T.A., Wischniewski, H., Wildt, S. & Gerngross, T.U. (2003) "Production of complex human glycoproteins in yeast". *Science*, 301, 1244-1246.

Harduin-Lepers, A., Stokes, D.C., Steelant, W.F., Samyn-Petit, B., Krzewinski-Recchi, M.A., Vallejo-Ruiz, V., Zanetta, J.P., Auge, C. & Delannoy, P. (2000) "Cloning, expression and gene organization of a human Neu5Ac alpha 2-3Gal beta 1-3GalNAc alpha2,6-sialyltransferase: hST6GalNAcIV". *Biochem. J.*, 352(Pt 1): 37-48.

Harduin-Lepers, A., Vallejo-Ruiz, V., Krzewinski-Recchi, M.A., Samyn-Petit, B., Julien, S. & Delannoy, P. (2001) "The human sialyltransferase family". *Biochimie*, 83(8): 727-37.

Helenius, A. & Aeby, M. (2001) "Intracellular functions of N-linked glycans". *Science*, 291 (5512): 2364-9.

Herscovics, A. & Orlean, P. (1993) "Glycoprotein biosynthesis in yeast". *FASEB J.*, 7(6): 540-50.

Hudgin, R.L., Pricer, W.E. Jr, Ashwell, G., Stockert, R.J. & Morell, A.G. (1974) "The isolation and properties of a rabbit liver binding protein specific for asialoglycoproteins". *J. Biol. Chem.*, 249(17): 5536-43.

Ikehara, Y., Shimizu, N., Kono, M., Nishihara, S., Nakanishi, H., Kitamura, T., Narimatsu, H., Tsuji, S. & Tatematsu, M. (1999) "A novel glycosyltransferase with a polyglutamine repeat; a new candidate for GD1 alpha synthase (ST6GalNAc V)(1)". *FEBS Lett.*, 463(1-2): 92-6.

Ishii, A., Ohta, M., Watanabe, Y., Matsuda, K., Ishiyama, K., Sakoe, K., Nakamura, M., Inokuchi, J., Sanai, Y. & Saito, M. (1998) "Expression cloning and functional characterization of human cDNA for ganglioside GM3 synthase". *J. Biol. Chem.*, 273(48): 31652-5.

Jarvis, D.L., Kowar, Z.S. & Hollister, J.R. (1998) "Engineering N-glycosylation pathways in the baculovirus-insect cell system". *Curr. Opin. Biotechnol.*, 9(5): 528-33.

Jeanneau, C. (2003) "Bioanalyse et ingénierie de glycosyltransferase". Thèse de l'université de Paris 7, pag. 179.

Jeanneau, C., Chazalet, V., Auge, C., Soumpasis, D.M., Harduin-Lepers, A., Delannoy, P., Imberty, A. & Breton, C. (2004) "Structure-function analysis of the human sialyltransferase ST3Gal I: role of n-glycosylation and a novel conserved sialylmotif". *J. Biol. Chem.*, 279(14): 13461-8.

Kawasaki, G. & Ashwell, T. (1976) "Carbohydrate structure of glycopeptides isolated from an hepatic membrane-binding protein specific for asialoglycoproteins". *J. Biol. Chem.*, 251(17): 5292-9.

Kim, Y.J., Kim, K.S., Kim, S.H., Kim, C.H., Ko, J.H., Choe, I.S., Tsuji, S. & Lee, Y.C. (1996) "Molecular cloning and expression of human Gal beta 1,3GalNAc alpha 2,3-sialyltransferase (hST3Gal II)". *Biochem. Biophys. Res. Commun.*, 228(2): 324-7.

Kitagawa, H. & Paulson, J.C. (1993) "Cloning and expression of human Gal beta 1,3(4)GlcNAc alpha 2,3-sialyltransferase". *Biochem. Biophys. Res. Commun.*, 194(1): 375-82.

Kitagawa, H. & Paulson, J.C. (1994a) "Differential expression of five sialyltransferase genes in human tissues". *J. Biol. Chem.*, 269(27): 17872-8.

Kitagawa, H. & Paulson, J.C. (1994b) "Cloning of a novel alpha 2,3-sialyltransferase that sialylates glycoprotein and glycolipid carbohydrate groups". *J. Biol. Chem.*, 269(2): 1394-401.

Kitagawa, H., Mattei, M.G. & Paulson, J.C. (1996) "Genomic organization and chromosomal mapping of the Gal beta 1,3GalNAc/Gal beta 1,4GlcNAc alpha 2,3-sialyltransferase". *J. Biol. Chem.*, 271(2): 931-8.

Kitazume-Kawaguchi, S., Dohmae, N., Takio, K., Tsuji, S. & Colley, K.J. (1999) "The relationship between ST6Gal I Golgi retention and its cleavage-secretion". *Glycobiology*, 9(12): 1397-406.

Kitazume-Kawaguchi, S., Kabata, S. & Arita, M. (2001) "Differential biosynthesis of polysialic or disialic acid Structure by ST8Sia II and ST8Sia IV". *J. Biol. Chem.*, 276(19): 15696-703.

Kono, M., Yoshida, Y., Kojima, N. & Tsuji, S. (1996) "Molecular cloning and expression of a fifth type of alpha2,8-sialyltransferase (ST8Sia V). Its substrate specificity is similar to that of SAT-V/III, which synthesize GD1c, GT1a, GQ1b and GT3". *J. Biol. Chem.*, 271(46): 29366-71.

Kono, M., Tsuda, T., Ogata S., Takashima, S., Liu, H., Hamamoto, T., Itzkowitz, S.H., Nishimura, S. & Tsuji, S. (2000) "Redefined substrate specificity of ST6GalNAc II: a second candidate sialyl-Tn synthase". *Biochem. Biophys. Res. Commun.*, 272(1): 94-7.

Kornfeld, R. & Kornfeld, S. (1985) "Assembly of asparagine-linked oligosaccharides". *Annu. Rev. Biochem.*, 54: 631-64.

Krzewinski-Recchi, M.A., Julien, S., Juliant, S., Teinturier-Lelievre, M., Samyn-Petit, B., Montiel, M.D., Mir, A.M., Cerutti, M., Harduin-Lepers, A. & Delannoy, P. (2003) "Identification and functional expression of a second human beta-galactoside alpha2,6-sialyltransferase, ST6Gal II". *Eur. J. Biochem.*, 270(5): 950-61.

Kurosawa, N., Kojima, N., Inoue, M., Hamamoto, T. & Tsuji, S. (1994) "Cloning and expression of Gal beta 1,3GalNAc-specific GalNAc alpha2,6-sialyltransferase". *J. Biol. Chem.*, 269(29): 9048-53.

Kurosawa, N., Inoue, M., Yoshida, Y. & Tsuji, S. (1996) "Molecular cloning and genomic analysis of mouse Galbeta1,3GalNAc-specific GalNAc alpha2,6-sialyltransferase". *J. Biol. Chem.*, 21 Jun 1996; 271(25): 15109-16.

Kurosawa, N., Takashima, S., Kono, M., Ikehara, Y., Inoue, M., Tachida, Y., Narimatsu, H. & Tsuji, S. (2000) "Molecular cloning and genomic analysis of mouse GalNAc alpha2,6-sialyltransferase (ST6GalNAc I)". *J. Biochem.*, 127(5): 845-54.

Laroy, W., Ameloot, P., & Contreras, R. (2001) "Characterization of sialyltransferase mutants using surface plasmon resonance". *Glycobiology*, 11(3): 175-82.

Lee, Y.C., Kim, Y.J., Lee, K.Y., Kim, K.S., Kim, B.U., Kim, H.N., Kim, C.H. & Do, S.I. (1998) "Cloning and expression of cDNA for a human Sia alpha 2,3Gal beta 1,4GlcNA: alpha2,8-sialyltransferase (hST8Sia III)". *Arch. Biochem. Biophys.*, 360(1): 41-6.

Lee, Y.C., Kaufmann, M., Kitazume-Kawaguchi, S., Kono, M., Takashima, S., Kurosawa, N., Liu, H., Pircher, H. & Tsuji, S. (1999) "Molecular cloning and functional expression of two members of mouse NeuAcalpha2,3Galbeta1,3GalNAc GalNAcalpha2,6-sialyltransferase family, ST6GalNAc III and IV". *J. Biol. Chem.*, 274(17): 11958-67.

Legaigreur, P., Brioude, B. & Ronin, C. (1999) "Antigènes glycaniques recombinants". *Immunoanal. Biol. Spec.*, 14: 297-307.

Legaigreur, P., Breton, C., El Battari, A., Guillemot, J.C., Auge, C., Malissard, M., Berger, E.G. & Ronin, C. (2001) "Exploring the acceptor substrate recognition of the human beta-galactoside alpha2,6-sialyltransferase". *J. Biol. Chem.*, 276(24): 21608-17.

Lerouge, P., Bardor, M., Pagny, S., Gomord, V. & Faye, L. (2000) "N-glycosylation of recombinant pharmaceutical glycoproteins produced in transgenic plants: towards an humanisation of plant N-glycans". *Curr. Pharm. Biotechnol.*, 1(4): 347-54.

Livingston, B.D. & Paulson, J.C. (1993) "Polymerase chain reaction cloning of a developmentally regulated member of the sialyltransferase gene family". *J. Biol. Chem.*, 268(16): 11504-7.

Ma, J., Qian, R., Rausa, F.M. 3rd & Colley, K.J. (1997) "Two naturally occurring alpha2,6-sialyltransferase forms with a single amino acid change in the catalytic domain differ in their catalytic activity and proteolytic processing". *J. Biol. Chem.*, 272(1): 672-9.

Machamer, C.E. (1991) "Golgi retention signals: do membranes hold the key?" *Trends Cell Biol.*, 1(6): 141-4.

Maras, M., Van Die, I., Contreras, R. & Van den Hondel, C.A.M.J.J. (1999) "Filamentous fungi as production organisms for glycoproteins of bio-medical interest". *Glycoconj. J.*, 16: 99-107.

Marchal, I., Jarvis, D.L., Cacan, R. & Verbert, A. (2001) "Glycoproteins from insect cells: sialylated or not?" *Biol. Chem.*, 382: 151-159.

Martina, J.A., Daniotti, J.L. & Maccioni, H.J. (1998) "Influence of N-glycosylation and N-glycan trimming on the

activity and intracellular traffic of GD3 synthase". *J. Biol. Chem.*, 273(6): 3725-31.

Meier, M., Bider, M.D., Malashkevich, V.N., Spiess, M. & Burkhard, P. (2000) "Crystal structure of the carbohydrate recognition domain of the H1 subunit of the asialoglycoprotein receptor". *J. Mol. Biol.*, 300(4): 857-65.

Morell, A.G., Gregoriadis, G., Scheinberg, I.H., Hickman, J. & Ashwell, G. (1971) "The role of sialic acid in determining the survival of glycoproteins in the circulation". *J. Biol. Chem.*, 246(5): 1461-7.

Muhlenhoff, M., Manegold, A., Windfuhr, M., Gotza, B. & Gerardy-Schahn, R. (2001) "The impact of N-glycosylation on the functions of polysialyltransferases". *J. Biol. Chem.*, 276(36): 34066-73.

Munro, S. (1991) "Sequences within and adjacent to the transmembrane segment of alpha-2,6-sialyltransferase specify Golgi retention". *EMBO J.*, 10(12): 3577-88.

Bretscher, M.S. & Munro, S. (1993) "Cholesterol and the Golgi apparatus". *Science*, 261(5126): 1280-1.

Munro, S. (1995) "An investigation of the role of transmembrane domains in Golgi protein retention". *EMBO J.*, 14(19): 4695-704.

Munro, S. (1998) "Localization of proteins to the Golgi apparatus". *Trends Cell Biol.*, 8(1): 11-5.

Nakayama, J., Fukuda, M.N., Fredette, B., Ranscht, B. & Fukuda, M. (1995) "Expression cloning of a human polysialyltransferase that forms the polysialylated neural cell adhesion molecule present in embryonic brain". *Proc. Natl. Acad. Sci. U.S.A.*, 92(15): 7031-5.

Nakayama, J., Fukuda, M.N., Hirabayashi, Y., Kanamori, A., Sasaki, K., Nishi, T. & Fukuda, M. (1996) "Expression cloning of a human GT3 synthase. GD3 AND GT3 are synthesized by a single enzyme". *J. Biol. Chem.*, 271(7): 3684-91.

Nara, K., Watanabe, Y., Kawashima, I., Tai, T., Nagai, Y. & Sanai, Y. (1996) "Acceptor substrate specificity of a cloned GD3 synthase that catalyzes the biosynthesis of both GD3 and GD1c/GT1a/GQ1b". *Eur. J. Biochem.*, 238(3): 647-52.

Narimatsu, H. (2004) "Construction of a human glycogene library and comprehensive functional analysis". *Glycoconj. J.*, 21(1-2): 17-24.

Okajima, T., Fukumoto, S., Miyazaki, H., Ishida, H., Kiso, M., Furukawa, K., Urano, T. & Furukawa, K. (1999) "Molecular cloning of a novel alpha2,3-sialyltransferase (ST3Gal VI) that sialylates type II lactosamine structures on glycoproteins and glycolipids". *J. Biol. Chem.*, 274(17): 11479-86.

Okajima, T., Chen, H.H., Ito, H., Kiso, M., Tai, T., Furukawa, K., Urano, T. & Furukawa, K. (2000) "Molecular cloning and expression of mouse GD1alpha/GT1alpha/GQ1balpha synthase (ST6GalNAc VI) gene". *J. Biol. Chem.*, 275(10): 6717-23.

Opat, A.S., van Vliet, C. & Gleeson, P.A. (2001) "Trafficking and localisation of resident Golgi glycosylation enzymes". *Biochimie*, 83(8): 763-73.

Paulson, J.C. (1989) "Glycoproteins: what are the sugar chains for?" *Trends Biochem. Sci.*, 14(7): 272-6.

Paulson, J.C., Weinstein, J., Ujita, E.L., Riggs, K.J. & Lai, PH. (1987) "The membrane-binding domain of a rat liver Golgi sialyltransferase". *Biochem. Soc. Trans.*, 15(4): 618-20.

Paulson, J.C. & Colley, K.J. (1989) "Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation". *J. Biol. Chem.*, 264(30): 17615-8.

Pilatte, Y., Bignon, J. & Lambre, C.R. (1993) "Sialic acids as important molecules in the regulation of the immune

system: pathophysiological implications of sialidases in immunity". *Glycobiology*, 3(3): 201-18.

Qian, R., Chen, C. & Colley, K.J. (2001) "Location and mechanism of alpha2,6-sialyltransferase dimer formation. Role of cysteine residues in enzyme dimerization, localization, activity, and processing". *J. Biol. Chem.*, 276(31): 28641-9.

Reid, M.E. & Lomas-Francis, C. (2002) "Molecular approaches to blood group identification". *Curr. Opin. Hematol.*, 9(2): 152-9.

Roth, J., Taatjes, D.J., Lucocq, J.M., Weinstein, J. & Paulson, J.C. (1985) "Demonstration of an extensive trans-tubular network continuous with the Golgi apparatus stack that may function in glycosylation". *Cell*, 43(1): 287-95.

Roth, J. (1987) "Subcellular organization of glycosylation in mammalian cells". *Biochim. Biophys. Acta*, 906(3): 405-36.

Roth, J. (1991) "Localization of glycosylation sites in the Golgi apparatus using immunolabeling and cytochemistry". *J. Electron. Microsc. Tech.*, 17(2): 121-31.

Roy, S. K., Chiba, Y. & Jigami, Y. (2000) "Production of Therapeutic Glycoproteins through the Engineering of Glycosylation Pathway in Yeast". *Biotechnol. Bioprocess Eng.*, 5: 219-226.

Sasaki, K. (1996) "Molecular cloning and characterization of sialyltransferases". *Trends Glycosci. Glycotechnol.*, 8: 195-215.

Sasaki, K., Kurata, K., Kojima, N., Kurosawa, N., Ohta, S., Hanai, N., Tsuji, S. & Nishi, T. (1994) "Expression cloning of a GM3-specific alpha-2,8-sialyltransferase (GD3 synthase)". *J. Biol. Chem.*, 269(22): 15950-6.

Scheidegger, E.P., Sternberg, L.R., Roth, J. & Lowe, JB. (1995) "A human STX cDNA confers polysialic acid expression in mammalian cells". *J. Biol. Chem.*, 270(39): 22685-8.

Sinnott, M.L. (1990) "Catalytic mechanism of enzymic glycosyl transfer". *Chem. Rev.*, 90. 1171-1202.

Sjoberg, E.R, Kitagawa, H., Glushka, J., van Halbeek, H. & Paulson, J.C. (1996) "Molecular cloning of a developmentally regulated N-acetylgalactosamine alpha2,6-sialyltransferase specific for sialylated glycoconjugates". *J. Biol. Chem.*, 271(13): 7450-9.

Sujino, K., Jackson, R.J., Chan, N.W., Tsuji, S. & Palcic, M.M. (2000) "A novel viral alpha2,3-sialyltransferase (v-ST3Gal I): transfer of sialic acid to fucosylated acceptors". *Glycobiology*, 10(3): 313-20.

Swartz, J.R. (2001) "Advances in Escherichia coli production of therapeutic proteins". *Curr. Spin. Biotechnol.*, 12. 195-201.

Takashima, S., Ishida, H.K., Inazu, T., Ando, T., Ishida, H., Kiso, M., Tsuji, S. & Tsujimoto, M. (2002) "Molecular cloning and expression of a sixth type of alpha 2,8-sialyltransferase (ST8Sia VI) that sialylates O-glycans". *J. Biol. Chem.*, 277(27): 24030-8.

Taniguchi, A., Kaneta, R., Morishita, K. & Matsumoto, K. (2001) "Gene structure and transcriptional regulation of human Gal beta-1,4(3) GlcNAc alpha2,3-sialyltransferase VI (hST3Gal VI) gene in prostate cancer cell line". *Biochem. Biophys. Res. Commun.*, (5): 1148-56.

Tanner, W. & Lehele, L. (1987) "Protein glycosylation in yeast". *Biochim. Biophys. Acta*, 906(1): 81-99.

Tetteroo, P.A., de Heij, H.T., Van den Eijnden, D.H., Visser, F.J., Schoenmarker, E. & Geurts van Kessel, AH. (1987) "A GDP-fucose: [Gal beta 1----4]GlcNAc alpha 1----3-fucosyltransferase activity is correlated with the presence of human chromosome 11 and the expression of the Lex, Ley, and sialyl-Lex antigens in human-mouse cell hybrids". *J. Biol. Chem.*, 262(33): 15984-9.

Vallejo-Ruiz, V., Haque R., Mir, A.M., Schwientek, T., Mandel, U., Cacan, R., Delannoy, P. & Harduin-Lepers, A. (2001) "Delineation of the minimal catalytic domain of human Galbetal-3GalNAc α 2,3-sialyltransferase (hST3Gal I)". *Biochim. Biophys. Acta*, 1549(2): 161-73.

van den Eijnden, D.H., Joziassse, D.H., Dorland, L., van Halbeek, H., Vliegenthart, J.F. & Schmid, K. (1980) "Specificity in the enzymic transfer of sialic acid to the oligosaccharide branches of b1- and triantennary glycopeptides of α 1-acid glycoprotein". *Biochem. Biophys. Res. Commun.*, 92(3): 839-45.

Van Den Hamer, C.J., Morell, A.G., Scheinberg, I.H., Hickman, J. & Ashwell, G. (1970) "Physical and chemical studies on ceruloplasmin. IX. The role of galactosyl residues in the clearance of ceruloplasmin from the circulation". *J. Biol. Chem.*, 245(17): 4397-402.

Vervecken, W., Kaigorodov, V., Callewaert, N., Geysens, S., De Vusser, K. & Contreras, R. (2004) "In vivo synthesis of mammalian-like, hybrid-type N-glycans in *Pichia pastoris*". *Appl. Environ. Microbiol.*, 70: 2639-2646.

Watanabe, Y., Nara, K., Takahashi, H., Nagai, Y. & Sanai, Y. (1996) "The molecular cloning and expression of α 2,8-sialyltransferase (GD3 synthase) in a rat brain". *J. Biochem.*, 120(5): 1020-7.

Weikert, S., Papac, D., Briggs, J., Cowfer, D., Tom, S., Gawlitzek, M., Lofgren, J., Mehta, S., Chisholm, V., Modi, N., Eppler, S., Carroll, K., Chamow, S., Peers, D., Berman, P. & Krummen, L. (1999) "Engineering Chinese hamster ovary cells to maximize sialic acid content of recombinant glycoproteins". *Nat. Biotechnol.*, 17(11): 1116-21.

Weinstein, J., de Souza-e-Silva, U. & Paulson, J.C. (1982) "Sialylation of glycoprotein oligosaccharides N-linked to asparagine. Enzymatic characterization of a Gal beta 1 to 3(4)GlcNAc α 2 to 3 sialyltransferase and a Gal beta 1 to 4GlcNAc α 2 to 6 sialyltransferase from rat liver". *J. Biol. Chem.*, 257(22): 13845-53.

Wickner, W.T. & Lodisch, H.F. (1985) "Multiple mechanisms of protein insertion into and across membranes". *Science*, 230(4724): 400-7.

Yang W., Pepperkok, R., Bender, P., Kreis, TE. & Storrie, B. (1996) "Modification of the cytoplasmic domain affects the subcellular localization of Golgi glycosyltransferases". *Eur. J. Cell Biol.*, 71(1): 53-61.

Yoshida, Y., Kojima, N. & Tsuji, S. (1995a) "Molecular cloning and characterization of a third type of N-glycan alpha 2,8-sialyltransferase from mouse lung". *J. Biochem.*, 118(3): 658-64.

Yoshida, Y., Kojima, N., Kurosawa, N., Hamamoto, T. & Tsuji, S. (1995b) "Molecular cloning of Sia alpha 2,3Gal beta 1,4GlcNAc alpha 2,8-sialyltransferase from mouse brain". *J. Biol. Chem.*, 270(24): 14628-33.

LISTAGEM DE SEQUÊNCIAS

<110> CNRS

Université de Provence

<120> PREPARAÇÃO E UTILIZAÇÕES DE GENES CODIFICANDO
GLICOSILTRANSFERASES QUIMÉRICAS COM ACTIVIDADE DE
GLICOSILAÇÃO OPTIMIZADA

<130> IOB 05 DI PRO CREA

<160> 176

<170> PatentIn versão 3.3

<210> 1

<211> 1221

<212> ADN

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1)..(1221)

<400> 1

atg att cac acc aac ctg aag aaa aag ttc agc tgc tgc gtc ctg gtc	48
Met Ile His Thr Asn Leu Lys Lys Lys Phe Ser Cys Cys Val Leu Val	
1 5 10 15	
ttt ctt ctg ttt gca gtc atc tgt gtg tgg aag gaa aag aag aaa ggg	96
Phe Leu Leu Phe Ala Val Ile Cys Val Trp Lys Glu Lys Lys Lys Gly	
20 25 30	
agt tac tat gat tcc ttt aaa ttg caa acc aag gaa ttc cag gtg tta	144
Ser Tyr Tyr Asp Ser Phe Lys Leu Gln Thr Lys Glu Phe Gln Val Leu	
35 40 45	
aag agt ctg ggg aaa ttg gcc atg ggg tct gat tcc cag tct gta tcc	192
Lys Ser Leu Gly Lys Leu Ala Met Gly Ser Asp Ser Gln Ser Val Ser	
50 55 60	
tca agc agc acc cag gac ccc cac agg ggc cgc cag acc ctc ggc agt	240
Ser Ser Ser Thr Gln Asp Pro His Arg Gly Arg Gln Thr Leu Gly Ser	
65 70 75 80	
ctc aga ggc cta gcc aag gcc aaa cca gag gcc tcc ttc cag gtg tgg	288
Leu Arg Gly Leu Ala Lys Ala Lys Pro Glu Ala Ser Phe Gln Val Trp	
85 90 95	
aac aag gac agc tct tcc aaa aac ctt atc cct agg ctg caa aag atc	336
Asn Lys Asp Ser Ser Ser Lys Asn Leu Ile Pro Arg Leu Gln Lys Ile	
100 105 110	
tgg aag aat tac cta agc atg aac aag tac aaa gtg tcc tac aag ggg	384
Trp Lys Asn Tyr Leu Ser Met Asn Lys Tyr Lys Val Ser Tyr Lys Gly	
115 120 125	
cca gga cca ggc atc aag ttc agt gca gag gcc ctg cgc tgc cac ctc	432
Pro Gly Pro Gly Ile Lys Phe Ser Ala Glu Ala Leu Arg Cys His Leu	
130 135 140	

cgg gac cat gtg aat gta tcc atg gta gag gtc aca gat ttt ccc ttc	480
Arg Asp His Val Asn Val Ser Met Val Glu Val Thr Asp Phe Pro Phe	
145 150 155 160	
aat acc tct gaa tgg gag ggt tat ctg ccc aag gag agc att agg acc	528
Asn Thr Ser Glu Trp Glu Gly Tyr Leu Pro Lys Glu Ser Ile Arg Thr	
165 170 175	
aag gct ggg cct tgg ggc agg tgt gct gtt gtg tgg tca gcg gga tct	576
Lys Ala Gly Pro Trp Gly Arg Cys Ala Val Val Ser Ser Ala Gly Ser	
180 185 190	
ctg aag tcc tcc caa cta ggc aga gaa atc gat gat cat gac gca gtc	624
Leu Lys Ser Ser Gln Leu Gly Arg Glu Ile Asp Asp His Asp Ala Val	
195 200 205	
ctg agg ttt aat ggg gca ccc aca gcc aac ttc caa caa gat gtg ggc	672
Leu Arg Phe Asn Gly Ala Pro Thr Ala Asn Phe Gln Gln Asp Val Gly	
210 215 220	
aca aaa act acc att cgc ctg atg aac tct cag ttg gtt acc aca gag	720
Thr Lys Thr Thr Ile Arg Leu Met Asn Ser Gln Leu Val Thr Thr Glu	
225 230 235 240	
aag cgc ttc ctc aaa gac agt ttg tac aat gaa gga atc cta att gta	768
Lys Arg Phe Leu Lys Asp Ser Leu Tyr Asn Glu Gly Ile Leu Ile Val	
245 250 255	
tgg gac cca tct gta tac cac tca gat atc cca aag tgg tac cag aat	816
Trp Asp Pro Ser Val Tyr His Ser Asp Ile Pro Lys Trp Tyr Gln Asn	
260 265 270	
ccg gat tat aat ttc ttt aac aac tac aag act tat cgt aag ctg cac	864
Pro Asp Tyr Asn Phe Phe Asn Asn Tyr Lys Thr Tyr Arg Lys Leu His	
275 280 285	
ccc aat cag ccc ttt tac atc ctc aag ccc cag atg cct tgg gag cta	912
Pro Asn Gln Pro Phe Tyr Ile Leu Lys Pro Gln Met Pro Trp Glu Leu	
290 295 300	
tgg gac att ctt caa gaa atc tcc cca gaa gag att cag cca aac ccc	960
Trp Asp Ile Leu Gln Glu Ile Ser Pro Glu Glu Ile Gln Pro Asn Pro	
305 310 315 320	
cca tcc tct ggg atg ctt ggt atc atc atc atg atg acg ctg tgt gac	1008
Pro Ser Ser Gly Met Leu Gly Ile Ile Ile Met Met Thr Leu Cys Asp	
325 330 335	
cag gtg gat att tat gag ttc ctc cca tcc aag cgc aag act gac gtg	1056
Gln Val Asp Ile Tyr Glu Phe Leu Pro Ser Lys Arg Lys Thr Asp Val	
340 345 350	
tgc tac tac tac cag aag ttc ttc gat agt gcc tgc acg atg ggt gcc	1104
Cys Tyr Tyr Tyr Gln Lys Phe Phe Asp Ser Ala Cys Thr Met Gly Ala	
355 360 365	
tac cac ccg ctg ctc tat gag aag aat ttg gtg aag cat ctc aac cag	1152
Tyr His Pro Leu Leu Tyr Glu Lys Asn Leu Val Lys His Leu Asn Gln	
370 375 380	
ggc aca gat gag gac atc tac ctg ctt gga aaa gcc aca ctg cct ggc	1200

Gly Thr Asp Glu Asp Ile Tyr Leu Leu Gly Lys Ala Thr Leu Pro Gly
385 390 395 400

ttc cgg acc att cac tgc taa
Phe Arg Thr Ile His Cys
405

1221

<210> 2

<211> 406

<212> PRT

<213> *Homo sapiens*

<400> 2

Met Ile His Thr Asn Leu Lys Lys Lys Phe Ser Cys Cys Val Leu Val
 1 5 10 15

Phe Leu Leu Phe Ala Val Ile Cys Val Trp Lys Glu Lys Lys Lys Gly
 20 25 30

Ser Tyr Tyr Asp Ser Phe Lys Leu Gln Thr Lys Glu Phe Gln Val Leu
 35 40 45

Lys Ser Leu Gly Lys Leu Ala Met Gly Ser Asp Ser Gln Ser Val Ser
 50 55 60

Ser Ser Ser Thr Gln Asp Pro His Arg Gly Arg Gln Thr Leu Gly Ser
 65 70 75 80

Leu Arg Gly Leu Ala Lys Ala Lys Pro Glu Ala Ser Phe Gln Val Trp
 85 90 95

Asn Lys Asp Ser Ser Ser Lys Asn Leu Ile Pro Arg Leu Gln Lys Ile
 100 105 110

Trp Lys Asn Tyr Leu Ser Met Asn Lys Tyr Lys Val Ser Tyr Lys Gly
 115 120 125

Pro Gly Pro Gly Ile Lys Phe Ser Ala Glu Ala Leu Arg Cys His Leu
 130 135 140

Arg Asp His Val Asn Val Ser Met Val Glu Val Thr Asp Phe Pro Phe
 145 150 155 160

Asn Thr Ser Glu Trp Glu Gly Tyr Leu Pro Lys Glu Ser Ile Arg Thr
 165 170 175

Lys Ala Gly Pro Trp Gly Arg Cys Ala Val Val Ser Ser Ala Gly Ser
 180 185 190

Leu Lys Ser Ser Gln Leu Gly Arg Glu Ile Asp Asp His Asp Ala Val
 195 200 205

Leu Arg Phe Asn Gly Ala Pro Thr Ala Asn Phe Gln Gln Asp Val Gly
 210 215 220

Thr Lys Thr Thr Ile Arg Leu Met Asn Ser Gln Leu Val Thr Thr Glu
 225 230 235 240

Lys Arg Phe Leu Lys Asp Ser Leu Tyr Asn Glu Gly Ile Leu Ile Val
 245 250 255

Trp Asp Pro Ser Val Tyr His Ser Asp Ile Pro Lys Trp Tyr Gln Asn
 260 265 270

Pro Asp Tyr Asn Phe Phe Asn Asn Tyr Lys Thr Tyr Arg Lys Leu His
 275 280 285

Pro Asn Gln Pro Phe Tyr Ile Leu Lys Pro Gln Met Pro Trp Glu Leu
 290 295 300

Trp Asp Ile Leu Gln Glu Ile Ser Pro Glu Glu Ile Gln Pro Asn Pro
 305 310 315 320

Pro Ser Ser Gly Met Leu Gly Ile Ile Ile Met Met Thr Leu Cys Asp
 325 330 335

Gln Val Asp Ile Tyr Glu Phe Leu Pro Ser Lys Arg Lys Thr Asp Val
 340 345 350

Cys Tyr Tyr Tyr Gln Lys Phe Phe Asp Ser Ala Cys Thr Met Gly Ala
 355 360 365

Tyr His Pro Leu Leu Tyr Glu Lys Asn Leu Val Lys His Leu Asn Gln
 370 375 380

Gly Thr Asp Glu Asp Ile Tyr Leu Leu Gly Lys Ala Thr Leu Pro Gly
 385 390 395 400

Phe Arg Thr Ile His Cys
 405

<210> 3
<211> 1590
<212> ADN
<213> *Homo sapiens*

<220>
<221> CDS
<222> (1) .. (1590)

<400> 3

atg aaa cca cac ttg aag caa tgg aga caa cga atg ctt ttc gga ata Met Lys Pro His Leu Lys Gln Trp Arg Gln Arg Met Leu Phe Gly Ile 1 5 10 15	48
ttc gct tgg ggg ctc ctc ttt ttg ctg att ttc atc tac ttc acc gac Phe Ala Trp Gly Leu Leu Phe Leu Leu Ile Phe Ile Tyr Phe Thr Asp 20 25 30	96
agc aac ccc gct gag cct gta ccc agc tcc ctc tcc ttc ctg gag acc Ser Asn Pro Ala Glu Pro Val Pro Ser Ser Leu Ser Phe Leu Glu Thr 35 40 45	144
agg agg ctc ctg ccg gtg cag ggg aag cag cgg gcc atc atg ggc gcc Arg Arg Leu Leu Pro Val Gln Gly Lys Gln Arg Ala Ile Met Gly Ala 50 55 60	192
gca cat gag ccc tcc ccg cct ggg ggc ctg gac gca cgc cag gcg ctg Ala His Glu Pro Ser Pro Pro Gly Gly Leu Asp Ala Arg Gln Ala Leu 65 70 75 80	240
ccc cgc gcc cac cca gcc ggt tcc ttt cat gcg ggg cct gga gac ctg Pro Arg Ala His Pro Ala Gly Ser Phe His Ala Gly Pro Gly Asp Leu 85 90 95	288
cag aaa tgg gcc cag tcc caa gat ggg ttt gaa cat aaa gag ttt ttt Gln Lys Trp Ala Gln Ser Gln Asp Gly Phe Glu His Lys Glu Phe Phe 100 105 110	336
tca tcc cag gtg ggg aga aaa tct caa agt gct ttc tac ccg gag gat Ser Ser Gln Val Gly Arg Lys Ser Gln Ser Ala Phe Tyr Pro Glu Asp 115 120 125	384
gac gac tac ttt ttt gct gct ggt cag cca ggg tgg cac agc cac act Asp Asp Tyr Phe Phe Ala Ala Gly Gln Pro Gly Trp His Ser His Thr 130 135 140	432
cag ggg aca ttg gga ttc cct tcc ccc ggg gag cca ggc cca cgg gag Gln Gly Thr Leu Gly Phe Pro Ser Pro Gly Glu Pro Gly Pro Arg Glu 145 150 155 160	480
ggg gct ttt ccg gct gca cag gtc cag agg agg cgg gtg aag aag agg Gly Ala Phe Pro Ala Ala Gln Val Gln Arg Arg Arg Val Lys Lys Arg 165 170 175	528
cac cgg agg cag aga agg agc cac gtg ttg gag gag ggc gac gac ggc His Arg Arg Gln Arg Arg Ser His Val Leu Glu Glu Gly Asp Asp Gly 180 185 190	576
gac agg ctg tac tcc tcc atg tcc agg gcc ttc ctg tac cgg ctc tgg Asp Arg Leu Tyr Ser Ser Met Ser Arg Ala Phe Leu Tyr Arg Leu Trp 195 200 205	624
aag ggg aac gtc tct tcc aaa atg ctg aac ccg cgc ctg cag aag gcg	672

Lys Gly Asn Val Ser Ser Lys Met Leu Asn Pro Arg Leu Gln Lys Ala	
210 215 220	
atg aag gat tac ctg acc gcc aac aag cac ggg gtg cgc ttc cgc ggg	720
Met Lys Asp Tyr Leu Thr Ala Asn Lys His Gly Val Arg Phe Arg Gly	
225 230 235 240	
aag cgg gag gcc ggg ctg agc agg gca cag ctg ctg tgc cag ctg cgg	768
Lys Arg Glu Ala Gly Leu Ser Arg Ala Gln Leu Leu Cys Gln Leu Arg	
245 250 255	
agc cgc gcg cgc gtg cgg acg ctg gac ggc acc gag gcg ccc ttt tct	816
Ser Arg Ala Arg Val Arg Thr Leu Asp Gly Thr Glu Ala Pro Phe Ser	
260 265 270	
gcg ctg ggc tgg cgg cgc ctg gtg ccc gcc gtg ccc ctg agc cag ctg	864
Ala Leu Gly Trp Arg Arg Leu Val Pro Ala Val Pro Leu Ser Gln Leu	
275 280 285	
cac ccc cgc ggc ctg cgc agc tgc gct gtc gtc atg tct gca ggc gca	912
His Pro Arg Gly Leu Arg Ser Cys Ala Val Val Met Ser Ala Gly Ala	
290 295 300	
atc ctc aac tct tcc ttg ggc gag gaa ata gat tct cat gat gcg gtt	960
Ile Leu Asn Ser Ser Leu Gly Glu Glu Ile Asp Ser His Asp Ala Val	
305 310 315 320	
ttg aga ttt aac tct gct cct aca cgt ggt tat gag aaa gat gtt ggg	1008
Leu Arg Phe Asn Ser Ala Pro Thr Arg Gly Tyr Glu Lys Asp Val Gly	
325 330 335	
aat aaa acc acc ata cgc atc att aat tcg cag att ctg acc aac ccc	1056
Asn Lys Thr Thr Ile Arg Ile Ile Asn Ser Gln Ile Leu Thr Asn Pro	
340 345 350	
agc cat cac ttc att gac agt tca ctg tat aaa gac gtc att ttg gtg	1104
Ser His His Phe Ile Asp Ser Ser Leu Tyr Lys Asp Val Ile Leu Val	
355 360 365	
gcc tgg gac cct gcc cca tat tcc gca aat ctt aac ctg tgg tac aaa	1152
Ala Trp Asp Pro Ala Pro Tyr Ser Ala Asn Leu Asn Leu Trp Tyr Lys	
370 375 380	
aaa ccg gat tac aac ctg ttc act cca tat att cag cat cgt cag aga	1200
Lys Pro Asp Tyr Asn Leu Phe Thr Pro Tyr Ile Gln His Arg Gln Arg	
385 390 395 400	
aac cca aat cag cca ttt tac att ctt cat cct aaa ttt ata tgg cag	1248
Asn Pro Asn Gln Pro Phe Tyr Ile Leu His Pro Lys Phe Ile Trp Gln	
405 410 415	
ctc tgg gat att atc cag gag aac act aaa gag aag att caa cca aac	1296
Leu Trp Asp Ile Ile Gln Glu Asn Thr Lys Glu Lys Ile Gln Pro Asn	
420 425 430	
cca cca tct tct ggt ttc att gga atc ctc atc atg atg tcc atg tgc	1344
Pro Pro Ser Ser Gly Phe Ile Gly Ile Leu Ile Met Met Ser Met Cys	
435 440 445	
aga gag gtg cac gtg tat gaa tat atc cca tcc gtg cgg cag acg gag	1392
Arg Glu Val His Val Tyr Glu Tyr Ile Pro Ser Val Arg Gln Thr Glu	

450	455	460	
ctg tgc cac tac cac gag ctg tac tac gac gca gcc tgc acc ctc ggg			1440
Leu Cys His Tyr His Glu Leu Tyr Tyr Asp Ala Ala Cys Thr Leu Gly			
465	470	475	480
ggg tac cac cca cta ctc tat gag aag ctc ctg gtg cag cgc ctg aac			1488
Ala Tyr His Pro Leu Leu Tyr Glu Lys Leu Leu Val Gln Arg Leu Asn			
	485	490	495
atg ggc acg cag ggg gat ttg cat cgc aag ggc aag gtg gtt ctt cct			1536
Met Gly Thr Gln Gly Asp Leu His Arg Lys Gly Lys Val Val Leu Pro			
	500	505	510
ggc ttc cag ggc gtg cac tgc cct gca cca agt cca gtc att cca cac			1584
Gly Phe Gln Ala Val His Cys Pro Ala Pro Ser Pro Val Ile Pro His			
	515	520	525
tct taa			1590
Ser			

<210> 4

<211> 529

<212> PRT

<213> *Homo sapiens*

<400> 4

Met Lys Pro His Leu Lys Gln Trp Arg Gln Arg Met Leu Phe Gly Ile
1 5 10 15

Phe Ala Trp Gly Leu Leu Phe Leu Leu Ile Phe Ile Tyr Phe Thr Asp
20 25 30

Ser Asn Pro Ala Glu Pro Val Pro Ser Ser Leu Ser Phe Leu Glu Thr
35 40 45

Arg Arg Leu Leu Pro Val Gln Gly Lys Gln Arg Ala Ile Met Gly Ala
50 55 60

Ala His Glu Pro Ser Pro Pro Gly Gly Leu Asp Ala Arg Gln Ala Leu
65 70 75 80

Pro Arg Ala His Pro Ala Gly Ser Phe His Ala Gly Pro Gly Asp Leu
85 90 95

Gln Lys Trp Ala Gln Ser Gln Asp Gly Phe Glu His Lys Glu Phe Phe
100 105 110

Ser Ser Gln Val Gly Arg Lys Ser Gln Ser Ala Phe Tyr Pro Glu Asp
115 120 125

Asp	Asp	Tyr	Phe	Phe	Ala	Ala	Gly	Gln	Pro	Gly	Trp	His	Ser	His	Thr
130						135					140				
Gln	Gly	Thr	Leu	Gly	Phe	Pro	Ser	Pro	Gly	Glu	Pro	Gly	Pro	Arg	Glu
145					150					155					160
Gly	Ala	Phe	Pro	Ala	Ala	Gln	Val	Gln	Arg	Arg	Arg	Val	Lys	Lys	Arg
				165					170					175	
His	Arg	Arg	Gln	Arg	Arg	Ser	His	Val	Leu	Glu	Glu	Gly	Asp	Asp	Gly
			180					185					190		
Asp	Arg	Leu	Tyr	Ser	Ser	Met	Ser	Arg	Ala	Phe	Leu	Tyr	Arg	Leu	Trp
		195					200					205			
Lys	Gly	Asn	Val	Ser	Ser	Lys	Met	Leu	Asn	Pro	Arg	Leu	Gln	Lys	Ala
	210					215					220				
Met	Lys	Asp	Tyr	Leu	Thr	Ala	Asn	Lys	His	Gly	Val	Arg	Phe	Arg	Gly
225					230					235					240
Lys	Arg	Glu	Ala	Gly	Leu	Ser	Arg	Ala	Gln	Leu	Leu	Cys	Gln	Leu	Arg
				245					250					255	
Ser	Arg	Ala	Arg	Val	Arg	Thr	Leu	Asp	Gly	Thr	Glu	Ala	Pro	Phe	Ser
			260					265					270		
Ala	Leu	Gly	Trp	Arg	Arg	Leu	Val	Pro	Ala	Val	Pro	Leu	Ser	Gln	Leu
		275					280						285		
His	Pro	Arg	Gly	Leu	Arg	Ser	Cys	Ala	Val	Val	Met	Ser	Ala	Gly	Ala
		290				295					300				
Ile	Leu	Asn	Ser	Ser	Leu	Gly	Glu	Glu	Ile	Asp	Ser	His	Asp	Ala	Val
305					310					315					320
Leu	Arg	Phe	Asn	Ser	Ala	Pro	Thr	Arg	Gly	Tyr	Glu	Lys	Asp	Val	Gly
				325					330					335	
Asn	Lys	Thr	Thr	Ile	Arg	Ile	Ile	Asn	Ser	Gln	Ile	Leu	Thr	Asn	Pro
			340					345					350		
Ser	His	His	Phe	Ile	Asp	Ser	Ser	Leu	Tyr	Lys	Asp	Val	Ile	Leu	Val
		355					360					365			

Ala Trp Asp Pro Ala Pro Tyr Ser Ala Asn Leu Asn Leu Trp Tyr Lys
 370 375 380

Lys Pro Asp Tyr Asn Leu Phe Thr Pro Tyr Ile Gln His Arg Gln Arg
 385 390 395 400

Asn Pro Asn Gln Pro Phe Tyr Ile Leu His Pro Lys Phe Ile Trp Gln
 405 410 415

Leu Trp Asp Ile Ile Gln Glu Asn Thr Lys Glu Lys Ile Gln Pro Asn
 420 425 430

Pro Pro Ser Ser Gly Phe Ile Gly Ile Leu Ile Met Met Ser Met Cys
 435 440 445

Arg Glu Val His Val Tyr Glu Tyr Ile Pro Ser Val Arg Gln Thr Glu
 450 455 460

Leu Cys His Tyr His Glu Leu Tyr Tyr Asp Ala Ala Cys Thr Leu Gly
 465 470 475 480

Ala Tyr His Pro Leu Leu Tyr Glu Lys Leu Leu Val Gln Arg Leu Asn
 485 490 495

Met Gly Thr Gln Gly Asp Leu His Arg Lys Gly Lys Val Val Leu Pro
 500 505 510

Gly Phe Gln Ala Val His Cys Pro Ala Pro Ser Pro Val Ile Pro His
 515 520 525

Ser

<210> 5

<211> 1803

<212> ADN

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1) .. (1803)

<400> 5

atg agg tcc tgc ctg tgg aga tgc agg cac ctg agc caa ggc gtc cag	48
Met Arg Ser Cys Leu Trp Arg Cys Arg His Leu Ser Gln Gly Val Gln	
1 5 10 15	
tgg tcc ttg ctt ctg gct gtc ctg gtc ttc ttt ctc ttc gcc ttg ccc	96
Trp Ser Leu Leu Leu Ala Val Leu Val Phe Phe Leu Phe Ala Leu Pro	

20	25	30	
tct ttt att aag gag cct caa aca aag cct tcc agg cat caa cgc aca			144
Ser Phe Ile Lys Glu Pro Gln Thr Lys Pro Ser Arg His Gln Arg Thr			
35	40	45	
gag aac att aaa gaa agg tct cta cag tcc ctg gca aag cct aag tcc			192
Glu Asn Ile Lys Glu Arg Ser Leu Gln Ser Leu Ala Lys Pro Lys Ser			
50	55	60	
cag gca ccc aca agg gca agg agg aca acc atc tat gca gag cca gtg			240
Gln Ala Pro Thr Arg Ala Arg Arg Thr Thr Ile Tyr Ala Glu Pro Val			
65	70	75	80
cca gag aac aat gcc ctc aac aca caa acc cag ccc aag gcc cac acc			288
Pro Glu Asn Asn Ala Leu Asn Thr Gln Thr Gln Pro Lys Ala His Thr			
85	90	95	
acc gga gac aga gga aag gag gcc aac cag gca ccg ccg gag gag cag			336
Thr Gly Asp Arg Gly Lys Glu Ala Asn Gln Ala Pro Pro Glu Glu Gln			
100	105	110	
gac aag gtg ccc cac aca gca cag agg gca gca tgg aag agc cca gaa			384
Asp Lys Val Pro His Thr Ala Gln Arg Ala Ala Trp Lys Ser Pro Glu			
115	120	125	
aaa gag aaa acc atg gtg aac aca ctg tca ccc aga ggg caa gat gca			432
Lys Glu Lys Thr Met Val Asn Thr Leu Ser Pro Arg Gly Gln Asp Ala			
130	135	140	
ggg atg gcc tct ggc agg aca gag gca caa tca tgg aag agc cag gac			480
Gly Met Ala Ser Gly Arg Thr Glu Ala Gln Ser Trp Lys Ser Gln Asp			
145	150	155	160
aca aag acg acc caa gga aat ggg ggc cag acc agg aag ctg acg gcc			528
Thr Lys Thr Thr Gln Gly Asn Gly Gly Gln Thr Arg Lys Leu Thr Ala			
165	170	175	
tcc agg acg gtg tca gag aag cac cag ggc aaa ggc gca acc aca gcc			576
Ser Arg Thr Val Ser Glu Lys His Gln Gly Lys Ala Ala Thr Thr Ala			
180	185	190	
aag acg ctc att ccc aaa agt cag cac aga atg ctg gct ccc aca gga			624
Lys Thr Leu Ile Pro Lys Ser Gln His Arg Met Leu Ala Pro Thr Gly			
195	200	205	
gca gtg tca aca agg acg aga cag aaa gga gtg acc aca gca gtc atc			672
Ala Val Ser Thr Arg Thr Arg Gln Lys Gly Val Thr Thr Ala Val Ile			
210	215	220	
cca cct aag gag aag aaa cct cag gcc acc cca ccc cct gcc cct ttc			720
Pro Pro Lys Glu Lys Lys Pro Gln Ala Thr Pro Pro Pro Ala Pro Phe			
225	230	235	240
cag agc ccc acg acg cag aga aac caa aga ctg aag gcc gcc aac ttc			768
Gln Ser Pro Thr Thr Gln Arg Asn Gln Arg Leu Lys Ala Ala Asn Phe			
245	250	255	
aaa tct gag cct cgg tgg gat ttt gag gaa aaa tac agc ttc gaa ata			816
Lys Ser Glu Pro Arg Trp Asp Phe Glu Glu Lys Tyr Ser Phe Glu Ile			
260	265	270	

gga ggc ctt cag acg act tgc cct gac tct gtg aag atc aaa gcc tcc Gly Gly Leu Gln Thr Thr Cys Pro Asp Ser Val Lys Ile Lys Ala Ser 275 280 285	864
aag tcg ctg tgg ctc cag aaa ctc ttt ctg ccc aac ctc act ctc ttc Lys Ser Leu Trp Leu Gln Lys Leu Phe Leu Pro Asn Leu Thr Leu Phe 290 295 300	912
ctg gac tcc aga cac ttc aac cag agt gag tgg gac cgc ctg gaa cac Leu Asp Ser Arg His Phe Asn Gln Ser Glu Trp Asp Arg Leu Glu His 305 310 315 320	960
ttt gca cca ccc ttt ggc ttc atg gag ctc aac tac tcc ttg gtg cag Phe Ala Pro Pro Phe Gly Phe Met Glu Leu Asn Tyr Ser Leu Val Gln 325 330 335	1008
aag gtc gtg aca cgc ttc cct cca gtg ccc cag cag cag ctg ctc ctg Lys Val Val Thr Arg Phe Pro Pro Val Pro Gln Gln Gln Leu Leu Leu 340 345 350	1056
gcc agc ctc ccc gct ggg agc ctc cgg tgc atc acc tgt gcc gtg gtg Ala Ser Leu Pro Ala Gly Ser Leu Arg Cys Ile Thr Cys Ala Val Val 355 360 365	1104
ggc aac ggg ggc atc ctg aac aac tcc cac atg ggc cag gag ata gac Gly Asn Gly Gly Ile Leu Asn Asn Ser His Met Gly Gln Glu Ile Asp 370 375 380	1152
agt cac gac tac gtg ttc cga ttg agc gga gct ctc att aaa ggc tac Ser His Asp Tyr Val Phe Arg Leu Ser Gly Ala Leu Ile Lys Gly Tyr 385 390 395 400	1200
gaa cag gat gtg ggg act cgg aca tcc ttc tac ggc ttt acc gcc ttc Glu Gln Asp Val Gly Thr Arg Thr Ser Phe Tyr Gly Phe Thr Ala Phe 405 410 415	1248
tcc ctg acc cag tca ctc ctt ata ttg ggc aat cgg ggt ttc aag aac Ser Leu Thr Gln Ser Leu Leu Ile Leu Gly Asn Arg Gly Phe Lys Asn 420 425 430	1296
gtg cct ctc ggg aag gac gtc cgc tac ttg cac ttc ctg gaa ggc acc Val Pro Leu Gly Lys Asp Val Arg Tyr Leu His Phe Leu Glu Gly Thr 435 440 445	1344
cgg gac tat gag tgg ctg gaa gca ctg ctt atg aat cag acg gtg atg Arg Asp Tyr Glu Trp Leu Glu Ala Leu Leu Met Asn Gln Thr Val Met 450 455 460	1392
tca aaa aac ctt ttc tgg ttc agg cac aga ccc cag gaa gct ttt cgg Ser Lys Asn Leu Phe Trp Phe Arg His Arg Pro Gln Glu Ala Phe Arg 465 470 475 480	1440
gaa gcc ctg cac atg gac agg tac ctg ttg ctg cac cca gac ttt ctc Glu Ala Leu His Met Asp Arg Tyr Leu Leu Leu His Pro Asp Phe Leu 485 490 495	1488
cga tac atg aag aac agg ttt ctg agg tct aag acc ctg gat ggt gcc Arg Tyr Met Lys Asn Arg Phe Leu Arg Ser Lys Thr Leu Asp Gly Ala 500 505 510	1536

cac tgg agg ata tac cgc ccc acc act ggg gcc ctc ctg ctg ctc act	1584
His Trp Arg Ile Tyr Arg Pro Thr Thr Gly Ala Leu Leu Leu Leu Thr	
515 520 525	
gcc ctt cag ctc tgt gac cag gtg agt gct tat ggc ttc atc act gag	1632
Ala Leu Gln Leu Cys Asp Gln Val Ser Ala Tyr Gly Phe Ile Thr Glu	
530 535 540	
ggc cat gag cgc ttt tct gat cac tac tat gat aca tca tgg aag cgg	1680
Gly His Glu Arg Phe Ser Asp His Tyr Tyr Asp Thr Ser Trp Lys Arg	
545 550 555 560	
ctg atc ttt tac ata aac cat gac ttc aag ctg gag aga gaa gtc tgg	1728
Leu Ile Phe Tyr Ile Asn His Asp Phe Lys Leu Glu Arg Glu Val Trp	
565 570 575	
aag cgg cta cac gat gaa ggg ata atc cgg ctg tac cag cgt cct ggt	1776
Lys Arg Leu His Asp Glu Gly Ile Ile Arg Leu Tyr Gln Arg Pro Gly	
580 585 590	
ccc gga act gcc aaa gcc aag aac tga	1803
Pro Gly Thr Ala Lys Ala Lys Asn	
595 600	

<210> 6

<211> 600

<212> PRT

<213> *Homo sapiens*

<400> 6

Met Arg Ser Cys Leu Trp Arg Cys Arg His Leu Ser Gln Gly Val Gln
1 5 10 15

Trp Ser Leu Leu Leu Ala Val Leu Val Phe Phe Leu Phe Ala Leu Pro
20 25 30

Ser Phe Ile Lys Glu Pro Gln Thr Lys Pro Ser Arg His Gln Arg Thr
35 40 45

Glu Asn Ile Lys Glu Arg Ser Leu Gln Ser Leu Ala Lys Pro Lys Ser
50 55 60

Gln Ala Pro Thr Arg Ala Arg Arg Thr Thr Ile Tyr Ala Glu Pro Val
65 70 75 80

Pro Glu Asn Asn Ala Leu Asn Thr Gln Thr Gln Pro Lys Ala His Thr
85 90 95

Thr Gly Asp Arg Gly Lys Glu Ala Asn Gln Ala Pro Pro Glu Glu Gln
100 105 110

Asp Lys Val Pro His Thr Ala Gln Arg Ala Ala Trp Lys Ser Pro Glu

115	120	125
Lys Glu Lys Thr Met Val Asn Thr Leu Ser Pro Arg Gly Gln Asp Ala		
130	135	140
Gly Met Ala Ser Gly Arg Thr Glu Ala Gln Ser Trp Lys Ser Gln Asp		
145	150	155
Thr Lys Thr Thr Gln Gly Asn Gly Gly Gln Thr Arg Lys Leu Thr Ala		
	165	170
Ser Arg Thr Val Ser Glu Lys His Gln Gly Lys Ala Ala Thr Thr Ala		
	180	185
Lys Thr Leu Ile Pro Lys Ser Gln His Arg Met Leu Ala Pro Thr Gly		
	195	200
Ala Val Ser Thr Arg Thr Arg Gln Lys Gly Val Thr Thr Ala Val Ile		
	210	215
Pro Pro Lys Glu Lys Lys Pro Gln Ala Thr Pro Pro Pro Ala Pro Phe		
225	230	235
Gln Ser Pro Thr Thr Gln Arg Asn Gln Arg Leu Lys Ala Ala Asn Phe		
	245	250
Lys Ser Glu Pro Arg Trp Asp Phe Glu Glu Lys Tyr Ser Phe Glu Ile		
	260	265
Gly Gly Leu Gln Thr Thr Cys Pro Asp Ser Val Lys Ile Lys Ala Ser		
	275	280
Lys Ser Leu Trp Leu Gln Lys Leu Phe Leu Pro Asn Leu Thr Leu Phe		
	290	295
Leu Asp Ser Arg His Phe Asn Gln Ser Glu Trp Asp Arg Leu Glu His		
305	310	315
Phe Ala Pro Pro Phe Gly Phe Met Glu Leu Asn Tyr Ser Leu Val Gln		
	325	330
Lys Val Val Thr Arg Phe Pro Pro Val Pro Gln Gln Gln Leu Leu Leu		
	340	345
Ala Ser Leu Pro Ala Gly Ser Leu Arg Cys Ile Thr Cys Ala Val Val		
	355	360
		365

Gly Asn Gly Gly Ile Leu Asn Asn Ser His Met Gly Gln Glu Ile Asp
370 375 380

Ser His Asp Tyr Val Phe Arg Leu Ser Gly Ala Leu Ile Lys Gly Tyr
385 390 395 400

Glu Gln Asp Val Gly Thr Arg Thr Ser Phe Tyr Gly Phe Thr Ala Phe
405 410 415

Ser Leu Thr Gln Ser Leu Leu Ile Leu Gly Asn Arg Gly Phe Lys Asn
420 425 430

Val Pro Leu Gly Lys Asp Val Arg Tyr Leu His Phe Leu Glu Gly Thr
435 440 445

Arg Asp Tyr Glu Trp Leu Glu Ala Leu Leu Met Asn Gln Thr Val Met
450 455 460

Ser Lys Asn Leu Phe Trp Phe Arg His Arg Pro Gln Glu Ala Phe Arg
465 470 475 480

Glu Ala Leu His Met Asp Arg Tyr Leu Leu Leu His Pro Asp Phe Leu
485 490 495

Arg Tyr Met Lys Asn Arg Phe Leu Arg Ser Lys Thr Leu Asp Gly Ala
500 505 510

His Trp Arg Ile Tyr Arg Pro Thr Thr Gly Ala Leu Leu Leu Leu Thr
515 520 525

Ala Leu Gln Leu Cys Asp Gln Val Ser Ala Tyr Gly Phe Ile Thr Glu
530 535 540

Gly His Glu Arg Phe Ser Asp His Tyr Tyr Asp Thr Ser Trp Lys Arg
545 550 555 560

Leu Ile Phe Tyr Ile Asn His Asp Phe Lys Leu Glu Arg Glu Val Trp
565 570 575

Lys Arg Leu His Asp Glu Gly Ile Ile Arg Leu Tyr Gln Arg Pro Gly
580 585 590

Pro Gly Thr Ala Lys Ala Lys Asn
595 600

<210> 7

<211> 1125

<212> ADN

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1) .. (1125)

<400> 7

atg ggg ctc ccg cgc ggg tcg ttc ttc tgg gtg ctg ctc ctg ctc acg Met Gly Leu Pro Arg Gly Ser Phe Phe Trp Val Leu Leu Leu Thr 1 5 10 15	48
gct gcc tgc tcg ggg ctc ctc ttt gcc ctg tac ttc tcg gcg gtg cag Ala Ala Cys Ser Gly Leu Leu Phe Ala Leu Tyr Phe Ser Ala Val Gln 20 25 30	96
cgg tac ccg ggg cca gcg gcc gga gcc agg gac acc aca tca ttt gaa Arg Tyr Pro Gly Pro Ala Ala Gly Ala Arg Asp Thr Thr Ser Phe Glu 35 40 45	144
gca ttc ttt caa tcc aag gca tcg aat tct tgg aca gga aag ggc cag Ala Phe Phe Gln Ser Lys Ala Ser Asn Ser Trp Thr Gly Lys Gly Gln 50 55 60	192
gcc tgc cga cac ctg ctt cac ctg gcc att cag cgg cac ccc cac ttc Ala Cys Arg His Leu Leu His Leu Ala Ile Gln Arg His Pro His Phe 65 70 75 80	240
cgt ggc ctg ttc aat ctc tcc att cca gtg ctg ctg tgg ggg gac ctc Arg Gly Leu Phe Asn Leu Ser Ile Pro Val Leu Leu Trp Gly Asp Leu 85 90 95	288
ttc acc cca gcg ctc tgg gac cgc ctg agc caa cac aaa gcc ccg tat Phe Thr Pro Ala Leu Trp Asp Arg Leu Ser Gln His Lys Ala Pro Tyr 100 105 110	336
ggc tgg cgg ggg ctc tct cac caa gtc atc gcc tcc acc ctg agc ctt Gly Trp Arg Gly Leu Ser His Gln Val Ile Ala Ser Thr Leu Ser Leu 115 120 125	384
ctg aac ggc tca gag agt gcc aag ctg ttt gcc ccg ccc agg gac acc Leu Asn Gly Ser Glu Ser Ala Lys Leu Phe Ala Pro Pro Arg Asp Thr 130 135 140	432
cct cca aag tgt atc cgg tgt gcc gtg gtg ggc aac gga ggc att ctg Pro Pro Lys Cys Ile Arg Cys Ala Val Val Gly Asn Gly Gly Ile Leu 145 150 155 160	480
aat ggg tcc cgc cag ggt ccc aac atc gat gcc cat gac tat gta ttc Asn Gly Ser Arg Gln Gly Pro Asn Ile Asp Ala His Asp Tyr Val Phe 165 170 175	528
aga ctc aat gga gct gtg atc aaa ggc ttc gag cgc gat gtg ggc acc Arg Leu Asn Gly Ala Val Ile Lys Gly Phe Glu Arg Asp Val Gly Thr 180 185 190	576
aag act tcc ttc tat ggt ttc act gtg aac acg atg aag aac tcc ctc	624

Lys Thr Ser Phe Tyr Gly Phe Thr Val Asn Thr Met Lys Asn Ser Leu	
195 200 205	
gtc tcc tac tgg aat ctg ggc ttc acc tcc gtg cca caa gga cag gac	672
Val Ser Tyr Trp Asn Leu Gly Phe Thr Ser Val Pro Gln Gly Gln Asp	
210 215 220	
ctg cag tat atc ttc atc ccc tca gac atc cgc gac tat gtg atg ctg	720
Leu Gln Tyr Ile Phe Ile Pro Ser Asp Ile Arg Asp Tyr Val Met Leu	
225 230 235 240	
aga tcg gcc att ctg ggc gtg cct gtc cct gag ggc cta gat aaa ggg	768
Arg Ser Ala Ile Leu Gly Val Pro Val Pro Glu Gly Leu Asp Lys Gly	
245 250 255	
gac agg ccg cac gcc tat ttt gga cca gaa gcc tct gcc agt aaa ttc	816
Asp Arg Pro His Ala Tyr Phe Gly Pro Glu Ala Ser Ala Ser Lys Phe	
260 265 270	
aag ctg cta cat ccg gac ttc atc agc tac ctg aca gaa agg ttc ttg	864
Lys Leu Leu His Pro Asp Phe Ile Ser Tyr Leu Thr Glu Arg Phe Leu	
275 280 285	
aaa tca aag ttg att aac aca cat ttt gga gac cta tat atg cct agt	912
Lys Ser Lys Leu Ile Asn Thr His Phe Gly Asp Leu Tyr Met Pro Ser	
290 295 300	
acc ggg gct ctc atg ctg ctg aca gct ttg cat acc tgt gac cag gtc	960
Thr Gly Ala Leu Met Leu Leu Thr Ala Leu His Thr Cys Asp Gln Val	
305 310 315 320	
agt gcc tat gga ttc atc aca agc aac tac tgg aaa ttt tcc gac cac	1008
Ser Ala Tyr Gly Phe Ile Thr Ser Asn Tyr Trp Lys Phe Ser Asp His	
325 330 335	
tat ttc gaa cga aaa atg aag cca ttg ata ttt tat gca aac cac gat	1056
Tyr Phe Glu Arg Lys Met Lys Pro Leu Ile Phe Tyr Ala Asn His Asp	
340 345 350	
ctg tcc ctg gaa gct gcc ctg tgg agg gac ctg cac aag gcc ggc atc	1104
Leu Ser Leu Glu Ala Ala Leu Trp Arg Asp Leu His Lys Ala Gly Ile	
355 360 365	
ctt cag ctg tac cag cgc tga	1125
Leu Gln Leu Tyr Gln Arg	
370	

<210> 8

<211> 374

<212> PRT

<213> *Homo sapiens*

<400> 8

Met Gly Leu Pro Arg Gly Ser Phe Phe Trp Val Leu Leu Leu Leu Thr
1 5 10 15

Ala Ala Cys Ser Gly Leu Leu Phe Ala Leu Tyr Phe Ser Ala Val Gln
20 25 30

Arg Tyr Pro Gly Pro Ala Ala Gly Ala Arg Asp Thr Thr Ser Phe Glu
 35 40 45

Ala Phe Phe Gln Ser Lys Ala Ser Asn Ser Trp Thr Gly Lys Gly Gln
 50 55 60

Ala Cys Arg His Leu Leu His Leu Ala Ile Gln Arg His Pro His Phe
 65 70 75 80

Arg Gly Leu Phe Asn Leu Ser Ile Pro Val Leu Leu Trp Gly Asp Leu
 85 90 95

Phe Thr Pro Ala Leu Trp Asp Arg Leu Ser Gln His Lys Ala Pro Tyr
 100 105 110

Gly Trp Arg Gly Leu Ser His Gln Val Ile Ala Ser Thr Leu Ser Leu
 115 120 125

Leu Asn Gly Ser Glu Ser Ala Lys Leu Phe Ala Pro Pro Arg Asp Thr
 130 135 140

Pro Pro Lys Cys Ile Arg Cys Ala Val Val Gly Asn Gly Gly Ile Leu
 145 150 155 160

Asn Gly Ser Arg Gln Gly Pro Asn Ile Asp Ala His Asp Tyr Val Phe
 165 170 175

Arg Leu Asn Gly Ala Val Ile Lys Gly Phe Glu Arg Asp Val Gly Thr
 180 185 190

Lys Thr Ser Phe Tyr Gly Phe Thr Val Asn Thr Met Lys Asn Ser Leu
 195 200 205

Val Ser Tyr Trp Asn Leu Gly Phe Thr Ser Val Pro Gln Gly Gln Asp
 210 215 220

Leu Gln Tyr Ile Phe Ile Pro Ser Asp Ile Arg Asp Tyr Val Met Leu
 225 230 235 240

Arg Ser Ala Ile Leu Gly Val Pro Val Pro Glu Gly Leu Asp Lys Gly
 245 250 255

Asp Arg Pro His Ala Tyr Phe Gly Pro Glu Ala Ser Ala Ser Lys Phe
 260 265 270

Lys Leu Leu His Pro Asp Phe Ile Ser Tyr Leu Thr Glu Arg Phe Leu
275 280 285

Lys Ser Lys Leu Ile Asn Thr His Phe Gly Asp Leu Tyr Met Pro Ser
290 295 300

Thr Gly Ala Leu Met Leu Leu Thr Ala Leu His Thr Cys Asp Gln Val
305 310 315 320

Ser Ala Tyr Gly Phe Ile Thr Ser Asn Tyr Trp Lys Phe Ser Asp His
325 330 335

Tyr Phe Glu Arg Lys Met Lys Pro Leu Ile Phe Tyr Ala Asn His Asp
340 345 350

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

Leu Gln Leu Tyr Gln Arg
370

<210> 9

<211> 918

<212> ADN

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1) .. (918)

<400> 9

atg gcc tgc atc ctg aag aga aag tct gtg att gct gtg agc ttc ata	48
Met Ala Cys Ile Leu Lys Arg Lys Ser Val Ile Ala Val Ser Phe Ile	
1 5 10 15	
gca gcg ttc ctt ttc ctg ctg gtt gtg cgt ctt gta aat gaa gtg aat	96
Ala Ala Phe Leu Phe Leu Leu Val Val Arg Leu Val Asn Glu Val Asn	
20 25 30	
ttc cca ttg cta cta aac tgc ttt gga caa cct ggt aca aag tgg ata	144
Phe Pro Leu Leu Leu Asn Cys Phe Gly Gln Pro Gly Thr Lys Trp Ile	
35 40 45	
cca ttc tcc tac aca tac agg cgg ccc ctt cga act cac tat gga tac	192
Pro Phe Ser Tyr Thr Tyr Arg Arg Pro Leu Arg Thr His Tyr Gly Tyr	
50 55 60	
ata aat gtg aag aca caa gag cct ttg caa ctg gac tgt gac ctt tgt	240
Ile Asn Val Lys Thr Gln Glu Pro Leu Gln Leu Asp Cys Asp Leu Cys	
65 70 75 80	
gcc ata gtg tca aac tca ggt cag atg gtt ggc cag aag gtg gga aat	288

Ala	Ile	Val	Ser	Asn	Ser	Gly	Gln	Met	Val	Gly	Gln	Lys	Val	Gly	Asn		
				85					90						95		
gag	ata	gat	cga	tcc	tcc	tgc	att	tgg	aga	atg	aac	aat	gcc	ccc	acc		336
Glu	Ile	Asp	Arg	Ser	Ser	Cys	Ile	Trp	Arg	Met	Asn	Asn	Ala	Pro	Thr		
			100					105					110				
aaa	ggt	tat	gaa	gaa	gat	gtc	ggc	cgc	atg	acc	atg	att	cga	gtt	gtg		384
Lys	Gly	Tyr	Glu	Glu	Asp	Val	Gly	Arg	Met	Thr	Met	Ile	Arg	Val	Val		
		115					120					125					
tcc	cat	acc	agc	gtt	cct	ctt	ttg	cta	aaa	aac	cct	gat	tat	ttt	ttc		432
Ser	His	Thr	Ser	Val	Pro	Leu	Leu	Leu	Lys	Asn	Pro	Asp	Tyr	Phe	Phe		
		130					135				140						
aag	gaa	gcg	aat	act	act	att	tat	gtt	att	tgg	gga	cct	ttc	cgc	aat		480
Lys	Glu	Ala	Asn	Thr	Thr	Ile	Tyr	Val	Ile	Trp	Gly	Pro	Phe	Arg	Asn		
145					150					155					160		
atg	agg	aaa	gat	ggc	aat	ggc	atc	att	tac	aac	atg	ttg	aaa	aag	aca		528
Met	Arg	Lys	Asp	Gly	Asn	Gly	Ile	Ile	Tyr	Asn	Met	Leu	Lys	Lys	Thr		
				165					170						175		
gtt	ggt	atc	tat	ccg	aat	gcc	caa	ata	tac	gtg	acc	aca	gag	aag	cgc		576
Val	Gly	Ile	Tyr	Pro	Asn	Ala	Gln	Ile	Tyr	Val	Thr	Thr	Glu	Lys	Arg		
			180					185					190				
atg	agt	tac	tgt	gat	gga	gtt	ttt	aag	aag	gaa	act	ggg	aag	gac	aga		624
Met	Ser	Tyr	Cys	Asp	Gly	Val	Phe	Lys	Lys	Glu	Thr	Gly	Lys	Asp	Arg		
		195					200					205					
gtc	cag	tct	ggc	tca	tat	ctc	agc	aca	ggg	tgg	ttt	acc	ttc	ctt	ctg		672
Val	Gln	Ser	Gly	Ser	Tyr	Leu	Ser	Thr	Gly	Trp	Phe	Thr	Phe	Leu	Leu		
		210					215					220					
gcc	atg	gac	gcc	tgt	tat	ggc	att	cac	gtc	tac	ggg	atg	ata	aat	gac		720
Ala	Met	Asp	Ala	Cys	Tyr	Gly	Ile	His	Val	Tyr	Gly	Met	Ile	Asn	Asp		
225					230					235					240		
acc	tac	tgc	aag	aca	gaa	ggg	tat	aga	aaa	gtc	ccc	tac	cat	tat	tat		768
Thr	Tyr	Cys	Lys	Thr	Glu	Gly	Tyr	Arg	Lys	Val	Pro	Tyr	His	Tyr	Tyr		
				245					250					255			
gaa	caa	gga	aga	gat	gag	tgt	gat	gaa	tat	ttt	ctt	cat	gaa	cat	gcc		816
Glu	Gln	Gly	Arg	Asp	Glu	Cys	Asp	Glu	Tyr	Phe	Leu	His	Glu	His	Ala		
			260					265					270				
cca	tat	ggg	ggt	cat	agg	ttt	atc	act	gaa	aag	aaa	gtg	ttt	gct	aaa		864
Pro	Tyr	Gly	Gly	His	Arg	Phe	Ile	Thr	Glu	Lys	Lys	Val	Phe	Ala	Lys		
		275					280					285					
tgg	gcc	aag	aag	cac	agg	ata	ata	ttt	aca	cat	cca	aac	tgg	aca	ttg		912
Trp	Ala	Lys	Lys	His	Arg	Ile	Ile	Phe	Thr	His	Pro	Asn	Trp	Thr	Leu		
		290					295					300					
tct	tga																918
Ser																	
305																	

<210> 10

<211> 305

<212> PRT

<213> *Homo sapiens*

<400> 10

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

Ala Met Asp Ala Cys Tyr Gly Ile His Val Tyr Gly Met Ile Asn Asp
225 230 235 240

Thr Tyr Cys Lys Thr Glu Gly Tyr Arg Lys Val Pro Tyr His Tyr Tyr
245 250 255

Glu Gln Gly Arg Asp Glu Cys Asp Glu Tyr Phe Leu His Glu His Ala
260 265 270

Pro Tyr Gly Gly His Arg Phe Ile Thr Glu Lys Lys Val Phe Ala Lys
275 280 285

Trp Ala Lys Lys His Arg Ile Ile Phe Thr His Pro Asn Trp Thr Leu
290 295 300

Ser
305

<210> 11
<211> 909
<212> ADN
<213> *Homo sapiens*

<220>
<221> CDS
<222> (1) .. (909)

<400> 11

atg aag gct ccg ggt cgg ctc gtg ctc atc atc ctg tgc tcc gtg gtc	48
Met Lys Ala Pro Gly Arg Leu Val Leu Ile Ile Leu Cys Ser Val Val	
1 5 10 15	
ttc tct gcc gtc tac atc ctc ctg tgc tgc tgg gcc ggc ctg ccc ctc	96
Phe Ser Ala Val Tyr Ile Leu Leu Cys Cys Trp Ala Gly Leu Pro Leu	
20 25 30	
tgc ctg gcc acc tgc ctg gac cac cac ttc ccc aca ggc tcc agg ccc	144
Cys Leu Ala Thr Cys Leu Asp His His Phe Pro Thr Gly Ser Arg Pro	
35 40 45	
act gtg ccg gga ccc ctg cac ttc agt gga tat agc agt gtg cca gat	192
Thr Val Pro Gly Pro Leu His Phe Ser Gly Tyr Ser Ser Val Pro Asp	
50 55 60	
ggg aag ccg ctg gtc cgc gag ccc tgc cgc agc tgt gcc gtg gtg tcc	240
Gly Lys Pro Leu Val Arg Glu Pro Cys Arg Ser Cys Ala Val Val Ser	
65 70 75 80	
agc tcc gcc caa atg ctg ggc tca ggc ctg ggt gct gag atc gac agt	288
Ser Ser Gly Gln Met Leu Gly Ser Gly Leu Gly Ala Glu Ile Asp Ser	
85 90 95	
gcc gag tgc gtg ttc cgc atg aac cag gcg ccc acc gtg ggc ttt gag	336

Ala	Glu	Cys	Val	Phe	Arg	Met	Asn	Gln	Ala	Pro	Thr	Val	Gly	Phe	Glu		
			100					105					110				
gcg	gat	gtg	ggc	cag	cgc	agc	acc	ctg	cgt	gtc	gtc	tca	cac	aca	agc	384	
Ala	Asp	Val	Gly	Gln	Arg	Ser	Thr	Leu	Arg	Val	Val	Ser	His	Thr	Ser		
		115					120					125					
gtg	ccg	ctg	ctg	ctg	cgc	aac	tat	tca	cac	tac	ttc	cag	aag	gcc	cga	432	
Val	Pro	Leu	Leu	Leu	Arg	Asn	Tyr	Ser	His	Tyr	Phe	Gln	Lys	Ala	Arg		
		130				135					140						
gac	acg	ctc	tac	atg	gtg	tgg	ggc	cag	ggc	agg	cac	atg	gac	cgg	gtg	480	
Asp	Thr	Leu	Tyr	Met	Val	Trp	Gly	Gln	Gly	Arg	His	Met	Asp	Arg	Val		
145					150					155					160		
ctc	ggc	ggc	cgc	acc	tac	cgc	acg	ctg	ctg	cag	ctc	acc	agg	atg	tac	528	
Leu	Gly	Gly	Arg	Thr	Tyr	Arg	Thr	Leu	Leu	Gln	Leu	Thr	Arg	Met	Tyr		
			165							170					175		
ccc	ggc	ctg	cag	gtg	tac	acc	ttc	acg	gag	cgc	atg	atg	gcc	tac	tgc	576	
Pro	Gly	Leu	Gln	Val	Tyr	Thr	Phe	Thr	Glu	Arg	Met	Met	Ala	Tyr	Cys		
			180					185					190				
gac	cag	atc	ttc	cag	gac	gag	acg	ggc	aag	aac	cgg	agg	cag	tcg	ggc	624	
Asp	Gln	Ile	Phe	Gln	Asp	Glu	Thr	Gly	Lys	Asn	Arg	Arg	Gln	Ser	Gly		
		195					200					205					
tcc	ttc	ctc	agc	acc	ggc	tgg	ttc	acc	atg	atc	ctc	gcg	ctg	gag	ctg	672	
Ser	Phe	Leu	Ser	Thr	Gly	Trp	Phe	Thr	Met	Ile	Leu	Ala	Leu	Glu	Leu		
	210					215					220						
tgt	gag	gag	atc	gtg	gtc	tat	ggg	atg	gtc	agc	gac	agc	tac	tgc	agg	720	
Cys	Glu	Glu	Ile	Val	Val	Tyr	Gly	Met	Val	Ser	Asp	Ser	Tyr	Cys	Arg		
225					230					235					240		
gag	aag	agc	cac	ccc	tca	gtg	cct	tac	cac	tac	ttt	gag	aag	ggc	cgg	768	
Glu	Lys	Ser	His	Pro	Ser	Val	Pro	Tyr	His	Tyr	Phe	Glu	Lys	Gly	Arg		
			245						250					255			
cta	gat	gag	tgt	cag	atg	tac	ctg	gca	cac	gag	cag	gcg	ccc	cga	agc	816	
Leu	Asp	Glu	Cys	Gln	Met	Tyr	Leu	Ala	His	Glu	Gln	Ala	Pro	Arg	Ser		
			260					265					270				
gcc	cac	cgc	ttc	atc	act	gag	aag	gcg	gtc	ttc	tcc	cgc	tgg	gcc	aag	864	
Ala	His	Arg	Phe	Ile	Thr	Glu	Lys	Ala	Val	Phe	Ser	Arg	Trp	Ala	Lys		
		275					280					285					
aag	agg	ccc	atc	gtg	ttc	gcc	cat	ccg	tcc	tgg	agg	act	gag	tag		909	
Lys	Arg	Pro	Ile	Val	Phe	Ala	His	Pro	Ser	Trp	Arg	Thr	Glu				
		290				295					300						

<210> 12

<211> 302

<212> PRT

<213> *Homo sapiens*

<400> 12

Met Lys Ala Pro Gly Arg Leu Val Leu Ile Ile Leu Cys Ser Val Val
1 5 10 15

Phe Ser Ala Val Tyr Ile Leu Leu Cys Cys Trp Ala Gly Leu Pro Leu
 20 25 30

Cys Leu Ala Thr Cys Leu Asp His His Phe Pro Thr Gly Ser Arg Pro
 35 40 45

Thr Val Pro Gly Pro Leu His Phe Ser Gly Tyr Ser Ser Val Pro Asp
 50 55 60

Gly Lys Pro Leu Val Arg Glu Pro Cys Arg Ser Cys Ala Val Val Ser
 65 70 75 80

Ser Ser Gly Gln Met Leu Gly Ser Gly Leu Gly Ala Glu Ile Asp Ser
 85 90 95

Ala Glu Cys Val Phe Arg Met Asn Gln Ala Pro Thr Val Gly Phe Glu
 100 105 110

Ala Asp Val Gly Gln Arg Ser Thr Leu Arg Val Val Ser His Thr Ser
 115 120 125

Val Pro Leu Leu Leu Arg Asn Tyr Ser His Tyr Phe Gln Lys Ala Arg
 130 135 140

Asp Thr Leu Tyr Met Val Trp Gly Gln Gly Arg His Met Asp Arg Val
 145 150 155 160

Leu Gly Gly Arg Thr Tyr Arg Thr Leu Leu Gln Leu Thr Arg Met Tyr
 165 170 175

Pro Gly Leu Gln Val Tyr Thr Phe Thr Glu Arg Met Met Ala Tyr Cys
 180 185 190

Asp Gln Ile Phe Gln Asp Glu Thr Gly Lys Asn Arg Arg Gln Ser Gly
 195 200 205

Ser Phe Leu Ser Thr Gly Trp Phe Thr Met Ile Leu Ala Leu Glu Leu
 210 215 220

Cys Glu Glu Ile Val Val Tyr Gly Met Val Ser Asp Ser Tyr Cys Arg
 225 230 235 240

Glu Lys Ser His Pro Ser Val Pro Tyr His Tyr Phe Glu Lys Gly Arg
 245 250 255

Leu Asp Glu Cys Gln Met Tyr Leu Ala His Glu Gln Ala Pro Arg Ser
260 265 270

Ala His Arg Phe Ile Thr Glu Lys Ala Val Phe Ser Arg Trp Ala Lys
275 280 285

Lys Arg Pro Ile Val Phe Ala His Pro Ser Trp Arg Thr Glu
290 295 300

<210> 13

<211> 1011

<212> ADN

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1)..(1011)

<400> 13

atg aag acc ctg atg cgc cat ggt ctg gca gtg tgt tta gcg ctc acc Met Lys Thr Leu Met Arg His Gly Leu Ala Val Cys Leu Ala Leu Thr 1 5 10 15	48
acg atg tgc acc agc ttg ttg cta gtg tac agc agc ctc ggc ggc cag Thr Met Cys Thr Ser Leu Leu Leu Val Tyr Ser Ser Leu Gly Gly Gln 20 25 30	96
aag gag cgg ccc ccg cag cag cag cag cag cag cag caa cag cag cag Lys Glu Arg Pro Pro Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln 35 40 45	144
cag gcg tgc gcc acc ggc agc tgc cag ccg gcg gcg gag agc agc acc Gln Ala Ser Ala Thr Gly Ser Ser Gln Pro Ala Ala Glu Ser Ser Thr 50 55 60	192
cag cag cgc ccc ggg gtc ccc gcg gga ccg cgg cca ctg gac gga tac Gln Gln Arg Pro Gly Val Pro Ala Gly Pro Arg Pro Leu Asp Gly Tyr 65 70 75 80	240
ctc gga gtg gcg gac cac aag ccc ctg aaa atg cac tgc agg gac tgt Leu Gly Val Ala Asp His Lys Pro Leu Lys Met His Cys Arg Asp Cys 85 90 95	288
gcc ctg gtg acc agc tca ggg cat ctg ctg cac agt cgg caa ggc tcc Ala Leu Val Thr Ser Ser Gly His Leu Leu His Ser Arg Gln Gly Ser 100 105 110	336
cag att gac cag aca gag tgt gtc atc cgc atg aat gac gcc ccc aca Gln Ile Asp Gln Thr Glu Cys Val Ile Arg Met Asn Asp Ala Pro Thr 115 120 125	384
cgc ggc tat ggg cgt gac gtg ggc aat cgc acc agc ctg agg gtc atc Arg Gly Tyr Gly Arg Asp Val Gly Asn Arg Thr Ser Leu Arg Val Ile 130 135 140	432
gcg cat tcc agc atc cag agg atc ctc cgc aac cgc cat gac ctg ctc	480

Ala	His	Ser	Ser	Ile	Gln	Arg	Ile	Leu	Arg	Asn	Arg	His	Asp	Leu	Leu		
145					150					155					160		
aac	gtg	agc	cag	ggc	acc	gtg	ttc	atc	ttc	tgg	ggc	ccc	agc	agc	tac	528	
Asn	Val	Ser	Gln	Gly	Thr	Val	Phe	Ile	Phe	Trp	Gly	Pro	Ser	Ser	Tyr		
			165					170						175			
atg	cgg	cgg	gac	ggc	aag	ggc	cag	gtc	tac	aac	aac	ctg	cat	ctc	ctg	576	
Met	Arg	Arg	Asp	Gly	Lys	Gly	Gln	Val	Tyr	Asn	Asn	Leu	His	Leu	Leu		
			180					185					190				
agc	cag	gtg	ctg	ccc	cgg	ctg	aag	gcc	ttc	atg	att	act	cgc	cac	aag	624	
Ser	Gln	Val	Leu	Pro	Arg	Leu	Lys	Ala	Phe	Met	Ile	Thr	Arg	His	Lys		
		195					200					205					
atg	ctg	cag	ttt	gat	gag	ctc	ttc	aag	cag	gag	act	ggc	aaa	gac	agg	672	
Met	Leu	Gln	Phe	Asp	Glu	Leu	Phe	Lys	Gln	Glu	Thr	Gly	Lys	Asp	Arg		
	210					215					220						
aag	ata	tcc	aac	act	tgg	ctc	agc	act	ggc	tgg	ttt	aca	atg	aca	att	720	
Lys	Ile	Ser	Asn	Thr	Trp	Leu	Ser	Thr	Gly	Trp	Phe	Thr	Met	Thr	Ile		
	225				230					235					240		
gca	ctg	gag	ctc	tgt	gac	agg	atc	aat	gtt	tat	ggc	atg	gtg	ccc	cca	768	
Ala	Leu	Glu	Leu	Cys	Asp	Arg	Ile	Asn	Val	Tyr	Gly	Met	Val	Pro	Pro		
				245					250					255			
gac	ttc	tgc	agg	gat	ccc	aat	cac	cct	tca	gta	cct	tat	cat	tat	tat	816	
Asp	Phe	Cys	Arg	Asp	Pro	Asn	His	Pro	Ser	Val	Pro	Tyr	His	Tyr	Tyr		
			260					265					270				
gaa	cct	ttt	gga	cct	gat	gaa	tgt	aca	atg	tac	ctc	tcc	cat	gag	cga	864	
Glu	Pro	Phe	Gly	Pro	Asp	Glu	Cys	Thr	Met	Tyr	Leu	Ser	His	Glu	Arg		
		275					280					285					
gga	cgc	aag	ggc	agt	cat	cac	cgc	ttt	atc	aca	gag	aaa	cga	gtc	ttt	912	
Gly	Arg	Lys	Gly	Ser	His	His	Arg	Phe	Ile	Thr	Glu	Lys	Arg	Val	Phe		
		290				295					300						
aag	aac	tgg	gca	cgg	aca	ttc	aat	att	cac	ttt	ttt	caa	cca	gac	tgg	960	
Lys	Asn	Trp	Ala	Arg	Thr	Phe	Asn	Ile	His	Phe	Phe	Gln	Pro	Asp	Trp		
	305				310					315					320		
aaa	cca	gaa	tca	ctt	gct	ata	aat	cat	cct	gag	aat	aaa	cct	gtg	ttc	1008	
Lys	Pro	Glu	Ser	Leu	Ala	Ile	Asn	His	Pro	Glu	Asn	Lys	Pro	Val	Phe		
			325						330					335			
taa																1011	

<210> 14

<211> 336

<212> PRT

<213> *Homo sapiens*

<400> 14

Met	Lys	Thr	Leu	Met	Arg	His	Gly	Leu	Ala	Val	Cys	Leu	Ala	Leu	Thr
1				5					10					15	

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

260	265	270
Glu Pro Phe Gly Pro Asp Glu Cys Thr Met Tyr Leu Ser His Glu Arg 275	280	285
Gly Arg Lys Gly Ser His His Arg Phe Ile Thr Glu Lys Arg Val Phe 290	295	300
Lys Asn Trp Ala Arg Thr Phe Asn Ile His Phe Phe Gln Pro Asp Trp 305	310	315
Lys Pro Glu Ser Leu Ala Ile Asn His Pro Glu Asn Lys Pro Val Phe 325	330	335

<210> 15

<211> 1002

<212> ADN

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1)..(1002)

<400> 15

atg gct tgc tcg agg ccc ccc agc cag tgt gaa ccc aca tcc ctg ccc	48
Met Ala Cys Ser Arg Pro Pro Ser Gln Cys Glu Pro Thr Ser Leu Pro	
1 5 10 15	
cca ggg cca cct gca gga cgc cga cac cta ccc ctc agc aga cgc cgg	96
Pro Gly Pro Pro Ala Gly Arg Arg His Leu Pro Leu Ser Arg Arg Arg	
20 25 30	
aga gaa atg agt agc aac aaa gag cag cgg tca gca gtg ttc gtg atc	144
Arg Glu Met Ser Ser Asn Lys Glu Gln Arg Ser Ala Val Phe Val Ile	
35 40 45	
ctc ttt gcc ctc atc acc atc ctc atc ctc tac agc tcc aac agt gcc	192
Leu Phe Ala Leu Ile Thr Ile Leu Ile Leu Tyr Ser Ser Asn Ser Ala	
50 55 60	
aat gag gtc ttc cat tac ggc tcc ctg cgg ggc cgt agc cgc cga cct	240
Asn Glu Val Phe His Tyr Gly Ser Leu Arg Gly Arg Ser Arg Arg Pro	
65 70 75 80	
gtc aac ctc aag aag tgg agc atc act gac ggc tat gtc ccc att ctc	288
Val Asn Leu Lys Lys Trp Ser Ile Thr Asp Gly Tyr Val Pro Ile Leu	
85 90 95	
ggc aac aag aca ctg ccc tct cgg tgc cac cag tgt gtg att gtc agc	336
Gly Asn Lys Thr Leu Pro Ser Arg Cys His Gln Cys Val Ile Val Ser	
100 105 110	
agc tcc agc cac ctg ctg ggc acc aag ctg ggc cct gag atc gag cgg	384
Ser Ser Ser His Leu Leu Gly Thr Lys Leu Gly Pro Glu Ile Glu Arg	
115 120 125	

gct gag tgt aca atc cgc atg aat gat gca ccc acc act ggc tac tca Ala Glu Cys Thr Ile Arg Met Asn Asp Ala Pro Thr Thr Gly Tyr Ser 130 135 140	432
gct gat gtg ggc aac aag acc acc tac cgc gtc gtg gcc cat tcc agt Ala Asp Val Gly Asn Lys Thr Thr Tyr Arg Val Val Ala His Ser Ser 145 150 155 160	480
gtg ttc cgc gtg ctg agg agg ccc cag gag ttt gtc aac cgg acc cct Val Phe Arg Val Leu Arg Arg Pro Gln Glu Phe Val Asn Arg Thr Pro 165 170 175	528
gaa acc gtg ttc atc ttc tgg ggg ccc ccg agc aag atg cag aag ccc Glu Thr Val Phe Ile Phe Trp Gly Pro Pro Ser Lys Met Gln Lys Pro 180 185 190	576
cag ggc agc ctc gtg cgt gtg atc cag cga gcg ggc ctg gtg ttc ccc Gln Gly Ser Leu Val Arg Val Ile Gln Arg Ala Gly Leu Val Phe Pro 195 200 205	624
aac atg gaa gca tat gcc gtc tct ccc ggc cgc atg cgg caa ttt gac Asn Met Glu Ala Tyr Ala Val Ser Pro Gly Arg Met Arg Gln Phe Asp 210 215 220	672
gac ctc ttc cgg ggt gag acg ggc aag gac agg gag aag tct cat tcg Asp Leu Phe Arg Gly Glu Thr Gly Lys Asp Arg Glu Lys Ser His Ser 225 230 235 240	720
tgg ttg agc aca ggc tgg ttt acc atg gtg atc gcg gtg gag ttg tgt Trp Leu Ser Thr Gly Trp Phe Thr Met Val Ile Ala Val Glu Leu Cys 245 250 255	768
gac cac gtg cat gtc tat ggc atg gtc ccc ccc aac tac tgc agc cag Asp His Val His Val Tyr Gly Met Val Pro Pro Asn Tyr Cys Ser Gln 260 265 270	816
cgg ccc cgc ctc cag cgc atg ccc tac cac tac tac gag ccc aag ggg Arg Pro Arg Leu Gln Arg Met Pro Tyr His Tyr Tyr Glu Pro Lys Gly 275 280 285	864
ccg gac gaa tgt gtc acc tac atc cag aat gag cac agt cgc aag ggc Pro Asp Glu Cys Val Thr Tyr Ile Gln Asn Glu His Ser Arg Lys Gly 290 295 300	912
aac cac cac cgc ttc atc acc gag aaa agg gtc ttc tca tcg tgg gcc Asn His His Arg Phe Ile Thr Glu Lys Arg Val Phe Ser Ser Trp Ala 305 310 315 320	960
cag ctg tat ggc atc acc ttc tcc cac ccc tcc tgg acc tag Gln Leu Tyr Gly Ile Thr Phe Ser His Pro Ser Trp Thr 325 330	1002

<210> 16

<211> 333

<212> PRT

<213> *Homo sapiens*

<400> 16

Met¹ Ala Cys Ser Arg Pro Pro Ser Gln Cys Glu Pro Thr Ser Leu Pro
1 5 10 15

Pro Gly Pro Pro Ala Gly Arg Arg His Leu Pro Leu Ser Arg Arg Arg
20 25 30

Arg Glu Met Ser Ser Asn Lys Glu Gln Arg Ser Ala Val Phe Val Ile
35 40 45

Leu Phe Ala Leu Ile Thr Ile Leu Ile Leu Tyr Ser Ser Asn Ser Ala
50 55 60

Asn Glu Val Phe His Tyr Gly Ser Leu Arg Gly Arg Ser Arg Arg Pro
65 70 75 80

Val Asn Leu Lys Lys Trp Ser Ile Thr Asp Gly Tyr Val Pro Ile Leu
85 90 95

Gly Asn Lys Thr Leu Pro Ser Arg Cys His Gln Cys Val Ile Val Ser
100 105 110

Ser Ser Ser His Leu Leu Gly Thr Lys Leu Gly Pro Glu Ile Glu Arg
115 120 125

Ala Glu Cys Thr Ile Arg Met Asn Asp Ala Pro Thr Thr Gly Tyr Ser
130 135 140

Ala Asp Val Gly Asn Lys Thr Thr Tyr Arg Val Val Ala His Ser Ser
145 150 155 160

Val Phe Arg Val Leu Arg Arg Pro Gln Glu Phe Val Asn Arg Thr Pro
165 170 175

Glu Thr Val Phe Ile Phe Trp Gly Pro Pro Ser Lys Met Gln Lys Pro
180 185 190

Gln Gly Ser Leu Val Arg Val Ile Gln Arg Ala Gly Leu Val Phe Pro
195 200 205

Asn Met Glu Ala Tyr Ala Val Ser Pro Gly Arg Met Arg Gln Phe Asp
210 215 220

Asp Leu Phe Arg Gly Glu Thr Gly Lys Asp Arg Glu Lys Ser His Ser
225 230 235 240

Trp Leu Ser Thr Gly Trp Phe Thr Met Val Ile Ala Val Glu Leu Cys

	245		250		255	
Asp His Val His Val Tyr Gly Met Val Pro Pro Asn Tyr Cys Ser Gln						
	260		265		270	
Arg Pro Arg Leu Gln Arg Met Pro Tyr His Tyr Tyr Glu Pro Lys Gly						
	275		280		285	
Pro Asp Glu Cys Val Thr Tyr Ile Gln Asn Glu His Ser Arg Lys Gly						
	290		295		300	
Asn His His Arg Phe Ile Thr Glu Lys Arg Val Phe Ser Ser Trp Ala						
	305		310		315	320
Gln Leu Tyr Gly Ile Thr Phe Ser His Pro Ser Trp Thr						
	325		330			

<210> 17

<211> 1023

<212> ADN

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1)..(1023)

<400> 17

atg gtg acc ctg cgg aag agg acc ctg aaa gtg ctc acc ttc ctc gtg Met Val Thr Leu Arg Lys Arg Thr Leu Lys Val Leu Thr Phe Leu Val 1 5 10 15	48
ctc ttc atc ttc ctc acc tcc ttc ttc ctg aac tac tcc cac acc atg Leu Phe Ile Phe Leu Thr Ser Phe Phe Leu Asn Tyr Ser His Thr Met 20 25 30	96
gtg gcc acc acc tgg ttc ccc aag cag atg gtc ctg gag ctc tcc gag Val Ala Thr Thr Trp Phe Pro Lys Gln Met Val Leu Glu Leu Ser Glu 35 40 45	144
aac ctg aag aga ctg atc aag cac agg cct tgc acc tgc acc cac tgc Asn Leu Lys Arg Leu Ile Lys His Arg Pro Cys Thr Cys Thr His Cys 50 55 60	192
atc ggg cag cgc aag ctc tcg gcc tgg ttc gat gag agg ttc aac cag Ile Gly Gln Arg Lys Leu Ser Ala Trp Phe Asp Glu Arg Phe Asn Gln 65 70 75 80	240
acc atg cag ccg ctg ctg acc gcc cag aac gcg ctc ttg gag gac gac Thr Met Gln Pro Leu Leu Thr Ala Gln Asn Ala Leu Leu Glu Asp Asp 85 90 95	288
acc tac cga tgg tgg ctg agg ctc cag cgg gag aag aag ccc aat aac Thr Tyr Arg Trp Trp Leu Arg Leu Gln Arg Glu Lys Lys Pro Asn Asn 100 105 110	336

ttg aat gac acc atc aag gag ctg ttc aga gtg gtg cct ggg aat gtg Leu Asn Asp Thr Ile Lys Glu Leu Phe Arg Val Val Pro Gly Asn Val 115 120 125	384
gac cct atg ctg gag aag agg tcg gtg ggc tgc cgg cgc tgc gcc gtt Asp Pro Met Leu Glu Lys Arg Ser Val Gly Cys Arg Arg Cys Ala Val 130 135 140	432
gtg ggc aac tcg ggc aac ctg agg gag tct tct tat ggg cct gag ata Val Gly Asn Ser Gly Asn Leu Arg Glu Ser Ser Tyr Gly Pro Glu Ile 145 150 155 160	480
gac agt cac gac ttt gtc ctc agg atg aac aag gcg ccc acg gca ggg Asp Ser His Asp Phe Val Leu Arg Met Asn Lys Ala Pro Thr Ala Gly 165 170 175	528
ttt gaa gct gat gtt ggg acc aag acc acc cac cat ctg gtg tac cct Phe Glu Ala Asp Val Gly Thr Lys Thr Thr His His Leu Val Tyr Pro 180 185 190	576
gag agc ttc cgg gag ctg gga gat aat gtc agc atg atc ctg gtg ccc Glu Ser Phe Arg Glu Leu Gly Asp Asn Val Ser Met Ile Leu Val Pro 195 200 205	624
ttc aag acc atc gac ttg gag tgg gtg gtg agc gcc atc acc acg ggc Phe Lys Thr Ile Asp Leu Glu Trp Val Val Ser Ala Ile Thr Thr Gly 210 215 220	672
acc att tcc cac acc tac atc cgg gtt cct gca aag atc aga gtg aaa Thr Ile Ser His Thr Tyr Ile Pro Val Pro Ala Lys Ile Arg Val Lys 225 230 235 240	720
cag gat aag atc ctg atc tac cac cca gcc ttc atc aag tat gtc ttt Gln Asp Lys Ile Leu Ile Tyr His Pro Ala Phe Ile Lys Tyr Val Phe 245 250 255	768
gac aac tgg ctg caa ggg cac ggg cga tac cca tct acc ggc atc ctc Asp Asn Trp Leu Gln Gly His Gly Arg Tyr Pro Ser Thr Gly Ile Leu 260 265 270	816
tcg gtc atc ttc tca atg cat gtc tgc gat gag gtg gac ttg tac ggc Ser Val Ile Phe Ser Met His Val Cys Asp Glu Val Asp Leu Tyr Gly 275 280 285	864
ttc ggg gca gac agc aaa ggg aac tgg cac cac tac tgg gag aac aac Phe Gly Ala Asp Ser Lys Gly Asn Trp His His Tyr Trp Glu Asn Asn 290 295 300	912
cca tcc gcg ggg gct ttt cgc aag acg ggg gtg cac gat gca gac ttt Pro Ser Ala Gly Ala Phe Arg Lys Thr Gly Val His Asp Ala Asp Phe 305 310 315 320	960
gag tct aac gtg acg gcc acc ttg gcc tcc atc aat aaa atc cgg atc Glu Ser Asn Val Thr Ala Thr Leu Ala Ser Ile Asn Lys Ile Arg Ile 325 330 335	1008
ttc aag ggg aga tga Phe Lys Gly Arg 340	1023

<210> 18

<211> 340

<212> PRT

<213> *Homo sapiens*

<400> 18

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

210				215				220							
Thr	Ile	Ser	His	Thr	Tyr	Ile	Pro	Val	Pro	Ala	Lys	Ile	Arg	Val	Lys
225					230					235					240
Gln	Asp	Lys	Ile	Leu	Ile	Tyr	His	Pro	Ala	Phe	Ile	Lys	Tyr	Val	Phe
				245					250					255	
Asp	Asn	Trp	Leu	Gln	Gly	His	Gly	Arg	Tyr	Pro	Ser	Thr	Gly	Ile	Leu
			260					265					270		
Ser	Val	Ile	Phe	Ser	Met	His	Val	Cys	Asp	Glu	Val	Asp	Leu	Tyr	Gly
		275					280					285			
Phe	Gly	Ala	Asp	Ser	Lys	Gly	Asn	Trp	His	His	Tyr	Trp	Glu	Asn	Asn
	290					295					300				
Pro	Ser	Ala	Gly	Ala	Phe	Arg	Lys	Thr	Gly	Val	His	Asp	Ala	Asp	Phe
305					310					315					320
Glu	Ser	Asn	Val	Thr	Ala	Thr	Leu	Ala	Ser	Ile	Asn	Lys	Ile	Arg	Ile
				325					330					335	
Phe	Lys	Gly	Arg												
			340												

```
<210> 19
<211> 1053
<212> ADN
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> (1)..(1053)
```

<400> 19

atg	aag	tgc	tcc	ctg	cgg	gtg	tgg	ttc	ctc	tcc	gtg	gcc	ttc	ctg	ctg	48
Met	Lys	Cys	Ser	Leu	Arg	Val	Trp	Phe	Leu	Ser	Val	Ala	Phe	Leu	Leu	
1				5				10						15		
gtg	ttc	atc	atg	tcc	ctg	ctc	ttc	acc	tac	tcg	cac	cac	agc	atg	gcc	96
Val	Phe	Ile	Met	Ser	Leu	Leu	Phe	Thr	Tyr	Ser	His	His	Ser	Met	Ala	
			20					25					30			
acg	ctc	ccc	tac	ctg	gac	tca	ggg	gcc	ctg	gat	ggg	acg	cac	cgg	gtg	144
Thr	Leu	Pro	Tyr	Leu	Asp	Ser	Gly	Ala	Leu	Asp	Gly	Thr	His	Arg	Val	
		35					40					45				
aag	ctg	gtg	ccc	ggc	tat	gcc	ggc	ctg	cag	cgc	ctc	agc	aag	gag	agg	192
Lys	Leu	Val	Pro	Gly	Tyr	Ala	Gly	Leu	Gln	Arg	Leu	Ser	Lys	Glu	Arg	
	50					55					60					

ctc tcg ggc aag agc tgt gcc tgt cgc cgc tgc atg ggc gat gcc ggt Leu Ser Gly Lys Ser Cys Ala Cys Arg Arg Cys Met Gly Asp Ala Gly 65 70 75 80	240
gcc tcc gac tgg ttt gac agc cac ttt gac ggt aac att tcc ccc gtc Ala Ser Asp Trp Phe Asp Ser His Phe Asp Gly Asn Ile Ser Pro Val 85 90 95	288
tgg acc cga gag aac atg gat ctt cca ccg gac gtc cag agg tgg tgg Trp Thr Arg Glu Asn Met Asp Leu Pro Pro Asp Val Gln Arg Trp Trp 100 105 110	336
atg atg ctg cag ccc cag ttc aag tca cac aac acc aat gag gtg ctg Met Met Leu Gln Pro Gln Phe Lys Ser His Asn Thr Asn Glu Val Leu 115 120 125	384
gag aag ctg ttc cag ata gtg cct ggc gag aac ccc tac cgc ttc cgg Glu Lys Leu Phe Gln Ile Val Pro Gly Glu Asn Pro Tyr Arg Phe Arg 130 135 140	432
gac ccc cac cag tgc cgg cgc tgt gcc gtg gtg ggg aac tcg ggc aac Asp Pro His Gln Cys Arg Arg Cys Ala Val Val Gly Asn Ser Gly Asn 145 150 155 160	480
ctg cgg ggc tct ggc tat ggg cag gac gtg gac ggg cac aac ttc atc Leu Arg Gly Ser Gly Tyr Gly Gln Asp Val Asp Gly His Asn Phe Ile 165 170 175	528
atg agg atg aat cag gcg cca acc gtg ggc ttt gag cag gat gtt ggc Met Arg Met Asn Gln Ala Pro Thr Val Gly Phe Glu Gln Asp Val Gly 180 185 190	576
agc cga acc acc cac cat ttc atg tac cct gag agt gcc aag aac ctg Ser Arg Thr Thr His His Phe Met Tyr Pro Glu Ser Ala Lys Asn Leu 195 200 205	624
ccc gcc aac gtc agc ttc gtg ctg gtg ccc ttc aag gtc ctg gac ctt Pro Ala Asn Val Ser Phe Val Leu Val Pro Phe Lys Val Leu Asp Leu 210 215 220	672
ctg tgg atc gcc agc gcc ttg tcc acg ggg cag atc cga ttc acc tac Leu Trp Ile Ala Ser Ala Leu Ser Thr Gly Gln Ile Arg Phe Thr Tyr 225 230 235 240	720
gcc cca gtg aag tcc ttc ctt cga gtg gat aaa gaa aag gtc cag atc Ala Pro Val Lys Ser Phe Leu Arg Val Asp Lys Glu Lys Val Gln Ile 245 250 255	768
tac aac cca gcc ttc ttc aag tat atc cac gac agg tgg aca gag cat Tyr Asn Pro Ala Phe Phe Lys Tyr Ile His Asp Arg Trp Thr Glu His 260 265 270	816
cac ggg cgg tac cct tcc acg ggg atg ctg gtg ctt ttc ttt gcc ctg His Gly Arg Tyr Pro Ser Thr Gly Met Leu Val Leu Phe Phe Ala Leu 275 280 285	864
cat gtg tgt gat gag gtg aac gtg tac ggg ttc ggg gcc gac agc cgg His Val Cys Asp Glu Val Asn Val Tyr Gly Phe Gly Ala Asp Ser Arg 290 295 300	912

ggc aac tgg cac cac tac tgg gag aac aac cgg tac gcg ggc gag ttc	960
Gly Asn Trp His His Tyr Trp Glu Asn Asn Arg Tyr Ala Gly Glu Phe	
305 310 315 320	
cgg aag act ggc gtg cac gac gcg gac ttc gag gcc cac atc atc gac	1008
Arg Lys Thr Gly Val His Asp Ala Asp Phe Glu Ala His Ile Ile Asp	
325 330 335	
atg ctg gcc aag gcc agc aag atc gaa gtc tac cgg ggc aac tga	1053
Met Leu Ala Lys Ala Ser Lys Ile Glu Val Tyr Arg Gly Asn	
340 345 350	

<210> 20

<211> 350

<212> PRT

<213> *Homo sapiens*

<400> 20

Met Lys Cys Ser Leu Arg Val Trp Phe Leu Ser Val Ala Phe Leu Leu
 1 5 10 15

Val Phe Ile Met Ser Leu Leu Phe Thr Tyr Ser His His Ser Met Ala
 20 25 30

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

Lys Leu Val Pro Gly Tyr Ala Gly Leu Gln Arg Leu Ser Lys Glu Arg
 50 55 60

Leu Ser Gly Lys Ser Cys Ala Cys Arg Arg Cys Met Gly Asp Ala Gly
 65 70 75 80

Ala Ser Asp Trp Phe Asp Ser His Phe Asp Gly Asn Ile Ser Pro Val
 85 90 95

Trp Thr Arg Glu Asn Met Asp Leu Pro Pro Asp Val Gln Arg Trp Trp
 100 105 110

Met Met Leu Gln Pro Gln Phe Lys Ser His Asn Thr Asn Glu Val Leu
 115 120 125

Glu Lys Leu Phe Gln Ile Val Pro Gly Glu Asn Pro Tyr Arg Phe Arg
 130 135 140

Asp Pro His Gln Cys Arg Arg Cys Ala Val Val Gly Asn Ser Gly Asn
 145 150 155 160

Leu Arg Gly Ser Gly Tyr Gly Gln Asp Val Asp Gly His Asn Phe Ile

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

```
<210> 21
<211> 1335
<212> ADN
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> (1) .. (1335)
```

<400> 21

atg gga ctc ttg gta ttt gtg cgc aat ctg ctg cta gcc ctc tgc ctc	48
Met Gly Leu Leu Val Phe Val Arg Asn Leu Leu Leu Ala Leu Cys Leu	
1 5 10 15	

ttt ctg gta ctg gga ttt ttg tat tat tct gcg tgg aag cta cac tta Phe Leu Val Leu Gly Phe Leu Tyr Tyr Ser Ala Trp Lys Leu His Leu 20 25 30	96
ctc cag tgg gag gag gac tcc agt aag tat agt cac tct agc tca ccc Leu Gln Trp Glu Glu Asp Ser Ser Lys Tyr Ser His Ser Ser Pro 35 40 45	144
cag gag aag cct gtt gca gat tca gtg gtt ctt tcc ttt gac tcc gct Gln Glu Lys Pro Val Ala Asp Ser Val Val Leu Ser Phe Asp Ser Ala 50 55 60	192
gga caa aca cta ggc tca gag tat gat cgg ttg ggc ttc ctc ctg aat Gly Gln Thr Leu Gly Ser Glu Tyr Asp Arg Leu Gly Phe Leu Leu Asn 65 70 75 80	240
ctg gac tct aaa ctc tca ccg agg act ctc tgc acg gtg gtt ttt ggc Leu Asp Ser Lys Leu Ser Pro Arg Thr Leu Cys Thr Val Val Phe Gly 85 90 95	288
ctt gac tgc ata ttg gaa tca cct gga gag cct aaa aaa tta ctg atg Leu Asp Cys Ile Leu Glu Ser Pro Gly Glu Pro Lys Lys Leu Leu Met 100 105 110	336
cct gca tcc cac cct cta gag att ttg aag tca ctg agc gag gac gca Pro Ala Ser His Pro Leu Glu Ile Leu Lys Ser Leu Ser Glu Asp Ala 115 120 125	384
gcc ttt gca tta gga ttt tta aag ctg ccc agg cct gct gaa tta gcc Ala Phe Ala Leu Gly Phe Leu Lys Leu Pro Arg Pro Ala Glu Leu Ala 130 135 140	432
acc aag tac gca aac ttt tca gag gga gct tgc aag cct ggc tat gct Thr Lys Tyr Ala Asn Phe Ser Glu Gly Ala Cys Lys Pro Gly Tyr Ala 145 150 155 160	480
tca gcc ttg atg acg gcc atc ttc ccc cgg ttc tcc aag cca gca ccc Ser Ala Leu Met Thr Ala Ile Phe Pro Arg Phe Ser Lys Pro Ala Pro 165 170 175	528
atg ttc ctg gat gac tcc ttt cgc aag tgg gct aga atc cgg gag ttc Met Phe Leu Asp Asp Ser Phe Arg Lys Trp Ala Arg Ile Arg Glu Phe 180 185 190	576
gtg ccg cct ttt ggg atc aaa ggt caa gac aat ctg atc aaa gcc atc Val Pro Pro Phe Gly Ile Lys Gly Gln Asp Asn Leu Ile Lys Ala Ile 195 200 205	624
ttg tca gtc acc aaa gag tac cgc ctg acc cct gcc ttg gac agc ctc Leu Ser Val Thr Lys Glu Tyr Arg Leu Thr Pro Ala Leu Asp Ser Leu 210 215 220	672
cgc tgc cgc cgc tgc atc atc gtg ggc aat gga ggc gtt ctt gcc aac Arg Cys Arg Arg Cys Ile Ile Val Gly Asn Gly Gly Val Leu Ala Asn 225 230 235 240	720
aag tct ctg ggg tca cga att gac gac tat gac att gtg gtg aga ctg Lys Ser Leu Gly Ser Arg Ile Asp Asp Tyr Asp Ile Val Val Arg Leu 245 250 255	768

aat tca gca cca gtg aaa ggc ttt gag aag gac gtg ggc agc aaa acg	816
Asn Ser Ala Pro Val Lys Gly Phe Glu Lys Asp Val Gly Ser Lys Thr	
260 265 270	
aca ctg cgc atc acc tac ccc gag ggc gcc atg cag cgg cct gag cag	864
Thr Leu Arg Ile Thr Tyr Pro Glu Gly Ala Met Gln Arg Pro Glu Gln	
275 280 285	
tac gag cgc gat tct ctc ttt gtc ctc gcc ggc ttc aag tgg cag gac	912
Tyr Glu Arg Asp Ser Leu Phe Val Leu Ala Gly Phe Lys Trp Gln Asp	
290 295 300	
ttt aag tgg ttg aaa tac atc gtc tac aag gag aga gtg agt gca tcg	960
Phe Lys Trp Leu Lys Tyr Ile Val Tyr Lys Glu Arg Val Ser Ala Ser	
305 310 315 320	
gat ggc ttc tgg aaa tct gtg gcc act cga gtg ccc aag gag ccc cct	1008
Asp Gly Phe Trp Lys Ser Val Ala Thr Arg Val Pro Lys Glu Pro Pro	
325 330 335	
gag att cga atc ctc aac cca tat ttc atc cag gag gcc gcc ttc acc	1056
Glu Ile Arg Ile Leu Asn Pro Tyr Phe Ile Gln Glu Ala Ala Phe Thr	
340 345 350	
ctc att ggc ctg ccc ttc aac aat ggc ctc atg ggc cgg ggg aac atc	1104
Leu Ile Gly Leu Pro Phe Asn Asn Gly Leu Met Gly Arg Gly Asn Ile	
355 360 365	
cct acc ctt ggc agt gtg gca gtg acc atg gca cta cac ggc tgt gac	1152
Pro Thr Leu Gly Ser Val Ala Val Thr Met Ala Leu His Gly Cys Asp	
370 375 380	
gag gtg gca gtc gca gga ttt ggc tat gac atg agc aca ccc aac gca	1200
Glu Val Ala Val Ala Gly Phe Gly Tyr Asp Met Ser Thr Pro Asn Ala	
385 390 395 400	
ccc ctg cac tac tat gag acc gtt cgc atg gca gcc atc aaa gag tcc	1248
Pro Leu His Tyr Tyr Glu Thr Val Arg Met Ala Ala Ile Lys Glu Ser	
405 410 415	
tgg acg cac aat atc cag cga gag aaa gag ttt ctg cgg aag ctg gtg	1296
Trp Thr His Asn Ile Gln Arg Glu Lys Glu Phe Leu Arg Lys Leu Val	
420 425 430	
aaa gct cgc gtc atc act gat cta agc agt ggc atc tga	1335
Lys Ala Arg Val Ile Thr Asp Leu Ser Ser Gly Ile	
435 440	

<210> 22

<211> 444

<212> PRT

<213> *Homo sapiens*

<400> 22

Met Gly Leu Leu Val Phe Val Arg Asn Leu Leu Leu Ala Leu Cys Leu
1 5 10 15

Phe Leu Val Leu Gly Phe Leu Tyr Tyr Ser Ala Trp Lys Leu His Leu

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

Thr Leu Arg Ile Thr Tyr Pro Glu Gly Ala Met Gln Arg Pro Glu Gln
 275 280 285

Tyr Glu Arg Asp Ser Leu Phe Val Leu Ala Gly Phe Lys Trp Gln Asp
 290 295 300

Phe Lys Trp Leu Lys Tyr Ile Val Tyr Lys Glu Arg Val Ser Ala Ser
 305 310 315 320

Asp Gly Phe Trp Lys Ser Val Ala Thr Arg Val Pro Lys Glu Pro Pro
 325 330 335

Glu Ile Arg Ile Leu Asn Pro Tyr Phe Ile Gln Glu Ala Ala Phe Thr
 340 345 350

Leu Ile Gly Leu Pro Phe Asn Asn Gly Leu Met Gly Arg Gly Asn Ile
 355 360 365

Pro Thr Leu Gly Ser Val Ala Val Thr Met Ala Leu His Gly Cys Asp
 370 375 380

Glu Val Ala Val Ala Gly Phe Gly Tyr Asp Met Ser Thr Pro Asn Ala
 385 390 395 400

Pro Leu His Tyr Tyr Glu Thr Val Arg Met Ala Ala Ile Lys Glu Ser
 405 410 415

Trp Thr His Asn Ile Gln Arg Glu Lys Glu Phe Leu Arg Lys Leu Val
 420 425 430

Lys Ala Arg Val Ile Thr Asp Leu Ser Ser Gly Ile
 435 440

<210> 23

<211> 990

<212> ADN

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1) .. (990)

<400> 23

```
atg gtc agc aag tcc cgc tgg aag ctc ctg gcc atg ttg gct ctg gtc      48
Met Val Ser Lys Ser Arg Trp Lys Leu Leu Ala Met Leu Ala Leu Val
1          5          10          15
```

ctg gtc gtc atg gtg tgg tat tcc atc tcc cgg gaa gac agt ttt tat	96
Leu Val Val Met Val Trp Tyr Ser Ile Ser Arg Glu Asp Ser Phe Tyr	
20 25 30	
ttt ccc atc cca gag aag aag gag ccg tgc ctc cag ggt gag gca gag	144
Phe Pro Ile Pro Glu Lys Lys Glu Pro Cys Leu Gln Gly Glu Ala Glu	
35 40 45	
agc aag gcc tct aag ctc ttt ggc aac tac tcc cgg gat cag ccc atc	192
Ser Lys Ala Ser Lys Leu Phe Gly Asn Tyr Ser Arg Asp Gln Pro Ile	
50 55 60	
ttc ctg cgg att gag gat tat ttc tgg gtc aag acg cca tct gct tac	240
Phe Leu Arg Leu Glu Asp Tyr Phe Trp Val Lys Thr Pro Ser Ala Tyr	
65 70 75 80	
gag ctg ccc tat ggg acc aag ggg agt gag gat ctg ctc ctc cgg gtg	288
Glu Leu Pro Tyr Thr Lys Gly Ser Glu Asp Leu Leu Leu Arg Val	
85 90 95	
cta gcc atc acc agc tcc tcc atc ccc aag aac atc cag agc ctc agg	336
Leu Ala Ile Thr Ser Ser Ser Ile Pro Lys Asn Ile Gln Ser Leu Arg	
100 105 110	
tgc cgc cgc tgt gtg gtc gtg ggg aac ggg cac cgg ctg cgg aac agc	384
Cys Arg Arg Cys Val Val Val Gly Asn Gly His Arg Leu Arg Asn Ser	
115 120 125	
tca ctg gga gat gcc atc aac aag tac gat gtg gtc atc aga ttg aac	432
Ser Leu Gly Asp Ala Ile Asn Lys Tyr Asp Val Val Ile Arg Leu Asn	
130 135 140	
aat gcc cca gtg gct ggc tat gag ggt gac gtg ggc tcc aag acc acc	480
Asn Ala Pro Val Ala Gly Tyr Glu Gly Asp Val Gly Ser Lys Thr Thr	
145 150 155 160	
atg cgt ctc ttc tac cct gaa tct gcc cac ttc gac ccc aaa gta gaa	528
Met Arg Leu Phe Tyr Pro Glu Ser Ala His Phe Asp Pro Lys Val Glu	
165 170 175	
aac aac cca gac aca ctc ctc gtc ctg gta gct ttc aag gca atg gac	576
Asn Asn Pro Asp Thr Leu Leu Val Leu Val Ala Phe Lys Ala Met Asp	
180 185 190	
ttc cac tgg att gag acc atc ctg agt gat aag aag cgg gtg cga aag	624
Phe His Trp Ile Glu Thr Ile Leu Ser Asp Lys Lys Arg Val Arg Lys	
195 200 205	
ggt ttc tgg aaa cag cct ccc ctc atc tgg gat gtc aat cct aaa cag	672
Gly Phe Trp Lys Gln Pro Pro Leu Ile Trp Asp Val Asn Pro Lys Gln	
210 215 220	
att cgg att ctc aac ccc ttc ttc atg gag att gca gct gac aaa ctg	720
Ile Arg Ile Leu Asn Pro Phe Phe Met Glu Ile Ala Ala Asp Lys Leu	
225 230 235 240	
ctg agc ctg cca atg caa cag cca cgg aag att aag cag aag ccc acc	768
Leu Ser Leu Pro Met Gln Gln Pro Arg Lys Ile Lys Gln Lys Pro Thr	
245 250 255	
acg ggc ctg ttg gcc atc acg ctg gcc ctc cac ctc tgt gac ttg gtg	816

Thr Gly Leu Leu Ala Ile Thr Leu Ala Leu His Leu Cys Asp Leu Val	
260 265 270	
cac att gcc ggc ttt ggc tac cca gac gcc tac aac aag aag cag acc	864
His Ile Ala Gly Phe Gly Tyr Pro Asp Ala Tyr Asn Lys Lys Gln Thr	
275 280 285	
att cac tac tat gag cag atc acg ctc aag tcc atg gcg ggg tca ggc	912
Ile His Tyr Tyr Glu Gln Ile Thr Leu Lys Ser Met Ala Gly Ser Gly	
290 295 300	
cat aat gtc tcc caa gag gcc ctg gcc att aag cgg atg ctg gag atg	960
His Asn Val Ser Gln Glu Ala Leu Ala Ile Lys Arg Met Leu Glu Met	
305 310 315 320	
gga gct atc aag aac ctc acg tcc ttc tga	990
Gly Ala Ile Lys Asn Leu Thr Ser Phe	
325	

<210> 24

<211> 329

<212> PRT

<213> *Homo sapiens*

<400> 24

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

Asn Ala Pro Val Ala Gly Tyr Glu Gly Asp Val Gly Ser Lys Thr Thr
145 150 155 160

Met Arg Leu Phe Tyr Pro Glu Ser Ala His Phe Asp Pro Lys Val Glu
165 170 175

Asn Asn Pro Asp Thr Leu Leu Val Leu Val Ala Phe Lys Ala Met Asp
180 185 190

Phe His Trp Ile Glu Thr Ile Leu Ser Asp Lys Lys Arg Val Arg Lys
195 200 205

Gly Phe Trp Lys Gln Pro Pro Leu Ile Trp Asp Val Asn Pro Lys Gln
210 215 220

Ile Arg Ile Leu Asn Pro Phe Phe Met Glu Ile Ala Ala Asp Lys Leu
225 230 235 240

Leu Ser Leu Pro Met Gln Gln Pro Arg Lys Ile Lys Gln Lys Pro Thr
245 250 255

Thr Gly Leu Leu Ala Ile Thr Leu Ala Leu His Leu Cys Asp Leu Val
260 265 270

His Ile Ala Gly Phe Gly Tyr Pro Asp Ala Tyr Asn Lys Lys Gln Thr
275 280 285

Ile His Tyr Tyr Glu Gln Ile Thr Leu Lys Ser Met Ala Gly Ser Gly
290 295 300

His Asn Val Ser Gln Glu Ala Leu Ala Ile Lys Arg Met Leu Glu Met
305 310 315 320

Gly Ala Ile Lys Asn Leu Thr Ser Phe
325

<210> 25
<211> 1089
<212> ADN
<213> *Homo sapiens*

<220>
<221> CDS
<222> (1)..(1089)

<400> 25

atg aga agg ccc agc ttg tta tta aaa gac atc ctc aaa tgt aca ttg Met Arg Arg Pro Ser Leu Leu Leu Lys Asp Ile Leu Lys Cys Thr Leu 1 5 10 15	48
ctt gtg ttt gga gtg tgg atc ctt tat atc ctc aag tta aat tat act Leu Val Phe Gly Val Trp Ile Leu Tyr Ile Leu Lys Leu Asn Tyr Thr 20 25 30	96
act gaa gaa tgt gac atg aaa aaa atg cat tat gtg gac cct gac cat Thr Glu Glu Cys Asp Met Lys Lys Met His Tyr Val Asp Pro Asp His 35 40 45	144
gta aag aga gct cag aaa tat gct cag caa gtc ttg cag aag gaa tgt Val Lys Arg Ala Gln Lys Tyr Ala Gln Gln Val Leu Gln Lys Glu Cys 50 55 60	192
cgt ccc aag ttt gcc aag aca tca atg gcg ctg tta ttt gag cac agg Arg Pro Lys Phe Ala Lys Thr Ser Met Ala Leu Leu Phe Glu His Arg 65 70 75 80	240
tat agc gtg gac tta ctc cct ttt gtg cag aag gcc ccc aaa gac agt Tyr Ser Val Asp Leu Leu Pro Phe Val Gln Lys Ala Pro Lys Asp Ser 85 90 95	288
gaa gct gag tcc aag tac gat cct cct ttt ggg ttc cgg aag ttc tcc Glu Ala Glu Ser Lys Tyr Asp Pro Pro Phe Gly Phe Arg Lys Phe Ser 100 105 110	336
agt aaa gtc cag acc ctc ttg gaa ctc ttg cca gag cac gac ctc cct Ser Lys Val Gln Thr Leu Leu Glu Leu Leu Pro Glu His Asp Leu Pro 115 120 125	384
gaa cac ttg aaa gcc aag acc tgt cgg cgc tgt gtg gtt att gga agc Glu His Leu Lys Ala Lys Thr Cys Arg Arg Cys Val Val Ile Gly Ser 130 135 140	432
gga gga ata ctg cac gga tta gaa ctg ggc cac acc ctg aac cag ttc Gly Gly Ile Leu His Gly Leu Glu Leu Gly His Thr Leu Asn Gln Phe 145 150 155 160	480
gat gtt gtg ata agg tta aac agt gca cca gtt gag gga tat tca gaa Asp Val Val Ile Arg Leu Asn Ser Ala Pro Val Glu Gly Tyr Ser Glu 165 170 175	528
cat gtt gga aat aaa act act ata agg atg act tat cca gag ggc gca His Val Gly Asn Lys Thr Thr Ile Arg Met Thr Tyr Pro Glu Gly Ala 180 185 190	576
cca ctg tct gac ctt gaa tat tat tcc aat gac tta ttt gtt gct gtt Pro Leu Ser Asp Leu Glu Tyr Tyr Ser Asn Asp Leu Phe Val Ala Val 195 200 205	624
tta ttt aag agt gtt gat ttc aac tgg ctt caa gca atg gta aaa aag Leu Phe Lys Ser Val Asp Phe Asn Trp Leu Gln Ala Met Val Lys Lys 210 215 220	672
gaa acc ctg cca ttc tgg gta cga ctc ttc ttt tgg aag cag gtg gca Glu Thr Leu Pro Phe Trp Val Arg Leu Phe Phe Trp Lys Gln Val Ala 225 230 235 240	720
gaa aaa atc cca ctg cag cca aaa cat ttc agg att ttg aat cca gtt	768

Glu Lys Ile Pro Leu Gln Pro Lys His Phe Arg Ile Leu Asn Pro Val	
245 250 255	
atc atc aaa gag act gcc ttt gac atc ctt cag tac tca gag cct cag	816
ile ile Lys Glu Thr Ala Phe Asp Ile Leu Gln Tyr Ser Glu Pro Gln	
260 265 270	
tca agg ttc tgg ggc cga gat aag aac gtc ccc aca atc ggt gtc att	864
Ser Arg Phe Trp Gly Arg Asp Lys Asn Val Pro Thr Ile Gly Val Ile	
275 280 285	
gcc gtt gtc tta gcc aca cat ctg tgc gat gaa gtc agt ttg gcg ggt	912
Ala Val Val Leu Ala Thr His Leu Cys Asp Glu Val Ser Leu Ala Gly	
290 295 300	
ttt gga tat gac ctc aat caa ccc aga aca cct ttg cac tac ttc gac	960
Phe Gly Tyr Asp Leu Asn Gln Pro Arg Thr Pro Leu His Tyr Phe Asp	
305 310 315 320	
agt caa tgc atg gct gct atg aac ttt cag acc atg cat aat gtg aca	1008
Ser Gln Cys Met Ala Ala Met Asn Phe Gln Thr Met His Asn Val Thr	
325 330 335	
acg gaa acc aag ttc ctc tta aag ctg gtc aaa gag gga gtg gtg aaa	1056
Thr Glu Thr Lys Phe Leu Leu Lys Leu Val Lys Glu Gly Val Val Lys	
340 345 350	
gat ctc agt gga ggc att gat cgt gaa ttt tga	1089
Asp Leu Ser Gly Gly Ile Asp Arg Glu Phe	
355 360	

<210> 26

<211> 362

<212> PRT

<213> *Homo sapiens*

<400> 26

Met Arg Arg Pro Ser Leu Leu Leu Lys Asp Ile Leu Lys Cys Thr Leu
1 5 10 15

Leu Val Phe Gly Val Trp Ile Leu Tyr Ile Leu Lys Leu Asn Tyr Thr
20 25 30

Thr Glu Glu Cys Asp Met Lys Lys Met His Tyr Val Asp Pro Asp His
35 40 45

Val Lys Arg Ala Gln Lys Tyr Ala Gln Gln Val Leu Gln Lys Glu Cys
50 55 60

Arg Pro Lys Phe Ala Lys Thr Ser Met Ala Leu Leu Phe Glu His Arg
65 70 75 80

Tyr Ser Val Asp Leu Leu Pro Phe Val Gln Lys Ala Pro Lys Asp Ser
85 90 95

Glu Ala Glu Ser Lys Tyr Asp Pro Pro Phe Gly Phe Arg Lys Phe Ser
 100 105 110

Ser Lys Val Gln Thr Leu Leu Glu Leu Leu Pro Glu His Asp Leu Pro
 115 120 125

Glu His Leu Lys Ala Lys Thr Cys Arg Arg Cys Val Val Ile Gly Ser
 130 135 140

Gly Gly Ile Leu His Gly Leu Glu Leu Gly His Thr Leu Asn Gln Phe
 145 150 155 160

Asp Val Val Ile Arg Leu Asn Ser Ala Pro Val Glu Gly Tyr Ser Glu
 165 170 175

His Val Gly Asn Lys Thr Thr Ile Arg Met Thr Tyr Pro Glu Gly Ala
 180 185 190

Pro Leu Ser Asp Leu Glu Tyr Tyr Ser Asn Asp Leu Phe Val Ala Val
 195 200 205

Leu Phe Lys Ser Val Asp Phe Asn Trp Leu Gln Ala Met Val Lys Lys
 210 215 220

Glu Thr Leu Pro Phe Trp Val Arg Leu Phe Phe Trp Lys Gln Val Ala
 225 230 235 240

Glu Lys Ile Pro Leu Gln Pro Lys His Phe Arg Ile Leu Asn Pro Val
 245 250 255

Ile Ile Lys Glu Thr Ala Phe Asp Ile Leu Gln Tyr Ser Glu Pro Gln
 260 265 270

Ser Arg Phe Trp Gly Arg Asp Lys Asn Val Pro Thr Ile Gly Val Ile
 275 280 285

Ala Val Val Leu Ala Thr His Leu Cys Asp Glu Val Ser Leu Ala Gly
 290 295 300

Phe Gly Tyr Asp Leu Asn Gln Pro Arg Thr Pro Leu His Tyr Phe Asp
 305 310 315 320

Ser Gln Cys Met Ala Ala Met Asn Phe Gln Thr Met His Asn Val Thr
 325 330 335

Thr Glu Thr Lys Phe Leu Leu Lys Leu Val Lys Glu Gly Val Val Lys
340 345 350

Asp Leu Ser Gly Gly Ile Asp Arg Glu Phe
355 360

<210> 27

<211> 996

<212> ADN

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1) .. (996)

<400> 27

atg aga ggg tat ctt gtg gcc ata ttc ctg agt gct gtc ttc ctc tat	48
Met Arg Gly Tyr Leu Val Ala Ile Phe Leu Ser Ala Val Phe Leu Tyr	
1 5 10 15	
tat gta ctg cat tgc ata tta tgg gga acg aat gtc tat tgg gtg gca	96
Tyr Val Leu His Cys Ile Leu Trp Gly Thr Asn Val Tyr Trp Val Ala	
20 25 30	
cct gtg gaa atg aaa cgg aga aat aag atc cag cct tgt tta tca aag	144
Pro Val Glu Met Lys Arg Arg Asn Lys Ile Gln Pro Cys Leu Ser Lys	
35 40 45	
cca gct ttt gcc tct ctg ctg agg ttt cat cag ttt cac cct ttt ctg	192
Pro Ala Phe Ala Ser Leu Leu Arg Phe His Gln Phe His Pro Phe Leu	
50 55 60	
tgt gcg gct gat ttt aga aag att gct tcc ttg tat ggt agc gat aag	240
Cys Ala Ala Asp Phe Arg Lys Ile Ala Ser Leu Tyr Gly Ser Asp Lys	
65 70 75 80	
ttt gat ttg ccc tat ggg atg aga aca tca gcg gaa tat ttt cga ctt	288
Phe Asp Leu Pro Tyr Gly Met Arg Thr Ser Ala Glu Tyr Phe Arg Leu	
85 90 95	
gct ctt tca aaa ctg cag agt tgt gat ctc ttt gat gag ttt gac aac	336
Ala Leu Ser Lys Leu Gln Ser Cys Asp Leu Phe Asp Glu Phe Asp Asn	
100 105 110	
ata ccc tgt aaa aag tgt gtg gtg gtt ggt aat gga gga gtt ttg aag	384
Ile Pro Cys Lys Lys Cys Val Val Val Gly Asn Gly Gly Val Leu Lys	
115 120 125	
aat aag aca tta gga gaa aaa atc gac tcc tat gat gta ata ata aga	432
Asn Lys Thr Leu Gly Glu Lys Ile Asp Ser Tyr Asp Val Ile Ile Arg	
130 135 140	
atg aat aat ggt cct gtt tta gga cat gaa gaa gaa gtt ggg aga agg	480
Met Asn Asn Gly Pro Val Leu Gly His Glu Glu Glu Val Gly Arg Arg	
145 150 155 160	
aca acc ttc cga ctt ttt tat cca gaa tct gtt ttt tca gat cct att	528

Thr Thr Phe Arg Leu Phe Tyr Pro Glu Ser Val Phe Ser Asp Pro Ile	
165 170 175	
cac aat gac cct aat acg aca gtg att ctc act gct ttt aag cca cat	576
His Asn Asp Pro Asn Thr Thr Val Ile Leu Thr Ala Phe Lys Pro His	
180 185 190	
gat tta agg tgg ctg ttg gaa ttg ttg atg ggt gac aaa ata aac act	624
Asp Leu Arg Trp Leu Leu Glu Leu Leu Met Gly Asp Lys Ile Asn Thr	
195 200 205	
aat ggt ttt tgg aag aaa cca gcc tta aac ctg att tat aaa cct tat	672
Asn Gly Phe Trp Lys Lys Pro Ala Leu Asn Leu Ile Tyr Lys Pro Tyr	
210 215 220	
caa atc cga ata tta gat cct ttc att atc aga aca gca gct tat gaa	720
Gln Ile Arg Ile Leu Asp Pro Phe Ile Ile Arg Thr Ala Ala Tyr Glu	
225 230 235 240	
ctg ctt cat ttt cca aaa gtg ttt ccc aaa aat cag aaa cct aaa cac	768
Leu Leu His Phe Pro Lys Val Phe Pro Lys Asn Gln Lys Pro Lys His	
245 250 255	
cca aca aca gga att att gcc atc aca ttg gcg ttt tac ata tgt cac	816
Pro Thr Thr Gly Ile Ile Ala Ile Thr Leu Ala Phe Tyr Ile Cys His	
260 265 270	
gaa gtt cac cta gct ggt ttt aaa tac aac ttt tct gac ctc aag agt	864
Glu Val His Leu Ala Gly Phe Lys Tyr Asn Phe Ser Asp Leu Lys Ser	
275 280 285	
cct ttg cac tac tat ggg aat gcc acc atg tct ttg atg aat aag aac	912
Pro Leu His Tyr Tyr Gly Asn Ala Thr Met Ser Leu Met Asn Lys Asn	
290 295 300	
gcg tat cac aat gtg act gca gag cag ctc ttt ttg aag gac att ata	960
Ala Tyr His Asn Val Thr Ala Glu Gln Leu Phe Leu Lys Asp Ile Ile	
305 310 315 320	
gaa aaa aac ctc gta atc aac ttg act caa gat tga	996
Glu Lys Asn Leu Val Ile Asn Leu Thr Gln Asp	
325 330	

<210> 28

<211> 331

<212> PRT

<213> *Homo sapiens*

<400> 28

Met Arg Gly Tyr Leu Val Ala Ile Phe Leu Ser Ala Val Phe Leu Tyr
1 5 10 15

Tyr Val Leu His Cys Ile Leu Trp Gly Thr Asn Val Tyr Trp Val Ala
20 25 30

Pro Val Glu Met Lys Arg Arg Asn Lys Ile Gln Pro Cys Leu Ser Lys
35 40 45

Pro Ala Phe Ala Ser Leu Leu Arg Phe His Gln Phe His Pro Phe Leu
50 55 60

Cys Ala Ala Asp Phe Arg Lys Ile Ala Ser Leu Tyr Gly Ser Asp Lys
65 70 75 80

Phe Asp Leu Pro Tyr Gly Met Arg Thr Ser Ala Glu Tyr Phe Arg Leu
85 90 95

Ala Leu Ser Lys Leu Gln Ser Cys Asp Leu Phe Asp Glu Phe Asp Asn
100 105 110

Ile Pro Cys Lys Lys Cys Val Val Val Gly Asn Gly Gly Val Leu Lys
115 120 125

Asn Lys Thr Leu Gly Glu Lys Ile Asp Ser Tyr Asp Val Ile Ile Arg
130 135 140

Met Asn Asn Gly Pro Val Leu Gly His Glu Glu Glu Val Gly Arg Arg
145 150 155 160

Thr Thr Phe Arg Leu Phe Tyr Pro Glu Ser Val Phe Ser Asp Pro Ile
165 170 175

His Asn Asp Pro Asn Thr Thr Val Ile Leu Thr Ala Phe Lys Pro His
180 185 190

Asp Leu Arg Trp Leu Leu Glu Leu Leu Met Gly Asp Lys Ile Asn Thr
195 200 205

Asn Gly Phe Trp Lys Lys Pro Ala Leu Asn Leu Ile Tyr Lys Pro Tyr
210 215 220

Gln Ile Arg Ile Leu Asp Pro Phe Ile Ile Arg Thr Ala Ala Tyr Glu
225 230 235 240

Leu Leu His Phe Pro Lys Val Phe Pro Lys Asn Gln Lys Pro Lys His
245 250 255

Pro Thr Thr Gly Ile Ile Ala Ile Thr Leu Ala Phe Tyr Ile Cys His
260 265 270

Glu Val His Leu Ala Gly Phe Lys Tyr Asn Phe Ser Asp Leu Lys Ser
275 280 285

Pro Leu His Tyr Tyr Gly Asn Ala Thr Met Ser Leu Met Asn Lys Asn
290 295 300

Ala Tyr His Asn Val Thr Ala Glu Gln Leu Phe Leu Lys Asp Ile Ile
305 310 315 320

Glu Lys Asn Leu Val Ile Asn Leu Thr Gln Asp
325 330

<210> 29

<211> 1125

<212> ADN

<213> *Rattus rattus*

<220>

<221> CDS

<222> (1) .. (1125)

<400> 29

atg gga ctc ttg gta ttt gta cgc aac ctg ctg cta gcc ctc tgc ctc	48
Met Gly Leu Leu Val Phe Val Arg Asn Leu Leu Leu Ala Leu Cys Leu	
1 5 10 15	
ttt ctg gtc ctg gga ttt ttg tat tat tct gcc tgg aag cta cac tta	96
Phe Leu Val Leu Gly Phe Leu Tyr Tyr Ser Ala Trp Lys Leu His Leu	
20 25 30	
ctc caa tgg gaa gac tcc aat tca ctg att ctt tcc ctt gac tcc got	144
Leu Gln Trp Glu Asp Ser Asn Ser Leu Ile Leu Ser Leu Asp Ser Ala	
35 40 45	
gga caa acc cta ggc aca gag tat gat agg ctg ggt ttc ctc ctg aag	192
Gly Gln Thr Leu Gly Thr Glu Tyr Asp Arg Leu Gly Phe Leu Leu Lys	
50 55 60	
ctg gac tct aaa ctg cct gca gag ctg gcc acc aag tac gct aac ttt	240
Leu Asp Ser Lys Leu Pro Ala Glu Leu Ala Thr Lys Tyr Ala Asn Phe	
65 70 75 80	
tcc gag gga gcc tgc aaa ccc ggc tac gct tca gcc atg atg act gcc	288
Ser Glu Gly Ala Cys Lys Pro Gly Tyr Ala Ser Ala Met Met Thr Ala	
85 90 95	
atc ttc ccc agg ttc tcc aag cca gca ccc atg ttc ctg gat gac tcc	336
Ile Phe Pro Arg Phe Ser Lys Pro Ala Pro Met Phe Leu Asp Asp Ser	
100 105 110	
ttt cgc aaa tgg gct agg att cgg gag ttt gtg cca ccc ttt ggg atc	384
Phe Arg Lys Trp Ala Arg Ile Arg Glu Phe Val Pro Pro Phe Gly Ile	
115 120 125	
aaa ggt caa gac aat ctg atc aaa gcc atc ttg tca gtc acc aaa gaa	432
Lys Gly Gln Asp Asn Leu Ile Lys Ala Ile Leu Ser Val Thr Lys Glu	
130 135 140	
tac cgc ctg acc cct gcc ttg gac agc ctc cac tgc cgc cgc tgc atc	480

Tyr Arg Leu Thr Pro Ala Leu Asp Ser Leu His Cys Arg Arg Cys Ile	
145 150 155 160	
atc gta ggc aat gga ggg gtc ctc gcc aac aag tct ctg ggg tca cga	528
Ile Val Gly Asn Gly Gly Val Leu Ala Asn Lys Ser Leu Gly Ser Arg	
165 170 175	
att gac gac tat gac att gtg atc aga ttg aac tca gca cct gtg aag	576
Ile Asp Asp Tyr Asp Ile Val Ile Arg Leu Asn Ser Ala Pro Val Lys	
180 185 190	
ggc ttt gag aag gac gtg ggc agc aag acc acc ctg cgc atc acc tac	624
Gly Phe Glu Lys Asp Val Gly Ser Lys Thr Thr Leu Arg Ile Thr Tyr	
195 200 205	
cct gaa ggt gcc atg cag cgg cct gag caa tat gaa cga gac tct ctc	672
Pro Glu Gly Ala Met Gln Arg Pro Glu Gln Tyr Glu Arg Asp Ser Leu	
210 215 220	
ttt gta cta gct ggc ttc aag tgg cag gac ttc aag tgg ctg aag tac	720
Phe Val Leu Ala Gly Phe Lys Trp Gln Asp Phe Lys Trp Leu Lys Tyr	
225 230 235 240	
atc gtc tac aag gag aga gtg agc gca tcc gat ggc ttc tgg aag tcc	768
Ile Val Tyr Lys Glu Arg Val Ser Ala Ser Asp Gly Phe Trp Lys Ser	
245 250 255	
gtg gcc acc cga gtg ccc aag gag ccc cct gag atc cgc atc ctc aac	816
Val Ala Thr Arg Val Pro Lys Glu Pro Pro Glu Ile Arg Ile Leu Asn	
260 265 270	
cag tac ttc atc cag gag gct gcc ttc acg ctc atc gga ctg ccc ttc	864
Pro Tyr Phe Ile Gln Glu Ala Ala Phe Thr Leu Ile Gly Leu Pro Phe	
275 280 285	
aac aat ggc ctc atg ggc aga ggg aac atc cca acc ctt ggc agt gtg	912
Asn Asn Gly Leu Met Gly Arg Gly Asn Ile Pro Thr Leu Gly Ser Val	
290 295 300	
gca gtg acc atg gca ctc gat ggc tgt gat gaa gtg gca gtc gcg ggc	960
Ala Val Thr Met Ala Leu Asp Gly Cys Asp Glu Val Ala Val Ala Gly	
305 310 315 320	
ttt ggc tat gac atg aac aca ccc aac gcc ccc ctg cac tac tat gaa	1008
Phe Gly Tyr Asp Met Asn Thr Pro Asn Ala Pro Leu His Tyr Tyr Glu	
325 330 335	
act gtg cgc atg gca gcc atc aaa gag tcc tgg aca cac aac atc cag	1056
Thr Val Arg Met Ala Ala Ile Lys Glu Ser Trp Thr His Asn Ile Gln	
340 345 350	
cga gag aaa gag ttt ctg cgg aag cta gtg aaa gca cgc gtc atc act	1104
Arg Glu Lys Glu Phe Leu Arg Lys Leu Val Lys Ala Arg Val Ile Thr	
355 360 365	
gac tta agc agt ggt atc tga	1125
Asp Leu Ser Ser Gly Ile	
370	

<210> 30

<211> 374

<212> PRT

<213> *Rattus rattus*

<400> 30

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

290	295	300
Ala Val Thr Met Ala Leu Asp Gly Cys Asp Glu Val Ala Val Ala Gly 305 310 315 320		
Phe Gly Tyr Asp Met Asn Thr Pro Asn Ala Pro Leu His Tyr Tyr Glu 325 330 335		
Thr Val Arg Met Ala Ala Ile Lys Glu Ser Trp Thr His Asn Ile Gln 340 345 350		
Arg Glu Lys Glu Phe Leu Arg Lys Leu Val Lys Ala Arg Val Ile Thr 355 360 365		
Asp Leu Ser Ser Gly Ile 370		

<210> 31
<211> 1071
<212> ADN
<213> *Homo sapiens*

<220>
<221> CDS
<222> (1)..(1071)

<400> 31

atg agc ccc tgc ggg cgg gcc cgg cga caa acg tcc aga ggg gcc atg	48
Met Ser Pro Cys Gly Arg Ala Arg Arg Gln Thr Ser Arg Gly Ala Met	
1 5 10 15	
gct gta ctg gcg tgg aag ttc ccg cgg acc cgg ctg ccc atg gga gcc	96
Ala Val Leu Ala Trp Lys Phe Pro Arg Thr Arg Leu Pro Met Gly Ala	
20 25 30	
agt gcc ctc tgt gtc gtg gtc ctc tgt tgg ctc tac atc ttc ccc gtc	144
Ser Ala Leu Cys Val Val Val Leu Cys Trp Leu Tyr Ile Phe Pro Val	
35 40 45	
tac cgg ctg ccc aac gag aaa gag atc gtg cag ggg gtg ctg caa cag	192
Tyr Arg Leu Pro Asn Glu Lys Glu Ile Val Gln Gly Val Leu Gln Gln	
50 55 60	
ggc acg gcg tgg agg agg aac cag acc gcg gcc aga gcg ttc agg aaa	240
Gly Thr Ala Trp Arg Arg Asn Gln Thr Ala Ala Arg Ala Phe Arg Lys	
65 70 75 80	
caa atg gaa gac tgc tgc gac cct gcc cat ctc ttt gct atg act aaa	288
Gln Met Glu Asp Cys Cys Asp Pro Ala His Leu Phe Ala Met Thr Lys	
85 90 95	
atg aat tcc cct atg ggg aag agc atg tgg tat gac ggg gag ttt tta	336
Met Asn Ser Pro Met Gly Lys Ser Met Trp Tyr Asp Gly Glu Phe Leu	
100 105 110	
tac tca ttc acc att gac aat tca act tac tct ctc ttc cca cag gca	384
Tyr Ser Phe Thr Ile Asp Asn Ser Thr Tyr Ser Leu Phe Pro Gln Ala	
115 120 125	

acc cca ttc cag ctg cca ttg aag aaa tgc gcg gtg gtg gga aat ggt	432
Thr Pro Phe Gln Leu Pro Leu Lys Lys Cys Ala Val Val Gly Asn Gly	
130 135 140	
ggg att ctg aag aag agt ggc tgt ggc cgt caa ata gat gaa gca aat	480
Gly Ile Leu Lys Lys Ser Gly Cys Gly Arg Gln Ile Asp Glu Ala Asn	
145 150 155 160	
ttt gtc atg cga tgc aat ctc cct cct ttg tca agt gaa tac act aag	528
Phe Val Met Arg Cys Asn Leu Pro Pro Leu Ser Ser Glu Tyr Thr Lys	
165 170 175	
gat gtt gga tcc aaa agt cag tta gtg aca gct aat ccc agc ata att	576
Asp Val Gly Ser Lys Ser Gln Leu Val Thr Ala Asn Pro Ser Ile Ile	
180 185 190	
cgg caa agg ttt cag aac ctt ctg tgg tcc aga aag aca ttt gtg gac	624
Arg Gln Arg Phe Gln Asn Leu Leu Trp Ser Arg Lys Thr Phe Val Asp	
195 200 205	
aac atg aaa att tat aac cac agt tac atc tac atg cct gcc ttt tct	672
Asn Met Lys Ile Tyr Asn His Ser Tyr Ile Tyr Met Pro Ala Phe Ser	
210 215 220	
atg aag aca gga aca gag cca tct ttg agg gtt tat tat aca ctg tca	720
Met Lys Thr Gly Thr Glu Pro Ser Leu Arg Val Tyr Tyr Thr Leu Ser	
225 230 235 240	
gat gtt ggt gcc aat caa' aca gtg ctg ttt gcc aac ccc aac ttt ctg	768
Asp Val Gly Ala Asn Gln Thr Val Leu Phe Ala Asn Pro Asn Phe Leu	
245 250 255	
cgt agc att gga aag ttc tgg aaa agt aga gga atc cat gcc aag cgc	816
Arg Ser Ile Gly Lys Phe Trp Lys Ser Arg Gly Ile His Ala Lys Arg	
260 265 270	
ctg tcc aca gga ctt ttt ctg gtg agc gca gct ctg ggt ctc tgt gaa	864
Leu Ser Thr Gly Leu Phe Leu Val Ser Ala Ala Leu Gly Leu Cys Glu	
275 280 285	
gag gtg gcc atc tat ggc ttc tgg ccc ttc tct gtg aat atg cat gag	912
Glu Val Ala Ile Tyr Gly Phe Trp Pro Phe Ser Val Asn Met His Glu	
290 295 300	
cag ccc atc agc cac cac tac tat gac aac gtc tta ccc ttt tct ggc	960
Gln Pro Ile Ser His His Tyr Tyr Asp Asn Val Leu Pro Phe Ser Gly	
305 310 315 320	
ttc cat gcc atg ccc gag gaa ttt ctc caa ctc tgg tat ctt cat aaa	1008
Phe His Ala Met Pro Glu Glu Phe Leu Gln Leu Trp Tyr Leu His Lys	
325 330 335	
atc ggt gca ctg aga atg cag ctg gac cca tgt gaa gat acc tca ctc	1056
Ile Gly Ala Leu Arg Met Gln Leu Asp Pro Cys Glu Asp Thr Ser Leu	
340 345 350	
cag ccc act tcc tag	1071
Gln Pro Thr Ser	
355	

<210> 32

<211> 356

<212> PRT

<213> *Homo sapiens*

<400> 32

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

Met Lys Thr Gly Thr Glu Pro Ser Leu Arg Val Tyr Tyr Thr Leu Ser
 225 230 235 240

Asp Val Gly Ala Asn Gln Thr Val Leu Phe Ala Asn Pro Asn Phe Leu
 245 250 255

Arg Ser Ile Gly Lys Phe Trp Lys Ser Arg Gly Ile His Ala Lys Arg
 260 265 270

Leu Ser Thr Gly Leu Phe Leu Val Ser Ala Ala Leu Gly Leu Cys Glu
 275 280 285

Glu Val Ala Ile Tyr Gly Phe Trp Pro Phe Ser Val Asn Met His Glu
 290 295 300

Gln Pro Ile Ser His His Tyr Tyr Asp Asn Val Leu Pro Phe Ser Gly
 305 310 315 320

Phe His Ala Met Pro Glu Glu Phe Leu Gln Leu Trp Tyr Leu His Lys
 325 330 335

Ile Gly Ala Leu Arg Met Gln Leu Asp Pro Cys Glu Asp Thr Ser Leu
 340 345 350

Gln Pro Thr Ser
 355

<210> 33

<211> 1128

<212> ADN

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1)..(1128)

<400> 33

atg	cag	ctg	cag	ttc	cgg	agc	tgg	atg	ctg	gcc	gcg	ctc	acg	ctg	ctc	48
Met	Gln	Leu	Gln	Phe	Arg	Ser	Trp	Met	Leu	Ala	Ala	Leu	Thr	Leu	Leu	
1				5				10					15			
gtg	gtc	ttc	ctc	atc	ttc	gca	gac	atc	tca	gag	atc	gaa	gaa	gaa	atc	96
Val	Val	Phe	Leu	Ile	Phe	Ala	Asp	Ile	Ser	Glu	Ile	Glu	Glu	Glu	Ile	
			20					25				30				
ggg	aat	tcg	gga	ggc	aga	ggc	aca	atc	aga	tca	gct	gtg	aac	agc	tta	144
Gly	Asn	Ser	Gly	Gly	Arg	Gly	Thr	Ile	Arg	Ser	Ala	Val	Asn	Ser	Leu	
	35						40					45				

cat agc aaa tct aat aga gct gaa gtt gta ata aac ggc tcc tca tca	192
His Ser Lys Ser Asn Arg Ala Glu Val Val Ile Asn Gly Ser Ser Ser	
50 55 60	
cca gct gtt gtt gac aga agt aat gaa agc atc aag cac aac atc cag	240
Pro Ala Val Val Asp Arg Ser Asn Glu Ser Ile Lys His Asn Ile Gln	
65 70 75 80	
cca gcc tcg tcc aaa tgg aga cat aac cag acg ctc tct ctg agg atc	288
Pro Ala Ser Ser Lys Trp Arg His Asn Gln Thr Leu Ser Leu Arg Ile	
85 90 95	
agg aag cag att tta aag ttc ttg gat gct gaa aag gac att tct gtc	336
Arg Lys Gln Ile Leu Lys Phe Leu Asp Ala Glu Lys Asp Ile Ser Val	
100 105 110	
cta aag gga acc ctg aag cct gga gat att att cat tac atc ttc gat	384
Leu Lys Gly Thr Leu Lys Pro Gly Asp Ile Ile His Tyr Ile Phe Asp	
115 120 125	
cga gac agc acc atg aat gtg tcc cag aac ctc tac gag ctc ctc ccc	432
Arg Asp Ser Thr Met Asn Val Ser Gln Asn Leu Tyr Glu Leu Leu Pro	
130 135 140	
agg act tcg cca ctg aag aat aag cac ttt ggg act tgt gcc atc gtg	480
Arg Thr Ser Pro Leu Lys Asn Lys His Phe Gly Thr Cys Ala Ile Val	
145 150 155 160	
ggc aac tcg ggg gtc ttg ctg aac agc ggc tgt ggg cag gag att gac	528
Gly Asn Ser Gly Val Leu Leu Asn Ser Gly Cys Gly Gln Glu Ile Asp	
165 170 175	
gcc cac agc ttc gtc atc agg tgc aac ctg gcc cca gta cag gag tat	576
Ala His Ser Phe Val Ile Arg Cys Asn Leu Ala Pro Val Gln Glu Tyr	
180 185 190	
gcc cgg gat gtg ggg ctc aag aca gac ctg gta acc atg aac ccc tcg	624
Ala Arg Asp Val Gly Leu Lys Thr Asp Leu Val Thr Met Asn Pro Ser	
195 200 205	
gtc atc cag cgg gcc ttt gag gac ttg gtc aat gcc acg tgg cgg gag	672
Val Ile Gln Arg Ala Phe Glu Asp Leu Val Asn Ala Thr Trp Arg Glu	
210 215 220	
aag ctg ctg caa cgg ctg cac agc ctc aat ggc agc atc ctg tgg atc	720
Lys Leu Leu Gln Arg Leu His Ser Leu Asn Gly Ser Ile Leu Trp Ile	
225 230 235 240	
cct gcc ttc atg gcc cgg ggc ggc aag gag cgt gtt gag tgg gtc aac	768
Pro Ala Phe Met Ala Arg Gly Gly Lys Glu Arg Val Glu Trp Val Asn	
245 250 255	
gag ctt atc ctg aag cac cac gtc aac gtg cgc act gca tac ccc tcg	816
Glu Leu Ile Leu Lys His His Val Asn Val Arg Thr Ala Tyr Pro Ser	
260 265 270	
ctg cgc ctg ctg cac gcc gtt cgc gga tac tgg ctg acc aac aaa gtc	864
Leu Arg Leu Leu His Ala Val Arg Gly Tyr Trp Leu Thr Asn Lys Val	
275 280 285	
cac atc aaa aga ccc acc acc ggc ctc ttg atg tat acc ctg gcc aca	912

His	Ile	Lys	Arg	Pro	Thr	Thr	Gly	Leu	Leu	Met	Tyr	Thr	Leu	Ala	Thr		
290						295					300						
cgt	ttc	tgc	aaa	caa	atc	tac	ctc	tac	ggc	ttc	tgg	ccc	ttt	ccg	ctg	960	
Arg	Phe	Cys	Lys	Gln	Ile	Tyr	Leu	Tyr	Gly	Phe	Trp	Pro	Phe	Pro	Leu		
305					310					315					320		
gat	cag	aac	cag	aac	cca	gtc	aag	tac	cac	tat	tat	gac	agc	ctc	aag	1008	
Asp	Gln	Asn	Gln	Asn	Pro	Val	Lys	Tyr	His	Tyr	Tyr	Asp	Ser	Leu	Lys		
				325					330					335			
tat	ggc	tac	acc	tcc	cag	gcc	agc	ccg	cat	acc	atg	ccc	ttg	gag	ttt	1056	
Tyr	Gly	Tyr	Thr	Ser	Gln	Ala	Ser	Pro	His	Thr	Met	Pro	Leu	Glu	Phe		
			340					345					350				
aag	gcc	ctc	aag	agc	cta	cat	gag	cag	ggg	gct	ttg	aaa	ctg	act	gtc	1104	
Lys	Ala	Leu	Lys	Ser	Leu	His	Glu	Gln	Gly	Ala	Leu	Lys	Leu	Thr	Val		
		355					360					365					
ggc	cag	tgc	gat	ggg	gcc	acg	tag									1128	
Gly	Gln	Cys	Asp	Gly	Ala	Thr											
		370				375											

<210> 34

<211> 375

<212> PRT

<213> *Homo sapiens*

<400> 34

Met Gln Leu Gln Phe Arg Ser Trp Met Leu Ala Ala Leu Thr Leu Leu
1 5 10 15

Val Val Phe Leu Ile Phe Ala Asp Ile Ser Glu Ile Glu Glu Glu Ile
20 25 30

Gly Asn Ser Gly Gly Arg Gly Thr Ile Arg Ser Ala Val Asn Ser Leu
35 40 45

His Ser Lys Ser Asn Arg Ala Glu Val Val Ile Asn Gly Ser Ser Ser
50 55 60

Pro Ala Val Val Asp Arg Ser Asn Glu Ser Ile Lys His Asn Ile Gln
65 70 75 80

Pro Ala Ser Ser Lys Trp Arg His Asn Gln Thr Leu Ser Leu Arg Ile
85 90 95

Arg Lys Gln Ile Leu Lys Phe Leu Asp Ala Glu Lys Asp Ile Ser Val
100 105 110

Leu Lys Gly Thr Leu Lys Pro Gly Asp Ile Ile His Tyr Ile Phe Asp
115 120 125

Arg	Asp	Ser	Thr	Met	Asn	Val	Ser	Gln	Asn	Leu	Tyr	Glu	Leu	Leu	Pro
130						135					140				
Arg	Thr	Ser	Pro	Leu	Lys	Asn	Lys	His	Phe	Gly	Thr	Cys	Ala	Ile	Val
145					150					155					160
Gly	Asn	Ser	Gly	Val	Leu	Leu	Asn	Ser	Gly	Cys	Gly	Gln	Glu	Ile	Asp
				165					170					175	
Ala	His	Ser	Phe	Val	Ile	Arg	Cys	Asn	Leu	Ala	Pro	Val	Gln	Glu	Tyr
			180					185					190		
Ala	Arg	Asp	Val	Gly	Leu	Lys	Thr	Asp	Leu	Val	Thr	Met	Asn	Pro	Ser
		195					200					205			
Val	Ile	Gln	Arg	Ala	Phe	Glu	Asp	Leu	Val	Asn	Ala	Thr	Trp	Arg	Glu
	210					215					220				
Lys	Leu	Leu	Gln	Arg	Leu	His	Ser	Leu	Asn	Gly	Ser	Ile	Leu	Trp	Ile
225					230					235					240
Pro	Ala	Phe	Met	Ala	Arg	Gly	Gly	Lys	Glu	Arg	Val	Glu	Trp	Val	Asn
				245					250					255	
Glu	Leu	Ile	Leu	Lys	His	His	Val	Asn	Val	Arg	Thr	Ala	Tyr	Pro	Ser
			260					265					270		
Leu	Arg	Leu	Leu	His	Ala	Val	Arg	Gly	Tyr	Trp	Leu	Thr	Asn	Lys	Val
		275					280					285			
His	Ile	Lys	Arg	Pro	Thr	Thr	Gly	Leu	Leu	Met	Tyr	Thr	Leu	Ala	Thr
	290					295					300				
Arg	Phe	Cys	Lys	Gln	Ile	Tyr	Leu	Tyr	Gly	Phe	Trp	Pro	Phe	Pro	Leu
305					310					315					320
Asp	Gln	Asn	Gln	Asn	Pro	Val	Lys	Tyr	His	Tyr	Tyr	Asp	Ser	Leu	Lys
				325					330					335	
Tyr	Gly	Tyr	Thr	Ser	Gln	Ala	Ser	Pro	His	Thr	Met	Pro	Leu	Glu	Phe
			340					345					350		
Lys	Ala	Leu	Lys	Ser	Leu	His	Glu	Gln	Gly	Ala	Leu	Lys	Leu	Thr	Val
	355						360					365			

Gly	Gln	Cys	Asp	Gly	Ala	Thr
370						375

<210> 35

<211> 1143

<212> ADN

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1) .. (1143)

<400> 35

atg aga aac tgc aaa atg gcc cgg gtc gcc agt gtg ctg ggg ctg gtc Met Arg Asn Cys Lys Met Ala Arg Val Ala Ser Val Leu Gly Leu Val 1 5 10 15	48
atg ctc agc gtc gcc ctg ctg att tta tcc ctc atc agc tac gtg tcc Met Leu Ser Val Ala Leu Leu Ile Leu Ser Leu Ile Ser Tyr Val Ser 20 25 30	96
ctg aaa aag gag aac atc ttc acc act ccc aag tac gcc agc ccg ggg Leu Lys Lys Glu Asn Ile Phe Thr Thr Pro Lys Tyr Ala Ser Pro Gly 35 40 45	144
gcg ccc cga atg tac atg ttc cac gcg gga ttc cgg tca caa ttt gcg Ala Pro Arg Met Tyr Met Phe His Ala Gly Phe Arg Ser Gln Phe Ala 50 55 60	192
ctg aag ttt cta gac ccg tca ttc gtg ccc att acg aat tct ctc acc Leu Lys Phe Leu Asp Pro Ser Phe Val Pro Ile Thr Asn Ser Leu Thr 65 70 75 80	240
cag gaa ctc caa gag aaa cct tct aag tgg aaa ttt aat cgg aca gcg Gln Glu Leu Gln Glu Lys Pro Ser Lys Trp Lys Phe Asn Arg Thr Ala 85 90 95	288
ttt tta cat caa agg caa gaa att ctt cag cat gtc gat gta ata aaa Phe Leu His Gln Arg Gln Glu Ile Leu Gln His Val Asp Val Ile Lys 100 105 110	336
aat ttt tct ttg acc aag aat agt gtt cgg att gga caa ctg atg cac Asn Phe Ser Leu Thr Lys Asn Ser Val Arg Ile Gly Gln Leu Met His 115 120 125	384
tat gat tat tcc agc cat aaa tat gtt ttc tct att agc aat aac ttc Tyr Asp Tyr Ser Ser His Lys Tyr Val Phe Ser Ile Ser Asn Asn Phe 130 135 140	432
cgg tca ctt ctt cca gat gtg tca ccc att atg aac aag cat tat aat Arg Ser Leu Leu Pro Asp Val Ser Pro Ile Met Asn Lys His Tyr Asn 145 150 155 160	480
att tgt gct gtg gtt gga aat agt ggg atc ctg aca ttt atc cag tgt Ile Cys Ala Val Val Gly Asn Ser Gly Ile Leu Thr Phe Ile Gln Cys 165 170 175	528
gga cga gaa ata gat aaa tca gat ttt gtt ttc cgt tgc aat ttc gcc	576

Gly	Arg	Glu	Ile	Asp	Lys	Ser	Asp	Phe	Val	Phe	Arg	Cys	Asn	Phe	Ala		
			180					185					190				
cca	tcg	gag	gct	ttc	caa	aga	gat	gtt	gga	agg	aaa	acc	aat	ctt	acc		624
Pro	Ser	Glu	Ala	Phe	Gln	Arg	Asp	Val	Gly	Arg	Lys	Thr	Asn	Leu	Thr		
		195					200					205					
acc	ttc	aac	ccc	agc	atc	ctg	gaa	aaa	tat	tac	aac	aat	ctc	ttg	act		672
Thr	Phe	Asn	Pro	Ser	Ile	Leu	Glu	Lys	Tyr	Tyr	Asn	Asn	Leu	Leu	Thr		
	210					215					220						
att	cag	gac	cgt	aac	aac	ttt	ttc	ctc	agt	tta	aaa	aag	ctt	gac	ggg		720
Ile	Gln	Asp	Arg	Asn	Asn	Phe	Phe	Leu	Ser	Leu	Lys	Lys	Leu	Asp	Gly		
225					230					235					240		
gcc	att	ctt	tgg	atc	cct	gca	ttt	ttc	ttc	cac	act	tca	gca	act	gtg		768
Ala	Ile	Leu	Trp	Ile	Pro	Ala	Phe	Phe	Phe	His	Thr	Ser	Ala	Thr	Val		
				245					250					255			
acc	agg	aca	tta	gtt	gac	ttt	ttt	gtt	gaa	cac	aga	ggt	cag	tta	aaa		816
Thr	Arg	Thr	Leu	Val	Asp	Phe	Phe	Val	Glu	His	Arg	Gly	Gln	Leu	Lys		
			260					265					270				
gtc	caa	ctg	gct	tgg	ccg	gga	aat	ata	atg	caa	cat	gtc	aac	agg	tac		864
Val	Gln	Leu	Ala	Trp	Pro	Gly	Asn	Ile	Met	Gln	His	Val	Asn	Arg	Tyr		
		275					280						285				
tgg	aaa	aac	aaa	cat	ttg	tca	cct	aaa	cgg	ctg	agc	aca	ggt	att	ctt		912
Trp	Lys	Asn	Lys	His	Leu	Ser	Pro	Lys	Arg	Leu	Ser	Thr	Gly	Ile	Leu		
	290					295					300						
atg	tac	acc	ctt	gca	tca	gca	ata	tgt	gaa	gag	atc	cac	ttg	tat	gga		960
Met	Tyr	Thr	Leu	Ala	Ser	Ala	Ile	Cys	Glu	Glu	Ile	His	Leu	Tyr	Gly		
305					310					315				320			
ttt	tgg	ccg	ttt	gga	ttt	gac	ccc	aac	aca	agg	gaa	gat	ctt	cca	tac		1008
Phe	Trp	Pro	Phe	Gly	Phe	Asp	Pro	Asn	Thr	Arg	Glu	Asp	Leu	Pro	Tyr		
				325					330					335			
cat	tac	tat	gac	aaa	aaa	gga	acc	aaa	ttt	acc	acc	aag	tgg	cag	gag		1056
His	Tyr	Tyr	Asp	Lys	Lys	Gly	Thr	Lys	Phe	Thr	Thr	Lys	Trp	Gln	Glu		
			340					345					350				
tcc	cac	cag	ctg	cct	gct	gag	ttt	cag	ctg	ctg	tac	cga	atg	cat	ggg		1104
Ser	His	Gln	Leu	Pro	Ala	Glu	Phe	Gln	Leu	Leu	Tyr	Arg	Met	His	Gly		
		355				360						365					
gaa	ggg	ctc	acc	aag	ctg	act	ctg	tca	cac	tgt	gcc	taa					1143
Glu	Gly	Leu	Thr	Lys	Leu	Thr	Leu	Ser	His	Cys	Ala						
	370					375					380						

<210> 36

<211> 380

<212> PRT

<213> *Homo sapiens*

<400> 36

Met	Arg	Asn	Cys	Lys	Met	Ala	Arg	Val	Ala	Ser	Val	Leu	Gly	Leu	Val
1				5					10					15	

Met Leu Ser Val Ala Leu Leu Ile Leu Ser Leu Ile Ser Tyr Val Ser
 20 25 30

Leu Lys Lys Glu Asn Ile Phe Thr Thr Pro Lys Tyr Ala Ser Pro Gly
 35 40 45

Ala Pro Arg Met Tyr Met Phe His Ala Gly Phe Arg Ser Gln Phe Ala
 50 55 60

Leu Lys Phe Leu Asp Pro Ser Phe Val Pro Ile Thr Asn Ser Leu Thr
 65 70 75 80

Gln Glu Leu Gln Glu Lys Pro Ser Lys Trp Lys Phe Asn Arg Thr Ala
 85 90 95

Phe Leu His Gln Arg Gln Glu Ile Leu Gln His Val Asp Val Ile Lys
 100 105 110

Asn Phe Ser Leu Thr Lys Asn Ser Val Arg Ile Gly Gln Leu Met His
 115 120 125

Tyr Asp Tyr Ser Ser His Lys Tyr Val Phe Ser Ile Ser Asn Asn Phe
 130 135 140

Arg Ser Leu Leu Pro Asp Val Ser Pro Ile Met Asn Lys His Tyr Asn
 145 150 155 160

Ile Cys Ala Val Val Gly Asn Ser Gly Ile Leu Thr Phe Ile Gln Cys
 165 170 175

Gly Arg Glu Ile Asp Lys Ser Asp Phe Val Phe Arg Cys Asn Phe Ala
 180 185 190

Pro Ser Glu Ala Phe Gln Arg Asp Val Gly Arg Lys Thr Asn Leu Thr
 195 200 205

Thr Phe Asn Pro Ser Ile Leu Glu Lys Tyr Tyr Asn Asn Leu Leu Thr
 210 215 220

Ile Gln Asp Arg Asn Asn Phe Phe Leu Ser Leu Lys Lys Leu Asp Gly
 225 230 235 240

Ala Ile Leu Trp Ile Pro Ala Phe Phe Phe His Thr Ser Ala Thr Val
 245 250 255

Thr Arg Thr Leu Val Asp Phe Phe Val Glu His Arg Gly Gln Leu Lys
 260 265 270

Val Gln Leu Ala Trp Pro Gly Asn Ile Met Gln His Val Asn Arg Tyr
 275 280 285

Trp Lys Asn Lys His Leu Ser Pro Lys Arg Leu Ser Thr Gly Ile Leu
 290 295 300

Met Tyr Thr Leu Ala Ser Ala Ile Cys Glu Glu Ile His Leu Tyr Gly
 305 310 315 320

Phe Trp Pro Phe Gly Phe Asp Pro Asn Thr Arg Glu Asp Leu Pro Tyr
 325 330 335

His Tyr Tyr Asp Lys Lys Gly Thr Lys Phe Thr Thr Lys Trp Gln Glu
 340 345 350

Ser His Gln Leu Pro Ala Glu Phe Gln Leu Leu Tyr Arg Met His Gly
 355 360 365

Glu Gly Leu Thr Lys Leu Thr Leu Ser His Cys Ala
 370 375 380

<210> 37

<211> 1080

<212> ADN

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1) .. (1080)

<400> 37

Gln	His	Asn	Val	Glu	Gly	Trp	Lys	Ile	Asn	Ser	Ser	Leu	Val	Leu	Glu	
65					70					75					80	
ata	agg	aag	aac	ata	ctt	cgt	ttc	tta	gat	gca	gaa	cga	gat	gtg	tca	288
Ile	Arg	Lys	Asn	Ile	Leu	Arg	Phe	Leu	Asp	Ala	Glu	Arg	Asp	Val	Ser	
			85						90					95		
gtg	gtc	aag	agc	agt	ttt	aag	cct	ggg	gat	gtc	ata	cac	tat	gtg	ctt	336
Val	Val	Lys	Ser	Ser	Phe	Lys	Pro	Gly	Asp	Val	Ile	His	Tyr	Val	Leu	
			100					105					110			
gac	agg	cgc	cgg	aca	cta	aac	att	tct	cat	gat	cta	cat	agc	ctc	cta	384
Asp	Arg	Arg	Arg	Thr	Leu	Asn	Ile	Ser	His	Asp	Leu	His	Ser	Leu	Leu	
			115				120					125				
cct	gaa	gtt	tca	cca	atg	aag	aat	cgc	agg	ttt	aag	acc	tgt	gca	gtt	432
Pro	Glu	Val	Ser	Pro	Met	Lys	Asn	Arg	Arg	Phe	Lys	Thr	Cys	Ala	Val	
	130					135					140					
gtt	gga	aat	tct	ggc	att	ctg	tta	gac	agt	gaa	tgt	gga	aag	gag	att	480
Val	Gly	Asn	Ser	Gly	Ile	Leu	Leu	Asp	Ser	Glu	Cys	Gly	Lys	Glu	Ile	
145				150					155					160		
gac	agt	cac	aat	ttt	gta	ata	agg	tgt	aat	cta	gct	cct	gtg	gtg	gag	528
Asp	Ser	His	Asn	Phe	Val	Ile	Arg	Cys	Asn	Leu	Ala	Pro	Val	Val	Glu	
				165				170					175			
ttt	gct	gca	gat	gtg	gga	act	aaa	tca	gat	ttt	att	acc	atg	aat	cca	576
Phe	Ala	Ala	Asp	Val	Gly	Thr	Lys	Ser	Asp	Phe	Ile	Thr	Met	Asn	Pro	
			180				185						190			
tca	gtt	gta	caa	aga	gca	ttt	gga	ggc	ttt	cga	aat	gag	agt	gac	aga	624
Ser	Val	Val	Gln	Arg	Ala	Phe	Gly	Gly	Phe	Arg	Asn	Glu	Ser	Asp	Arg	
		195				200					205					
gaa	aaa	ttt	gtg	cat	aga	ctt	toc	atg	ctg	aat	gac	agt	gtc	ctt	tgg	672
Glu	Lys	Phe	Val	His	Arg	Leu	Ser	Met	Leu	Asn	Asp	Ser	Val	Leu	Trp	
	210					215					220					
att	cct	gct	ttc	atg	gtc	aaa	gga	gga	gag	aag	cac	gtg	gag	tgg	gtt	720
Ile	Pro	Ala	Phe	Met	Val	Lys	Gly	Gly	Glu	Lys	His	Val	Glu	Trp	Val	
225				230					235					240		
aat	gca	tta	atc	ctt	aag	aat	aaa	ctg	aaa	gtg	cga	act	gcc	tat	ccg	768
Asn	Ala	Leu	Ile	Leu	Lys	Asn	Lys	Leu	Lys	Val	Arg	Thr	Ala	Tyr	Pro	
			245					250					255			
tca	ttg	aga	ctt	att	cat	gct	gtc	aga	ggg	tac	tgg	ctg	acc	aac	aaa	816
Ser	Leu	Arg	Leu	Ile	His	Ala	Val	Arg	Gly	Tyr	Trp	Leu	Thr	Asn	Lys	
			260				265						270			
gtt	cct	atc	aaa	aga	ccc	agc	aca	ggg	ctt	ctc	atg	tat	aca	ctt	gcc	864
Val	Pro	Ile	Lys	Arg	Pro	Ser	Thr	Gly	Leu	Leu	Met	Tyr	Thr	Leu	Ala	
		275				280					285					
aca	aga	ttc	tgt	gat	gaa	att	cac	ctg	tat	gga	ttc	tgg	ccc	ttc	cct	912
Thr	Arg	Phe	Cys	Asp	Glu	Ile	His	Leu	Tyr	Gly	Phe	Trp	Pro	Phe	Pro	
	290					295				300						
aag	gat	tta	aat	gga	aaa	gcg	gtc	aaa	tat	cat	tat	tat	gat	gac	tta	960
Lys	Asp	Leu	Asn	Gly	Lys	Ala	Val	Lys	Tyr	His	Tyr	Tyr	Asp	Asp	Leu	

305	310	315	320	
aaa tat agg tac ttt tcc aat gca agc cct cac aga atg cca tta gaa				1008
Lys Tyr Arg Tyr Phe Ser Asn Ala Ser Pro His Arg Met Pro Leu Glu				
	325	330	335	
ttc aaa aca tta aat gtg cta cat aat aga gga gct cta aaa ctg aca				1056
Phe Lys Thr Leu Asn Val Leu His Asn Arg Gly Ala Leu Lys Leu Thr				
	340	345	350	
aca gga aag tgt gta aag caa taa				1080
Thr Gly Lys Cys Val Lys Gln				
	355			

<210> 38

<211> 359

<212> PRT

<213> *Homo sapiens*

<400> 38

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

Asp Ser His Asn Phe Val Ile Arg Cys Asn Leu Ala Pro Val Val Glu
 165 170 175

Phe Ala Ala Asp Val Gly Thr Lys Ser Asp Phe Ile Thr Met Asn Pro
 180 185 190

Ser Val Val Gln Arg Ala Phe Gly Gly Phe Arg Asn Glu Ser Asp Arg
 195 200 205

Glu Lys Phe Val His Arg Leu Ser Met Leu Asn Asp Ser Val Leu Trp
 210 215 220

Ile Pro Ala Phe Met Val Lys Gly Gly Glu Lys His Val Glu Trp Val
 225 230 235 240

Asn Ala Leu Ile Leu Lys Asn Lys Leu Lys Val Arg Thr Ala Tyr Pro
 245 250 255

Ser Leu Arg Leu Ile His Ala Val Arg Gly Tyr Trp Leu Thr Asn Lys
 260 265 270

Val Pro Ile Lys Arg Pro Ser Thr Gly Leu Leu Met Tyr Thr Leu Ala
 275 280 285

Thr Arg Phe Cys Asp Glu Ile His Leu Tyr Gly Phe Trp Pro Phe Pro
 290 295 300

Lys Asp Leu Asn Gly Lys Ala Val Lys Tyr His Tyr Tyr Asp Asp Leu
 305 310 315 320

Lys Tyr Arg Tyr Phe Ser Asn Ala Ser Pro His Arg Met Pro Leu Glu
 325 330 335

Phe Lys Thr Leu Asn Val Leu His Asn Arg Gly Ala Leu Lys Leu Thr
 340 345 350

Thr Gly Lys Cys Val Lys Gln
 355

<210> 39

<211> 1131

<212> ADN

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1) .. (1131)

<400> 39

atg cgc tac gcg gac ccc tcg gcc aac cgg gat ttg ttg ggg agc cga Met Arg Tyr Ala Asp Pro Ser Ala Asn Arg Asp Leu Leu Gly Ser Arg 1 5 10 15	48
act ttg ctc ttc atc ttc atc tgc gcc ttt gcc ttg gtg acc ttg ctg Thr Leu Leu Phe Ile Phe Ile Cys Ala Phe Ala Leu Val Thr Leu Leu 20 25 30	96
caa cag atc ctg tat ggc agg aac tac att aag agg tac ttt gaa ttt Gln Gln Ile Leu Tyr Gly Arg Asn Tyr Ile Lys Arg Tyr Phe Glu Phe 35 40 45	144
tat gag ggg cct ttt gaa tat aac tcc aca aga tgc ctg gag ctg agg Tyr Glu Gly Pro Phe Glu Tyr Asn Ser Thr Arg Cys Leu Glu Leu Arg 50 55 60	192
cac gaa ata ttg gaa gtg aag gtg ctg tcc atg gtg aag cag tca gag His Glu Ile Leu Glu Val Lys Val Leu Ser Met Val Lys Gln Ser Glu 65 70 75 80	240
ctg ttc gac agg tgg aag agc ctc cag atg tgc aaa tgg gcg atg aac Leu Phe Asp Arg Trp Lys Ser Leu Gln Met Cys Lys Trp Ala Met Asn 85 90 95	288
atc tct gag gcc aac cag ttc aag tct act ctg tcc agg tgc tgc aac Ile Ser Glu Ala Asn Gln Phe Lys Ser Thr Leu Ser Arg Cys Cys Asn 100 105 110	336
gcc cct gcc ttt ctc ttc acc acc cag aag aac act ccc ctg ggg aca Ala Pro Ala Phe Leu Phe Thr Thr Gln Lys Asn Thr Pro Leu Gly Thr 115 120 125	384
aag ctc aag tat gag gtg gac acc agt ggc atc tac cac atc aac cag Lys Leu Lys Tyr Glu Val Asp Thr Ser Gly Ile Tyr His Ile Asn Gln 130 135 140	432
gag atc ttc cgc atg ttt ccc aag gac atg ccc tac tac cgg tcc cag Glu Ile Phe Arg Met Phe Pro Lys Asp Met Pro Tyr Tyr Arg Ser Gln 145 150 155 160	480
ttt aag aag tgt gct gta gtg ggc aac gga ggc atc ttg aag aac agc Phe Lys Lys Cys Ala Val Val Gly Asn Gly Gly Ile Leu Lys Asn Ser 165 170 175	528
cgc tgc ggg agg gag atc aac agc gcc gac ttc gtc ttc cgg tgc aac Arg Cys Gly Arg Glu Ile Asn Ser Ala Asp Phe Val Phe Arg Cys Asn 180 185 190	576
ctg ccc ccc atc tca gag aag tac acc atg gat gtg ggg gtg aag acg Leu Pro Pro Ile Ser Glu Lys Tyr Thr Met Asp Val Gly Val Lys Thr 195 200 205	624
gat gtg gtc act gtg aac ccc agc atc atc aca gag agg ttc cac aag Asp Val Val Thr Val Asn Pro Ser Ile Ile Thr Glu Arg Phe His Lys 210 215 220	672
ctg gag aag tgg cgg cgg ccg ttc tat cgc gtg ctg cag gtg tac gag Leu Glu Lys Trp Arg Arg Pro Phe Tyr Arg Val Leu Gln Val Tyr Glu	720

225	230	235	240	
aac gcg tcg gtg ctg ctg cct gcc ttc tac aac acg cgc aac acc gac				768
Asn Ala Ser Val Leu Leu Pro Ala Phe Tyr Asn Thr Arg Asn Thr Asp				
245		250	255	
gtg tcc atc cgc gtc aag tac gtg ctg gac gac ttc gaa tcg ccg caa				816
Val Ser Ile Arg Val Lys Tyr Val Leu Asp Asp Phe Glu Ser Pro Gln				
260		265	270	
gct gtc tac tac ttc cat ccg cag tac ctg gtc aac gtg tcg cgc tac				864
Ala Val Tyr Tyr Phe His Pro Gln Tyr Leu Val Asn Val Ser Arg Tyr				
275		280	285	
tgg ctc agc ctg ggg gtg cgc gcc aag cgc atc agc acc ggc ctc att				912
Trp Leu Ser Leu Gly Val Arg Ala Lys Arg Ile Ser Thr Gly Leu Ile				
290		295	300	
ctg gtc act gcg gcg ctg gag ctc tgt gag gag gtg cac ctc ttt ggc				960
Leu Val Thr Ala Ala Leu Glu Leu Cys Glu Glu Val His Leu Phe Gly				
305		310	315	320
ttc tgg gcc ttc ccc atg aac ccc tcg ggc ctc tac atc act cac cac				1008
Phe Trp Ala Phe Pro Met Asn Pro Ser Gly Leu Tyr Ile Thr His His				
325		330	335	
tac tat gac aac gtc aag ccg cgt ccc ggc ttc cac gcc atg ccc tct				1056
Tyr Tyr Asp Asn Val Lys Pro Arg Pro Gly Phe His Ala Met Pro Ser				
340		345	350	
gag atc ttc aac ttc ctg cac ttg cac agc cga ggc atc ctc cgc gtg				1104
Glu Ile Phe Asn Phe Leu His Leu His Ser Arg Gly Ile Leu Arg Val				
355		360	365	
cac acg ggc acc tgc agc tgc tgc tga				1131
His Thr Gly Thr Cys Ser Cys Cys				
370		375		

<210> 40

<211> 376

<212> PRT

<213> *Homo sapiens*

<400> 40

Met Arg Tyr Ala Asp Pro Ser Ala Asn Arg Asp Leu Leu Gly Ser Arg
1 5 10 15

Thr Leu Leu Phe Ile Phe Ile Cys Ala Phe Ala Leu Val Thr Leu Leu
20 25 30

Gln Gln Ile Leu Tyr Gly Arg Asn Tyr Ile Lys Arg Tyr Phe Glu Phe
35 40 45

Tyr Glu Gly Pro Phe Glu Tyr Asn Ser Thr Arg Cys Leu Glu Leu Arg
50 55 60

His Glu Ile Leu Glu Val Lys Val Leu Ser Met Val Lys Gln Ser Glu
65 70 75 80

Leu Phe Asp Arg Trp Lys Ser Leu Gln Met Cys Lys Trp Ala Met Asn
85 90 95

Ile Ser Glu Ala Asn Gln Phe Lys Ser Thr Leu Ser Arg Cys Cys Asn
100 105 110

Ala Pro Ala Phe Leu Phe Thr Thr Gln Lys Asn Thr Pro Leu Gly Thr
115 120 125

Lys Leu Lys Tyr Glu Val Asp Thr Ser Gly Ile Tyr His Ile Asn Gln
130 135 140

Glu Ile Phe Arg Met Phe Pro Lys Asp Met Pro Tyr Tyr Arg Ser Gln
145 150 155 160

Phe Lys Lys Cys Ala Val Val Gly Asn Gly Gly Ile Leu Lys Asn Ser
165 170 175

Arg Cys Gly Arg Glu Ile Asn Ser Ala Asp Phe Val Phe Arg Cys Asn
180 185 190

Leu Pro Pro Ile Ser Glu Lys Tyr Thr Met Asp Val Gly Val Lys Thr
195 200 205

Asp Val Val Thr Val Asn Pro Ser Ile Ile Thr Glu Arg Phe His Lys
210 215 220

Leu Glu Lys Trp Arg Arg Pro Phe Tyr Arg Val Leu Gln Val Tyr Glu
225 230 235 240

Asn Ala Ser Val Leu Leu Pro Ala Phe Tyr Asn Thr Arg Asn Thr Asp
245 250 255

Val Ser Ile Arg Val Lys Tyr Val Leu Asp Asp Phe Glu Ser Pro Gln
260 265 270

Ala Val Tyr Tyr Phe His Pro Gln Tyr Leu Val Asn Val Ser Arg Tyr
275 280 285

Trp Leu Ser Leu Gly Val Arg Ala Lys Arg Ile Ser Thr Gly Leu Ile
290 295 300

Leu Val Thr Ala Ala Leu Glu Leu Cys Glu Glu Val His Leu Phe Gly
305 310 315 320

Phe Trp Ala Phe Pro Met Asn Pro Ser Gly Leu Tyr Ile Thr His His
325 330 335

Tyr Tyr Asp Asn Val Lys Pro Arg Pro Gly Phe His Ala Met Pro Ser
340 345 350

Glu Ile Phe Asn Phe Leu His Leu His Ser Arg Gly Ile Leu Arg Val
355 360 365

His Thr Gly Thr Cys Ser Cys Cys
370 375

<210> 41

<211> 1197

<212> ADN

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1) .. (1197)

<400> 41

atg cgg ccg ggg ggc gca ctg ctc gcc ctg ctc gcc agc ctg ctg ctg	48
Met Arg Pro Gly Gly Ala Leu Leu Ala Leu Leu Ala Ser Leu Leu Leu	
1 5 10 15	
ctg ctg ctg ctg cgc ctg ctc tgg tgc ccg gca gac gcg ccc ggc cgc	96
Leu Leu Leu Leu Arg Leu Leu Trp Cys Pro Ala Asp Ala Pro Gly Arg	
20 25 30	
gcc agg att ctg gtg gag gaa agc agg gag gcc acc cac ggc acc ccc	144
Ala Arg Ile Leu Val Glu Glu Ser Arg Glu Ala Thr His Gly Thr Pro	
35 40 45	
gca gcg ctg agg acg ctc cgg agc ccg gcg acc gcg gta ccg cgc gcc	192
Ala Ala Leu Arg Thr Leu Arg Ser Pro Ala Thr Ala Val Pro Arg Ala	
50 55 60	
act aac agc aca tat ctg aat gag aag tcg ctc caa ctg acg gag aaa	240
Thr Asn Ser Thr Tyr Leu Asn Glu Lys Ser Leu Gln Leu Thr Glu Lys	
65 70 75 80	
tgc aaa aat ctg caa tat ggc att gag tct ttc tct aac aaa acg aaa	288
Cys Lys Asn Leu Gln Tyr Gly Ile Glu Ser Phe Ser Asn Lys Thr Lys	
85 90 95	
ggg tat tca gag aac gac tac ctt cag att atc aca gat ata cag agt	336
Gly Tyr Ser Glu Asn Asp Tyr Leu Gln Ile Ile Thr Asp Ile Gln Ser	
100 105 110	
tgt cca tgg aaa cgg caa gca gaa gaa tat gca aat ttt aga gcc aaa	384
Cys Pro Trp Lys Arg Gln Ala Glu Glu Tyr Ala Asn Phe Arg Ala Lys	

115	120	125	
ctt gct tcc tgc tgt gat gct gtt caa aac ttt gtt gtt tct cag aat Leu Ala Ser Cys Cys Asp Ala Val Gln Asn Phe Val Val Ser Gln Asn 130 135 140			432
aac act cca gtt ggg act aat atg agt tac gag gtg gaa agc aaa aaa Asn Thr Pro Val Gly Thr Asn Met Ser Tyr Glu Val Glu Ser Lys Lys 145 150 155 160			480
gaa atc cca att aag aag aac att ttt cat atg ttt cca gtg tcc cag Glu Ile Pro Ile Lys Lys Asn Ile Phe His Met Phe Pro Val Ser Gln 165 170 175			528
cct ttt gtg gac tac cct tat aat cag tgt gca gtg gtc gga aat ggg Pro Phe Val Asp Tyr Pro Tyr Asn Gln Cys Ala Val Val Gly Asn Gly 180 185 190			576
gga att ctg aat aag tct ctc tgt gga act gaa ata gat aaa tcc gac Gly Ile Leu Asn Lys Ser Leu Cys Gly Thr Glu Ile Asp Lys Ser Asp 195 200 205			624
ttc gtt ttt agg tgt aac cta ccc cca acc aca gga gat gtt agt aaa Phe Val Phe Arg Cys Asn Leu Pro Pro Thr Thr Gly Asp Val Ser Lys 210 215 220			672
gat gtt ggc agt aaa aca aat ctt gtg act ata aat cca agc atc ata Asp Val Gly Ser Lys Lys Asn Leu Val Thr Ile Asn Pro Ser Ile Ile 225 230 235 240			720
act ctg aaa tat ggg aac tta aag gaa aaa aaa gcc cta ttt ctg gag Thr Leu Lys Tyr Gly Asn Leu Lys Glu Lys Lys Ala Leu Phe Leu Glu 245 250 255			768
gac att gca acc tat gga gat gca ttt ttt ctt ctg cca gca ttt tcc Asp Ile Ala Thr Tyr Gly Asp Ala Phe Phe Leu Leu Pro Ala Phe Ser 260 265 270			816
ttc agg gcc aac acg ggt acc tct ttc aaa gta tac tac acg ctc gaa Phe Arg Ala Asn Thr Gly Thr Ser Phe Lys Val Tyr Tyr Thr Leu Glu 275 280 285			864
gag tct aaa gca aga caa aag gtt cta ttt ttc cat ccc aag tac ctg Glu Ser Lys Ala Arg Gln Lys Val Leu Phe Phe His Pro Lys Tyr Leu 290 295 300			912
aaa gat ctg gcc ctt ttc tgg aga act aaa ggt gtg act gca tac cgc Lys Asp Leu Ala Leu Phe Trp Arg Thr Lys Gly Val Thr Ala Tyr Arg 305 310 315 320			960
ttg tcc acc ggc ttg atg atc aca agt gtt gca gtg gaa ctg tgt aaa Leu Ser Thr Gly Leu Met Ile Thr Ser Val Ala Val Glu Leu Cys Lys 325 330 335			1008
aat gtg aag ctg tat gga ttc tgg ccc ttc tct aaa act gta gaa gac Asn Val Lys Leu Tyr Gly Phe Trp Pro Phe Ser Lys Thr Val Glu Asp 340 345 350			1056
ata cct gtc agc cat cac tat tat gac aac aag cta cct aaa cat ggt Ile Pro Val Ser His His Tyr Tyr Asp Asn Lys Leu Pro Lys His Gly 355 360 365			1104

ttc cat cag atg ccc aaa gaa tac agc cag atc ctc caa ctt cac atg 1152
Phe His Gln Met Pro Lys Glu Tyr Ser Gln Ile Leu Gln Leu His Met
370 375 380

aaa gga atc ctc aaa ctg caa ttt agc aaa tgt gaa gtc gcc taa 1197
Lys Gly Ile Leu Lys Leu Gln Phe Ser Lys Cys Glu Val Ala
385 390 395

<210> 42

<211> 398

<212> PRT

<213> *Homo sapiens*

<400> 42

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

Pro Phe Val Asp Tyr Pro Tyr Asn Gln Cys Ala Val Val Gly Asn Gly
 180 185 190

Gly Ile Leu Asn Lys Ser Leu Cys Gly Thr Glu Ile Asp Lys Ser Asp
 195 200 205

Phe Val Phe Arg Cys Asn Leu Pro Pro Thr Thr Gly Asp Val Ser Lys
 210 215 220

Asp Val Gly Ser Lys Thr Asn Leu Val Thr Ile Asn Pro Ser Ile Ile
 225 230 235 240

Thr Leu Lys Tyr Gly Asn Leu Lys Glu Lys Lys Ala Leu Phe Leu Glu
 245 250 255

Asp Ile Ala Thr Tyr Gly Asp Ala Phe Phe Leu Leu Pro Ala Phe Ser
 260 265 270

Phe Arg Ala Asn Thr Gly Thr Ser Phe Lys Val Tyr Tyr Thr Leu Glu
 275 280 285

Glu Ser Lys Ala Arg Gln Lys Val Leu Phe Phe His Pro Lys Tyr Leu
 290 295 300

Lys Asp Leu Ala Leu Phe Trp Arg Thr Lys Gly Val Thr Ala Tyr Arg
 305 310 315 320

Leu Ser Thr Gly Leu Met Ile Thr Ser Val Ala Val Glu Leu Cys Lys
 325 330 335

Asn Val Lys Leu Tyr Gly Phe Trp Pro Phe Ser Lys Thr Val Glu Asp
 340 345 350

Ile Pro Val Ser His His Tyr Tyr Asp Asn Lys Leu Pro Lys His Gly
 355 360 365

Phe His Gln Met Pro Lys Glu Tyr Ser Gln Ile Leu Gln Leu His Met
 370 375 380

Lys Gly Ile Leu Lys Leu Gln Phe Ser Lys Cys Glu Val Ala
 385 390 395

<211> 951

<212> ADN

<213> Sequência artificial

<220>

<223> Nucleótidos 268-1218 de SEQ ID NO:1

<220>

<221> CDS

<222> (1)..(951)

<400> 43

gag gcc tcc ttc cag gtg tgg aac aag gac agc tct tcc aaa aac ctt	48
Glu Ala Ser Phe Gln Val Trp Asn Lys Asp Ser Ser Ser Lys Asn Leu	
1 5 10 15	
atc cct agg ctg caa aag atc tgg aag aat tac cta agc atg aac aag	96
Ile Pro Arg Leu Gln Lys Ile Trp Lys Asn Tyr Leu Ser Met Asn Lys	
20 25 30	
tac aaa gtg tcc tac aag ggg cca gga cca ggc atc aag ttc agt gca	144
Tyr Lys Val Ser Tyr Lys Gly Pro Gly Pro Gly Ile Lys Phe Ser Ala	
35 40 45	
gag gcc ctg cgc tgc cac ctg cgg gac cat gtg aat gta tcc atg gta	192
Glu Ala Leu Arg Cys His Leu Arg Asp His Val Asn Val Ser Met Val	
50 55 60	
gag gtc aca gat ttt ccc ttc aat acc tct gaa tgg gag ggt tat ctg	240
Glu Val Thr Asp Phe Pro Phe Asn Thr Ser Glu Trp Glu Gly Tyr Leu	
65 70 75 80	
ccc aag gag agc att agg acc aag gct ggg cct tgg ggc agg tgt gct	288
Pro Lys Glu Ser Ile Arg Thr Lys Ala Gly Pro Trp Gly Arg Cys Ala	
85 90 95	
gtt gtg tgc tca gcg gga tct ctg aag tcc tcc caa cta ggc aga gaa	336
Val Val Ser Ser Ala Gly Ser Leu Lys Ser Ser Gln Leu Gly Arg Glu	
100 105 110	
atc gat gat cat gac gca gtc ctg agg ttt aat ggg gca ccc aca gcc	384
Ile Asp Asp His Asp Ala Val Leu Arg Phe Asn Gly Ala Pro Thr Ala	
115 120 125	
aac ttc caa caa gat gtg ggc aca aaa act acc att cgc ctg atg aac	432
Asn Phe Gln Gln Asp Val Gly Thr Lys Thr Thr Ile Arg Leu Met Asn	
130 135 140	
tct cag ttg gtt acc aca gag aag cgc ttc ctg aaa gac agt ttg tac	480
Ser Gln Leu Val Thr Thr Glu Lys Arg Phe Leu Lys Asp Ser Leu Tyr	
145 150 155 160	
aat gaa gga atc cta att gta tgg gac cca tct gta tac cac tca gat	528
Asn Glu Gly Ile Leu Ile Val Trp Asp Pro Ser Val Tyr His Ser Asp	
165 170 175	
atc cca aag tgg tac cag aat ccg gat tat aat ttc ttt aac aac tac	576
Ile Pro Lys Trp Tyr Gln Asn Pro Asp Tyr Asn Phe Phe Asn Asn Tyr	
180 185 190	
aag act tat cgt aag ctg cac ccc aat cag ccc ttt tac atc ctg aag	624
Lys Thr Tyr Arg Lys Leu His Pro Asn Gln Pro Phe Tyr Ile Leu Lys	
195 200 205	

ccc cag atg cct tgg gag cta tgg gac att ctt caa gaa atc tcc cca Pro Gln Met Pro Trp Glu Leu Trp Asp Ile Leu Gln Glu Ile Ser Pro 210 215 220	672
gaa gag att cag cca aac ccc sca tcc tct ggg atg ctt ggt atc atc Glu Glu Ile Gln Pro Asn Pro Pro Ser Ser Gly Met Leu Gly Ile Ile 225 230 235 240	720
atc atg atg acg ctg tgt gac cag gtg gat att tat gag ttc ctc cca Ile Met Met Thr Leu Cys Asp Gln Val Asp Ile Tyr Glu Phe Leu Pro 245 250 255	768
tcc aag cgc aag act gac gtg tgc tac tac tac cag aag ttc ttc gat Ser Lys Arg Lys Thr Asp Val Cys Tyr Tyr Tyr Gln Lys Phe Phe Asp 260 265 270	816
agt gcc tgc acg atg ggt gcc tac cac ccg ctg ctc tat gag aag aat Ser Ala Cys Thr Met Gly Ala Tyr His Pro Leu Leu Tyr Glu Lys Asn 275 280 285	864
ttg gtg aag cat ctc aac cag ggc aca gat gag gac atc tac ctg ctt Leu Val Lys His Leu Asn Gln Gly Thr Asp Glu Asp Ile Tyr Leu Leu 290 295 300	912
gga aaa gcc aca ctg cct ggc ttc cgg acc att cac tgc Gly Lys Ala Thr Leu Pro Gly Phe Arg Thr Ile His Cys 305 310 315	951

<210> 44

<211> 317

<212> PRT

<213> Sequência artificial

<220>

<223> Construção Sintética

<400> 44

Glu Ala Ser Phe Gln Val Trp Asn Lys Asp Ser Ser Ser Lys Asn Leu
1 5 10 15

Ile Pro Arg Leu Gln Lys Ile Trp Lys Asn Tyr Leu Ser Met Asn Lys
20 25 30

Tyr Lys Val Ser Tyr Lys Gly Pro Gly Pro Gly Ile Lys Phe Ser Ala
35 40 45

Glu Ala Leu Arg Cys His Leu Arg Asp His Val Asn Val Ser Met Val
50 55 60

Glu Val Thr Asp Phe Pro Phe Asn Thr Ser Glu Trp Glu Gly Tyr Leu
65 70 75 80

Pro Lys Glu Ser Ile Arg Thr Lys Ala Gly Pro Trp Gly Arg Cys Ala
85 90 95

Val Val Ser Ser Ala Gly Ser Leu Lys Ser Ser Gln Leu Gly Arg Glu
 100 105 110

Ile Asp Asp His Asp Ala Val Leu Arg Phe Asn Gly Ala Pro Thr Ala
 115 120 125

Asn Phe Gln Gln Asp Val Gly Thr Lys Thr Thr Ile Arg Leu Met Asn
 130 135 140

Ser Gln Leu Val Thr Thr Glu Lys Arg Phe Leu Lys Asp Ser Leu Tyr
 145 150 155 160

Asn Glu Gly Ile Leu Ile Val Trp Asp Pro Ser Val Tyr His Ser Asp
 165 170 175

Ile Pro Lys Trp Tyr Gln Asn Pro Asp Tyr Asn Phe Phe Asn Asn Tyr
 180 185 190

Lys Thr Tyr Arg Lys Leu His Pro Asn Gln Pro Phe Tyr Ile Leu Lys
 195 200 205

Pro Gln Met Pro Trp Glu Leu Trp Asp Ile Leu Gln Glu Ile Ser Pro
 210 215 220

Glu Glu Ile Gln Pro Asn Pro Pro Ser Ser Gly Met Leu Gly Ile Ile
 225 230 235 240

Ile Met Met Thr Leu Cys Asp Gln Val Asp Ile Tyr Glu Phe Leu Pro
 245 250 255

Ser Lys Arg Lys Thr Asp Val Cys Tyr Tyr Tyr Gln Lys Phe Phe Asp
 260 265 270

Ser Ala Cys Thr Met Gly Ala Tyr His Pro Leu Leu Tyr Glu Lys Asn
 275 280 285

Leu Val Lys His Leu Asn Gln Gly Thr Asp Glu Asp Ile Tyr Leu Leu
 290 295 300

Gly Lys Ala Thr Leu Pro Gly Phe Arg Thr Ile His Cys
 305 310 315

<210> 45

<211> 27

<212> ADN

<213> Sequência artificial

<220>

<223> Nucleótidos 1-27 de SEQ ID NO:1

<220>

<221> CDS

<222> (1)..(27)

<400> 45

atg att cac acc aac ctg aag aaa aag
Met Ile His Thr Asn Leu Lys Lys Lys
1 5

27

<210> 46

<211> 9

<212> PRT

<213> Sequência artificial

<220>

<223> Construção Sintética

<400> 46

Met Ile His Thr Asn Leu Lys Lys Lys
1 5

<210> 47

<211> 30

<212> ADN

<213> Sequência artificial

<220>

<223> Nucleótidos 1-30 de SEQ ID NO:3

<220>

<221> CDS

<222> (1)..(30)

<400> 47

atg aaa cca cac ttg aag caa tgg aga caa
Met Lys Pro His Leu Lys Gln Trp Arg Gln
1 5 10

30

<210> 48
<211> 10
<212> PRT
<213> Sequência artificial

<220>
<223> Construção Sintética

<400> 48

Met	Lys	Pro	His	Leu	Lys	Gln	Trp	Arg	Gln
1				5					10

<210> 49
<211> 42
<212> ADN
<213> Sequência artificial

<220>
<223> Nucleótidos 1-42 de SEQ ID NO:5

<220>
<221> CDS
<222> (1)..(42)

<400> 49

atg	agg	tcc	tgc	ctg	tgg	aga	tgc	agg	cac	ctg	agc	caa	ggc
Met	Arg	Ser	Cys	Leu	Trp	Arg	Cys	Arg	His	Leu	Ser	Gln	Gly
1				5					10				

42

<210> 50
<211> 14
<212> PRT
<213> Sequência artificial

<220>
<223> Construção Sintética

<400> 50

Met	Arg	Ser	Cys	Leu	Trp	Arg	Cys	Arg	His	Leu	Ser	Gln	Gly
1				5					10				

<210> 51
<211> 21
<212> ADN
<213> Sequência artificial

<220>
<223> Nucleótidos 1-21 de SEQ ID NO:7

<220>
<221> CDS
<222> (1) .. (21)

<400> 51
atg ggg ctc ccg cgc ggg tcg
Met Gly Leu Pro Arg Gly Ser
1 5

21

<210> 52
<211> 7
<212> PRT
<213> Sequência artificial

<220>
<223> Construção Sintética

<400> 52
Met Gly Leu Pro Arg Gly Ser
1 5

<210> 53
<211> 24
<212> ADN
<213> Artificial

<220>
<223> Nucleótidos 1-24 de SEQ ID NO:9

<220>
<221> CDS
<222> (1) .. (24)

<400> 53

atg gcc tgc atc ctg aag aga aag
Met Ala Cys Ile Leu Lys Arg Lys
1 5

24

<210> 54

<211> 8

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 54

Met Ala Cys Ile Leu Lys Arg Lys
1 5

<210> 55

<211> 18

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-18 de SEQ ID NO:11

<220>

<221> CDS

<222> (1)..(18)

<400> 55

atg aag gct ccg ggt cgg
Met Lys Ala Pro Gly Arg
1 5

18

<210> 56

<211> 6

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 56

Met Lys Ala Pro Gly Arg
1 5

<210> 57

<211> 24

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-24 de SEQ ID NO:13

<220>

<221> CDS

<222> (1)..(21)

<400> 57

atg aag acc ctg atg cgc cat ggt
Met Lys Thr Leu Met Arg His
1 5

24

<210> 58

<211> 7

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 58

Met Lys Thr Leu Met Arg His
1 5

<210> 59

<211> 129

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-129 de SEQ ID NO:15

<220>

<221> CDS

<222> (1)..(129)

<400> 59

```

atg gct tgc tcg agg ccc ccc agc cag tgt gaa ccc aca tcc ctg ccc      48
Met Ala Cys Ser Arg Pro Pro Ser Gln Cys Glu Pro Thr Ser Leu Pro

```

```

      1             5             10             15
cca ggg cca cct gca gga cgc cga cac cta ccc ctc agc aga cgc cgg      96
Pro Gly Pro Pro Ala Gly Arg Arg His Leu Pro Leu Ser Arg Arg Arg
      20             25             30

aga gaa atg agt agc aac aaa gag cag cgg tca      129
Arg Glu Met Ser Ser Asn Lys Glu Gln Arg Ser
      35             40

```

<210> 60

<211> 43

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 60

```

Met Ala Cys Ser Arg Pro Pro Ser Gln Cys Glu Pro Thr Ser Leu Pro
1             5             10             15

Pro Gly Pro Pro Ala Gly Arg Arg His Leu Pro Leu Ser Arg Arg Arg
      20             25             30

Arg Glu Met Ser Ser Asn Lys Glu Gln Arg Ser
      35             40

```

<210> 61

<211> 39

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-39 de SEQ ID NO:17

<220>

<221> CDS

<222> (1) .. (30)

<400> 61

```
atg gtg acc ctg cgg aag agg acc ctg aaa gtgctcacc
Met Val Thr Leu Arg Lys Arg Thr Leu Lys
1           5           10
```

39

<210> 62

<211> 10

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 62

```
Met Val Thr Leu Arg Lys Arg Thr Leu Lys
1           5           10
```

<210> 63

<211> 18

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-18 de SEQ ID NO:19

<220>

<221> CDS

<222> (1) .. (18)

<400> 63

```
atg aag tgc tcc ctg cgg
Met Lys Cys Ser Leu Arg
1           5
```

18

<210> 64

<211> 6

<212> PRT

<213> Artificial

<220>

<223> Construção sintética

<400> 64

Met Lys Cys Ser Leu Arg
1 5

<210> 65

<211> 24

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-24 de SEQ ID NO:21

<220>

<221> CDS

<222> (1)..(24)

<400> 65

atg gga ctc ttg gta ttt gtg cgc
Met Gly Leu Leu Val Phe Val Arg
1 5

24

<210> 66

<211> 8

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 66

Met Gly Leu Leu Val Phe Val Arg
1 5

<210> 67

<211> 24

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-24 de SEQ ID NO:23

<220>

<221> CDS

<222> (1)..(24)

<400> 67

atg gtc agc aag tcc cgc tgg aag
Met Val Ser Lys Ser Arg Trp Lys
1 5

24

<210> 68

<211> 8

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 68

Met Val Ser Lys Ser Arg Trp Lys
1 5

<210> 69

<211> 15

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-15 de SEQ ID NO:25

<220>

<221> CDS

<222> (1)..(15)

<400> 69

atg aga agg ccc agc
Met Arg Arg Pro Ser
1 5

15

<210> 70

<211> 5

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 70

Met	Arg	Arg	Pro	Ser
1			5	

<210> 71

<211> 12

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-12 de SEQ ID NO:27

<220>

<221> CDS

<222> (1)..(12)

<400> 71

atg	aga	ggg	tat
Met	Arg	Gly	Tyr
1			

12

<210> 72

<211> 4

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 72

Met	Arg	Gly	Tyr
1			

<210> 73

<211> 24

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-24 de SEQ ID NO:29

<220>

<221> CDS

<222> (1) .. (24)

<400> 73

atg gga ctc ttg gta ttt gta cgc
Met Gly Leu Leu Val Phe Val Arg
1 5

24

<210> 74

<211> 8

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 74

Met Gly Leu Leu Val Phe Val Arg
1 5

<210> 75

<211> 87

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-87 de SEQ ID NO:31

<220>

<221> CDS

<222> (1) .. (87)

<400> 75

atg agc ccc tgc ggg cgg gcc cgg cga caa acg tcc aga ggg gcc atg 48
 Met Ser Pro Cys Gly Arg Ala Arg Arg Gln Thr Ser Arg Gly Ala Met
 1 5 10 15

gct gta ctg gcg tgg aag ttc ccg cgg acc cgg ctg ccc 87
 Ala Val Leu Ala Trp Lys Phe Pro Arg Thr Arg Leu Pro
 20 25

<210> 76

<211> 29

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 76

Met Ser Pro Cys Gly Arg Ala Arg Arg Gln Thr Ser Arg Gly Ala Met
 1 5 10 15

Ala Val Leu Ala Trp Lys Phe Pro Arg Thr Arg Leu Pro
 20 25

<210> 77

<211> 18

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-18 de SEQ ID NO:33

<220>

<221> CDS

<222> (1)..(18)

<400> 77

atg cag ctg cag ttc cgg
 Met Gln Leu Gln Phe Arg
 1 5

18

<210> 78

<211> 6

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 78

Met	Gln	Leu	Gln	Phe	Arg
1				5	

<210> 79

<211> 27

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-27 de SEQ ID NO:35

<220>

<221> CDS

<222> (1)..(27)

<400> 79

atg	aga	aac	tgc	aaa	atg	gcc	cgg	gtc
Met	Arg	Asn	Cys	Lys	Met	Ala	Arg	Val
1				5				

27

<210> 80

<211> 9

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 80

Met	Arg	Asn	Cys	Lys	Met	Ala	Arg	Val
1				5				

<210> 81

<211> 21

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-21 de SEQ ID NO:37

<220>

<221> CDS

<222> (1)..(21)

<400> 81

atg cgc tcc att agg aag agg
Met Arg Ser Ile Arg Lys Arg
1 5

21

<210> 82

<211> 7

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 82

Met Arg Ser ile Arg Lys Arg
1 5

<210> 83

<211> 51

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-51 de SEQ ID NO:39

<220>

<221> CDS

<222> (1)..(51)

<400> 83

```

atg cgc tac gcg gac ccc tcg gcc aac cgg gat ttg ttg ggg agc cga      48
Met Arg Tyr Ala Asp Pro Ser Ala Asn Arg Asp Leu Leu Gly Ser Arg
1           5           10           15

```

```

act      51
Thr

```

<210> 84

<211> 17

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 84

```

Met Arg Tyr Ala Asp Pro Ser Ala Asn Arg Asp Leu Leu Gly Ser Arg
1           5           10           15

```

Thr

<210> 85

<211> 9

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-9 de SEQ ID NO:41

<220>

<221> CDS

<222> (1)..(9)

<400> 85

```

atg cgg ccg
Met Arg Pro
1

```

9

<210> 86

<211> 3

<212> PRT

<213> Artificial

<220>

<223> Aminoácidos 1-3 de SEQ ID NO:42

<400> 86

Met Arg Pro
1

<210> 87

<211> 51

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 28-78 de SEQ ID NO:1

<220>

<221> CDS

<222> (1)..(51)

<400> 87

ttc	agc	tgc	tgc	gtc	ctg	gtc	ttt	ctt	ctg	ttt	gca	gtc	atc	tgt	gtg	48
Phe	Ser	Cys	Cys	Val	Leu	Val	Phe	Leu	Leu	Phe	Ala	Val	Ile	Cys	Val	
1				5				10					15			
tgg																51
Trp																

<210> 88

<211> 17

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 88

Phe	Ser	Cys	Cys	Val	Leu	Val	Phe	Leu	Leu	Phe	Ala	Val	Ile	Cys	Val
1				5				10					15		
Trp															

<210> 89

<211> 60

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 31-90 de SEQ ID NO:3

<220>

<221> CDS

<222> (1)..(60)

<400> 89

cga atg ctt ttc gga ata ttc gct tgg ggg ctc ctc ttt ttg ctg att	48
Arg Met Leu Phe Gly Ile Phe Ala Trp Gly Leu Leu Phe Leu Leu Ile	
1 5 10 15	
ttc atc tac ttc	60
Phe Ile Tyr Phe	
20	

<210> 90

<211> 20

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 90

Arg Met Leu Phe Gly Ile Phe Ala Trp Gly Leu Leu Phe Leu Leu Ile	
1 5 10 15	
Phe Ile Tyr Phe	
20	

<210> 91

<211> 63

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 43-105 de SEQ ID NO:5

<220>

<221> CDS

<222> (1)..(63)

<400> 91

gtc	cag	tgg	tcc	ttg	ctt	ctg	gct	gtc	ctg	gtc	ttc	ttt	ctc	ttc	gcc	48
Val	Gln	Trp	Ser	Leu	Leu	Leu	Ala	Val	Leu	Val	Phe	Phe	Leu	Phe	Ala	
1				5				10					15			
ttg	ccc	tct	ttt	att												63
Leu	Pro	Ser	Phe	Ile												
				20												

<210> 92

<211> 21

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 92

Val	Gln	Trp	Ser	Leu	Leu	Leu	Ala	Val	Leu	Val	Phe	Phe	Leu	Phe	Ala
1				5				10					15		
Leu	Pro	Ser	Phe	Ile											
				20											

<210> 93

<211> 63

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 22-84 de SEQ ID NO:7

<220>

<221> CDS

<222> (1)..(63)

<400> 93

ttc ttc tgg gtg ctg ctc ctg ctc acg gct gcc tgc tcg ggg ctc ctc 48
 Phe Phe Trp Val Leu Leu Leu Leu Thr Ala Ala Cys Ser Gly Leu Leu
 1 5 10 15

ttt gcc ctg tac ttc 63
 Phe Ala Leu Tyr Phe
 20

<210> 94

<211> 21

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 94

Phe Phe Trp Val Leu Leu Leu Leu Thr Ala Ala Cys Ser Gly Leu Leu
 1 5 10 15

Phe Ala Leu Tyr Phe
 20

<210> 95

<211> 60

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 25-84 de SEQ ID NO:9

<220>

<221> CDS

<222> (1)..(60)

<400> 95

tct gtg att gct gtg agc ttc ata gca gcg ttc ctt ttc ctg ctg gtt 48
 Ser Val Ile Ala Val Ser Phe Ile Ala Ala Phe Leu Phe Leu Leu Val
 1 5 10 15

gtg cgt ctt gta 60
 Val Arg Leu Val
 20

<210> 96

<211> 20

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 96

Ser	Val	Ile	Ala	Val	Ser	Phe	Ile	Ala	Ala	Phe	Leu	Phe	Leu	Leu	Val
1				5					10					15	

Val	Arg	Leu	Val
			20

<210> 97

<211> 63

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 19-81 de SEQ ID NO:11

<220>

<221> CDS

<222> (1)..(63)

<400> 97

ctc	gtg	ctc	atc	atc	ctg	tgc	tcc	gtg	gtc	ttc	tct	gcc	gtc	tac	atc	48
Leu	Val	Leu	Ile	Ile	Leu	Cys	Ser	Val	Val	Phe	Ser	Ala	Val	Tyr	Ile	
1				5					10					15		

ctc	ctg	tgc	tgc	tgg												63
Leu	Leu	Cys	Cys	Trp												
				20												

<210> 98

<211> 21

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 98

Leu Val Leu Ile Ile Leu Cys Ser Val Val Phe Ser Ala Val Tyr Ile
 1 5 10 15

Leu Leu Cys Cys Trp
 20

<210> 99

<211> 63

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 25-87 de SEQ ID NO:13

<220>

<221> CDS

<222> (1)..(63)

<400> 99

ctg gca gtg tgt tta gcg ctc acc acc atg tgc acc agc ttg ttg cta 48
 Leu Ala Val Cys Leu Ala Leu Thr Thr Met Cys Thr Ser Leu Leu Leu
 1 5 10 15

gtg tac agc agc ctc 63
 Val Tyr Ser Ser Leu
 20

<210> 100

<211> 21

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 100

Leu Ala Val Cys Leu Ala Leu Thr Thr Met Cys Thr Ser Leu Leu Leu
 1 5 10 15

Val Tyr Ser Ser Leu
 20

<210> 101

<211> 48

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 130-177 de SEQ ID NO:15

<220>

<221> CDS

<222> (1)..(48)

<400> 101

gca	gtg	ttc	gtg	atc	ctc	ttt	gcc	ctc	atc	acc	atc	ctc	atc	ctc	tac	48
Ala	Val	Phe	Val	Ile	Leu	Phe	Ala	Leu	Ile	Thr	Ile	Leu	Ile	Leu	Tyr	
1				5				10					15			

<210> 102

<211> 16

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 102

Ala	Val	Phe	Val	Ile	Leu	Phe	Ala	Leu	Ile	Thr	Ile	Leu	Ile	Leu	Tyr
1				5				10					15		

<210> 103

<211> 63

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 40-102 de SEQ ID NO:17

<220>

<221> CDS

<222> (1)..(63)

<400> 103

```
ttc ctc gty ctc ttc atc ttc ctc acc tcc ttc ttc ctg aac tac tcc      48
Phe Leu Val Leu Phe Ile Phe Leu Thr Ser Phe Phe Leu Asn Tyr Ser
1           5           10           15
```

```
cac acc atg gty gcc      63
His Thr Met Val Ala
           20
```

<210> 104

<211> 21

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 104

```
Phe Leu Val Leu Phe Ile Phe Leu Thr Ser Phe Phe Leu Asn Tyr Ser
1           5           10           15
```

```
His Thr Met Val Ala
           20
```

<210> 105

<211> 63

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 19-81 de SEQ ID NO:19

<220>

<221> CDS

<222> (1)..(63)

<400> 105

gtg tgg ttc ctc tcc gtg gcc ttc ctg ctg gtg ttc atc atg tcc ctg 48

Val Trp Phe Leu Ser Val Ala Phe Leu Leu Val Phe Ile Met Ser Leu
1 5 10 15

ctc ttc acc tac tcg 63
Leu Phe Thr Tyr Ser
20

<210> 106

<211> 21

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 106

Val Trp Phe Leu Ser Val Ala Phe Leu Leu Val Phe Ile Met Ser Leu
1 5 10 15

Leu Phe Thr Tyr Ser
20

<210> 107

<211> 60

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 25-84 de SEQ ID NO:21

<220>

<221> CDS

<222> (1)..(60)

<400> 107

aat ctg ctg cta gcc ctc tgc ctc ttt ctg gta ctg gga ttt ttg tat 48
Asn Leu Leu Leu Ala Leu Cys Leu Phe Leu Val Leu Gly Phe Leu Tyr
1 5 10 15

tat tct gcg tgg 60
Tyr Ser Ala Trp
20

<210> 108

<211> 20

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 108

Asn	Leu	Leu	Leu	Ala	Leu	Cys	Leu	Phe	Leu	Val	Leu	Gly	Phe	Leu	Tyr
1				5					10					15	

Tyr	Ser	Ala	Trp
			20

<210> 109

<211> 54

<212> ADN

<213> Artificial

<220>

<223> nucleótidos 25-78 de SEQ ID NO:23

<220>

<221> CDS

<222> (1)..(54)

<400> 109

ctc	ctg	gcc	atg	ttg	gct	ctg	gtc	ctg	gtc	gtc	atg	gtg	tgg	tat	tcc	48
Leu	Leu	Ala	Met	Leu	Ala	Leu	Val	Leu	Val	Val	Met	Val	Trp	Tyr	Ser	
1				5				10					15			
atc	tcc															54
Ile	Ser															

<210> 110

<211> 18

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 110

Leu	Leu	Ala	Met	Leu	Ala	Leu	Val	Leu	Val	Val	Met	Val	Trp	Tyr	Ser
1				5				10						15	

Ile Ser

<210> 111

<211> 63

<212> ADN

<213> Artificial

<220>

<223> nucleótidos 16-78 de SEQ ID NO:25

<220>

<221> CDS

<222> (1)..(63)

<400> 111

ttg	tta	tta	aaa	gac	atc	ctc	aaa	tgt	aca	ttg	ctt	gtg	ttt	gga	gtg	48
Leu	Leu	Leu	Lys	Asp	Ile	Leu	Lys	Cys	Thr	Leu	Leu	Val	Phe	Gly	Val	
1			5					10					15			
tgg	atc	ctt	tat	atc												63
Trp	Ile	Leu	Tyr	Ile												
				20												

<210> 112

<211> 21

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 112

Leu Leu Leu Lys Asp Ile Leu Lys Cys Thr Leu Leu Val Phe Gly Val
 1 5 10 15

Trp Ile Leu Tyr Ile
 20

<210> 113

<211> 63

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 13-75 de SEQ ID NO:27

<220>

<221> CDS

<222> (1)..(63)

<400> 113

ctt gtc gcc ata ttc ctg agt gct gtc ttc ctc tat tat gta ctg cat	48
Leu Val Ala Ile Phe Leu Ser Ala Val Phe Leu Tyr Tyr Val Leu His	
1 5 10 15	
tgc ata tta tgg gga	63
Cys Ile Leu Trp Gly	
20	

<210> 114

<211> 21

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 114

Leu Val Ala Ile Phe Leu Ser Ala Val Phe Leu Tyr Tyr Val Leu His

1 5 10 15

Cys Ile Leu Trp Gly
20

<210> 115

<211> 60

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 25-84 de SEQ ID NO:29

<220>

<221> CDS

<222> (1)..(60)

<400> 115

aac ctg ctg cta gcc ctc tgc ctc ttt ctg gtc ctg gga ttt ttg tat 48
Asn Leu Leu Leu Ala Leu Cys Leu Phe Leu Val Leu Gly Phe Leu Tyr
1 5 10 15

tat tct gcc tgg 60
Tyr Ser Ala Trp
20

<210> 116

<211> 20

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 116

Asn Leu Leu Leu Ala Leu Cys Leu Phe Leu Val Leu Gly Phe Leu Tyr
1 5 10 15

Tyr Ser Ala Trp
20

<210> 117

<211> 57

<212> ADN

<213> Artificial

 $\langle 220 \rangle$

<223> Nucleótidos 88-144 de SEQ ID NO:31

 $\langle 220 \rangle$

<221> CDS

<222> (1) .. (57)

<400> 117

atg gga gcc agt gcc ctc tgt gtc gtg gtc ctc tgt tgg ctc tac atc 48
Met Gly Ala Ser Ala Leu Cys Val Val Val Leu Cys Trp Leu Tyr Ile
1 5 10 15

ttc ccc gtc	57
Phe Pro Val	

<210> 118

<211> 19

<212> PRT

<213> Artificial

 $\langle 220 \rangle$

<223> Construção Sintética

<400> 118

Met Gly Ala Ser Ala Leu Cys Val Val Val Leu Cys Trp Leu Tyr Ile
1 5 10 15

Phe Pro Val

<210> 119

<211> 51

<212> ADN

<213> Artificial

 $\langle 220 \rangle$

<223> Nucleótidos 19-69 de SEQ ID NO:33

 $\langle 220 \rangle$

<221> CDS

<222> (1)..(51)

<400> 119

agc	tgg	atg	ctg	gcc	gcg	ctc	acg	ctg	ctc	gtg	gtc	ttc	ctc	atc	ttc	48
Ser	Trp	Met	Leu	Ala	Ala	Leu	Thr	Leu	Leu	Val	Val	Phe	Leu	Ile	Phe	
1				5					10				15			

gca																51
Ala																

<210> 120

<211> 17

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 120

Ser	Trp	Met	Leu	Ala	Ala	Leu	Thr	Leu	Leu	Val	Val	Phe	Leu	Ile	Phe
1				5					10				15		

Ala

<210> 121

<211> 72

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 28-99 de SEQ ID NO:35

<220>

<221> CDS

<222> (1)..(72)

<400> 121

```

gcc agt gtg ctg ggg ctg gtc atg ctc agc gtc gcc ctg ctg att tta      48
Ala Ser Val Leu Gly Leu Val Met Leu Ser Val Ala Leu Leu Ile Leu
1           5           10           15

```

```

tcg ctc atc agc tac gtg tcc ctg      72
Ser Leu Ile Ser Tyr Val Ser Leu
          20

```

<210> 122

<211> 24

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 122

```

Ala Ser Val Leu Gly Leu Val Met Leu Ser Val Ala Leu Leu Ile Leu
1           5           10           15

```

```

Ser Leu Ile Ser Tyr Val Ser Leu
          20

```

<210> 123

<211> 39

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 22-60 de SEQ ID NO:37

<220>

<221> CDS

<222> (1)..(39)

<400> 123

```

tgg acg atc tgc aca ata agt ctg ctc ctg atc ttt tat      39
Trp Thr Ile Cys Thr Ile Ser Leu Leu Leu Ile Phe Tyr
1           5           10

```

<210> 124

<211> 13

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 124

```

Trp Thr Ile Cys Thr Ile Ser Leu Leu Leu Ile Phe Tyr
1           5           10

```

<210> 125

<211> 63

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 52-114 de SEQ ID NO:39

<220>

<221> CDS

<222> (1)..(63)

<400> 125

```

ttg ctc ttc atc ttc atc tgc gcc ttt gcc ttg gtg acc ttg ctg caa      48
Leu Leu Phe Ile Phe Ile Cys Ala Phe Ala Leu Val Thr Leu Leu Gln
1           5           10           15

cag atc ctg tat ggc      63
Gln Ile Leu Tyr Gly
                20

```

<210> 126

<211> 21

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 126

```

Leu Leu Phe Ile Phe Ile Cys Ala Phe Ala Leu Val Thr Leu Leu Gln
1           5           10           15
Gln Ile Leu Tyr Gly
                20

```

<210> 127

<211> 63

<212> ADN

<213> Artificial

 $\langle 220 \rangle$

<223> Nucleótidos 10-72 de SEQ ID NO:41

 $\langle 220 \rangle$

<221> CDS

 $\langle 222 \rangle \quad (1) \dots (63)$

<400> 127

ggg ggc gca ctg ctc gcc ctg ctc gcc agc ctg ctg ctg ctg ctg ctg 48
 Gly Gly Ala Leu Leu Ala Leu Leu Ala Ser Leu Leu Leu Leu Leu Leu
 1 5 10 15

 ctg cgc ctg ctc tgg 63
 Leu Arg Leu Leu Trp
 20

$\langle 210 \rangle$ 128

 $\langle 211 \rangle$ 21

<212> PRT

<213> Artificial

 $\langle 220 \rangle$

<223> Construção Sintética

<400> 128

Gly Gly Ala Leu Leu Ala Leu Leu Ala Ser Leu Leu Leu Leu Leu Leu
1 5 10 15
Leu Arg Leu Leu Trp
20

<210> 129

<211> 117

<212> ADN

<213> Artificial

 $\langle 220 \rangle$

<223> Nucleótidos 106-222 de SEQ ID NO:5

<220>

<221> CDS

<222> (1)..(117)

<400> 129

aag	gag	cct	caa	aca	aag	cct	tcc	agg	cat	caa	cgc	aca	gag	aac	att	48
Lys	Glu	Pro	Gln	Thr	Lys	Pro	Ser	Arg	His	Gln	Arg	Thr	Glu	Asn	Ile	
1				5				10					15			

aaa	gaa	agg	tct	cta	cag	tcc	ctg	gca	aag	cct	aag	tcc	cag	gca	ccc	96

Lys	Glu	Arg	Ser	Leu	Gln	Ser	Leu	Ala	Lys	Pro	Lys	Ser	Gln	Ala	Pro	
			20					25					30			

aca	agg	gca	agg	agg	aca	acc	117
Thr	Arg	Ala	Arg	Arg	Thr	Thr	
			35				

<210> 130

<211> 39

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 130

Lys	Glu	Pro	Gln	Thr	Lys	Pro	Ser	Arg	His	Gln	Arg	Thr	Glu	Asn	Ile	
1				5				10					15			

Lys	Glu	Arg	Ser	Leu	Gln	Ser	Leu	Ala	Lys	Pro	Lys	Ser	Gln	Ala	Pro	
			20					25					30			

Thr	Arg	Ala	Arg	Arg	Thr	Thr	
			35				

<210> 131

<211> 114

<212> ADN

<213> Artificial

<220>

<220>

<221> CDS

<222> (1)..(315)

<400> 133

aag gag cct caa aca aag cct tcc agg cat caa cgc aca gag aac att	48
Lys Glu Pro Gln Thr Lys Pro Ser Arg His Gln Arg Thr Glu Asn Ile	
1 5 10 15	
aaa gaa agg tct cta cag tcc ctg gca aag cct aag tcc cag gca ccc	96
Lys Glu Arg Ser Leu Gln Ser Leu Ala Lys Pro Lys Ser Gln Ala Pro	
20 25 30	
aca agg gca agg agg aca acc atc tat gca gag cca gtg cca gag aac	144
Thr Arg Ala Arg Arg Thr Thr Ile Tyr Ala Glu Pro Val Pro Glu Asn	
35 40 45	
aat gcc ctc aac aca caa acc cag ccc aag gcc cac acc acc gga gac	192
Asn Ala Leu Asn Thr Gln Thr Gln Pro Lys Ala His Thr Thr Gly Asp	
50 55 60	
aga gga aag gag gcc aac cag gca ccg ccg gag gag cag gac aag gtg	240
Arg Gly Lys Glu Ala Asn Gln Ala Pro Pro Glu Glu Gln Asp Lys Val	
65 70 75 80	
ccc cac aca gca cag agg gca gca tgg aag agc cca gaa aaa gag aaa	288
Pro His Thr Ala Gln Arg Ala Ala Trp Lys Ser Pro Glu Lys Glu Lys	
85 90 95	
acc atg gtg aac aca ctg tca ccc aga	315
Thr Met Val Asn Thr Leu Ser Pro Arg	
100 105	

<210> 134

<211> 105

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 134

Lys Glu Pro Gln Thr Lys Pro Ser Arg His Gln Arg Thr Glu Asn Ile
1 5 10 15

Lys Glu Arg Ser Leu Gln Ser Leu Ala Lys Pro Lys Ser Gln Ala Pro
20 25 30

Thr Arg Ala Arg Arg Thr Thr Ile Tyr Ala Glu Pro Val Pro Glu Asn
35 40 45

Asn Ala Leu Asn Thr Gln Thr Gln Pro Lys Ala His Thr Thr Gly Asp
50 55 60

Arg Gly Lys Glu Ala Asn Gln Ala Pro Pro Glu Glu Gln Asp Lys Val
65 70 75 80

Pro His Thr Ala Gln Arg Ala Ala Trp Lys Ser Pro Glu Lys Glu Lys
85 90 95

Thr Met Val Asn Thr Leu Ser Pro Arg
100 105

<210> 135

<211> 669

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 106-774 de SEQ ID NO:5

<220>

<221> CDS

<222> (1)..(669)

<400> 135

aag gag cct caa aca aag cct tcc agg cat caa cgc aca gag aac att	48
Lys Glu Pro Gln Thr Lys Pro Ser Arg His Gln Arg Thr Glu Asn Ile	
1 5 10 15	
aaa gaa agg tct cta cag tcc ctg gca aag cct aag tcc cag gca ccc	96
Lys Glu Arg Ser Leu Gln Ser Leu Ala Lys Pro Lys Ser Gln Ala Pro	
20 25 30	
aca agg gca agg agg aca acc atc tat gca gag cca gtg cca gag aac	144
Thr Arg Ala Arg Arg Thr Thr Ile Tyr Ala Glu Pro Val Pro Glu Asn	
35 40 45	
aat gcc ctc aac aca caa acc cag ccc aag gcc cac acc acc gga gac	192
Asn Ala Leu Asn Thr Gln Thr Gln Pro Lys Ala His Thr Thr Gly Asp	
50 55 60	
aga gga aag gag gcc aac cag gca ccg ccg gag gag cag gac aag gtg	240
Arg Gly Lys Glu Ala Asn Gln Ala Pro Pro Glu Glu Gln Asp Lys Val	
65 70 75 80	
ccc cac aca gca cag agg gca gca tgg aag agc cca gaa aaa gag aaa	288
Pro His Thr Ala Gln Arg Ala Ala Trp Lys Ser Pro Glu Lys Glu Lys	
85 90 95	
acc atg gtg aac aca ctg tca ccc aga ggg caa gat gca ggg atg gcc	336
Thr Met Val Asn Thr Leu Ser Pro Arg Gly Gln Asp Ala Gly Met Ala	
100 105 110	
tct ggc agg aca gag gca caa tca tgg aag agc cag gac aca aag acg	384
Ser Gly Arg Thr Glu Ala Gln Ser Trp Lys Ser Gln Asp Thr Lys Thr	
115 120 125	
acc caa gga aat ggg ggc cag acc agg aag ctg acg gcc tcc agg acg	432
Thr Gln Gly Asn Gly Gly Gln Thr Arg Lys Leu Thr Ala Ser Arg Thr	
130 135 140	
gtg tca gag aag cac cag ggc aaa gcg gca acc aca gcc aag acg ctc	480
Val Ser Glu Lys His Gln Gly Lys Ala Ala Thr Thr Ala Lys Thr Leu	
145 150 155 160	
att ccc aaa agt cag cac aga atg ctg gct ccc aca gga gca gtg tca	528
Ile Pro Lys Ser Gln His Arg Met Leu Ala Pro Thr Gly Ala Val Ser	
165 170 175	
aca agg acg aga cag aaa gga gtg acc aca gca gtc atc cca cct aag	576
Thr Arg Thr Arg Gln Lys Gly Val Thr Thr Ala Val Ile Pro Pro Lys	
180 185 190	
gag aag aaa cct cag gcc acc cca ccc cct gcc cct ttc cag agc ccc	624
Glu Lys Lys Pro Gln Ala Thr Pro Pro Pro Ala Pro Phe Gln Ser Pro	
195 200 205	
acg acg cag aga aac caa aga ctg aag gcc gcc aac ttc aaa tct	669
Thr Thr Gln Arg Asn Gln Arg Leu Lys Ala Ala Asn Phe Lys Ser	
210 215 220	

<211> 223

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 136

Lys Glu Pro Gln Thr Lys Pro Ser Arg His Gln Arg Thr Glu Asn Ile
 1 5 10 15

Lys Glu Arg Ser Leu Gln Ser Leu Ala Lys Pro Lys Ser Gln Ala Pro
 20 25 30

Thr Arg Ala Arg Arg Thr Thr Ile Tyr Ala Glu Pro Val Pro Glu Asn
 35 40 45

Asn Ala Leu Asn Thr Gln Thr Gln Pro Lys Ala His Thr Thr Gly Asp
 50 55 60

Arg Gly Lys Glu Ala Asn Gln Ala Pro Pro Glu Glu Gln Asp Lys Val
 65 70 75 80

Pro His Thr Ala Gln Arg Ala Ala Trp Lys Ser Pro Glu Lys Glu Lys
 85 90 95

Thr Met Val Asn Thr Leu Ser Pro Arg Gly Gln Asp Ala Gly Met Ala
 100 105 110

Ser Gly Arg Thr Glu Ala Gln Ser Trp Lys Ser Gln Asp Thr Lys Thr
 115 120 125

Thr Gln Gly Asn Gly Gly Gln Thr Arg Lys Leu Thr Ala Ser Arg Thr
 130 135 140

Val Ser Glu Lys His Gln Gly Lys Ala Ala Thr Thr Ala Lys Thr Leu
 145 150 155 160

Ile Pro Lys Ser Gln His Arg Met Leu Ala Pro Thr Gly Ala Val Ser
 165 170 175

Thr Arg Thr Arg Gln Lys Gly Val Thr Thr Ala Val Ile Pro Pro Lys
 180 185 190

Glu Lys Lys Pro Gln Ala Thr Pro Pro Pro Ala Pro Phe Gln Ser Pro
 195 200 205

Thr Thr Gln Arg Asn Gln Arg Leu Lys Ala Ala Asn Phe Lys Ser
 210 215 220

<210> 137

<211> 54

<212> ADN

<213> Artificial

 $\langle 220 \rangle$

<223> Nucleótidos 85-138 de SEQ ID NO:21

 $\langle 220 \rangle$

<221> CDS

<222> (1) . . (54)

<400> 137

aag cta cac tta ctc cag tgg gag gag gac tcc agt aag tat agt cac 48
Lys Leu His Leu Leu Gln Trp Glu Glu Asp Ser Ser Lys Tyr Ser His
1 5 10 15

Lys Leu His Leu Leu Gln Trp Glu Glu Asp Ser Ser Lys Tyr Ser His
1 5 10 15

1

5

10

15

tct agc 54
Ser Ser

Ser Ser

<210> 138

<211> 18

<212> PRT

<213> Artificial

 $\langle 220 \rangle$

<223> Construção Sintética

<400> 138

Lys Leu His Leu Leu Gln Trp Glu Glu Asp Ser Ser Lys Tyr Ser His
1 5 10 15

1

5

10

15

Ser Ser

<210> 139

<211> 54

<212> ADN

<213> Artificial

 $\langle 220 \rangle$

<223> Nucleótidos 85-138 de SEQ ID NO:29

 $\langle 220 \rangle$

<221> CDS

<222> (1) . . (54)

<400> 139

aag	cta	cac	tta	ctc	caa	tgg	gaa	gac	tcc	aat	tca	ctg	att	ctt	tcc	48
Lys	Leu	His	Leu	Leu	Gln	Trp	Glu	Asp	Ser	Asn	Ser	Leu	Ile	Leu	Ser	
1				5				10						15		
ctt gac																54
Leu Asp																

<210> 140

<211> 18

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 140

Lys	Leu	His	Leu	Leu	Gln	Trp	Glu	Asp	Ser	Asn	Ser	Leu	Ile	Leu	Ser
1				5				10						15	
Leu Asp															

<210> 141

<211> 168

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 70-237 de SEQ ID NO:33

<220>

<221> CDS

<222> (1)..(168)

<400> 141

```

gac atc tca gag atc gaa gaa gaa atc ggg aat tcg gga ggc aga ggt      48
Asp Ile Ser Glu Ile Glu Glu Glu Ile Gly Asn Ser Gly Gly Arg Gly
1           5           10           15

aca atc aga tca gct gtg aac agc tta cat agc aaa tct aat aga gct      96
Thr Ile Arg Ser Ala Val Asn Ser Leu His Ser Lys Ser Asn Arg Ala
           20           25           30

gaa gtt gta ata aac ggc tcc tca tca cca gct gtt gtt gac aga agt      144
Glu Val Val Ile Asn Gly Ser Ser Ser Pro Ala Val Val Asp Arg Ser
           35           40           45

aat gaa agc atc aag cac aac atc      168
Asn Glu Ser Ile Lys His Asn Ile
           50           55

```

<210> 142

<211> 56

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 142

```

Asp Ile Ser Glu Ile Glu Glu Glu Ile Gly Asn Ser Gly Gly Arg Gly
1           5           10           15

Thr Ile Arg Ser Ala Val Asn Ser Leu His Ser Lys Ser Asn Arg Ala
           20           25           30

Glu Val Val Ile Asn Gly Ser Ser Ser Pro Ala Val Val Asp Arg Ser
           35           40           45

Asn Glu Ser Ile Lys His Asn Ile
           50           55

```

<210> 143

<211> 141

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 61-201 de SEQ ID NO:37

<220>

<221> CDS

<222> (1)..(141)

<400> 143

```

aag aca aaa gaa ata gca aga act gag gag cac cag gag acg caa ctc      48
Lys Thr Lys Glu Ile Ala Arg Thr Glu Glu His Gln Glu Thr Gln Leu
1          5          10          15

atc gga gat ggt gaa ttg tct ttg agt cgg tca ctt gtc aat agc tct      96
Ile Gly Asp Gly Glu Leu Ser Leu Ser Arg Ser Leu Val Asn Ser Ser
          20          25          30

gat aaa atc att cga aag gct ggc tct tca atc ttc cag cac aat      141
Asp Lys Ile Ile Arg Lys Ala Gly Ser Ser Ile Phe Gln His Asn
          35          40          45

```

<210> 144

<211> 47

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 144

```

Lys Thr Lys Glu Ile Ala Arg Thr Glu Glu His Gln Glu Thr Gln Leu
1          5          10          15

Ile Gly Asp Gly Glu Leu Ser Leu Ser Arg Ser Leu Val Asn Ser Ser
          20          25          30

Asp Lys Ile Ile Arg Lys Ala Gly Ser Ser Ile Phe Gln His Asn
          35          40          45

```

<210> 145

<211> 222

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-222 de SEQ ID NO:5

<220>

<221> CDS

<222> (1)..(222)

<400> 145

atg agg tcc tgc ctg tgg aga tgc agg cac ctg agc caa ggc gtc cag	48
Met Arg Ser Cys Leu Trp Arg Cys Arg His Leu Ser Gln Gly Val Gln	
1 5 10 15	
tggtcc ttgtctt ctgtgt gctgtc ctgtgt tctttt ctcttc gcc ttgtcc	96
Trp Ser Leu Leu Leu Ala Val Leu Val Phe Phe Leu Phe Ala Leu Pro	
20 25 30	
tctttt att aag gag cct caa aca aag cct tcc agg cat caa cgc aca	144
Ser Phe Ile Lys Glu Pro Gln Thr Lys Pro Ser Arg His Gln Arg Thr	
35 40 45	
gag aac att aaa gaa agg tct cta cag tcc ctg gca aag cct aag tcc	192
Glu Asn Ile Lys Glu Arg Ser Leu Gln Ser Leu Ala Lys Pro Lys Ser	
50 55 60	
cag gca ccc aca agg gca agg agg aca acc	222
Gln Ala Pro Thr Arg Ala Arg Arg Thr Thr	
65 70	

<210> 146

<211> 74

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 146

Met Arg Ser Cys Leu Trp Arg Cys Arg His Leu Ser Gln Gly Val Gln
1 5 10 15

Trp Ser Leu Leu Leu Ala Val Leu Val Phe Phe Leu Phe Ala Leu Pro
20 25 30

Ser Phe Ile Lys Glu Pro Gln Thr Lys Pro Ser Arg His Gln Arg Thr
35 40 45

Glu Asn Ile Lys Glu Arg Ser Leu Gln Ser Leu Ala Lys Pro Lys Ser
50 55 60

Gln Ala Pro Thr Arg Ala Arg Arg Thr Thr
65 70

<210> 147

<211> 420

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-420 de SEQ ID NO:5

<220>

<221> CDS

<222> (1)..(420)

<400> 147

atg agg tcc tgc ctg tgg aga tgc agg cac ctg agc caa ggc gtc cag	48
Met Arg Ser Cys Leu Trp Arg Cys Arg His Leu Ser Gln Gly Val Gln	
1 5 10 15	
tgg tcc ttg ctt ctg gct gtc ctg gtc ttc ttt ctc ttc gcc ttg ccc	96
Trp Ser Leu Leu Leu Ala Val Leu Val Phe Phe Leu Phe Ala Leu Pro	
20 25 30	
tct ttt att aag gag cct caa aca aag cct tcc agg cat caa cgc aca	144
Ser Phe Ile Lys Glu Pro Gln Thr Lys Pro Ser Arg His Gln Arg Thr	
35 40 45	
gag aac att aaa gaa agg tct cta cag tcc ctg gca aag cct aag tcc	192
Glu Asn Ile Lys Glu Arg Ser Leu Gln Ser Leu Ala Lys Pro Lys Ser	
50 55 60	
cag gca ccc aca agg gca agg agg aca acc atc tat gca gag cca gtg	240
Gln Ala Pro Thr Arg Ala Arg Arg Thr Thr Ile Tyr Ala Glu Pro Val	
65 70 75 80	
cca gag aac aat gcc ctc aac aca caa acc cag ccc aag gcc cac acc	288
Pro Glu Asn Asn Ala Leu Asn Thr Gln Thr Gln Pro Lys Ala His Thr	
85 90 95	
acc gga gac aga gga aag gag gcc aac cag gca ccg ccg gag gag cag	336
Thr Gly Asp Arg Gly Lys Glu Ala Asn Gln Ala Pro Pro Glu Glu Gln	
100 105 110	
gac aag gtg ccc cac aca gca cag agg gca gca tgg aag agc cca gaa	384
Asp Lys Val Pro His Thr Ala Gln Arg Ala Ala Trp Lys Ser Pro Glu	
115 120 125	
aaa gag aaa acc atg gtg aac aca ctg tca ccc aga	420
Lys Glu Lys Thr Met Val Asn Thr Leu Ser Pro Arg	
130 135 140	

<210> 148

<211> 140

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 148

Met Arg Ser Cys Leu Trp Arg Cys Arg His Leu Ser Gln Gly Val Gln
1 5 10 15

Trp Ser Leu Leu Leu Ala Val Leu Val Phe Phe Leu Phe Ala Leu Pro
20 25 30

Ser Phe Ile Lys Glu Pro Gln Thr Lys Pro Ser Arg His Gln Arg Thr
35 40 45

Glu Asn Ile Lys Glu Arg Ser Leu Gln Ser Leu Ala Lys Pro Lys Ser
50 55 60

Gln Ala Pro Thr Arg Ala Arg Arg Thr Thr Ile Tyr Ala Glu Pro Val
65 70 75 80

Pro Glu Asn Asn Ala Leu Asn Thr Gln Thr Gln Pro Lys Ala His Thr
85 90 95

Thr Gly Asp Arg Gly Lys Glu Ala Asn Gln Ala Pro Pro Glu Glu Gln
100 105 110

Asp Lys Val Pro His Thr Ala Gln Arg Ala Ala Trp Lys Ser Pro Glu
115 120 125

Lys Glu Lys Thr Met Val Asn Thr Leu Ser Pro Arg
130 135 140

<210> 149

<211> 774

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-774 de SEQ ID NO:5

<220>

<221> CDS

<222> (1)..(774)

<400> 149

```

      . . . . .
atg agg tcc tgc ctg tgg aga tgc agg cac ctg agc caa ggc gtc cag      48
Met Arg Ser Cys Leu Trp Arg Cys Arg His Leu Ser Gln Gly Val Gln
1          5          10          15

tgg tcc ttg ctt ctg gct gtc ctg gtc ttc ttt ctc ttc gcc ttg ccc      96
Trp Ser Leu Leu Leu Ala Val Leu Val Phe Phe Leu Phe Ala Leu Pro
          20          25          30

tct ttt att aag gag cct caa aca aag cct tcc agg cat caa cgc aca      144
Ser Phe Ile Lys Glu Pro Gln Thr Lys Pro Ser Arg His Gln Arg Thr

```

35	40	45	
gag aac att aaa gaa agg tct cta cag tcc ctg gca aag cct aag tcc			192
Glu Asn Ile Lys Glu Arg Ser Leu Gln Ser Leu Ala Lys Pro Lys Ser			
50	55	60	
cag gca ccc aca agg gca agg agg aca acc atc tat gca gag cca gtg			240
Gln Ala Pro Thr Arg Ala Arg Arg Thr Thr Ile Tyr Ala Glu Pro Val			
65	70	75	80
cca gag aac aat gcc ctc aac aca caa acc cag ccc aag gcc cac acc			288
Pro Glu Asn Asn Ala Leu Asn Thr Gln Thr Gln Pro Lys Ala His Thr			
85	90	95	
acc gga gac aga gga aag gag gcc aac cag gca ccg ccg gag gag cag			336
Thr Gly Asp Arg Gly Lys Glu Ala Asn Gln Ala Pro Pro Glu Glu Gln			
100	105	110	
gac aag gtg ccc cac aca gca cag agg gca gca tgg aag agc cca gaa			384
Asp Lys Val Pro His Thr Ala Gln Arg Ala Ala Trp Lys Ser Pro Glu			
115	120	125	
aaa gag aaa acc atg gtg aac aca ctg tca ccc aga ggg caa gat gca			432
Lys Glu Lys Thr Met Val Asn Thr Leu Ser Pro Arg Gly Gln Asp Ala			
130	135	140	
ggg atg gcc tct ggc agg aca gag gca caa tca tgg aag agc cag gac			480
Gly Met Ala Ser Gly Arg Thr Glu Ala Gln Ser Trp Lys Ser Gln Asp			
145	150	155	160
aca aag acg acc caa gga aat ggg ggc cag acc agg aag ctg acg gcc			528
Thr Lys Thr Thr Gln Gly Asn Gly Gly Gln Thr Arg Lys Leu Thr Ala			
165	170	175	
tcc agg acg gtg tca gag aag cac cag gcc aaa gcg gca acc aca gcc			576
Ser Arg Thr Val Ser Glu Lys His Gln Gly Lys Ala Ala Thr Thr Ala			
180	185	190	
aag acg ctc att ccc aaa agt cag cac aga atg ctg gct ccc aca gga			624
Lys Thr Leu Ile Pro Lys Ser Gln His Arg Met Leu Ala Pro Thr Gly			
195	200	205	
gca gtg tca aca agg acg aga cag aaa gga gtg acc aca gca gtc atc			672
Ala Val Ser Thr Arg Thr Arg Gln Lys Gly Val Thr Thr Ala Val Ile			
210	215	220	
cca cct aag gag aag aaa cct cag gcc acc cca ccc cct gcc cct ttc			720
Pro Pro Lys Glu Lys Lys Pro Gln Ala Thr Pro Pro Pro Ala Pro Phe			
225	230	235	240
cag agc ccc acg acg cag aga aac caa aga ctg aag gcc gcc aac ttc			768
Gln Ser Pro Thr Thr Gln Arg Asn Gln Arg Leu Lys Ala Ala Asn Phe			
245	250	255	
aaa tct			774
Lys Ser			

<210> 150

<211> 258

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 150

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

210	215	220
Pro Pro Lys Glu Lys Lys Pro Gln Ala Thr Pro Pro Pro Ala Pro Phe		
225	230	235 240
Gln Ser Pro Thr Thr Gln Arg Asn Gln Arg Leu Lys Ala Ala Asn Phe		
	245	250 255

Lys Ser

<210> 151

<211> 138

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-138 de SEQ ID NO:21,

<220>

<221> CDS

<222> (1)..(138)

<400> 151

atg gga ctc ttg gta ttt gtg cgc aat ctg ctg cta gcc ctc tgc ctc	48
Met Gly Leu Leu Val Phe Val Arg Asn Leu Leu Leu Ala Leu Cys Leu	
1 5 10 15	
ttt ctg gta ctg gga ttt ttg tat tat tct gcg tgg aag cta cac tta	96
Phe Leu Val Leu Gly Phe Leu Tyr Tyr Ser Ala Trp Lys Leu His Leu	
20 25 30	
ctc cag tgg gag gag gac tcc agt aag tat agt cac tct agc	138
Leu Gln Trp Glu Glu Asp Ser Ser Lys Tyr Ser His Ser Ser	
35 40 45	

<210> 152

<211> 46

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 152

Met	Gly	Leu	Leu	Val	Phe	Val	Arg	Asn	Leu	Leu	Leu	Ala	Leu	Cys	Leu
1				5				10						15	

Phe	Leu	Val	Leu	Gly	Phe	Leu	Tyr	Tyr	Ser	Ala	Trp	Lys	Leu	His	Leu
			20				25						30		

Leu	Gln	Trp	Glu	Glu	Asp	Ser	Ser	Lys	Tyr	Ser	His	Ser	Ser

35

40

45

<210> 153

<211> 138

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-138 de SEQ ID NO:29

<220>

<221> CDS

<222> (1)..(138)

<400> 153

atg	gga	ctc	ttg	gta	ttt	gta	cgc	aac	ctg	ctg	cta	gcc	ctc	tgc	ctc	48
Met	Gly	Leu	Leu	Val	Phe	Val	Arg	Asn	Leu	Leu	Leu	Ala	Leu	Cys	Leu	
1				5				10						15		

ttt	ctg	gtc	ctg	gga	ttt	ttg	tat	tat	tct	gcc	tgg	aag	cta	cac	tta	96
Phe	Leu	Val	Leu	Gly	Phe	Leu	Tyr	Tyr	Ser	Ala	Trp	Lys	Leu	His	Leu	
			20				25						30			

ctc	caa	tgg	gaa	gac	tcc	aat	tca	ctg	att	ctt	tcc	ctt	gac	138
Leu	Gln	Trp	Glu	Asp	Ser	Asn	Ser	Leu	Ile	Leu	Ser	Leu	Asp	
			35				40					45		

<210> 154

<211> 46

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 154

Met	Gly	Leu	Leu	Val	Phe	Val	Arg	Asn	Leu	Leu	Leu	Ala	Leu	Cys	Leu
1				5					10					15	

Phe	Leu	Val	Leu	Gly	Phe	Leu	Tyr	Tyr	Ser	Ala	Trp	Lys	Leu	His	Leu
			20					25					30		

Leu	Gln	Trp	Glu	Asp	Ser	Asn	Ser	Leu	Ile	Leu	Ser	Leu	Asp
		35					40					45	

<210> 155

<211> 237

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-237 de SEQ ID NO:33

<220>

<221> CDS

<222> (1)..(237)

<400> 155

atg	cag	ctg	cag	ttc	cgg	agc	tgg	atg	ctg	gcc	gcg	ctc	acg	ctg	ctc	48
Met	Gln	Leu	Gln	Phe	Arg	Ser	Trp	Met	Leu	Ala	Ala	Leu	Thr	Leu	Leu	
1				5					10					15		
gtg	gtc	ttc	ctc	atc	ttc	gca	gac	atc	tca	gag	atc	gaa	gaa	gaa	atc	96
Val	Val	Phe	Leu	Ile	Phe	Ala	Asp	Ile	Ser	Glu	Ile	Glu	Glu	Glu	Ile	
			20					25					30			
ggg	aat	tcg	gga	ggc	aga	ggt	aca	atc	aga	tca	gct	gtg	aac	agc	tta	144
Gly	Asn	Ser	Gly	Gly	Arg	Gly	Thr	Ile	Arg	Ser	Ala	Val	Asn	Ser	Leu	
			35				40					45				
cat	agc	aaa	tct	aat	aga	gct	gaa	gtt	gta	ata	aac	ggc	tcc	tca	tca	192
His	Ser	Lys	Ser	Asn	Arg	Ala	Glu	Val	Val	Ile	Asn	Gly	Ser	Ser	Ser	
			50				55				60					
cca	gct	gtt	gtt	gac	aga	agt	aat	gaa	agc	atc	aag	cac	aac	atc		237
Pro	Ala	Val	Val	Asp	Arg	Ser	Asn	Glu	Ser	Ile	Lys	His	Asn	Ile		
65						70				75						

<210> 156

<211> 79

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 156

Met	Gln	Leu	Gln	Phe	Arg	Ser	Trp	Met	Leu	Ala	Ala	Leu	Thr	Leu	Leu
1				5					10					15	

Val	Val	Phe	Leu	Ile	Phe	Ala	Asp	Ile	Ser	Glu	Ile	Glu	Glu	Glu	Ile
			20					25					30		

Gly	Asn	Ser	Gly	Gly	Arg	Gly	Thr	Ile	Arg	Ser	Ala	Val	Asn	Ser	Leu
		35					40						45		

His	Ser	Lys	Ser	Asn	Arg	Ala	Glu	Val	Val	Ile	Asn	Gly	Ser	Ser	Ser
	50					55					60				

Pro	Ala	Val	Val	Asp	Arg	Ser	Asn	Glu	Ser	Ile	Lys	His	Asn	Ile
65					70					75				

<210> 157

<211> 201

<212> ADN

<213> Artificial

<220>

<223> Nucleótidos 1-201 de SEQ ID NO:37

<220>

<221> CDS

<222> (1)..(201)

<400> 157

```

---
atg cgc tcc att agg aag agg tgg acg atc tgc aca ata agt ctg ctc      48
Met Arg Ser Ile Arg Lys Arg Trp Thr Ile Cys Thr Ile Ser Leu Leu
1           5           10           15

ctg atc ttt tat aag aca aaa gaa ata gca aga act gag gag cac cag      96
Leu Ile Phe Tyr Lys Thr Lys Glu Ile Ala Arg Thr Glu Glu His Gln
          20           25           30

gag acg caa ctc atc gga gat ggt gaa ttg tct ttg agt cgg tca ctt      144
Glu Thr Gln Leu Ile Gly Asp Gly Glu Leu Ser Leu Ser Arg Ser Leu
          35           40           45

gtc aat agc tct gat aaa atc att cga aag gct ggc tct tca atc ttc      192
Val Asn Ser Ser Asp Lys Ile Ile Arg Lys Ala Gly Ser Ser Ile Phe
          50           55           60

cag cac aat
Gln His Asn
65

```

<210> 158

<211> 67

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 158

```

Met Arg Ser Ile Arg Lys Arg Trp Thr Ile Cys Thr Ile Ser Leu Leu
1           5           10           15

Leu Ile Phe Tyr Lys Thr Lys Glu Ile Ala Arg Thr Glu Glu His Gln
          20           25           30

Glu Thr Gln Leu Ile Gly Asp Gly Glu Leu Ser Leu Ser Arg Ser Leu
          35           40           45

Val Asn Ser Ser Asp Lys Ile Ile Arg Lys Ala Gly Ser Ser Ile Phe
          50           55           60

Gln His Asn
65

```

<210> 159

<211> 198

<212> ADN

Met Gly Leu Leu Val Phe Val Arg Asn Leu Leu Leu Ala Leu Cys Leu
1 5 10 15

Phe Leu Val Leu Gly Phe Leu Tyr Tyr Ser Ala Trp Glu Pro Gln Thr
20 25 30

Lys Pro Ser Arg His Gln Arg Thr Glu Asn Ile Lys Glu Arg Ser Leu
35 40 45

Gln Ser Leu Ala Lys Pro Lys Ser Gln Ala Pro Thr Arg Ala Arg Arg
50 55 60

Thr Thr
65

<210> 161

<211> 1179

<212> ADN

<213> Artificial

<220>

<223> Fusão da sequência nucleotídica SEQ ID NO:145 com a
sequência nucleotídica SEQ ID NO:43

<220>

<221> CDS

<222> (1)..(1179)

<400> 161

atg agg tcc tgc ctg tgg aga tgc agg cac ctg agc caa ggc gtc cag Met Arg Ser Cys Leu Trp Arg Cys Arg His Leu Ser Gln Gly Val Gln 1 5 10 15	48
tgg tcc ttg ctt ctg gct gtc ctg gtc ttc ttt ctc ttc gcc ttg ccc Trp Ser Leu Leu Leu Ala Val Leu Val Phe Phe Leu Phe Ala Leu Pro 20 25 30	96
tct ttt att aag gag cct caa aca aag cct tcc agg cat caa cgc aca Ser Phe Ile Lys Glu Pro Gln Thr Lys Pro Ser Arg His Gln Arg Thr 35 40 45	144
gag aac att aaa gaa agg tct cta cag tcc ctg gca aag cct aag tcc Glu Asn Ile Lys Glu Arg Ser Leu Gln Ser Leu Ala Lys Pro Lys Ser 50 55 60	192
cag gca ccc aca agg gca agg agg aca acc gga tcc gag gcc tcc ttc Gln Ala Pro Thr Arg Ala Arg Arg Thr Thr Gly Ser Glu Ala Ser Phe 65 70 75 80	240
cag gtg tgg aac aag gac agc tct tcc aaa aac ctt atc cct agg ctg Gln Val Trp Asn Lys Asp Ser Ser Ser Lys Asn Leu Ile Pro Arg Leu 85 90 95	288
caa aag atc tgg aag aat tac cta agc atg aac aag tac aaa gtg tcc Gln Lys Ile Trp Lys Asn Tyr Leu Ser Met Asn Lys Tyr Lys Val Ser 100 105 110	336
tac aag ggg cca gga cca ggc atc aag ttc agt gca gag gcc ctg cgc Tyr Lys Gly Pro Gly Pro Gly Ile Lys Phe Ser Ala Glu Ala Leu Arg 115 120 125	384
tgc cac ctc cgg gac cat gtg aat gta tcc atg gta gag gtc aca gat Cys His Leu Arg Asp His Val Asn Val Ser Met Val Glu Val Thr Asp 130 135 140	432
ttt ccc ttc aat acc tct gaa tgg gag ggt tat ctg ccc aag gag agc Phe Pro Phe Asn Thr Ser Glu Trp Glu Gly Tyr Leu Pro Lys Glu Ser 145 150 155 160	480
att agg acc aag gct ggg cct tgg ggc agg tgt gct gtt gtg tcg tca Ile Arg Thr Lys Ala Gly Pro Trp Gly Arg Cys Ala Val Val Ser Ser	528

165	170	175	
gcg gga tct ctg aag tcc tcc caa cta ggc aga gaa atc gat gat cat Ala Gly Ser Leu Lys Ser Ser Gln Leu Gly Arg Glu Ile Asp Asp His 180 185 190			576
gac gca gtc ctg agg ttt aat ggg gca ccc aca gcc aac ttc caa caa Asp Ala Val Leu Arg Phe Asn Gly Ala Pro Thr Ala Asn Phe Gln Gln 195 200 205			624
gat gtg ggc aca aaa act acc att cgc ctg atg aac tct cag ttg gtt Asp Val Gly Thr Lys Thr Thr Ile Arg Leu Met Asn Ser Gln Leu Val 210 215 220			672
acc aca gag aag cgc ttc ctc aaa gac agt ttg tac aat gaa gga atc Thr Thr Glu Lys Arg Phe Leu Lys Asp Ser Leu Tyr Asn Glu Gly Ile 225 230 235 240			720
cta att gta tgg gac cca tct gta tac cac tca gat atc cca aag tgg Leu Ile Val Trp Asp Pro Ser Val Tyr His Ser Asp Ile Pro Lys Trp 245 250 255			768
tac cag aat ccg gat tat aat ttc ttt aac aac tac aag act tat cgt Tyr Gln Asn Pro Asp Tyr Asn Phe Phe Asn Asn Tyr Lys Thr Tyr Arg 260 265 270			816
aag ctg cac ccc aat cag ccc ttt tac atc ctc aag ccc cag atg cct Lys Leu His Pro Asn Gln Pro Phe Tyr Ile Leu Lys Pro Gln Met Pro 275 280 285			864
tgg gag cta tgg gac att ctt caa gaa atc tcc cca gaa gag att cag Trp Glu Leu Trp Asp Ile Leu Gln Glu Ile Ser Pro Glu Glu Ile Gln 290 295 300			912
cca aac ccc cca tcc tct ggg atg ctt ggt atc atc atc atg atg acg Pro Asn Pro Pro Ser Ser Gly Met Leu Gly Ile Ile Ile Met Met Thr 305 310 315 320			960
ctg tgt gac cag gtg gat att tat gag ttc ctc cca tcc aag cgc aag Leu Cys Asp Gln Val Asp Ile Tyr Glu Phe Leu Pro Ser Lys Arg Lys 325 330 335			1008
act gac gtg tgc tac tac tac cag aag ttc ttc gat agt gcc tgc acg Thr Asp Val Cys Tyr Tyr Tyr Gln Lys Phe Phe Asp Ser Ala Cys Thr 340 345 350			1056
atg ggt gcc tac cac ccg ctg ctc tat gag aag aat ttg gtg aag cat Met Gly Ala Tyr His Pro Leu Leu Tyr Glu Lys Asn Leu Val Lys His 355 360 365			1104
ctc aac cag ggc aca gat gag gac atc tac ctg ctt gga aaa gcc aca Leu Asn Gln Gly Thr Asp Glu Asp Ile Tyr Leu Leu Gly Lys Ala Thr 370 375 380			1152
ctg cct ggc ttc cgg acc att cac tgc Leu Pro Gly Phe Arg Thr Ile His Cys 385 390			1179

<211> 393

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 162

Met Arg Ser Cys Leu Trp Arg Cys Arg His Leu Ser Gln Gly Val Gln
1 5 10 15

Trp Ser Leu Leu Leu Ala Val Leu Val Phe Phe Leu Phe Ala Leu Pro
20 25 30

Ser Phe Ile Lys Glu Pro Gln Thr Lys Pro Ser Arg His Gln Arg Thr
35 40 45

Glu Asn Ile Lys Glu Arg Ser Leu Gln Ser Leu Ala Lys Pro Lys Ser
50 55 60

Gln Ala Pro Thr Arg Ala Arg Arg Thr Thr Gly Ser Glu Ala Ser Phe
65 70 75 80

Gln Val Trp Asn Lys Asp Ser Ser Ser Lys Asn Leu Ile Pro Arg Leu
85 90 95

Gln Lys Ile Trp Lys Asn Tyr Leu Ser Met Asn Lys Tyr Lys Val Ser
100 105 110

Tyr Lys Gly Pro Gly Pro Gly Ile Lys Phe Ser Ala Glu Ala Leu Arg
115 120 125

Cys His Leu Arg Asp His Val Asn Val Ser Met Val Glu Val Thr Asp
130 135 140

Phe Pro Phe Asn Thr Ser Glu Trp Glu Gly Tyr Leu Pro Lys Glu Ser
145 150 155 160

Ile Arg Thr Lys Ala Gly Pro Trp Gly Arg Cys Ala Val Val Ser Ser
165 170 175

Ala Gly Ser Leu Lys Ser Ser Gln Leu Gly Arg Glu Ile Asp Asp His
180 185 190

Asp Ala Val Leu Arg Phe Asn Gly Ala Pro Thr Ala Asn Phe Gln Gln
195 200 205

Asp Val Gly Thr Lys Thr Thr Ile Arg Leu Met Asn Ser Gln Leu Val

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

<210> 163

<211> 1377

<212> ADN

<213> Artificial

<220>

<223> Fusão da sequência nucleotídica SEQ ID NO:147 com a
sequência nucleotídica SEQ ID NO:43

<220>

<221> CDS

<222> (1) .. (1377)

<400> 163

atg agg tcc tgc ctg tgg aga tgc agg cac ctg agc caa ggc gtc cag Met Arg Ser Cys Leu Trp Arg Cys Arg His Leu Ser Gln Gly Val Gln 1 5 10 15 48
tgg tcc ttg ctt ctg gct gtc ctg gtc ttc ttt ctc ttc gcc ttg ccc Trp Ser Leu Leu Leu Ala Val Leu Val Phe Phe Leu Phe Ala Leu Pro 20 25 30 96
tct ttt att aag gag cct caa aca aag cct tcc agg cat caa cgc aca Ser Phe Ile Lys Glu Pro Gln Thr Lys Pro Ser Arg His Gln Arg Thr 35 40 45 144
gag aac att aaa gaa agg tct cta cag tcc ctg gca aag cct aag tcc Glu Asn Ile Lys Glu Arg Ser Leu Gln Ser Leu Ala Lys Pro Lys Ser 50 55 60 192
cag gca ccc aca agg gca agg agg aca acc atc tat gca gag cca gtg Gln Ala Pro Thr Arg Ala Arg Arg Thr Thr Ile Tyr Ala Glu Pro Val 65 70 75 80 240
cca gag aac aat gcc ctc aac aca caa acc cag ccc aag gcc cac acc Pro Glu Asn Asn Ala Leu Asn Thr Gln Thr Gln Pro Lys Ala His Thr 85 90 95 288
acc gga gac aga gga aag gag gcc aac cag gca ccg ccg gag gag cag Thr Gly Asp Arg Gly Lys Glu Ala Asn Gln Ala Pro Pro Glu Glu Gln 100 105 110 336
gac aag gtg ccc cac aca gca cag agg gca gca tgg aag agc cca gaa Asp Lys Val Pro His Thr Ala Gln Arg Ala Ala Trp Lys Ser Pro Glu 115 120 125 384
aaa gag aaa acc atg gtg aac aca ctg tca ccc aga gga tcc gag gcc Lys Glu Lys Thr Met Val Asn Thr Leu Ser Pro Arg Gly Ser Glu Ala 130 135 140 432
tcc ttc cag gtg tgg aac aag gac agc tct tcc aaa aac ctt atc cct Ser Phe Gln Val Trp Asn Lys Asp Ser Ser Ser Lys Asn Leu Ile Pro 145 150 155 160 480
agg ctg caa aag atc tgg aag aat tac cta agc atg aac aag tac aaa Arg Leu Gln Lys Ile Trp Lys Asn Tyr Leu Ser Met Asn Lys Tyr Lys 165 170 175 528
gtg tcc tac aag ggg cca gga cca ggc atc aag ttc agt gca gag gcc Val Ser Tyr Lys Gly Pro Gly Pro Gly Ile Lys Phe Ser Ala Glu Ala 180 185 190 576
ctg cgc tgc cac ctc cgg gac cat gtg aat gta tcc atg gta gag gtc Leu Arg Cys His Leu Arg Asp His Val Asn Val Ser Met Val Glu Val 195 200 205 624
aca gat ttt ccc ttc aat acc tct gaa tgg gag ggt tat ctg ccc aag Thr Asp Phe Pro Phe Asn Thr Ser Glu Trp Glu Gly Tyr Leu Pro Lys 210 215 220 672
gag agc att agg acc aag gct ggg cct tgg ggc agg tgt gct gtt gtg Glu Ser Ile Arg Thr Lys Ala Gly Pro Trp Gly Arg Cys Ala Val Val 225 230 235 240 720

tcg tca gcg gga tct ctg aag tcc tcc caa cta ggc aga gaa atc gat Ser Ser Ala Gly Ser Leu Lys Ser Ser Gln Leu Gly Arg Glu Ile Asp 245 250 255	768
gat cat gac gca gtc ctg agg ttt aat ggg gca ccc aca gcc aac ttc Asp His Asp Ala Val Leu Arg Phe Asn Gly Ala Pro Thr Ala Asn Phe 260 265 270	816
caa caa gat gtg ggc aca aaa act acc att cgc ctg atg aac tct cag Gln Gln Asp Val Gly Thr Lys Thr Thr Ile Arg Leu Met Asn Ser Gln 275 280 285	864
ttg gtt acc aca gag aag cgc ttc ctc aaa gac agt ttg tac aat gaa Leu Val Thr Thr Glu Lys Arg Phe Leu Lys Asp Ser Leu Tyr Asn Glu 290 295 300	912
gga atc cta att gta tgg gac cca tct gta tac cac tca gat atc cca Gly Ile Leu Ile Val Trp Asp Pro Ser Val Tyr His Ser Asp Ile Pro 305 310 315 320	960
aag tgg tac cag aat ccg gat tat aat ttc ttt aac aac tac aag act Lys Trp Tyr Gln Asn Pro Asp Tyr Asn Phe Phe Asn Asn Tyr Lys Thr 325 330 335	1008
tat cgt aag ctg cac ccc aat cag ccc ttt tac atc ctc aag ccc cag Tyr Arg Lys Leu His Pro Asn Gln Pro Phe Tyr Ile Leu Lys Pro Gln 340 345 350	1056
atg cct tgg gag cta tgg gac att ctt caa gaa atc tcc cca gaa gag Met Pro Trp Glu Leu Trp Asp Ile Leu Gln Glu Ile Ser Pro Glu Glu 355 360 365	1104
att cag cca aac ccc cca tcc tct ggg atg ctt ggt atc atc atc atg Ile Gln Pro Asn Pro Pro Ser Ser Gly Met Leu Gly Ile Ile Ile Met 370 375 380	1152
atg acg ctg tgt gac cag gtg gat att tat gag ttc ctc cca tcc aag Met Thr Leu Cys Asp Gln Val Asp Ile Tyr Glu Phe Leu Pro Ser Lys 385 390 395 400	1200
cgc aag act gac gtg tgc tac tac tac cag aag ttc ttc gat agt gcc Arg Lys Thr Asp Val Cys Tyr Tyr Tyr Gln Lys Phe Phe Asp Ser Ala 405 410 415	1248
tgc acg atg ggt gcc tac cac ccg ctg ctc tat gag aag aat ttg gtg Cys Thr Met Gly Ala Tyr His Pro Leu Leu Tyr Glu Lys Asn Leu Val 420 425 430	1296
aag cat ctc aac cag ggc aca gat gag gac atc tac ctg ctt gga aaa Lys His Leu Asn Gln Gly Thr Asp Glu Asp Ile Tyr Leu Leu Gly Lys 435 440 445	1344
gcc aca ctg cct ggc ttc cgg acc att cac tgc Ala Thr Leu Pro Gly Phe Arg Thr Ile His Cys 450 455	1377

<210> 164

<211> 459

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 164

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

Glu Ser Ile Arg Thr Lys Ala Gly Pro Trp Gly Arg Cys Ala Val Val
225 230 235 240

Ser Ser Ala Gly Ser Leu Lys Ser Ser Gln Leu Gly Arg Glu Ile Asp
245 250 255

Asp His Asp Ala Val Leu Arg Phe Asn Gly Ala Pro Thr Ala Asn Phe
260 265 270

Gln Gln Asp Val Gly Thr Lys Thr Thr Ile Arg Leu Met Asn Ser Gln
275 280 285

Leu Val Thr Thr Glu Lys Arg Phe Leu Lys Asp Ser Leu Tyr Asn Glu
290 295 300

Gly Ile Leu Ile Val Trp Asp Pro Ser Val Tyr His Ser Asp Ile Pro
305 310 315 320

Lys Trp Tyr Gln Asn Pro Asp Tyr Asn Phe Phe Asn Asn Tyr Lys Thr
325 330 335

Tyr Arg Lys Leu His Pro Asn Gln Pro Phe Tyr Ile Leu Lys Pro Gln
340 345 350

Met Pro Trp Glu Leu Trp Asp Ile Leu Gln Glu Ile Ser Pro Glu Glu
355 360 365

Ile Gln Pro Asn Pro Pro Ser Ser Gly Met Leu Gly Ile Ile Ile Met
370 375 380

Met Thr Leu Cys Asp Gln Val Asp Ile Tyr Glu Phe Leu Pro Ser Lys
385 390 395 400

Arg Lys Thr Asp Val Cys Tyr Tyr Tyr Gln Lys Phe Phe Asp Ser Ala
405 410 415

Cys Thr Met Gly Ala Tyr His Pro Leu Leu Tyr Glu Lys Asn Leu Val
420 425 430

Lys His Leu Asn Gln Gly Thr Asp Glu Asp Ile Tyr Leu Leu Gly Lys
435 440 445

Ala Thr Leu Pro Gly Phe Arg Thr Ile His Cys
450 455

<210> 165

<211> 1731

<212> ADN

<213> Artificial

<220>

<223> Fusão da sequência nucleotídica SEQ ID NO:149 com a
sequência nucleotídica SEQ ID NO:43

<220>

<221> CDS

<222> (1) .. (1731)

<400> 165

atg agg tcc tgc ctg tgg aga tgc agg cac ctg agc caa ggc gtc cag Met Arg Ser Cys Leu Trp Arg Cys Arg His Leu Ser Gln Gly Val Gln 1 5 10 15	48
tgg tcc ttg ctt ctg gct gtc ctg gtc ttc ttt ctc ttc gcc ttg ccc Trp Ser Leu Leu Leu Ala Val Leu Val Phe Phe Leu Phe Ala Leu Pro 20 25 30	96
tct ttt att aag gag cct caa aca aag cct tcc agg cat caa cgc aca Ser Phe Ile Lys Glu Pro Gln Thr Lys Pro Ser Arg His Gln Arg Thr 35 40 45	144
gag aac att aaa gaa agg tct cta cag tcc ctg gca aag cct aag tcc Glu Asn Ile Lys Glu Arg Ser Leu Gln Ser Leu Ala Lys Pro Lys Ser 50 55 60	192
cag gca ccc aca agg gca agg agg aca acc atc tat gca gag cca gtg Gln Ala Pro Thr Arg Ala Arg Arg Thr Thr Ile Tyr Ala Glu Pro Val 65 70 75 80	240
cca gag aac aat gcc ctc aac aca caa acc cag ccc aag gcc cac acc Pro Glu Asn Asn Ala Leu Asn Thr Gln Thr Gln Pro Lys Ala His Thr 85 90 95	288
acc gga gac aga gga aag gag gcc aac cag gca ccg ccg gag gag cag Thr Gly Asp Arg Gly Lys Glu Ala Asn Gln Ala Pro Pro Glu Glu Gln 100 105 110	336
gac aag gtg ccc cac aca gca cag agg gca gca tgg aag agc cca gaa Asp Lys Val Pro His Thr Ala Gln Arg Ala Ala Trp Lys Ser Pro Glu 115 120 125	384
aaa gag aaa acc atg gtg aac aca ctg tca ccc aga ggg caa gat gca Lys Glu Lys Thr Met Val Asn Thr Leu Ser Pro Arg Gly Gln Asp Ala 130 135 140	432
ggg atg gcc tct ggc agg aca gag gca caa tca tgg aag agc cag gac Gly Met Ala Ser Gly Arg Thr Glu Ala Gln Ser Trp Lys Ser Gln Asp 145 150 155 160	480
aca aag acg acc caa gga aat ggg ggc cag acc agg aag ctg acg gcc Thr Lys Thr Thr Gln Gly Asn Gly Gly Gln Thr Arg Lys Leu Thr Ala 165 170 175	528
tcc agg acg gtg tca gag aag cac cag ggc aaa gcg gca acc aca gcc Ser Arg Thr Val Ser Glu Lys His Gln Gly Lys Ala Ala Thr Thr Ala	576

180	185	190	
aag acg ctc att ccc aaa agt cag cac aga atg ctg gct ccc aca gga Lys Thr Leu Ile Pro Lys Ser Gln His Arg Met Leu Ala Pro Thr Gly 195 200 205			624
gca gtg tca aca agg acg aga cag aaa gga gtg acc aca gca gtc atc Ala Val Ser Thr Arg Thr Arg Gln Lys Gly Val Thr Thr Ala Val Ile 210 215 220			672
cca cct aag gag aag aaa cct cag gcc acc cca ccc cct gcc cct ttc Pro Pro Lys Glu Lys Lys Pro Gln Ala Thr Pro Pro Pro Ala Pro Phe 225 230 235 240			720
cag agc ccc acg acg cag aga aac caa aga ctg aag gcc gcc aac ttc Gln Ser Pro Thr Thr Gln Arg Asn Gln Arg Leu Lys Ala Ala Asn Phe 245 250 255			768
aaa tct tct aga gag gcc tcc ttc cag gtg tgg aac aag gac agc tct Lys Ser Ser Arg Glu Ala Ser Phe Gln Val Trp Asn Lys Asp Ser Ser 260 265 270			816
tcc aaa aac ctt atc cct agg ctg caa aag atc tgg aag aat tac cta Ser Lys Asn Leu Ile Pro Arg Leu Gln Lys Ile Trp Lys Asn Tyr Leu 275 280 285			864
agc atg aac aag tac aaa gtg tcc tac aag ggg cca gga cca gcc atc Ser Met Asn Lys Tyr Lys Val Ser Tyr Lys Gly Pro Gly Pro Gly Ile 290 295 300			912
aag ttc agt gca gag gcc ctg cgc tgc cac ctc cgg gac cat gtg aat Lys Phe Ser Ala Glu Ala Leu Arg Cys His Leu Arg Asp His Val Asn 305 310 315 320			960
gta tcc atg gta gag gtc aca gat ttt ccc ttc aat acc tct gaa tgg Val Ser Met Val Glu Val Thr Asp Phe Pro Phe Asn Thr Ser Glu Trp 325 330 335			1008
gag ggt tat ctg ccc aag gag agc att agg acc aag gct ggg cct tgg Glu Gly Tyr Leu Pro Lys Glu Ser Ile Arg Thr Lys Ala Gly Pro Trp 340 345 350			1056
ggc agg tgt gct gtt gtg tgg tca gcg gga tct ctg aag tcc tcc caa Gly Arg Cys Ala Val Val Ser Ser Ala Gly Ser Leu Lys Ser Ser Gln 355 360 365			1104
cta ggc aga gaa atc gat gat cat gac gca gtc ctg agg ttt aat ggg Leu Gly Arg Glu Ile Asp Asp His Asp Ala Val Leu Arg Phe Asn Gly 370 375 380			1152
gca ccc aca gcc aac ttc caa caa gat gtg ggc aca aaa act acc att Ala Pro Thr Ala Asn Phe Gln Gln Asp Val Gly Thr Lys Thr Thr Ile 385 390 395 400			1200
cgc ctg atg aac tct cag ttg gtt acc aca gag aag cgc ttc ctc aaa Arg Leu Met Asn Ser Gln Leu Val Thr Thr Glu Lys Arg Phe Leu Lys 405 410 415			1248
gac agt ttg tac aat gaa gga atc cta att gta tgg gac cca tct gta Asp Ser Leu Tyr Asn Glu Gly Ile Leu Ile Val Trp Asp Pro Ser Val 420 425 430			1296

tac cac tca gat atc cca aag tgg tac cag aat ccg gat tat aat ttc	1344
Tyr His Ser Asp Ile Pro Lys Trp Tyr Gln Asn Pro Asp Tyr Asn Phe	
435 440 445	
ttt aac aac tac aag act tat cgt aag ctg cac ccc aat cag ccc ttt	1392
Phe Asn Asn Tyr Lys Thr Tyr Arg Lys Leu His Pro Asn Gln Pro Phe	
450 455 460	
tac atc ctc aag ccc cag atg cct tgg gag cta tgg gac att ctt caa	1440
Tyr Ile Leu Lys Pro Gln Met Pro Trp Glu Leu Trp Asp Ile Leu Gln	
465 470 475 480	
gaa atc tcc cca gaa gag att cag cca aac ccc cca tcc tct ggg atg	1488
Glu Ile Ser Pro Glu Glu Ile Gln Pro Asn Pro Pro Ser Ser Gly Met	
485 490 495	
ctt ggt atc atc atc atg atg acg ctg tgt gac cag gtg gat att tat	1536
Leu Gly Ile Ile Ile Met Met Thr Leu Cys Asp Gln Val Asp Ile Tyr	
500 505 510	
gag ttc ctc cca tcc aag cgc aag act gac gtg tgc tac tac tac cag	1584
Glu Phe Leu Pro Ser Lys Arg Lys Thr Asp Val Cys Tyr Tyr Tyr Gln	
515 520 525	
aag ttc ttc gat agt gcc tgc acg atg ggt gcc tac cac ccg ctg ctc	1632
Lys Phe Phe Asp Ser Ala Cys Thr Met Gly Ala Tyr His Pro Leu Leu	
530 535 540	
tat gag aag aat ttg gtg aag cat ctc aac cag ggc aca gat gag gac	1680
Tyr Glu Lys Asn Leu Val Lys His Leu Asn Gln Gly Thr Asp Glu Asp	
545 550 555 560	
atc tac ctg ctt gga aaa gcc aca ctg cct ggc ttc cgg acc att cac	1728
Ile Tyr Leu Leu Gly Lys Ala Thr Leu Pro Gly Phe Arg Thr Ile His	
565 570 575	
tgc	1731
Cys	

<210> 166

<211> 577

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 166

Met Arg Ser Cys Leu Trp Arg Cys Arg His Leu Ser Gln Gly Val Gln
1 5 10 15

Trp Ser Leu Leu Leu Ala Val Leu Val Phe Phe Leu Phe Ala Leu Pro
20 25 30

Ser Phe Ile Lys Glu Pro Gln Thr Lys Pro Ser Arg His Gln Arg Thr

35	40	45
Glu Asn Ile Lys Glu Arg Ser Leu Gln Ser Leu Ala Lys Pro Lys Ser		
50	55	60
Gln Ala Pro Thr Arg Ala Arg Arg Thr Thr Ile Tyr Ala Glu Pro Val		
65	70	75
Pro Glu Asn Asn Ala Leu Asn Thr Gln Thr Gln Pro Lys Ala His Thr		
	85	90
Thr Gly Asp Arg Gly Lys Glu Ala Asn Gln Ala Pro Pro Glu Glu Gln		
	100	105
Asp Lys Val Pro His Thr Ala Gln Arg Ala Ala Trp Lys Ser Pro Glu		
	115	120
Lys Glu Lys Thr Met Val Asn Thr Leu Ser Pro Arg Gly Gln Asp Ala		
	130	135
Gly Met Ala Ser Gly Arg Thr Glu Ala Gln Ser Trp Lys Ser Gln Asp		
145	150	155
Thr Lys Thr Thr Gln Gly Asn Gly Gly Gln Thr Arg Lys Leu Thr Ala		
	165	170
Ser Arg Thr Val Ser Glu Lys His Gln Gly Lys Ala Ala Thr Thr Ala		
	180	185
Lys Thr Leu Ile Pro Lys Ser Gln His Arg Met Leu Ala Pro Thr Gly		
	195	200
Ala Val Ser Thr Arg Thr Arg Gln Lys Gly Val Thr Thr Ala Val Ile		
	210	215
Pro Pro Lys Glu Lys Lys Pro Gln Ala Thr Pro Pro Pro Ala Pro Phe		
225	230	235
Gln Ser Pro Thr Thr Gln Arg Asn Gln Arg Leu Lys Ala Ala Asn Phe		
	245	250
Lys Ser Ser Arg Glu Ala Ser Phe Gln Val Trp Asn Lys Asp Ser Ser		
	260	265
Ser Lys Asn Leu Ile Pro Arg Leu Gln Lys Ile Trp Lys Asn Tyr Leu		
	275	280

Ser Met Asn Lys Tyr Lys Val Ser Tyr Lys Gly Pro Gly Pro Gly Ile
 290 295 300

Lys Phe Ser Ala Glu Ala Leu Arg Cys His Leu Arg Asp His Val Asn
 305 310 315 320

Val Ser Met Val Glu Val Thr Asp Phe Pro Phe Asn Thr Ser Glu Trp
 325 330 335

Glu Gly Tyr Leu Pro Lys Glu Ser Ile Arg Thr Lys Ala Gly Pro Trp
 340 345 350

Gly Arg Cys Ala Val Val Ser Ser Ala Gly Ser Leu Lys Ser Ser Gln
 355 360 365

Leu Gly Arg Glu Ile Asp Asp His Asp Ala Val Leu Arg Phe Asn Gly
 370 375 380

Ala Pro Thr Ala Asn Phe Gln Gln Asp Val Gly Thr Lys Thr Thr Ile
 385 390 395 400

Arg Leu Met Asn Ser Gln Leu Val Thr Thr Glu Lys Arg Phe Leu Lys
 405 410 415

Asp Ser Leu Tyr Asn Glu Gly Ile Leu Ile Val Trp Asp Pro Ser Val
 420 425 430

Tyr His Ser Asp Ile Pro Lys Trp Tyr Gln Asn Pro Asp Tyr Asn Phe
 435 440 445

Phe Asn Asn Tyr Lys Thr Tyr Arg Lys Leu His Pro Asn Gln Pro Phe
 450 455 460

Tyr Ile Leu Lys Pro Gln Met Pro Trp Glu Leu Trp Asp Ile Leu Gln
 465 470 475 480

Glu Ile Ser Pro Glu Glu Ile Gln Pro Asn Pro Pro Ser Ser Gly Met
 485 490 495

Leu Gly Ile Ile Ile Met Met Thr Leu Cys Asp Gln Val Asp Ile Tyr
 500 505 510

Glu Phe Leu Pro Ser Lys Arg Lys Thr Asp Val Cys Tyr Tyr Tyr Gln
 515 520 525

Lys Phe Phe Asp Ser Ala Cys Thr Met Gly Ala Tyr His Pro Leu Leu
530 535 540

Tyr Glu Lys Asn Leu Val Lys His Leu Asn Gln Gly Thr Asp Glu Asp
545 550 555 560

Ile Tyr Leu Leu Gly Lys Ala Thr Leu Pro Gly Phe Arg Thr Ile His
565 570 575

Cys

<210> 167

<211> 1095

<212> ADN

<213> Artificial

<220>

<223> Fusão da sequência nucleotídica SEQ ID NO:151 com a
sequência nucleotídica SEQ ID NO:43

<220>

<221> CDS

<222> (1)..(1095)

<400> 167

atg gga ctc ttg gta ttt gtg cgc aat ctg ctg cta gcc ctc tgc ctc	48
Met Gly Leu Leu Val Phe Val Arg Asn Leu Leu Leu Ala Leu Cys Leu	
1 5 10 15	
ttt ctg gta ctg gga ttt ttg tat tat tct gcg tgg aag cta cac tta	96
Phe Leu Val Leu Gly Phe Leu Tyr Tyr Ser Ala Trp Lys Leu His Leu	
20 25 30	
ctc cag tgg gag gag gac tcc agt aag tat agt cac tct agc gga tcc	144
Leu Gln Trp Glu Glu Asp Ser Ser Lys Tyr Ser His Ser Ser Gly Ser	
35 40 45	
gag gcc tcc ttc cag gtg tgg aac aag gac agc tct tcc aaa aac ctt	192
Glu Ala Ser Phe Gln Val Trp Asn Lys Asp Ser Ser Ser Lys Asn Leu	
50 55 60	
atc cct agg ctg caa aag atc tgg aag aat tac cta agc atg aac aag	240
Ile Pro Arg Leu Gln Lys Ile Trp Lys Asn Tyr Leu Ser Met Asn Lys	
65 70 75 80	
tac aaa gtg tcc tac aag ggg cca gga cca ggc atc aag ttc agt gca	288
Tyr Lys Val Ser Tyr Lys Gly Pro Gly Pro Gly Ile Lys Phe Ser Ala	
85 90 95	
gag gcc ctg cgc tgc cac ctc cgg gac cat gtg aat gta tcc atg gta	336
Glu Ala Leu Arg Cys His Leu Arg Asp His Val Asn Val Ser Met Val	
100 105 110	
gag gtc aca gat ttt ccc ttc aat acc tct gaa tgg gag ggt tat ctg	384

Glu Val Thr Asp Phe Pro Phe Asn Thr Ser Glu Trp Glu Gly Tyr Leu	
115 120 125	
ccc aag gag agc att agg acc aag gct ggg cct tgg ggc agg tgt gct	432
Pro Lys Glu Ser Ile Arg Thr Lys Ala Gly Pro Trp Gly Arg Cys Ala	
130 135 140	
gtt gtg tcg tca gcg gga tct ctg aag tcc tcc caa cta ggc aga gaa	480
Val Val Ser Ser Ala Gly Ser Leu Lys Ser Ser Gln Leu Gly Arg Glu	
145 150 155 160	
atc gat gat cat gac gca gtc ctg agg ttt aat ggg gca ccc aca gcc	528
Ile Asp Asp His Asp Ala Val Leu Arg Phe Asn Gly Ala Pro Thr Ala	
165 170 175	
aac ttc caa caa gat gtg ggc aca aaa act acc att cgc ctg atg aac	576
Asn Phe Gln Gln Asp Val Gly Thr Lys Thr Thr Ile Arg Leu Met Asn	
180 185 190	
tct cag ttg gtt acc aca gag aag cgc ttc ctc aaa gac agt ttg tac	624
Ser Gln Leu Val Thr Thr Glu Lys Arg Phe Leu Lys Asp Ser Leu Tyr	
195 200 205	
aat gaa gga atc cta att gta tgg gac cca tct gta tac cac tca gat	672
Asn Glu Gly Ile Leu Ile Val Trp Asp Pro Ser Val Tyr His Ser Asp	
210 215 220	
atc cca aag tgg tac cag aat ccg gat tat aat ttc ttt aac aac tac	720
Ile Pro Lys Trp Tyr Gln Asn Pro Asp Tyr Ash Phe Phe Asn Asn Tyr	
225 230 235 240	
aag act tat cgt aag ctg cac ccc aat cag ccc ttt tac atc ctc aag	768
Lys Thr Tyr Arg Lys Leu His Pro Asn Gln Pro Phe Tyr Ile Leu Lys	
245 250 255	
ccc cag atg cct tgg gag cta tgg gac att ctt caa gaa atc tcc cca	816
Pro Gln Met Pro Trp Glu Leu Trp Asp Ile Leu Gln Glu Ile Ser Pro	
260 265 270	
gaa gag att cag cca aac ccc cca tcc tct ggg atg ctt ggt atc atc	864
Glu Glu Ile Gln Pro Asn Pro Pro Ser Ser Gly Met Leu Gly Ile Ile	
275 280 285	
atc atg atg acg ctg tgt gac cag gtg gat att tat gag ttc ctc cca	912
Ile Met Met Thr Leu Cys Asp Gln Val Asp Ile Tyr Glu Phe Leu Pro	
290 295 300	
tcc aag cgc aag act gac gtg tgc tac tac tac cag aag ttc ttc gat	960
Ser Lys Arg Lys Thr Asp Val Cys Tyr Tyr Tyr Gln Lys Phe Phe Asp	
305 310 315 320	
agt gcc tgc acg atg ggt gcc tac cac ccg ctg ctc tat gag aag aat	1008
Ser Ala Cys Thr Met Gly Ala Tyr His Pro Leu Leu Tyr Glu Lys Asn	
325 330 335	
ttg gtg aag cat ctc aac cag ggc aca gat gag gac atc tac ctg ctt	1056
Leu Val Lys His Leu Asn Gln Gly Thr Asp Glu Asp Ile Tyr Leu Leu	
340 345 350	
gga aaa gcc aca ctg cct ggc ttc cgg acc att cac tgc	1095
Gly Lys Ala Thr Leu Pro Gly Phe Arg Thr Ile His Cys	

355

360

365

<210> 168

<211> 365

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 168

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

Ser Gln Leu Val Thr Thr Glu Lys Arg Phe Leu Lys Asp Ser Leu Tyr
 195 200 205

Asn Glu Gly Ile Leu Ile Val Trp Asp Pro Ser Val Tyr His Ser Asp
 210 215 220

Ile Pro Lys Trp Tyr Gln Asn Pro Asp Tyr Asn Phe Phe Asn Asn Tyr
 225 230 235 240

Lys Thr Tyr Arg Lys Leu His Pro Asn Gln Pro Phe Tyr Ile Leu Lys
 245 250 255

Pro Gln Met Pro Trp Glu Leu Trp Asp Ile Leu Gln Glu Ile Ser Pro
 260 265 270

Glu Glu Ile Gln Pro Asn Pro Pro Ser Ser Gly Met Leu Gly Ile Ile
 275 280 285

Ile Met Met Thr Leu Cys Asp Gln Val Asp Ile Tyr Glu Phe Leu Pro
 290 295 300

Ser Lys Arg Lys Thr Asp Val Cys Tyr Tyr Tyr Gln Lys Phe Phe Asp
 305 310 315 320

Ser Ala Cys Thr Met Gly Ala Tyr His Pro Leu Leu Tyr Glu Lys Asn
 325 330 335

Leu Val Lys His Leu Asn Gln Gly Thr Asp Glu Asp Ile Tyr Leu Leu
 340 345 350

Gly Lys Ala Thr Leu Pro Gly Phe Arg Thr Ile His Cys
 355 360 365

<210> 169

<211> 1095

<212> ADN

<213> Artificial

<220>

<223> Fusão da sequência nucleotídica SEQ ID NO:153 com a
 sequência nucleotídica SEQ ID NO:43

<220>

<221> CDS

<222> (1)..(1095)

<400> 169

```
atg gga ctc ttg gta ttt gta cgc aac ctg ctg cta gcc ctc tgc ctc    48
Met Gly Leu Leu Val Phe Val Arg Asn Leu Leu Leu Ala Leu Cys Leu
```

1	5	10	15	
ttt ctg gtc ctg gga ttt ttg tat tat tct gcc tgg aag cta cac tta				96
Phe Leu Val Leu Gly Phe Leu Tyr Tyr Ser Ala Trp Lys Leu His Leu	20	25	30	
ctc caa tgg gaa gac tcc aat tca ctg att ctt tcc ctt gac tct aga				144
Leu Gln Trp Glu Asp Ser Asn Ser Leu Ile Leu Ser Leu Asp Ser Arg	35	40	45	
gag gcc tcc ttc cag gtg tgg aac aag gac agc tct tcc aaa aac ctt				192
Glu Ala Ser Phe Gln Val Trp Asn Lys Asp Ser Ser Ser Lys Asn Leu	50	55	60	
atc cct agg ctg caa aag atc tgg aag aat tac cta agc atg aac aag				240
Ile Pro Arg Leu Gln Lys Ile Trp Lys Asn Tyr Leu Ser Met Asn Lys	65	70	75	80
tac aaa gtg tcc tac aag ggg cca gga cca ggc atc aag ttc agt gca				288
Tyr Lys Val Ser Tyr Lys Gly Pro Gly Pro Gly Ile Lys Phe Ser Ala	85	90	95	
gag gcc ctg cgc tgc cac ctg cgg gac cat gtg aat gta tcc atg gta				336
Glu Ala Leu Arg Cys His Leu Arg Asp His Val Asn Val Ser Met Val	100	105	110	
gag gtc aca gat ttt ccc ttc aat acc tct gaa tgg gag ggt tat ctg				384
Glu Val Thr Asp Phe Pro Phe Asn Thr Ser Glu Trp Glu Gly Tyr Leu	115	120	125	
ccc aag gag agc att agg acc aag gct ggg cct tgg ggc agg tgt gct				432
Pro Lys Glu Ser Ile Arg Thr Lys Ala Gly Pro Trp Gly Arg Cys Ala	130	135	140	
gtt gtg tgg tca gcg gga tct ctg aag tcc tcc caa cta ggc aga gaa				480
Val Val Ser Ser Ala Gly Ser Leu Lys Ser Ser Gln Leu Gly Arg Glu	145	150	155	160
atc gat gat cat gac gca gtc ctg agg ttt aat ggg gca ccc aca gcc				528
Ile Asp Asp His Asp Ala Val Leu Arg Phe Asn Gly Ala Pro Thr Ala	165	170	175	
aac ttc caa caa gat gtg ggc aca aaa act acc att cgc ctg atg aac				576
Asn Phe Gln Gln Asp Val Gly Thr Lys Thr Thr Ile Arg Leu Met Asn	180	185	190	
tct cag ttg gtt acc aca gag aag cgc ttc ctc aaa gac agt ttg tac				624
Ser Gln Leu Val Thr Thr Glu Lys Arg Phe Leu Lys Asp Ser Leu Tyr	195	200	205	
aat gaa gga atc cta att gta tgg gac cca tct gta tac cac tca gat				672
Asn Glu Gly Ile Leu Ile Val Trp Asp Pro Ser Val Tyr His Ser Asp	210	215	220	
atc cca aag tgg tac cag aat ccg gat tat aat ttc ttt aac aac tac				720
Ile Pro Lys Trp Tyr Gln Asn Pro Asp Tyr Asn Phe Phe Asn Asn Tyr	225	230	235	240
aag act tat cgt aag ctg cac ccc aat cag ccc ttt tac atc ctc aag				768
Lys Thr Tyr Arg Lys Leu His Pro Asn Gln Pro Phe Tyr Ile Leu Lys	245	250	255	

ccc cag atg cct tgg gag cta tgg gac att ctt caa gaa atc tcc cca	816
Pro Gln Met Pro Trp Glu Leu Trp Asp Ile Leu Gln Glu Ile Ser Pro	
260 265 270	
gaa gag att cag cca aac ccc cca tcc tct ggg atg ctt ggt atc atc	864
Glu Glu Ile Gln Pro Asn Pro Pro Ser Ser Gly Met Leu Gly Ile Ile	
275 280 285	
atc atg atg acg ctg tgt gac cag gtg gat att tat gag ttc ctc cca	912
Ile Met Met Thr Leu Cys Asp Gln Val Asp Ile Tyr Glu Phe Leu Pro	
290 295 300	
tcc aag cgc aag act gac gtg tgc tac tac tac cag aag ttc ttc gat	960
Ser Lys Arg Lys Thr Asp Val Cys Tyr Tyr Tyr Gln Lys Phe Phe Asp	
305 310 315 320	
agt gcc tgc acg atg ggt gcc tac cac ccg ctg ctc tat gag aag aat	1008
Ser Ala Cys Thr Met Gly Ala Tyr His Pro Leu Leu Tyr Glu Lys Asn	
325 330 335	
ttg gtg aag cat ctc aac cag ggc aca gat gag gac atc tac ctg ctt	1056
Leu Val Lys His Leu Asn Gln Gly Thr Asp Glu Asp Ile Tyr Leu Leu	
340 345 350	
gga aaa gcc aca ctg cct ggc ttc cgg acc att cac tgc	1095
Gly Lys Ala Thr Leu Pro Gly Phe Arg Thr Ile His Cys	
355 360 365	

<210> 170

<211> 365

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 170

Met Gly Leu Leu Val Phe Val Arg Asn Leu Leu Leu Ala Leu Cys Leu
1 5 10 15

Phe Leu Val Leu Gly Phe Leu Tyr Tyr Ser Ala Trp Lys Leu His Leu
20 25 30

Leu Gln Trp Glu Asp Ser Asn Ser Leu Ile Leu Ser Leu Asp Ser Arg
35 40 45

Glu Ala Ser Phe Gln Val Trp Asn Lys Asp Ser Ser Ser Lys Asn Leu
50 55 60

Ile Pro Arg Leu Gln Lys Ile Trp Lys Asn Tyr Leu Ser Met Asn Lys
65 70 75 80

Tyr Lys Val Ser Tyr Lys Gly Pro Gly Pro Gly Ile Lys Phe Ser Ala

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

Leu Val Lys His Leu Asn Gln Gly Thr Asp Glu Asp Ile Tyr Leu Leu
340 345 350

Gly Lys Ala Thr Leu Pro Gly Phe Arg Thr Ile His Cys
355 360 365

<210> 171

<211> 1194

<212> ADN

<213> Artificial

<220>

<223> Fusão da sequência nucleotídica SEQ ID NO:155 com a
sequência nucleotídica SEQ ID NO:43

<220>

<221> CDS

<222> (1)..(1194)

<400> 171

atg cag ctg cag ttc cgg agc tgg atg ctg gcc gcg ctc acg ctg ctc	48
Met Gln Leu Gln Phe Arg Ser Trp Met Leu Ala Ala Leu Thr Leu Leu	
1 5 10 15	
gtg gtc ttc ctc atc ttc gca gac atc tca gag atc gaa gaa gaa atc	96
Val Val Phe Leu Ile Phe Ala Asp Ile Ser Glu Ile Glu Glu Glu Ile	
20 25 30	
ggg aat tcg gga ggc aga ggt aca atc aga tca gct gtg aac agc tta	144
Gly Asn Ser Gly Gly Arg Gly Thr Ile Arg Ser Ala Val Asn Ser Leu	
35 40 45	
cat agc aaa tct aat aga gct gaa gtt gta ata aac ggc tcc tca tca	192
His Ser Lys Ser Asn Arg Ala Glu Val Val Ile Asn Gly Ser Ser Ser	
50 55 60	
cca gct gtt gtt gac aga agt aat gaa agc atc aag cac aac atc aag	240
Pro Ala Val Val Asp Arg Ser Asn Glu Ser Ile Lys His Asn Ile Lys	
65 70 75 80	
ctt gag gcc tcc ttc cag gtg tgg aac aag gac agc tct tcc aaa aac	288
Leu Glu Ala Ser Phe Gln Val Trp Asn Lys Asp Ser Ser Ser Lys Asn	
85 90 95	
ctt atc cct agg ctg caa aag atc tgg aag aat tac cta agc atg aac	336
Leu Ile Pro Arg Leu Gln Lys Ile Trp Lys Asn Tyr Leu Ser Met Asn	
100 105 110	
aag tac aaa gtg tcc tac aag ggg cca gga cca ggc atc aag ttc agt	384
Lys Tyr Lys Val Ser Tyr Lys Gly Pro Gly Pro Gly Ile Lys Phe Ser	
115 120 125	
gca gag gcc ctg cgc tgc cac ctc cgg gac cat gtg aat gta tcc atg	432
Ala Glu Ala Leu Arg Cys His Leu Arg Asp His Val Asn Val Ser Met	
130 135 140	

gta gag gtc aca gat ttt ccc ttc aat acc tct gaa tgg gag ggt tat Val Glu Val Thr Asp Phe Pro Phe Asn Thr Ser Glu Trp Glu Gly Tyr 145 150 155 160	480
ctg ccc aag gag agc att agg acc aag gct ggg cct tgg ggc agg tgt Leu Pro Lys Glu Ser Ile Arg Thr Lys Ala Gly Pro Trp Gly Arg Cys 165 170 175	528
gct gtt gtg tgg tca gcg gga tct ctg aag tcc tcc caa cta ggc aga Ala Val Val Ser Ser Ala Gly Ser Leu Lys Ser Ser Gln Leu Gly Arg 180 185 190	576
gaa atc gat gat cat gac gca gtc ctg agg ttt aat ggg gca ccc aca Glu Ile Asp Asp His Asp Ala Val Leu Arg Phe Asn Gly Ala Pro Thr 195 200 205	624
gcc aac ttc caa caa gat gtg ggc aca aaa act acc att cgc ctg atg Ala Asn Phe Gln Gln Asp Val Gly Thr Lys Thr Thr Ile Arg Leu Met 210 215 220	672
aac tct cag ttg gtt acc aca gag aag cgc ttc ctc aaa gac agt ttg Asn Ser Gln Leu Val Thr Thr Glu Lys Arg Phe Leu Lys Asp Ser Leu 225 230 235 240	720
tac aat gaa gga atc cta att gta tgg gac cca tct gta tac cac tca Tyr Asn Glu Gly Ile Leu Ile Val Trp Asp Pro Ser Val Tyr His Ser 245 250 255	768
gat atc cca aag tgg tad cag aat cgg gat tat aat ttc ttt aac aac Asp Ile Pro Lys Trp Tyr Gln Asn Pro Asp Tyr Asn Phe Phe Asn Asn 260 265 270	816
tac aag act tat cgt aag ctg cac ccc aat cag ccc ttt tac atc ctc Tyr Lys Thr Tyr Arg Lys Leu His Pro Asn Gln Pro Phe Tyr Ile Leu 275 280 285	864
aag ccc cag atg cct tgg gag cta tgg gac att ctt caa gaa atc tcc Lys Pro Gln Met Pro Trp Glu Leu Trp Asp Ile Leu Gln Glu Ile Ser 290 295 300	912
cca gaa gag att cag cca aac ccc cca tcc tct ggg atg ctt ggt atc Pro Glu Glu Ile Gln Pro Asn Pro Pro Ser Gly Met Leu Gly Ile 305 310 315 320	960
atc atc atg atg acg ctg tgt gac cag gtg gat att tat gag ttc ctc Ile Ile Met Met Thr Leu Cys Asp Gln Val Asp Ile Tyr Glu Phe Leu 325 330 335	1008
cca tcc aag cgc aag act gac gtg tgc tac tac tac cag aag ttc ttc Pro Ser Lys Arg Lys Thr Asp Val Cys Tyr Tyr Tyr Gln Lys Phe Phe 340 345 350	1056
gat agt gcc tgc acg atg ggt gcc tac cac ccg ctg ctc tat gag aag Asp Ser Ala Cys Thr Met Gly Ala Tyr His Pro Leu Leu Tyr Glu Lys 355 360 365	1104
aat ttg gtg aag cat ctc aac cag ggc aca gat gag gac atc tac ctg Asn Leu Val Lys His Leu Asn Gln Gly Thr Asp Glu Asp Ile Tyr Leu 370 375 380	1152
ctt gga aaa gcc aca ctg cct ggc ttc cgg acc att cac tgc	1194

Leu	Gly	Lys	Ala	Thr	Leu	Pro	Gly	Phe	Arg	Thr	Ile	His	Cys
385					390					395			

<210> 172

<211> 398

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 172

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

Glu Ile Asp Asp His Asp Ala Val Leu Arg Phe Asn Gly Ala Pro Thr
 195 200 205

Ala Asn Phe Gln Gln Asp Val Gly Thr Lys Thr Thr Ile Arg Leu Met
 210 215 220

Asn Ser Gln Leu Val Thr Thr Glu Lys Arg Phe Leu Lys Asp Ser Leu
 225 230 235 240

Tyr Asn Glu Gly Ile Leu Ile Val Trp Asp Pro Ser Val Tyr His Ser
 245 250 255

Asp Ile Pro Lys Trp Tyr Gln Asn Pro Asp Tyr Asn Phe Phe Asn Asn
 260 265 270

Tyr Lys Thr Tyr Arg Lys Leu His Pro Asn Gln Pro Phe Tyr Ile Leu
 275 280 285

Lys Pro Gln Met Pro Trp Glu Leu Trp Asp Ile Leu Gln Glu Ile Ser
 290 295 300

Pro Glu Glu Ile Gln Pro Asn Pro Pro Ser Ser Gly Met Leu Gly Ile
 305 310 315 320

Ile Ile Met Met Thr Leu Cys Asp Gln Val Asp Ile Tyr Glu Phe Leu
 325 330 335

Pro Ser Lys Arg Lys Thr Asp Val Cys Tyr Tyr Tyr Gln Lys Phe Phe
 340 345 350

Asp Ser Ala Cys Thr Met Gly Ala Tyr His Pro Leu Leu Tyr Glu Lys
 355 360 365

Asn Leu Val Lys His Leu Asn Gln Gly Thr Asp Glu Asp Ile Tyr Leu
 370 375 380

Leu Gly Lys Ala Thr Leu Pro Gly Phe Arg Thr Ile His Cys
 385 390 395

<210> 173

<211> 1158

<212> ADN

<213> Artificial

<220>

<223> Fusão da sequência nucleotídica SEQ ID NO:157 com a
sequência nucleotídica SEQ ID NO:43

<220>

<221> CDS

<222> (1)..(1158)

<400> 173

atg cgc tcc att agg aag agg tgg acg atc tgc aca ata agt ctg ctc	48
Met Arg Ser Ile Arg Lys Arg Trp Thr Ile Cys Thr Ile Ser Leu Leu	
1 5 10 15	
ctg atc ttt tat aag aca aaa gaa ata gca aga act gag gag cac cag	96
Leu Ile Phe Tyr Lys Thr Lys Glu Ile Ala Arg Thr Glu Glu His Gln	
20 25 30	
gag acg caa ctc atc gga gat ggt gaa ttg tct ttg agt cgg tca ctt	144
Glu Thr Gln Leu Ile Gly Asp Gly Glu Leu Ser Leu Ser Arg Ser Leu	
35 40 45	
gtc aat agc tct gat aaa atc att cga aag gct ggc tct tca atc ttc	192
Val Asn Ser Ser Asp Lys Ile Ile Arg Lys Ala Gly Ser Ser Ile Phe	
50 55 60	
cag cac aat aag ctt gag gcc tcc ttc cag gtg tgg aac aag gac agc	240
Gln His Asn Lys Leu Glu Ala Ser Phe Gln Val Trp Asn Lys Asp Ser	
65 70 75 80	
tct tcc aaa aac ctt atc cct agg ctg caa aag atc tgg aag aat tac	288
Ser Ser Lys Asn Leu Ile Pro Arg Leu Gln Lys Ile Trp Lys Asn Tyr	
85 90 95	
cta agc atg aac aag tac aaa gtg tcc tac aag ggg cca gga cca ggc	336
Leu Ser Met Asn Lys Tyr Lys Val Ser Tyr Lys Gly Pro Gly Pro Gly	
100 105 110	
atc aag ttc agt gca gag gcc ctg cgc tgc cac ctc cgg gac cat gtg	384
Ile Lys Phe Ser Ala Glu Ala Leu Arg Cys His Leu Arg Asp His Val	
115 120 125	
aat gta tcc atg gta gag gtc aca gat ttt ccc ttc aat acc tct gaa	432
Asn Val Ser Met Val Glu Val Thr Asp Phe Pro Phe Asn Thr Ser Glu	
130 135 140	
tgg gag ggt tat ctg ccc aag gag agc att agg acc aag gct ggg cct	480
Trp Glu Gly Tyr Leu Pro Lys Glu Ser Ile Arg Thr Lys Ala Gly Pro	
145 150 155 160	
tgg ggc agg tgt gct gtt gtg tcg tca gcg gga tct ctg aag tcc tcc	528
Trp Gly Arg Cys Ala Val Val Ser Ser Ala Gly Ser Leu Lys Ser Ser	
165 170 175	
caa cta ggc aga gaa atc gat gat cat gac gca gtc ctg agg ttt aat	576
Gln Leu Gly Arg Glu Ile Asp Asp His Asp Ala Val Leu Arg Phe Asn	
180 185 190	
ggg gca ccc aca gcc aac ttc caa caa gat gtg ggc aca aaa act acc	624
Gly Ala Pro Thr Ala Asn Phe Gln Gln Asp Val Gly Thr Lys Thr Thr	
195 200 205	
att cgc ctg atg aac tct cag ttg gtt acc aca gag aag cgc ttc ctc	672
Ile Arg Leu Met Asn Ser Gln Leu Val Thr Thr Glu Lys Arg Phe Leu	

210	215	220	
aaa gac agt ttg tac aat gaa gga atc cta att gta tgg gac cca tct			720
Lys Asp Ser Leu Tyr Asn Glu Gly Ile Leu Ile Val Trp Asp Pro Ser			
225	230	235	240
gta tac cac tca gat atc cca aag tgg tac cag aat ccg gat tat aat			768
Val Tyr His Ser Asp Ile Pro Lys Trp Tyr Gln Asn Pro Asp Tyr Asn			
	245	250	255
ttc ttt aac aac tac aag act tat cgt aag ctg cac ccc aat cag ccc			816
Phe Phe Asn Asn Tyr Lys Thr Tyr Arg Lys Leu His Pro Asn Gln Pro			
	260	265	270
ttt tac atc ctc aag ccc cag atg cct tgg gag cta tgg gac att ctt			864
Phe Tyr Ile Leu Lys Pro Gln Met Pro Trp Glu Leu Trp Asp Ile Leu			
	275	280	285
caa gaa atc tcc cca gaa gag att cag cca aac ccc cca tcc tct ggg			912
Gln Glu Ile Ser Pro Glu Glu Ile Gln Pro Asn Pro Pro Ser Ser Gly			
	290	295	300
atg ctt ggt atc atc atc atg atg acg ctg tgt gac cag gtg gat att			960
Met Leu Gly Ile Ile Ile Met Met Thr Leu Cys Asp Gln Val Asp Ile			
	305	310	315
tat gag ttc ctc cca tcc aag cgc aag act gac gtg tgc tac tac tac			1008
Tyr Glu Phe Leu Pro Ser Lys Arg Lys Thr Asp Val Cys Tyr Tyr Tyr			
	325	330	335
cag aag ttc ttc gat agt gcc tgc acg atg ggt gcc tac cac ccg ctg			1056
Gln Lys Phe Phe Asp Ser Ala Cys Thr Met Gly Ala Tyr His Pro Leu			
	340	345	350
ctc tat gag aag aat ttg gtg aag cat ctc aac cag ggc aca gat gag			1104
Leu Tyr Glu Lys Asn Leu Val Lys His Leu Asn Gln Gly Thr Asp Glu			
	355	360	365
gac atc tac ctg ctt gga aaa gcc aca ctg cct ggc ttc cgg acc att			1152
Asp Ile Tyr Leu Leu Gly Lys Ala Thr Leu Pro Gly Phe Arg Thr Ile			
	370	375	380
cac tgc			1158
His Cys			
385			

<210> 174

<211> 386

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 174

Met	Arg	Ser	Ile	Arg	Lys	Arg	Trp	Thr	Ile	Cys	Thr	Ile	Ser	Leu	Leu
1				5					10					15	

Leu Ile Phe Tyr Lys Thr Lys Glu Ile Ala Arg Thr Glu Glu His Gln
 20 25 30

Glu Thr Gln Leu Ile Gly Asp Gly Glu Leu Ser Leu Ser Arg Ser Leu
 35 40 45

Val Asn Ser Ser Asp Lys Ile Ile Arg Lys Ala Gly Ser Ser Ile Phe
 50 55 60

Gln His Asn Lys Leu Glu Ala Ser Phe Gln Val Trp Asn Lys Asp Ser
 65 70 75 80

Ser Ser Lys Asn Leu Ile Pro Arg Leu Gln Lys Ile Trp Lys Asn Tyr
 85 90 95

Leu Ser Met Asn Lys Tyr Lys Val Ser Tyr Lys Gly Pro Gly Pro Gly
 100 105 110

Ile Lys Phe Ser Ala Glu Ala Leu Arg Cys His Leu Arg Asp His Val
 115 120 125

Asn Val Ser Met Val Glu Val Thr Asp Phe Pro Phe Asn Thr Ser Glu
 130 135 140

Trp Glu Gly Tyr Leu Pro Lys Glu Ser Ile Arg Thr Lys Ala Gly Pro
 145 150 155 160

Trp Gly Arg Cys Ala Val Val Ser Ser Ala Gly Ser Leu Lys Ser Ser
 165 170 175

Gln Leu Gly Arg Glu Ile Asp Asp His Asp Ala Val Leu Arg Phe Asn
 180 185 190

Gly Ala Pro Thr Ala Asn Phe Gln Gln Asp Val Gly Thr Lys Thr Thr
 195 200 205

Ile Arg Leu Met Asn Ser Gln Leu Val Thr Thr Glu Lys Arg Phe Leu
 210 215 220

Lys Asp Ser Leu Tyr Asn Glu Gly Ile Leu Ile Val Trp Asp Pro Ser
 225 230 235 240

Val Tyr His Ser Asp Ile Pro Lys Trp Tyr Gln Asn Pro Asp Tyr Asn
 245 250 255

Phe Phe Asn Asn Tyr Lys Thr Tyr Arg Lys Leu His Pro Asn Gln Pro

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

<210> 175

<211> 1155

<212> ADN

<213> Artificial

<220>

<223> Fusão da sequência nucleotídica SEQ ID NO:159 com a
sequência nucleotídica SEQ ID NO:43

<220>

<221> CDS

<222> (1)..(1155)

<400> 175

cag tcc ctg gca aag cct aag tcc cag gca ccc aca agg gca agg agg Gln Ser Leu Ala Lys Pro Lys Ser Gln Ala Pro Thr Arg Ala Arg Arg 50 55 60	192
aca acc gga tcc gag gcc tcc ttc cag gtg tgg aac aag gac agc tct Thr Thr Gly Ser Glu Ala Ser Phe Gln Val Trp Asn Lys Asp Ser Ser 65 70 75 80	240
tcc aaa aac ctt atc cct agg ctg caa aag atc tgg aag aat tac cta Ser Lys Asn Leu Ile Pro Arg Leu Gln Lys Ile Trp Lys Asn Tyr Leu 85 90 95	288
agc atg aac aag tac aaa gtg tcc tac aag ggg cca gga cca ggc atc Ser Met Asn Lys Tyr Lys Val Ser Tyr Lys Gly Pro Gly Pro Gly Ile 100 105 110	336
aag ttc agt gca gag gcc ctg cgc tgc cac ctc cgg gac cat gtg aat Lys Phe Ser Ala Glu Ala Leu Arg Cys His Leu Arg Asp His Val Asn 115 120 125	384
gta tcc atg gta gag gtc aca gat ttt ccc ttc aat acc tct gaa tgg Val Ser Met Val Glu Val Thr Asp Phe Pro Phe Asn Thr Ser Glu Trp 130 135 140	432
gag ggt tat ctg ccc aag gag agc att agg acc aag gct ggg cct tgg Glu Gly Tyr Leu Pro Lys Glu Ser Ile Arg Thr Lys Ala Gly Pro Trp 145 150 155 160	480
ggc agg tgt gct gtt gtg tgc tca gcg gga tct ctg aag tcc tcc caa Gly Arg Cys Ala Val Val Ser Ser Ala Gly Ser Leu Lys Ser Ser Gln 165 170 175	528
cta ggc aga gaa atc gat gat cat gac gca gtc ctg agg ttt aat ggg Leu Gly Arg Glu Ile Asp Asp His Asp Ala Val Leu Arg Phe Asn Gly 180 185 190	576
gca ccc aca gcc aac ttc caa caa gat gtg ggc aca aaa act acc att Ala Pro Thr Ala Asn Phe Gln Gln Asp Val Gly Thr Lys Thr Thr Ile 195 200 205	624
cgc ctg atg aac tct cag ttg gtt acc aca gag aag cgc ttc ctc aaa Arg Leu Met Asn Ser Gln Leu Val Thr Thr Glu Lys Arg Phe Leu Lys 210 215 220	672
gac agt ttg tac aat gaa gga atc cta att gta tgg gac cca tct gta Asp Ser Leu Tyr Asn Glu Gly Ile Leu Ile Val Trp Asp Pro Ser Val 225 230 235 240	720
tac cac tca gat atc cca aag tgg tac cag aat ccg gat tat aat ttc Tyr His Ser Asp Ile Pro Lys Trp Tyr Gln Asn Pro Asp Tyr Asn Phe 245 250 255	768
ttt aac aac tac aag act tat cgt aag ctg cac ccc aat cag ccc ttt Phe Asn Asn Tyr Lys Thr Tyr Arg Lys Leu His Pro Asn Gln Pro Phe 260 265 270	816
tac atc ctc aag ccc cag atg cct tgg gag cta tgg gac att ctt caa Tyr Ile Leu Lys Pro Gln Met Pro Trp Glu Leu Trp Asp Ile Leu Gln 275 280 285	864

gaa atc tcc cca gaa gag att cag cca aac ccc cca tcc tct ggg atg	912
Glu Ile Ser Pro Glu Glu Ile Gln Pro Asn Pro Pro Ser Ser Gly Met	
290 295 300	
ctt ggt atc atc atc atg atg acg ctg tgt gac cag gtg gat att tat	960
Leu Gly Ile Ile Ile Met Met Thr Leu Cys Asp Gln Val Asp Ile Tyr	
305 310 315 320	
gag ttc ctc cca tcc aag cgc aag act gac gtg tgc tac tac tac cag	1008
Glu Phe Leu Pro Ser Lys Arg Lys Thr Asp Val Cys Tyr Tyr Tyr Gln	
325 330 335	
aag ttc ttc gat agt gcc tgc acg atg ggt gcc tac cac ccg ctg ctc	1056
Lys Phe Phe Asp Ser Ala Cys Thr Met Gly Ala Tyr His Pro Leu Leu	
340 345 350	
tat gag aag aat ttg gtg aag cat ctc aac cag ggc aca gat gag gac	1104
Tyr Glu Lys Asn Leu Val Lys His Leu Asn Gln Gly Thr Asp Glu Asp	
355 360 365	
atc tac ctg ctt gga aaa gcc aca ctg cct ggc ttc cgg acc att cac	1152
Ile Tyr Leu Leu Gly Lys Ala Thr Leu Pro Gly Phe Arg Thr Ile His	
370 375 380	
tgc	1155
Cys	
385	

<210> 176

<211> 385

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 176

Met Gly Leu Leu Val Phe Val Arg Asn Leu Leu Leu Ala Leu Cys Leu
1 5 10 15

Phe Leu Val Leu Gly Phe Leu Tyr Tyr Ser Ala Trp Glu Pro Gln Thr
20 25 30

Lys Pro Ser Arg His Gln Arg Thr Glu Asn Ile Lys Glu Arg Ser Leu
35 40 45

Gln Ser Leu Ala Lys Pro Lys Ser Gln Ala Pro Thr Arg Ala Arg Arg
50 55 60

Thr Thr Gly Ser Glu Ala Ser Phe Gln Val Trp Asn Lys Asp Ser Ser
65 70 75 80

Ser Lys Asn Leu Ile Pro Arg Leu Gln Lys Ile Trp Lys Asn Tyr Leu
85 90 95

Ser Met Asn Lys Tyr Lys Val Ser Tyr Lys Gly Pro Gly Pro Gly Ile
 100 105 110

Lys Phe Ser Ala Glu Ala Leu Arg Cys His Leu Arg Asp His Val Asn
 115 120 125

Val Ser Met Val Glu Val Thr Asp Phe Pro Phe Asn Thr Ser Glu Trp
 130 135 140

Glu Gly Tyr Leu Pro Lys Glu Ser Ile Arg Thr Lys Ala Gly Pro Trp
 145 150 155 160

Gly Arg Cys Ala Val Val Ser Ser Ala Gly Ser Leu Lys Ser Ser Gln
 165 170 175

Leu Gly Arg Glu Ile Asp Asp His Asp Ala Val Leu Arg Phe Asn Gly
 180 185 190

Ala Pro Thr Ala Asn Phe Gln Gln Asp Val Gly Thr Lys Thr Thr Ile
 195 200 205

Arg Leu Met Asn Ser Gln Leu Val Thr Thr Glu Lys Arg Phe Leu Lys
 210 215 220

Asp Ser Leu Tyr Asn Glu Gly Ile Leu Ile Val Trp Asp Pro Ser Val
 225 230 235 240

Tyr His Ser Asp Ile Pro Lys Trp Tyr Gln Asn Pro Asp Tyr Asn Phe
 245 250 255

Phe Asn Asn Tyr Lys Thr Tyr Arg Lys Leu His Pro Asn Gln Pro Phe
 260 265 270

Tyr Ile Leu Lys Pro Gln Met Pro Trp Glu Leu Trp Asp Ile Leu Gln
 275 280 285

Glu Ile Ser Pro Glu Glu Ile Gln Pro Asn Pro Pro Ser Ser Gly Met
 290 295 300

Leu Gly Ile Ile Ile Met Met Thr Leu Cys Asp Gln Val Asp Ile Tyr
 305 310 315 320

Glu Phe Leu Pro Ser Lys Arg Lys Thr Asp Val Cys Tyr Tyr Tyr Gln
 325 330 335

Lys Phe Phe Asp Ser Ala Cys Thr Met Gly Ala Tyr His Pro Leu Leu
 340 345 350

Tyr Glu Lys Asn Leu Val Lys His Leu Asn Gln Gly Thr Asp Glu Asp
 355 360 365

Ile Tyr Leu Leu Gly Lys Ala Thr Leu Pro Gly Phe Arg Thr Ile His
 370 375 380

Cys
 385

<210> 177

<211> 84

<212> ADN

<213> Artificial

<220>

<223> Fusão da sequência nucleotídica SEQ ID NO:65 com a
 sequência nucleotídica SEQ ID NO:107

<220>

<221> CDS

<222> (1)..(84)

<400> 177

atg gga ctc ttg gta ttt gtg cgc aat ctg ctg cta gcc ctc tgc ctc	48
Met Gly Leu Leu Val Phe Val Arg Asn Leu Leu Leu Ala Leu Cys Leu	
1 5 10 15	

ttt ctg gta ctg gga ttt ttg tat tat tct gcg tgg	84
Phe Leu Val Leu Gly Phe Leu Tyr Tyr Ser Ala Trp	
1 20 25	

<210> 178

<211> 28

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 178

Met Gly Leu Leu Val Phe Val Arg Asn Leu Leu Leu Ala Leu Cys Leu
 1 5 10 15

Phe Leu Val Leu Gly Phe Leu Tyr Tyr Ser Ala Trp
 20 25

<210> 179

<211> 192

<212> ADN

<213> Artificial

<220>

<223> Fusão da sequência nucleotídica SEQ ID NO:65 com a
 sequência nucleotídica SEQ ID NO:119 e com a sequência
 nucleotídica SEQ ID NO:129

<220>

<221> CDS

<222> (1) .. (192)

<400> 179

atg gga ctc ttg gta ttt gtg cgc agc tgg atg ctg gcc gcg ctc acg	48
Met Gly Leu Leu Val Phe Val Arg Ser Trp Met Leu Ala Ala Leu Thr	
1 5 10 15	
ctg ctc gtg gtc ttc ctc atc ttc gca aag gag cct caa aca aag cct	96
Leu Leu Val Val Phe Leu Ile Phe Ala Lys Glu Pro Gln Thr Lys Pro	
20 25 30	
tcc agg cat caa cgc aca gag aac att aaa gaa agg tct cta cag tcc	144
Ser Arg His Gln Arg Thr Glu Asn Ile Lys Glu Arg Ser Leu Gln Ser	
35 40 45	
ctg gca aag cct aag tcc cag gca ccc aca agg gca agg agg aca acc	192
Leu Ala Lys Pro Lys Ser Gln Ala Pro Thr Arg Ala Arg Arg Thr Thr	
50 55 60	

<210> 180

<211> 64

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 180

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

<210> 181

<211> 1143

<212> ADN

<213> Artificial

<220>

<223> Fusão da sequência nucleotídica SEQ ID NO:179 com a
sequência nucleotídica SEQ ID NO:43

<220>

<221> CDS

<222> (1)..(1143)

<400> 181

atg gga ctc ttg gta ttt gtg cgc agc tgg atg ctg gcc gcg ctc acg Met Gly Leu Leu Val Phe Val Arg Ser Trp Met Leu Ala Ala Leu Thr 1 5 10 15	48
ctg ctc gtg gtc ttc ctc atc ttc gca aag gag cct caa aca aag cct Leu Leu Val Val Phe Leu Ile Phe Ala Lys Glu Pro Gln Thr Lys Pro 20 25 30	96
tcc agg cat caa cgc aca gag aac att aaa gaa agg tct cta cag tcc Ser Arg His Gln Arg Thr Glu Asn Ile Lys Glu Arg Ser Leu Gln Ser 35 40 45	144
ctg gca aag cct aag tcc cag gca ccc aca agg gca agg agg aca acc Leu Ala Lys Pro Lys Ser Gln Ala Pro Thr Arg Ala Arg Arg Thr Thr 50 55 60	192
gag gcc tcc ttc cag gtg tgg aac aag gac agc tct tcc aaa aac ctt Glu Ala Ser Phe Gln Val Trp Asn Lys Asp Ser Ser Ser Lys Asn Leu 65 70 75 80	240
atc cct agg ctg caa aag atc tgg aag aat tac cta agc atg aac aag Ile Pro Arg Leu Gln Lys Ile Trp Lys Asn Tyr Leu Ser Met Asn Lys 85 90 95	288
tac aaa gtg tcc tac aag ggg cca gga cca ggc atc aag ttc agt gca Tyr Lys Val Ser Tyr Lys Gly Pro Gly Pro Gly Ile Lys Phe Ser Ala 100 105 110	336
gag gcc ctg cgc tgc cac ctc cgg gac cat gtg aat gta tcc atg gta Glu Ala Leu Arg Cys His Leu Arg Asp His Val Asn Val Ser Met Val 115 120 125	384
gag gtc aca gat ttt ccc ttc aat acc tct gaa tgg gag ggt tat ctg Glu Val Thr Asp Phe Pro Phe Asn Thr Ser Glu Trp Glu Gly Tyr Leu 130 135 140	432
ccc aag gag agc att agg acc aag gct ggg cct tgg ggc agg tgt gct Pro Lys Glu Ser Ile Arg Thr Lys Ala Gly Pro Trp Gly Arg Cys Ala 145 150 155 160	480
gtt gtg tcg tca gcg gga tct ctg aag tcc tcc caa cta ggc aga gaa Val Val Ser Ser Ala Gly Ser Leu Lys Ser Ser Gln Leu Gly Arg Glu 165 170 175	528
atc gat gat cat gac gca gtc ctg agg ttt aat ggg gca ccc aca gcc Ile Asp Asp His Asp Ala Val Leu Arg Phe Asn Gly Ala Pro Thr Ala 180 185 190	576
aac ttc caa caa gat gtg ggc aca aaa act acc att cgc ctg atg aac Asn Phe Gln Gln Asp Val Gly Thr Lys Thr Thr Ile Arg Leu Met Asn 195 200 205	624
tct cag ttg gtt acc aca gag aag cgc ttc ctc aaa gac agt ttg tac Ser Gln Leu Val Thr Thr Glu Lys Arg Phe Leu Lys Asp Ser Leu Tyr 210 215 220	672

aat gaa gga atc cta att gta tgg gac cca tct gta tac cac tca gat Asn Glu Gly Ile Leu Ile Val Trp Asp Pro Ser Val Tyr His Ser Asp 225 230 235 240	720
atc cca aag tgg tac cag aat ccg gat tat aat ttc ttt aac aac tac Ile Pro Lys Trp Tyr Gln Asn Pro Asp Tyr Asn Phe Phe Asn Asn Tyr 245 250 255	768
aag act tat cgt aag ctg cac ccc aat cag ccc ttt tac atc ctc aag Lys Thr Tyr Arg Lys Leu His Pro Asn Gln Pro Phe Tyr Ile Leu Lys 260 265 270	816
ccc cag atg cct tgg gag cta tgg gac att ctt caa gaa atc tcc cca Pro Gln Met Pro Trp Glu Leu Trp Asp Ile Leu Gln Glu Ile Ser Pro 275 280 285	864
gaa gag att cag cca aac ccc cca tcc tct ggg atg ctt ggt atc atc Glu Glu Ile Gln Pro Asn Pro Pro Ser Ser Gly Met Leu Gly Ile Ile 290 295 300	912
atc atg atg acg ctg tgt gac cag gtg gat att tat gag ttc ctc cca Ile Met Met Thr Leu Cys Asp Gln Val Asp Ile Tyr Glu Phe Leu Pro 305 310 315 320	960
tcc aag cgc aag act gac gtg tgc tac tac tac cag aag ttc ttc gat Ser Lys Arg Lys Thr Asp Val Cys Tyr Tyr Tyr Gln Lys Phe Phe Asp 325 330 335	1008
agt gcc tgc acg atg ggt gcc tac cac ccg ctg ctg tat gag aag aat Ser Ala Cys Thr Met Gly Ala Tyr His Pro Leu Leu Tyr Glu Lys Asn 340 345 350	1056
ttg gtg aag cat ctc aac cag ggc aca gat gag gac atc tac ctg ctt Leu Val Lys His Leu Asn Gln Gly Thr Asp Glu Asp Ile Tyr Leu Leu 355 360 365	1104
gga aaa gcc aca ctg cct ggc ttc cgg acc att cac tgc Gly Lys Ala Thr Leu Pro Gly Phe Arg Thr Ile His Cys 370 375 380	1143

<210> 182

<211> 381

<212> PRT

<213> Artificial

<220>

<223> Construção Sintética

<400> 182

Met Gly Leu Leu Val Phe Val Arg Ser Trp Met Leu Ala Ala Leu Thr
1 5 10 15
Leu Leu Val Val Phe Leu Ile Phe Ala Lys Glu Pro Gln Thr Lys Pro
 20 25 30
Ser Arg His Gln Arg Thr Glu Asn Ile Lys Glu Arg Ser Leu Gln Ser
 35 40 45
Leu Ala Lys Pro Lys Ser Gln Ala Pro Thr Arg Ala Arg Arg Thr Thr

50	55	60
Glu Ala Ser Phe Gln Val Trp Asn Lys Asp Ser Ser Ser Lys Asn Leu 65 70 75 80		
Ile Pro Arg Leu Gln Lys Ile Trp Lys Asn Tyr Leu Ser Met Asn Lys 85 90 95		
Tyr Lys Val Ser Tyr Lys Gly Pro Gly Pro Gly Ile Lys Phe Ser Ala 100 105 110		
Glu Ala Leu Arg Cys His Leu Arg Asp His Val Asn Val Ser Met Val 115 120 125		
Glu Val Thr Asp Phe Pro Phe Asn Thr Ser Glu Trp Glu Gly Tyr Leu 130 135 140		
Pro Lys Glu Ser Ile Arg Thr Lys Ala Gly Pro Trp Gly Arg Cys Ala 145 150 155 160		
Val Val Ser Ser Ala Gly Ser Leu Lys Ser Ser Gln Leu Gly Arg Glu 165 170 175		
Ile Asp Asp His Asp Ala Val Leu Arg Phe Asn Gly Ala Pro Thr Ala 180 185 190		
Asn Phe Gln Gln Asp Val Gly Thr Lys Thr Thr Ile Arg Leu Met Asn 195 200 205		
Ser Gln Leu Val Thr Thr Glu Lys Arg Phe Leu Lys Asp Ser Leu Tyr 210 215 220		
Asn Glu Gly Ile Leu Ile Val Trp Asp Pro Ser Val Tyr His Ser Asp 225 230 235 240		
Ile Pro Lys Trp Tyr Gln Asn Pro Asp Tyr Asn Phe Phe Asn Asn Tyr 245 250 255		
Lys Thr Tyr Arg Lys Leu His Pro Asn Gln Pro Phe Tyr Ile Leu Lys 260 265 270		
Pro Gln Met Pro Trp Glu Leu Trp Asp Ile Leu Gln Glu Ile Ser Pro 275 280 285		
Glu Glu Ile Gln Pro Asn Pro Pro Ser Ser Gly Met Leu Gly Ile Ile 290 295 300		
Ile Met Met Thr Leu Cys Asp Gln Val Asp Ile Tyr Glu Phe Leu Pro 305 310 315 320		
Ser Lys Arg Lys Thr Asp Val Cys Tyr Tyr Tyr Gln Lys Phe Phe Asp 325 330 335		
Ser Ala Cys Thr Met Gly Ala Tyr His Pro Leu Leu Tyr Glu Lys Asn 340 345 350		
Leu Val Lys His Leu Asn Gln Gly Thr Asp Glu Asp Ile Tyr Leu Leu 355 360 365		
Gly Lys Ala Thr Leu Pro Gly Phe Arg Thr Ile His Cys 370 375 380		

REIVINDICAÇÕES

1. Processo para produção de sequências génicas codificando glicosiltransferases membranares quiméricas apresentando uma actividade de glicosilação otimizada em células transformadas com as referidas sequências, quando comparada com a actividade de glicosilação das correspondentes glicosiltransferases nativas para o substrato aceitador, sendo a referida actividade de glicosilação otimizada pelo menos 30 vezes maior que a actividade inicial das glicosiltransferases inteiras nativas,

o referido processo compreendendo a fusão:

- de um primeiro ácido nucleico codificando para um fragmento mínimo C-terminal do domínio catalítico (CD) da glicosiltransferase inteira nativa,

sendo a referida primeira sequência nucleica obtida através de remoção dos nucleótidos codificando para um ou vários aminoácidos contíguos estendendo-se desde o primeiro aminoácido da extremidade N-terminal do referido CD, e selecção do ácido nucleico codificando o referido fragmento de CD mínimo C-terminal que é tal que se n representa o número de aminoácidos contíguos tal como definido acima que foram suprimidos, então o fragmento obtido quando se suprime pelo menos $n+1$ aminoácidos contíguos tal como definido acima, não tem actividade de transferase substancial,

- com um segundo ácido nucleico de sequência variável codificando para uma cadeia peptídica transmembranar especificando a ancoragem das glicosiltransferases em compartimentos intracelulares e compreendendo na sua região N-terminal a região da cauda citoplasmática (CT) localizada a montante de um domínio transmembranar (TMD), ele próprio localizado a montante de uma região de caule (SR) ou de um fragmento de pelo menos 3 aminoácidos contíguos da SR, sendo a referida SR ou parte desta ligada ao referido domínio catalítico, através de um péptido ligador ou de conexão de pelo menos 2 aminoácidos codificados por um local de restrição que não existe na sequência nucleotídica codificando para o CD nativo mencionado acima,

desde que pelo menos um destes péptidos CT, TMD, SR seja diferente da estrutura primária dos péptidos de ocorrência natural correspondentes presentes na glicosiltransferase

nativa da qual é derivado o fragmento de CD com actividade de glicosiltransferase óptima tal como definido acima,

sendo esta fusão realizada de modo a que o primeiro ácido nucleico fique localizado a jusante do segundo ácido nucleico e proporcione um produto proteico no qual o CD esteja na metade C-terminal.

2. Processo de acordo com a reivindicação 1, caracterizado por o primeiro e segundo ácidos nucleicos serem derivados de sequências nucleotídicas codificando os domínios CD ou CT, TMD, e as regiões SR, respectivamente, em glicosiltransferases de origem eucariótica, de preferência de mamíferos ou humanos, estando as referidas glicosiltransferases envolvidas em:

- O-glicosilação de proteínas em células eucarióticas, tais como as N-acetilgalactosaminil-, N-acetilglucosaminil-, fucosil-, galactosil- ou sialiltransferases,
- N-glicosilação de proteínas em células eucarióticas tais como glucosaminil-, galactosil-, fucosil- ou sialiltransferases,
- Glicosilação de lípidos tais como glucosil-, N-acetilgalactosaminil-, glucosaminil-, fucosil-, galactosil- ou sialiltransferases.

3. Processo de acordo com a reivindicação 1 ou 2, caracterizado por o primeiro e segundo ácidos nucleicos serem derivados de sequências nucleotídicas codificando domínios CD ou CT, TMD, e regiões SR, respectivamente em glucosiltransferases, N-acetilgalactosaminiltransferases, N-acetilglucosaminiltransferases, galactosiltransferases, fucosiltransferases ou sialiltransferases.

4. Processo de acordo com qualquer uma das reivindicações 1 a 3, caracterizado por o primeiro e segundo ácidos nucleicos serem derivados de sequências nucleotídicas codificando os domínios CD ou CT, TMD, e regiões SR, respectivamente, em sialiltransferases.

5. Processo de acordo com qualquer uma das reivindicações 1 a 4, caracterizado por o primeiro e segundo ácidos nucleicos serem derivados de sequências nucleotídicas codificando domínios CD ou CT, TMD, e regiões SR,

respectivamente, em $\alpha 2,6$ -sialiltransferases, $\alpha 2,3$ -sialiltransferases ou $\alpha 2,8$ -sialiltransferases.

6. Processo de acordo com qualquer uma das reivindicações 1 a 5, caracterizado por o primeiro e segundo ácidos nucleicos serem derivados de sequências nucleotídicas codificando domínios CD ou CT, TMD, e regiões SR, respectivamente, em:

- $\alpha 2,6$ -sialiltransferases escolhidas de entre:

* as $\beta 1,4$ -galactósido- $\alpha 2,6$ -sialiltransferases humanas I e II (hST6Gal I e hST6Gal II) representadas por SEQ ID NO:2, e SEQ ID NO:4, respectivamente, codificadas pelas sequências nucleotídicas SEQ ID NO:1, e SEQ ID NO:3, respectivamente, ou

* as N-acetilgalactosaminida- $\alpha 2,6$ -sialiltransferases humanas I a VI (hST6GalNAc I a VI) representadas por SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, e SEQ ID NO:16, respectivamente, codificadas pelas sequências nucleotídicas SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, e SEQ ID NO:15, respectivamente,

- $\alpha 2,3$ -sialiltransferases escolhidas de entre as galactósido- $\alpha 2,3$ -sialiltransferases humanas I a VI (hST3Gal I a VI) representadas por SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, e SEQ ID NO:28, respectivamente, codificadas pelas sequências nucleotídicas SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, e SEQ ID NO:27, respectivamente, ou as galactósido- $\alpha 2,3$ -sialiltransferases de rato I a VI (ratST3Gal I a VI), tais como a rST3Gal III representada por SEQ ID NO:30 codificada pela sequência nucleotídica SEQ ID NO:29 ou qualquer ST de outra origem animal desde que partilhe pelo menos 85% de homologia com a enzima humana,

- $\alpha 2,8$ -sialiltransferases escolhidas de entre as ácido siálico- $\alpha 2,8$ -sialiltransferases humanas I a VI (hST8Sia I a VI) representadas por SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, e SEQ ID NO:42, respectivamente, codificadas pelas sequências nucleotídicas SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, e SEQ ID NO:41, respectivamente.

7. Processo de acordo com qualquer uma das reivindicações 1 a 6, caracterizado por a sequência nucleotídica codificando o domínio CD compreendida no primeiro ácido nucleico ser escolhida de entre as sequências constituídas por, ou compreendendo:

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 268 a 330 e na sua extremidade 3' pelo nucleótido localizado na posição 1218 de SEQ ID NO:1, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST6Gal I delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 90 a 110 e na sua extremidade C-terminal pelo aminoácido localizado na posição 406 de SEQ ID NO:2,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 307 a 369 e na sua extremidade 3' pelo nucleótido localizado na posição 1587 de SEQ ID NO:3, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST6Gal II delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 103 a 123 e na sua extremidade C-terminal pelo aminoácido localizado na posição 529 de SEQ ID NO:4,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 814 a 876 e na sua extremidade 3' pelo nucleótido localizado na posição 1800 de SEQ ID NO:5, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST6GalNAc I delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 272 a 292 e na sua extremidade C-terminal pelo aminoácido localizado na posição 600 de SEQ ID NO:6,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 172 a 234 e na sua extremidade 3' pelo nucleótido localizado na posição 1122 de SEQ ID NO:7, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST6GalNAc II delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 58 a 78 e na sua extremidade C-terminal pelo aminoácido localizado na posição 374 de SEQ ID NO:8,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 73 a 135 e na sua extremidade 3' pelo nucleótido localizado na posição 915 de SEQ ID NO:9, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST6GalNAc III delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 25 a 45 e na sua extremidade C-terminal pelo aminoácido localizado na posição 305 de SEQ ID NO:10,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 61 a 123 e na sua extremidade 3' pelo nucleótido localizado na posição 906 de SEQ ID NO:11, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST6GalNAc IV delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 21 a 41 e na sua extremidade C-terminal pelo aminoácido localizado na posição 302 de SEQ ID NO:12,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 121 a 183 e na sua extremidade 3' pelo nucleótido localizado na posição 1008 de SEQ ID NO:13, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST6GalNAc V delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 41 a 61 e na sua extremidade C-terminal pelo aminoácido localizado na posição 336 de SEQ ID NO:14,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 154 a 216 e na sua extremidade 3' pelo nucleótido localizado na posição 999 de SEQ ID NO:15, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST6GalNAc VI delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 52 a 72 e na sua extremidade C-terminal pelo aminoácido localizado na posição 333 de SEQ ID NO:16,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 145 a 207 e na sua extremidade 3' pelo nucleótido localizado na posição 1020 de SEQ ID NO:17, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST3Gal I delimitada na sua extremidade N-terminal pelo

aminoácido localizado na posição 49 a 69 e na sua extremidade C-terminal pelo aminoácido localizado na posição 340 de SEQ ID NO:18,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 175 a 237 e na sua extremidade 3' pelo nucleótido localizado na posição 1050 de SEQ ID NO:19, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST3Gal II delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 59 a 79 e na sua extremidade C-terminal pelo aminoácido localizado na posição 350 de SEQ ID NO:20,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 199 a 261 e na sua extremidade 3' pelo nucleótido localizado na posição 1332 de SEQ ID NO:21, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST3Gal III delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 67 a 87 e na sua extremidade C-terminal pelo aminoácido localizado na posição 444 de SEQ ID NO:22,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 79 a 141 e na sua extremidade 3' pelo nucleótido localizado na posição 987 de SEQ ID NO:23, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST3Gal IV delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 27 a 47 e na sua extremidade C-terminal pelo aminoácido localizado na posição 329 de SEQ ID NO:24,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 136 a 198 e na sua extremidade 3' pelo nucleótido localizado na posição 1086 de SEQ ID NO:25, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST3Gal V delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 46 a 66 e na sua extremidade C-terminal pelo aminoácido localizado na posição 362 de SEQ ID NO:26,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 73 a 135 e na sua extremidade 3' pelo nucleótido localizado na posição 993 de

SEQ ID NO:27, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST3Gal VI delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 25 a 45 e na sua extremidade C-terminal pelo aminoácido localizado na posição 331 de SEQ ID NO:28,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 103 a 165 e na sua extremidade 3' pelo nucleótido localizado na posição 1122 de SEQ ID NO:29, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de ratST3Gal III delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 35 a 55 e na sua extremidade C-terminal pelo aminoácido localizado na posição 374 de SEQ ID NO:30,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 133 a 195 e na sua extremidade 3' pelo nucleótido localizado na posição 1068 de SEQ ID NO:31, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST8Sia I delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 45 a 65 e na sua extremidade C-terminal pelo aminoácido localizado na posição 356 de SEQ ID NO:32,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 190 a 252 e na sua extremidade 3' pelo nucleótido localizado na posição 1125 de SEQ ID NO:33, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST8Sia II delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 64 a 84 e na sua extremidade C-terminal pelo aminoácido localizado na posição 375 de SEQ ID NO:34,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 205 a 267 e na sua extremidade 3' pelo nucleótido localizado na posição 1140 de SEQ ID NO:35, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hSTBSia III delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 69 a 89 e na sua extremidade C-terminal pelo aminoácido localizado na posição 380 de SEQ ID NO:36,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 145 a 207 e na sua extremidade 3' pelo nucleótido localizado na posição 1077 de SEQ ID NO:37, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST8Sia IV delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 49 a 69 e na sua extremidade C-terminal pelo aminoácido localizado na posição 359 de SEQ ID NO:38,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 211 a 273 e na sua extremidade 3' pelo nucleótido localizado na posição 1128 de SEQ ID NO:39, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST8Sia V delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 71 a 91 e na sua extremidade C-terminal pelo aminoácido localizado na posição 376 de SEQ ID NO:40,

- a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 277 a 339 e na sua extremidade 3' pelo nucleótido localizado na posição 1194 de SEQ ID NO:41, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo ao domínio CD de hST8Sia VI delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 93 a 113 e na sua extremidade C-terminal pelo aminoácido localizado na posição 398 de SEQ ID NO:42.

8. Processo de acordo com qualquer uma das reivindicações 1 a 7, caracterizado por o primeiro ácido nucleico ser a sequência nucleotídica SEQ ID NO:43 correspondendo à sequência delimitada pelos nucleótidos localizados na posição 268 e 1218 de SEQ ID NO:1, a referida sequência nucleotídica SEQ ID NO:43 codificando a sequência polipeptídica SEQ ID NO:44 correspondendo ao fragmento mínimo C-terminal do domínio CD de hST6Gal I delimitada pelos aminoácidos localizados nas posições 90 a 406 de SEQ ID NO:2.

9. Processo de acordo com qualquer uma das reivindicações 1 a 6, caracterizado por:

- a sequência nucleotídica codificando a região CT compreendida no segundo ácido nucleico ser escolhida de entre:

* a sequência nucleotídica SEQ ID NO:45 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 27 de SEQ ID NO:1, a referida sequência nucleotídica SEQ ID NO:45 codificando a sequência polipeptídica SEQ ID NO:46 correspondendo à região CT de hST6Gal I delimitada pelos aminoácidos localizados nas posições 1 a 9 de SEQ ID NO:2,

* a sequência nucleotídica SEQ ID NO 47 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 30 de SEQ ID NO:3, a referida sequência nucleotídica SEQ ID NO:47 codificando a sequência polipeptídica SEQ ID NO:48 correspondendo à região CT de hST6Gal II delimitada pelos aminoácidos localizados nas posições 1 a 10 de SEQ ID NO:4,

* a sequência nucleotídica SEQ ID NO:49 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 42 de SEQ ID NO:5, a referida sequência nucleotídica SEQ ID NO:49 codificando a sequência polipeptídica SEQ ID NO:50 correspondendo à região CT de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 1 a 14 de SEQ ID NO:6,

* a sequência nucleotídica SEQ ID NO:51 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 21 de SEQ ID NO:7, a referida sequência nucleotídica SEQ ID NO:51 codificando a sequência polipeptídica SEQ ID NO:52 correspondendo à região CT de hST6GalNAc II delimitada pelos aminoácidos localizados nas posições 1 a 7 de SEQ ID NO:8,

* a sequência nucleotídica SEQ ID NO:53 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 24 de SEQ ID NO:9, a referida sequência nucleotídica SEQ ID NO:53 codificando a sequência polipeptídica SEQ ID NO:54 correspondendo à região CT de hST6GalNAc III delimitada pelos aminoácidos localizados nas posições 1 a 8 de SEQ ID NO:10,

* a sequência nucleotídica SEQ ID NO:55 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 18 de SEQ ID NO:11, a referida sequência nucleotídica SEQ ID NO:55 codificando a

sequência polipeptídica SEQ ID NO:56 correspondendo à região CT de hST6GalNAc IV delimitada pelos aminoácidos localizados nas posições 1 a 6 de SEQ ID NO:12,

* a sequência nucleotídica SEQ ID NO:57 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 24 de SEQ ID NO:13, a referida sequência nucleotídica SEQ ID NO:57 codificando a sequência polipeptídica SEQ ID NO:58 correspondendo à região CT de hST6GalNAc V delimitada pelos aminoácidos localizados nas posições 1 a 8 de SEQ ID NO:14,

* a sequência nucleotídica SEQ ID NO:59 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 129 de SEQ ID NO:15, a referida sequência nucleotídica SEQ ID NO:59 codificando a sequência polipeptídica SEQ ID NO:60 correspondendo à região CT de hST6GalNAc VI delimitada pelos aminoácidos localizados nas posições 1 a 43 de SEQ ID NO:16,

* a sequência nucleotídica SEQ ID NO:61 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 39 de SEQ ID NO:17, a referida sequência nucleotídica SEQ ID NO:61 codificando a sequência polipeptídica SEQ ID NO:62 correspondendo à região CT de hST3Gal I delimitada pelos aminoácidos localizados nas posições 1 a 13 de SEQ ID NO:18,

* a sequência nucleotídica SEQ ID NO:63 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 18 de SEQ ID NO:19, a referida sequência nucleotídica SEQ ID NO:63 codificando a sequência polipeptídica SEQ ID NO:64 correspondendo à região CT de hST3Gal II delimitada pelos aminoácidos localizados nas posições 1 a 6 de SEQ ID NO:20,

* a sequência nucleotídica SEQ ID NO:65 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 24 de SEQ ID NO:21, a referida sequência nucleotídica SEQ ID NO:65 codificando a sequência polipeptídica SEQ ID NO:66 correspondendo à região CT de hST3Gal III delimitada pelos aminoácidos localizados nas posições 1 a 8 de SEQ ID NO:22,

* a sequência nucleotídica SEQ ID NO:67 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 24 de SEQ ID NO:23, a referida sequência nucleotídica SEQ ID NO:67 codificando a

sequência polipeptídica SEQ ID NO:68 correspondendo à região CT de hST3Gal IV delimitada pelos aminoácidos localizados nas posições 1 a 8 de SEQ ID NO:24,

* a sequência nucleotídica SEQ ID NO:69 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 15 de SEQ ID NO:25, a referida sequência nucleotídica SEQ ID NO:69 codificando a sequência polipeptídica SEQ ID NO:70 correspondendo à região CT de hST3Gal V delimitada pelos aminoácidos localizados nas posições 1 a 5 de SEQ ID NO:26,

* a sequência nucleotídica SEQ ID NO:71 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 12 de SEQ ID NO:27, a referida sequência nucleotídica SEQ ID NO:71 codificando a sequência polipeptídica SEQ ID NO:72 correspondendo à região CT de hST3Gal VI delimitada pelos aminoácidos localizados nas posições 1 a 4 de SEQ ID NO:28,

* a sequência nucleotídica SEQ ID NO:73 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 24 de SEQ ID NO:29, a referida sequência nucleotídica SEQ ID NO:73 codificando a sequência polipeptídica SEQ ID NO:74 correspondendo à região CT de ratST3Gal III delimitada pelos aminoácidos localizados nas posições 1 a 8 de SEQ ID NO:30,

* a sequência nucleotídica SEQ ID NO:75 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 87 de SEQ ID NO:31, a referida sequência nucleotídica SEQ ID NO:75 codificando a sequência polipeptídica SEQ ID NO:76 correspondendo à região CT de hST8Sia I delimitada pelos aminoácidos localizados nas posições 1 a 29 de SEQ ID NO:32,

* a sequência nucleotídica SEQ ID NO:77 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 18 de SEQ ID NO:33, a referida sequência nucleotídica SEQ ID NO:77 codificando a sequência polipeptídica SEQ ID NO:78 correspondendo à região CT de hST8Sia II delimitada pelos aminoácidos localizados nas posições 1 a 6 de SEQ ID NO:34,

* a sequência nucleotídica SEQ ID NO:79 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 27 de SEQ ID NO:35, a referida sequência nucleotídica SEQ ID NO:79 codificando a

sequência polipeptídica SEQ ID NO:80 correspondendo à região CT de hST8Sia III delimitada pelos aminoácidos localizados nas posições 1 a 9 de SEQ ID NO:36,

* a sequência nucleotídica SEQ ID NO:81 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 21 de SEQ ID NO:37, a referida sequência nucleotídica SEQ ID NO:81 codificando a sequência polipeptídica SEQ ID NO:82 correspondendo à região CT de hST8Sia IV delimitada pelos aminoácidos localizados nas posições 1 a 7 de SEQ ID NO:38,

* a sequência nucleotídica SEQ ID NO:83 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 51 de SEQ ID NO:39, a referida sequência nucleotídica SEQ ID NO:83 codificando a sequência polipeptídica SEQ ID NO:84 correspondendo à região CT de hST8Sia V delimitada pelos aminoácidos localizados nas posições 1 a 17 de SEQ ID NO:40,

* a sequência nucleotídica SEQ ID NO:85 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 9 de SEQ ID NO:41, a referida sequência nucleotídica SEQ ID NO:85 codificando a sequência polipeptídica SEQ ID NO:86 correspondendo à região CT de hST8Sia VI delimitada pelos aminoácidos localizados nas posições 1 a 3 de SEQ ID NO:42,

- a sequência nucleotídica codificando a região TMD compreendida no segundo ácido nucleico ser escolhida de entre:

* a sequência nucleotídica SEQ ID NO:87 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 28 e 78 de SEQ ID NO:1, a referida sequência nucleotídica SEQ ID NO:87 codificando a sequência polipeptídica SEQ ID NO:88 correspondendo à região TMD de hST6Gal I delimitada pelos aminoácidos localizados nas posições 10 a 26 de SEQ ID NO:2,

* a sequência nucleotídica SEQ ID NO:89 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 31 e 90 de SEQ ID NO:3, a referida sequência nucleotídica SEQ ID NO:89 codificando a sequência polipeptídica SEQ ID NO:90 correspondendo à região TMD de hST6Gal II delimitada pelos aminoácidos localizados nas posições 11 a 30 de SEQ ID NO:4,

* a sequência nucleotídica SEQ ID NO:91 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 43 e 105 de SEQ ID NO:5, a referida sequência nucleotídica SEQ ID NO:91 codificando a sequência polipeptídica SEQ ID NO:92 correspondendo à região TMD de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 15 a 35 de SEQ ID NO:6,

* a sequência nucleotídica SEQ ID NO:93 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 22 e 84 de SEQ ID NO:7, a referida sequência nucleotídica SEQ ID NO:91 codificando a sequência polipeptídica SEQ ID NO:92 correspondendo à região TMD de hST6GalNAc II delimitada pelos aminoácidos localizados nas posições 8 a 28 de SEQ ID NO:8,

* a sequência nucleotídica SEQ ID NO:95 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 25 e 84 de SEQ ID NO:9, a referida sequência nucleotídica SEQ ID NO:95 codificando a sequência polipeptídica SEQ ID NO:96 correspondendo à região TMD de hST6GalNAc III delimitada pelos aminoácidos localizados nas posições 9 a 28 de SEQ ID NO:10,

* a sequência nucleotídica SEQ ID NO:97 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 19 e 81 de SEQ ID NO:11, a referida sequência nucleotídica SEQ ID NO:97 codificando a sequência polipeptídica SEQ ID NO:98 correspondendo à região TMD de hST6GalNAc IV delimitada pelos aminoácidos localizados nas posições 7 a 27 de SEQ ID NO:12,

* a sequência nucleotídica SEQ ID NO:99 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 25 e 87 de SEQ ID NO:13, a referida sequência nucleotídica SEQ ID NO:99 codificando a sequência polipeptídica SEQ ID NO:100 correspondendo à região TMD de hST6GalNAc V delimitada pelos aminoácidos localizados nas posições 9 a 29 de SEQ ID NO:14,

* a sequência nucleotídica SEQ ID NO:101 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 130 e 177 de SEQ ID NO:15, a referida sequência nucleotídica SEQ ID NO:101 codificando a sequência polipeptídica SEQ ID NO:102 correspondendo à região TMD de hST6GalNAc VI delimitada pelos aminoácidos localizados nas posições 44 a 59 de SEQ ID NO:16,

* a sequência nucleotídica SEQ ID NO:103 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 40 e 102 de SEQ ID NO:17, a referida sequência nucleotídica SEQ ID NO:103 codificando a sequência polipeptídica SEQ ID NO:104 correspondendo à região TMD de hST3Gal I delimitada pelos aminoácidos localizados nas posições 14 a 34 de SEQ ID NO:18,

* a sequência nucleotídica SEQ ID NO:105 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 19 e 81 de SEQ ID NO:19, a referida sequência nucleotídica SEQ ID NO:105 codificando a sequência polipeptídica SEQ ID NO:106 correspondendo à região TMD de hST3Gal II delimitada pelos aminoácidos localizados nas posições 7 a 27 de SEQ ID NO:20,

* a sequência nucleotídica SEQ ID NO:107 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 25 e 84 de SEQ ID NO:21, a referida sequência nucleotídica SEQ ID NO:107 codificando a sequência polipeptídica SEQ ID NO:108 correspondendo à região TMD de hST3Gal III delimitada pelos aminoácidos localizados nas posições 9 a 28 de SEQ ID NO:22,

* a sequência nucleotídica SEQ ID NO:109 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 25 e 78 de SEQ ID NO:23, a referida sequência nucleotídica SEQ ID NO:109 codificando a sequência polipeptídica SEQ ID NO:110 correspondendo à região TMD de hST3Gal IV delimitada pelos aminoácidos localizados nas posições 9 a 26 de SEQ ID NO:24,

* a sequência nucleotídica SEQ ID NO:111 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 16 e 78 de SEQ ID NO:25, a referida sequência nucleotídica SEQ ID NO:111 codificando a sequência polipeptídica SEQ ID NO:112 correspondendo à região TMD de hST3Gal V delimitada pelos aminoácidos localizados nas posições 6 a 26 de SEQ ID NO:26,

* a sequência nucleotídica SEQ ID NO:113 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 13 e 75 de SEQ ID NO:27, a referida sequência nucleotídica SEQ ID NO:113 codificando a sequência polipeptídica SEQ ID NO:114 correspondendo à região TMD de hST3Gal VI delimitada pelos aminoácidos localizados nas posições 5 a 25 de SEQ ID NO:28,

* a sequência nucleotídica SEQ ID NO:115 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 25 e 84 de SEQ ID NO:29, a referida sequência nucleotídica SEQ ID NO:115 codificando a sequência polipeptídica SEQ ID NO:116 correspondendo à região TMD de ratST3Gal III delimitada pelos aminoácidos localizados nas posições 9 a 28 de SEQ ID NO:30,

* a sequência nucleotídica SEQ ID NO:117 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 88 e 144 de SEQ ID NO:31, a referida sequência nucleotídica SEQ ID NO:117 codificando a sequência polipeptídica SEQ ID NO:118 correspondendo à região TMD de hST8Sia I delimitada pelos aminoácidos localizados nas posições 30 a 48 de SEQ ID NO:32,

* a sequência nucleotídica SEQ ID NO:119 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 19 e 69 de SEQ ID NO:33, a referida sequência nucleotídica SEQ ID NO:119 codificando a sequência polipeptídica SEQ ID NO:120 correspondendo à região TMD de hST8Sia II delimitada pelos aminoácidos localizados nas posições 7 a 23 de SEQ ID NO:34,

* a sequência nucleotídica SEQ ID NO:121 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 28 e 99 de SEQ ID NO:35, a referida sequência nucleotídica SEQ ID NO:121 codificando a sequência polipeptídica SEQ ID NO:122 correspondendo à região TMD de hST8Sia III delimitada pelos aminoácidos localizados nas posições 10 a 33 de SEQ ID NO:36,

* a sequência nucleotídica SEQ ID NO:123 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 22 e 60 de SEQ ID NO:37, a referida sequência nucleotídica SEQ ID NO:123 codificando a sequência polipeptídica SEQ ID NO:124 correspondendo à região TMD de hST8Sia IV delimitada pelos aminoácidos localizados nas posições 8 a 20 de SEQ ID NO:38,

* a sequência nucleotídica SEQ ID NO:125 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 52 e 114 de SEQ ID NO:39, a referida sequência nucleotídica SEQ ID NO:125 codificando a sequência polipeptídica SEQ ID NO:126 correspondendo à região TMD de hST8Sia V delimitada pelos aminoácidos localizados nas posições 18 a 38 de SEQ ID NO:40,

* a sequência nucleotídica SEQ ID NO:127 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 10 e 72 de SEQ ID NO:41, a referida sequência nucleotídica SEQ ID NO:127 codificando a sequência polipeptídica SEQ ID NO:128 correspondendo à região TMD de hST8Sia VI delimitada pelos aminoácidos localizados nas posições 4 a 24 de SEQ ID NO:42,

- a sequência nucleotídica codificando a região SR compreendida no segundo ácido nucleico ou codificando um fragmento de pelo menos 3 aminoácidos deste, ser escolhida de entre as sequências constituídas por, ou compreendendo:

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 79 e na sua extremidade 3' pelo nucleótido localizado na posição 267 a 327 de SEQ ID NO:1, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST6Gal I delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 27 e na sua extremidade C-terminal pelo aminoácido localizado na posição 89 a 109 de SEQ ID NO:2,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 91 e na sua extremidade 3' pelo nucleótido localizado na posição 306 a 336 de SEQ ID NO:3, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST6Gal II delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 31 e na sua extremidade C-terminal pelo aminoácido localizado na posição 102 a 112 de SEQ ID NO:4,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 106 e na sua extremidade 3' pelo nucleótido localizado na posição 813 a 873 de SEQ ID NO:5, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST6GalNAc I delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 36 e na sua extremidade C-terminal pelo aminoácido localizado na posição 271 a 291 de SEQ ID NO:6,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 85 e na sua extremidade 3' pelo nucleótido localizado na posição 171 a 231 de SEQ ID NO:7, a referida sequência nucleotídica

codificando a sequência polipeptídica correspondendo à região SR de hST6GalNAc II delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 29 e na sua extremidade C-terminal pelo aminoácido localizado na posição 57 a 77 de SEQ ID NO:8,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 85 e na sua extremidade 3' pelo nucleótido localizado na posição 102 a 132 de SEQ ID NO:9, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST6GalNAc III delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 29 e na sua extremidade C-terminal pelo aminoácido localizado na posição 34 a 44 de SEQ ID NO:10,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 82 e na sua extremidade 3' pelo nucleótido localizado na posição 90 a 120 de SEQ ID NO:11, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST6GalNAc IV delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 28 e na sua extremidade C-terminal pelo aminoácido localizado na posição 30 a 40 de SEQ ID NO:12,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 88 e na sua extremidade 3' pelo nucleótido localizado na posição 120 a 180 de SEQ ID NO:13, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST6GalNAc V delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 30 e na sua extremidade C-terminal pelo aminoácido localizado na posição 40 a 60 de SEQ ID NO:14,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 178 e na sua extremidade 3' pelo nucleótido localizado na posição 183 de SEQ ID NO:15, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST6GalNAc VI delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 60 e na sua extremidade C-terminal pelo aminoácido localizado na posição 61 de SEQ ID NO:16,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 103 e na sua extremidade 3' pelo nucleótido localizado na posição 144 a 204 de SEQ ID NO:17, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST3Gal I delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 35 e na sua extremidade C-terminal pelo aminoácido localizado na posição 48 a 68 de SEQ ID NO:18,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 82 e na sua extremidade 3' pelo nucleótido localizado na posição 174 a 234 de SEQ ID NO:19, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST3Gal II delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 28 e na sua extremidade C-terminal pelo aminoácido localizado na posição 58 a 78 de SEQ ID NO:20,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 85 e na sua extremidade 3' pelo nucleótido localizado na posição 198 a 258 de SEQ ID NO:21, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST3Gal III delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 29 e na sua extremidade C-terminal pelo aminoácido localizado na posição 66 a 86 de SEQ ID NO:22,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 79 e na sua extremidade 3' pelo nucleótido localizado na posição 108 a 138 de SEQ ID NO:23, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST3Gal IV delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 27 e na sua extremidade C-terminal pelo aminoácido localizado na posição 36 a 46 de SEQ ID NO:24,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 79 e na sua extremidade 3' pelo nucleótido localizado na posição 135 a 195 de SEQ ID NO:25, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST3Gal V delimitada na sua extremidade N-terminal pelo

aminoácido localizado na posição 27 e na sua extremidade C-terminal pelo aminoácido localizado na posição 45 a 65 de SEQ ID NO:26,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 76 e na sua extremidade 3' pelo nucleótido localizado na posição 102 a 132 de SEQ ID NO:27, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST3Gal VI delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 26 e na sua extremidade C-terminal pelo aminoácido localizado na posição 34 a 44 de SEQ ID NO:28,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 85 e na sua extremidade 3' pelo nucleótido localizado na posição 105 a 165 de SEQ ID NO:29, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de ratST3Gal III delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 29 e na sua extremidade C-terminal pelo aminoácido localizado na posição 35 a 55 de SEQ ID NO:30,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 145 e na sua extremidade 3' pelo nucleótido localizado na posição 162 a 192 de SEQ ID NO:31, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST8Sia I delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 49 e na sua extremidade C-terminal pelo aminoácido localizado na posição 54 a 64 de SEQ ID NO:32,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 70 e na sua extremidade 3' pelo nucleótido localizado na posição 189 a 249 de SEQ ID NO:33, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST8Sia II delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 24 e na sua extremidade C-terminal pelo aminoácido localizado na posição 63 a 83 de SEQ ID NO:34,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 100 e na sua extremidade 3' pelo nucleótido localizado na posição 204

a 264 de SEQ ID NO:35, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST8Sia III delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 34 e na sua extremidade C-terminal pelo aminoácido localizado na posição 68 a 88 de SEQ ID NO:36,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 61 e na sua extremidade 3' pelo nucleótido localizado na posição 144 a 204 de SEQ ID NO:37, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST8Sia IV delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 21 e na sua extremidade C-terminal pelo aminoácido localizado na posição 48 a 68 de SEQ ID NO:38,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 115 e na sua extremidade 3' pelo nucleótido localizado na posição 210 a 270 de SEQ ID NO:39, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST8Sia V delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 39 e na sua extremidade C-terminal pelo aminoácido localizado na posição 70 a 90 de SEQ ID NO:40,

* a sequência nucleotídica delimitada na sua extremidade 5' pelo nucleótido localizado na posição 73 e na sua extremidade 3' pelo nucleótido localizado na posição 276 a 336 de SEQ ID NO:41, a referida sequência nucleotídica codificando a sequência polipeptídica correspondendo à região SR de hST8Sia VI delimitada na sua extremidade N-terminal pelo aminoácido localizado na posição 25 e na sua extremidade C-terminal pelo aminoácido localizado na posição 92 a 112 de SEQ ID NO:42,

* ou qualquer fragmento de pelo menos 6 nucleótidos das sequências nucleotídicas codificando as sequências polipeptídicas correspondendo às regiões SR mencionadas acima, e codificando pelo menos 2 aminoácidos contíguos das referidas regiões SR, tais como:

** o fragmento SEQ ID NO:129 delimitado pelos nucleótidos localizados nas posições 106 a 222 de SEQ ID NO:5, codificando a sequência polipeptídica SEQ ID NO:130 correspondendo ao fragmento da região SR de hST6GalNAc I

delimitada pelos aminoácidos localizados nas posições 36 a 74 de SEQ ID NO:6,

** o fragmento SEQ ID NO:131 delimitado pelos nucleótidos localizados nas posições 109 a 222 de SEQ ID NO:5, codificando a sequência polipeptídica SEQ ID NO:132 correspondendo ao fragmento da região SR de hST6GaINAc I delimitada pelos aminoácidos localizados nas posições 37 a 74 de SEQ ID NO:6,

** o fragmento SEQ ID NO:133 delimitado pelos nucleótidos localizados nas posições 106 a 420 de SEQ ID NO:5, codificando a sequência polipeptídica SEQ ID NO:134 correspondendo ao fragmento da região SR de hST6GalNac I delimitada pelos aminoácidos localizados nas posições 36 a 140 de SEQ ID NO:6,

** o fragmento SEQ ID NO:135 delimitado pelos nucleótidos localizados nas posições 106 a 774 de SEQ ID NO:5, codificando a sequência polipeptídica SEQ ID NO:136 correspondendo ao fragmento da região SR de hST6GaINAc I delimitada pelos aminoácidos localizados nas posições 36 a 258 de SEQ ID NO:6,

** o fragmento SEQ ID NO:137 delimitado pelos nucleótidos localizados nas posições 85 a 138 de SEQ ID NO:21, codificando a sequência polipeptídica SEQ ID NO:138 correspondendo ao fragmento da região SR de hST3Gal III delimitada pelos aminoácidos localizados nas posições 29 a 46 de SEQ ID NO:22,

** o fragmento SEQ ID NO:139 delimitado pelos nucleótidos localizados nas posições 85 a 138 de SEQ ID NO:29, codificando a sequência polipeptídica SEQ ID NO:140 correspondendo ao fragmento da região SR de rST3Gal III delimitada pelos aminoácidos localizados nas posições 29 a 46 de SEQ ID NO:30,

** o fragmento SEQ ID NO:141 delimitado pelos nucleótidos localizados nas posições 70 a 237 de SEQ ID NO:33, codificando a sequência polipeptídica SEQ ID NO:142 correspondendo ao fragmento da região SR de hST8Sia II delimitada pelos aminoácidos localizados nas posições 24 a 79 de SEQ ID NO:34,

** o fragmento SEQ ID NO:143 delimitado pelos nucleótidos localizados nas posições 61 a 201 de SEQ ID NO:37, codificando a sequência polipeptídica SEQ ID NO:144 correspondendo ao fragmento da região SR de hST8Sia IV

delimitada pelos aminoácidos localizados nas posições 21 a 67 de SEQ ID NO:38.

10. Processo de acordo com a reivindicação 9, caracterizado por:

- a sequência nucleotídica codificando a região CT compreendida no segundo ácido nucleico ser escolhida de entre:

* a sequência nucleotídica SEQ ID NO:49 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 42 de SEQ ID NO:5, a referida sequência nucleotídica SEQ ID NO:49 codificando a sequência polipeptídica SEQ ID NO:50 correspondendo à região CT de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 1 a 14 de SEQ ID NO:6,

* a sequência nucleotídica SEQ ID NO:65 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 24 de SEQ ID NO:21, a referida sequência nucleotídica SEQ ID NO:65 codificando a sequência polipeptídica SEQ ID NO:66 correspondendo à região CT de hST3Gal III delimitada pelos aminoácidos localizados nas posições 1 a 8 de SEQ ID NO:22,

* a sequência nucleotídica SEQ ID NO:73 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 24 de SEQ ID NO:29, a referida sequência nucleotídica SEQ ID NO:73 codificando a sequência polipeptídica SEQ ID NO:74 correspondendo à região CT de ratST3Gal III delimitada pelos aminoácidos localizados nas posições 1 a 8 de SEQ ID NO:30,

* a sequência nucleotídica SEQ ID NO:77 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 18 de SEQ ID NO:33, a referida sequência nucleotídica SEQ ID NO:77 codificando a sequência polipeptídica SEQ ID NO:78 correspondendo à região CT de hST8Sia II delimitada pelos aminoácidos localizados nas posições 1 a 6 de SEQ ID NO:34,

* a sequência nucleotídica SEQ ID NO:81 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 1 e 21 de SEQ ID NO:37, a referida sequência nucleotídica SEQ ID NO:81 codificando a sequência polipeptídica SEQ ID NO:82 correspondendo à região

CT de hST8Sia IV delimitada pelos aminoácidos localizados nas posições 1 a 7 de SEQ ID NO:38,

- a sequência nucleotídica codificando a região TMD compreendida no segundo ácido nucleico ser escolhida de entre:

* a sequência nucleotídica SEQ ID NO:91 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 43 e 105 de SEQ ID NO:5, a referida sequência nucleotídica SEQ ID NO:91 codificando a sequência polipeptídica SEQ ID NO:92 correspondendo à região TMD de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 15 a 35 de SEQ ID NO:6,

* a sequência nucleotídica SEQ ID NO:107 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 25 e 84 de SEQ ID NO:21, a referida sequência nucleotídica SEQ ID NO:107 codificando a sequência polipeptídica SEQ ID NO:108 correspondendo à região TMD de hST3Gal III delimitada pelos aminoácidos localizados nas posições 9 a 28 de SEQ ID NO:22,

* a sequência nucleotídica SEQ ID NO:115 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 25 e 84 de SEQ ID NO:29, a referida sequência nucleotídica SEQ ID NO:115 codificando a sequência polipeptídica SEQ ID NO:116 correspondendo à região TMD de ratST3Gal III delimitada pelos aminoácidos localizados nas posições 9 a 28 de SEQ ID NO:30,

* a sequência nucleotídica SEQ ID NO:119 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 19 e 69 de SEQ ID NO:33, a referida sequência nucleotídica SEQ ID NO:119 codificando a sequência polipeptídica SEQ ID NO:120 correspondendo à região TMD de hST8Sia II delimitada pelos aminoácidos localizados nas posições 7 a 23 de SEQ ID NO:34,

* a sequência nucleotídica SEQ ID NO:123 correspondendo à sequência nucleotídica delimitada pelos nucleótidos localizados nas posições 22 e 60 de SEQ ID NO:37, a referida sequência nucleotídica SEQ ID NO:123 codificando a sequência polipeptídica SEQ ID NO:124 correspondendo à região TMD de hST8Sia IV delimitada pelos aminoácidos localizados nas posições 8 a 20 de SEQ ID NO:38,

- a sequência nucleotídica codificando a região SR, ou um fragmento desta, compreendida no segundo ácido nucleico, ser escolhida de entre:

* a sequência SEQ ID NO:129 delimitada pelos nucleótidos localizados nas posições 106 a 222 de SEQ ID NO:5, codificando a sequência polipeptídica SEQ ID NO:130 correspondendo ao fragmento da região SR de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 36 a 74 de SEQ ID NO:6,

* a sequência SEQ ID NO:131 delimitada pelos nucleótidos localizados nas posições 109 a 222 de SEQ ID NO:5, codificando a sequência polipeptídica SEQ ID NO:132 correspondendo ao fragmento da região SR de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 37 a 74 de SEQ ID NO:6,

* a sequência SEQ ID NO:133 delimitada pelos nucleótidos localizados nas posições 106 a 420 de SEQ ID NO:5, codificando a sequência polipeptídica SEQ ID NO:134 correspondendo ao fragmento da região SR de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 36 a 140 de SEQ ID NO:6,

* a sequência SEQ ID NO:135 delimitada pelos nucleótidos localizados nas posições 106 a 774 de SEQ ID NO:5, codificando a sequência polipeptídica SEQ ID NO:136 correspondendo ao fragmento da região SR de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 36 a 258 de SEQ ID NO:6,

* a sequência SEQ ID NO:137 delimitada pelos nucleótidos localizados nas posições 85 a 138 de SEQ ID NO:21, codificando a sequência polipeptídica SEQ ID NO:138 correspondendo ao fragmento da região SR de hST3Gal III delimitada pelos aminoácidos localizados nas posições 29 a 46 de SEQ ID NO:22,

* a sequência SEQ ID NO:139 delimitada pelos nucleótidos localizados nas posições 85 a 138 de SEQ ID NO:29, codificando a sequência polipeptídica SEQ ID NO:140 correspondendo ao fragmento da região SR de rST3Gal III delimitada pelos aminoácidos localizados nas posições 29 a 46 de SEQ ID NO:30,

* a sequência SEQ ID NO:141 delimitada pelos nucleótidos localizados nas posições 70 a 237 de SEQ ID NO:33, codificando a sequência polipeptídica SEQ ID NO:142

correspondendo ao fragmento da região SR de hST8Sia II delimitada pelos aminoácidos localizados nas posições 24 a 79 de SEQ ID NO:34,

* a sequência SEQ ID NO:143 delimitada pelos nucleótidos localizados nas posições 61 a 201 de SEQ ID NO:37, codificando a sequência polipeptídica SEQ ID NO:144 correspondendo ao fragmento da região SR de hST8Sia IV delimitada pelos aminoácidos localizados nas posições 21 a 67 de SEQ ID NO:38.

11. Processo de acordo com a reivindicação 9 ou 10, caracterizado por os péptidos CT, TMD, SR ou fragmento de SR compreendidos no segundo ácido nucleico, serem sequências homólogas derivando da mesma glicosiltransferase nativa, sendo estes últimos diferente dos péptidos na glicosiltransferase nativa da qual é derivado o fragmento CD com actividade de glicosiltransferase óptima tal como definido na reivindicação 1.

12. Processo de acordo com qualquer uma das reivindicações 9 a 11, caracterizado por o segundo ácido nucleico ser escolhido de entre as seguintes sequências:

- a sequência SEQ ID NO:145 delimitada pelos nucleótidos localizados nas posições 1 a 222 de SEQ ID NO:5, contendo SEQ ID NO:49, 91 e 129, e codificando a sequência polipeptídica SEQ ID NO:146 correspondendo ao fragmento de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 1 a 74 de SEQ ID NO:6 e contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 130, respectivamente,

- a sequência SEQ ID NO:147 delimitada pelos nucleótidos localizados nas posições 1 a 420 de SEQ ID NO:5, contendo SEQ ID NO:49, 91 e 133, e codificando a sequência polipeptídica SEQ ID NO:148 correspondendo ao fragmento da região SR de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 1 a 140 de SEQ ID NO:6 e contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 134, respectivamente,

- a sequência SEQ ID NO:149 delimitada pelos nucleótidos localizados nas posições 1 a 774 de SEQ ID NO:5, contendo SEQ ID NO:49, 91 e 135, e codificando a sequência polipeptídica SEQ ID NO:150 correspondendo ao fragmento da região SR de

hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 1 a 258 de SEQ ID NO:6 e contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 136, respectivamente,

- a sequência SEQ ID NO:151 delimitada pelos nucleótidos localizados nas posições 1 a 138 de SEQ ID NO:21, contendo SEQ ID NO:65, 107, e 137, e codificando a sequência polipeptídica SEQ ID NO:152 correspondendo ao fragmento da região SR de hST3Gal III delimitada pelos aminoácidos localizados nas posições 1 a 46 de SEQ ID NO:22 e contendo as regiões CT, TMD e fragmento de SR de hST3Gal III correspondendo a SEQ ID NO:66, 108 e 138, respectivamente,

- a sequência SEQ ID NO:153 delimitada pelos nucleótidos localizados nas posições 1 a 138 de SEQ ID NO:29, contendo SEQ ID NO:73, 115 e 139, e codificando a sequência polipeptídica SEQ ID NO:154 correspondendo ao fragmento da região SR de ratST3Gal III delimitada pelos aminoácidos localizados nas posições 1 a 46 de SEQ ID NO:30 e contendo as regiões CT, TMD e fragmento de SR de hST3Gal III correspondendo a SEQ ID NO:74, 116 e 140, respectivamente,

- a sequência SEQ ID NO:155 delimitada pelos nucleótidos localizados nas posições 1 a 237 de SEQ ID NO:33, contendo SEQ ID NO:77, 119 e 141, e codificando a sequência polipeptídica SEQ ID NO:156 correspondendo ao fragmento da região SR de hST8Sia II delimitada pelos aminoácidos localizados nas posições 1 a 79 de SEQ ID NO:34 e contendo as regiões CT, TMD e SR de hST8Sia II correspondendo a SEQ ID NO:78, 120 e 142, respectivamente,

- a sequência SEQ ID NO:157 delimitada pelos nucleótidos localizados nas posições 1 a 201 de SEQ ID NO:37, contendo SEQ ID NO:81, 123 e 143, e codificando a sequência polipeptídica SEQ ID NO:158 correspondendo ao fragmento da região SR de hST8Sia IV delimitada pelos aminoácidos localizados nas posições 1 a 67 de SEQ ID NO:38 e contendo as regiões CT, TMD e SR de hST8Sia IV correspondendo a SEQ ID NO:82, 124 e 144, respectivamente.

13. Processo de acordo com a reivindicação 9 ou 10, caracterizado por os péptidos CT, TMD, SR ou fragmento de SR compreendidos no segundo ácido nucleico, serem sequências heterólogas derivando de diferentes glicosiltransferases nativas.

14. Processo de acordo com a reivindicação 13, caracterizado por o segundo ácido nucleico ser a sequência SEQ ID NO:159 correspondendo à fusão da sequência nucleotídica SEQ ID NO:177 contendo SEQ ID NO:65 e 107 codificando as regiões CT e TMD de hST3Gal III correspondendo a SEQ ID NO:66 e 108, respectivamente, com a sequência nucleotídica SEQ ID NO:131 codificando a sequência polipeptídica SEQ ID NO:132 correspondendo ao fragmento da região SR de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 37 a 74 de SEQ ID NO:6, a referida sequência SEQ ID NO:159 codificando o polipéptido de fusão SEQ ID NO:160 entre as regiões CT e TMD de hST3Gal III, por um lado, e o fragmento 37-74 da região SR de hST6GalNAc I, por outro.

15. Processo de acordo com a reivindicação 13, caracterizado por o segundo ácido nucleico ser a sequência SEQ ID NO:179 correspondendo à fusão da sequência nucleotídica SEQ ID NO:65 codificando a CT de hST3Gal III correspondendo a SEQ ID NO:66, com a sequência nucleotídica SEQ ID NO:119 codificando a região TMD de hST8Sia II correspondendo a SEQ ID NO:120 e com a sequência nucleotídica SEQ ID NO:129 codificando a sequência polipeptídica SEQ ID NO:130 correspondendo ao fragmento da região SR de hST6GalNAc I delimitada pelos aminoácidos localizados nas posições 36 a 74 de SEQ ID NO:6, a referida sequência SEQ ID NO:179 codificando o polipéptido de fusão SEQ ID NO:180 entre a região CT de hST3Gal III, a região TMD de hST8Sia II e o fragmento 36-74 da região SR de hST6GalNAc I.

16. Processo de acordo com qualquer uma das reivindicações 1 a 15, caracterizado por compreender a fusão da sequência SEQ ID NO:43 como primeiro ácido nucleico, com um segundo ácido nucleico escolhido de entre:

- a sequência SEQ ID NO:145, conduzindo à sequência nucleotídica SEQ ID NO:161 codificando a proteína de fusão SEQ ID NO:162 contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 130, respectivamente, ligadas através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I,

- a sequência SEQ ID NO:147, conduzindo à sequência nucleotídica SEQ ID NO:163 codificando a proteína de fusão SEQ ID NO:164 contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 134, respectivamente, ligadas através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I,

- a sequência SEQ ID NO:149, conduzindo à sequência nucleotídica SEQ ID NO:165 codificando a proteína de fusão SEQ ID NO:166 contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 136, respectivamente, ligadas através de um ligador SR ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I,

- a sequência SEQ ID NO:151, conduzindo à sequência nucleotídica SEQ ID NO:167 codificando a proteína de fusão SEQ ID NO:168 contendo as regiões CT, TMD e fragmento de SR de hST3Gal III correspondendo a SEQ ID NO:66, 108 e 138, respectivamente, ligadas através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I,

- a sequência SEQ ID NO:153, conduzindo à sequência nucleotídica SEQ ID NO:169 codificando a proteína de fusão SEQ ID NO:170 contendo as regiões CT, TMD e fragmento de SR de ratST3Gal III correspondendo a SEQ ID NO:74, 116 e 140, respectivamente, ligadas através de um ligador SR ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I,

- a sequência SEQ ID NO:155, conduzindo à sequência nucleotídica SEQ ID NO:171 1 codificando a proteína de fusão SEQ ID NO:172 contendo as regiões CT, TMD e SR de hST8Sia II correspondendo a SEQ ID NO:78, 120 e 142, respectivamente, ligadas através de um ligador KL ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I,

- a sequência SEQ ID NO:157, conduzindo à sequência nucleotídica SEQ ID NO:173 codificando a proteína de fusão SEQ ID NO:174 contendo as regiões CT, TMD e SR de hST8Sia IV correspondendo a SEQ ID NO:82, 124 e 144, respectivamente, ligadas através de um ligador KL ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I,

- a sequência SEQ ID NO:159, conduzindo à sequência nucleotídica SEQ ID NO:175 codificando a proteína de fusão

SEQ ID NO:176 contendo as regiões CT e TMD de hST3Gal III correspondendo a SEQ ID NO:66 e 108, respectivamente, e o fragmento 37-74 da região SR de hST6GalNAc I correspondendo a SEQ ID NO:132, ligados através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I,

- a sequência SEQ ID NO:179, conduzindo à sequência nucleotídica SEQ ID NO:181 codificando a proteína de fusão SEQ ID NO:182 contendo a região CT de hST3Gal III, a região TMD de hST8Sia II e o fragmento 36-74 da região SR de hST6GalNAc I, correspondendo a SEQ ID NO:66, 120 e 130, respectivamente, ligadas através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I.

17. Sequências génicas codificando glicosiltransferases quiméricas obtidas de acordo com o processo definido em qualquer uma das reivindicações 1 a 16.

18. Sequências génicas de acordo com a reivindicação 17, escolhidas de entre:

- a sequência SEQ ID NO:161 codificando a proteína de fusão SEQ ID NO:162 contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 130, respectivamente, ligadas através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I, a referida sequência SEQ ID NO:161 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:145,

- a sequência SEQ ID NO:163 codificando a proteína de fusão SEQ ID NO:164 contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 134, respectivamente, ligadas através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I, a referida sequência SEQ ID NO:163 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:147,

- a sequência SEQ ID NO:165 codificando a proteína de fusão SEQ ID NO:166 contendo as regiões CT, TMD e fragmento de SR de hST6GalNAc I correspondendo a SEQ ID NO:50, 92 e 136, respectivamente, ligadas através de um ligador SR ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I, a referida sequência SEQ ID NO:165 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:149,

- a sequência SEQ ID NO:167 codificando a proteína de fusão SEQ ID NO:168 contendo as regiões CT, TMD e fragmento de SR de hST3Gal III correspondendo a SEQ ID NO:66, 108 e 138, respectivamente, ligadas através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I, a referida sequência SEQ ID NO:167 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:151,

- a sequência SEQ ID NO:169 codificando a proteína de fusão SEQ ID NO:170 contendo as regiões CT, TMD e fragmento de SR de rST3Gal III correspondendo a SEQ ID NO:74, 116 e 140, respectivamente, ligadas através de um ligador SR ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I, a referida sequência SEQ ID NO:169 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:153,

- a sequência SEQ ID NO:171 codificando a proteína de fusão SEQ ID NO:172 contendo as regiões CT, TMD e SR de hST8Sia II correspondendo a SEQ ID NO:78, 120 e 142, respectivamente, ligadas através de um ligador KL ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I, a referida sequência SEQ ID NO:171 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:155,

- a sequência SEQ ID NO:173 codificando a proteína de fusão SEQ ID NO:174 contendo as regiões CT, TMD e SR de hST8Sia IV correspondendo a SEQ ID NO:82, 124 e 144, respectivamente, ligadas através de um ligador KL ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I, a referida sequência SEQ ID NO:173 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:157,

- a sequência SEQ ID NO:175 codificando a proteína de fusão SEQ ID NO:176 contendo as regiões CT e TMD de hST3Gal III correspondendo a SEQ ID NO:66 e 108, respectivamente, e o fragmento 37-74 da região SR de hST6GalNAc I correspondendo a SEQ ID NO:132, ligados através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal I, a referida sequência SEQ ID NO:175 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:159,

- a sequência SEQ ID NO:181 codificando a proteína de fusão SEQ ID NO:182 contendo a região CT de hST3Gal III, a região TMD de hST8Sia II e o fragmento 36-74 da região SR de hST6GalNAc I, correspondendo a SEQ ID NO:66, 120 e 130, respectivamente, ligados através de um ligador GS ao fragmento mínimo C-terminal 90-406 do domínio CD de hST6Gal

I, a referida sequência SEQ ID NO:181 correspondendo à fusão da sequência SEQ ID NO:43 com a sequência SEQ ID NO:179.

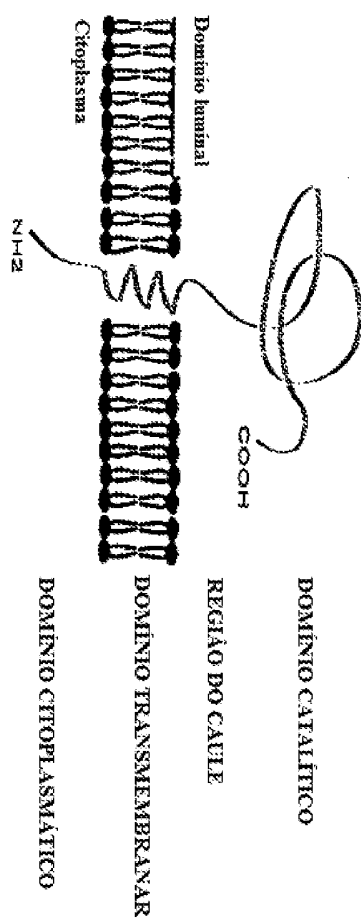
19. Vectores, tais como plasmídeos, vectores virais ou bacterianos, compreendendo pelo menos uma sequência génica de acordo com a reivindicação 17 ou 18.

20. Células hospedeiras, tais como células de levedura, de fungos, vegetais, de insecto, de ave, de mamífero ou humanas, transformadas com pelo menos uma sequência génica de acordo com a reivindicação 17 ou 18, utilizando um vector de acordo com a reivindicação 19.

21. Utilização de pelo menos uma sequência génica de acordo com a reivindicação 17 ou 18, ou de um vector de acordo com a reivindicação 19, para a transformação de células de acordo com a reivindicação 20, no quadro da preparação de proteínas recombinantes de interesse, sendo as referidas proteínas de interesse escolhidas de entre hormonas, enzimas, factores de coagulação, antigénios de hidratos de carbono/biomarcadores séricos, citocinas, factores de crescimento, anticorpos ou receptores.

22. Método para a preparação de proteínas recombinantes de interesse compreendendo a transformação de células de acordo com a reivindicação 20, com um vector contendo a sequência nucleotídica codificando as referidas proteínas recombinantes de interesse.

Lisboa, 2011-08-30

**Figura 1**

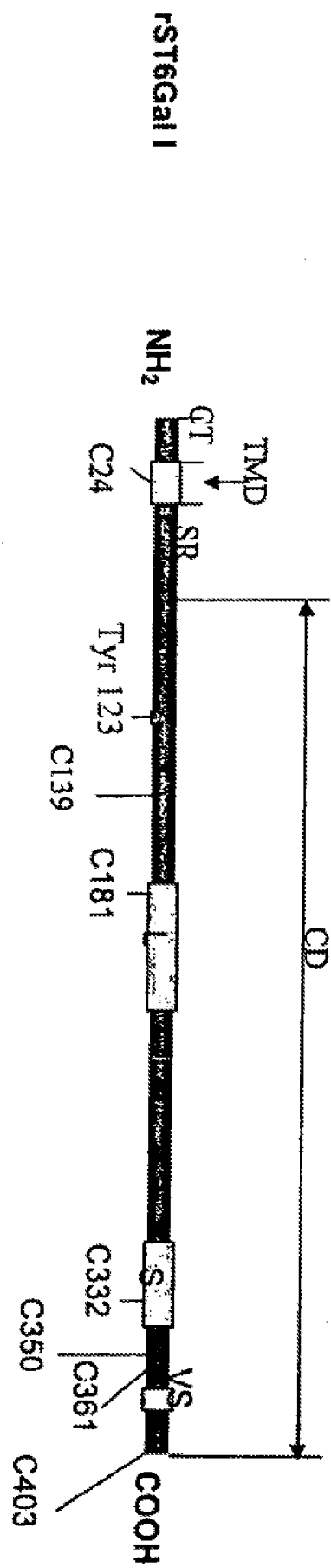
TCAIVGNSGVLLNSGCGQEIDASHFVIRCNLAPV-QEYARDVGLKT
TCAVVGNSGILLDSECGKEIDSHNFVIRCNLAPV-VEFAADVGTKS
RCAVVGNSGNLRESSYGPEIDSHDFVLRMNKAPT-AGFEADVGTKT
RCAVVGNSGNLRGSGYGQDQDVGNHFMNMNAPT-VGFEADVGSRT
LCAIVSNSGQMVGQKVGNIEDRSSCIWMMNAPT-KGYEEDVGRMT
SCAVVSSSGMVLGSGGLGAEIDSAECVFRMNQAPT-VGFEADVQRS
DCALVTSSGHLHLSRQGSQIDQTECVIRMNDAPT-RGYGRDVGNR
QCVIVSSSSSHLLGTLGLPEIERAECTIRMNDAPT-TGYSADVGNKT
ICAVVGNSGILTFIQCGREIDKSDFVFRCNFAPS-EAFQRDVGRKT
KCAVVGNGGILKKSGCGRQIDEANFVMRCNLPLLSSEYTKDVGSKS
QCAVVGNGGILNKS LCGTEIDKSDFVFRCNLPPTTGDVSKDVGSKT
KCAVVGNGGILKNSRCGREINSADFVFRCNLPPISEKYTMDVGVKT
RCIIVGNGGVLANKSLGSRIDDYDIVRLNSAPV-KGFEKDVGSKT
RCVVISSGGILHGLELGLHTLNQFDVVIRLNSAPV-EGYSEHVGNKT
RCVVVGNGHRLNRSSSLGDAINKYDVVIRLNNAPV-AGYEGDVGSKT
KCVVVGNGGVLKNTLGEKIDSYDVII RMNNGPV-LGHEEEVGRRT
TCAVVGNGGILNNSHMGQEIDSHDYVFRLSGALI-KGYEQDVGTRT
RCAVVGNGGILNGSRQGPNI DAHDYVFRNLNGAVI-KGFERDVGTKT
RCAVVSSAGSLKSSQLGREIDDHDAVLRFNAGAPT-ANFQQDVGTKT
SCAVVMSAGAILNSSLGEEIDSHDAVLRFNAPT-RGYEKDVGNKT

* * *

PTTGLLMYTLATRFCKQIYLYGF
PSTGLLMYTLATRFCD EIHLYGF
PSTGILSVIFSMHVCDEVDLYGF
PSTGMLVLFFALHVCDEVNVYGF
LSTGWFTMLLAMDACYGIHVYGF
LSTGWFTMILALELCEEIVVYGM
LSTGWFTMTIALELCDRINVYGM
LSTGWFTMVIAVELCCHVHVYGM
LSTGILMYTLASAICEEIHLYGF
LSTGLFLVSAAAGLCEEVALYGF
LSTGLMITSVAVELCKNVKLYGF
ISTGLILVTALELCEEVHLFGF
PTLGSVAVTMALHGCD EAVAGF
PTIGVIAVVLATHLCDEVSLAGF
PTTGLLAITLALHLCDLVHIAGF
PTTGLIAITLAFYICHEVHLAGF
PTTGALLLLTALQLCDQVSAYGF
PSTGALMLLTALHTCDQVSAYGF
PSSGMLGIIIMMTLCDQVDIYEF
PSSGFGIILIMMSMCREVHYEY
:
*
*
:

ASPHMTMPLEFKA
ASPHRMPLEFKT
TGVHDADFESNV
TGVHDADFEAHV
YGGHRFITEKKV
RSAHRFITEKAV
GSHHRFITEKRV
GNHHRFITEKRV
QESHQLPAEFQL
SGFHAMPEEFLO
HGFHQMFKKEYSQ
EGGHAMPSEIFN
SWTHNIQREKEF
QTMHNVTETKF
GSGMNVYSQEALA
NAYHNVTAEQLF
YINHDFKLEREV
YANNHDSLEAL
GAYHPLLYEKNL
GAYHPLLYEKLL

Figura 2

**Figura 3**

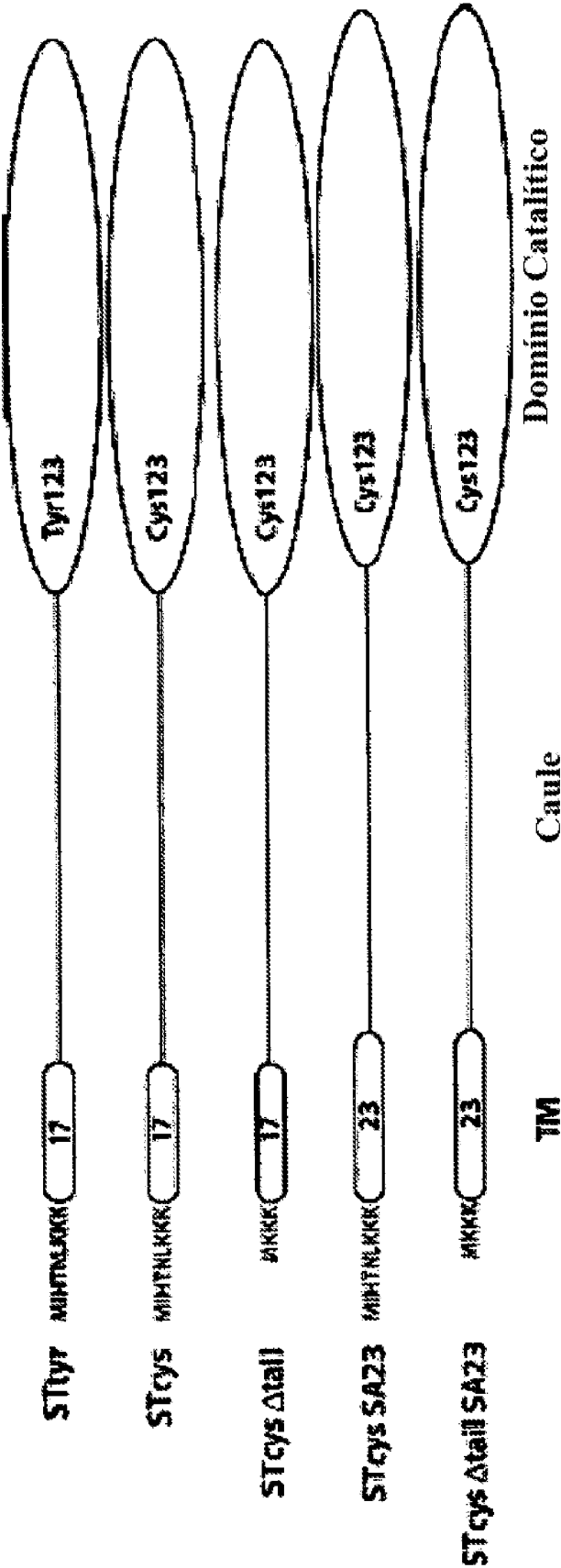


Figura 4

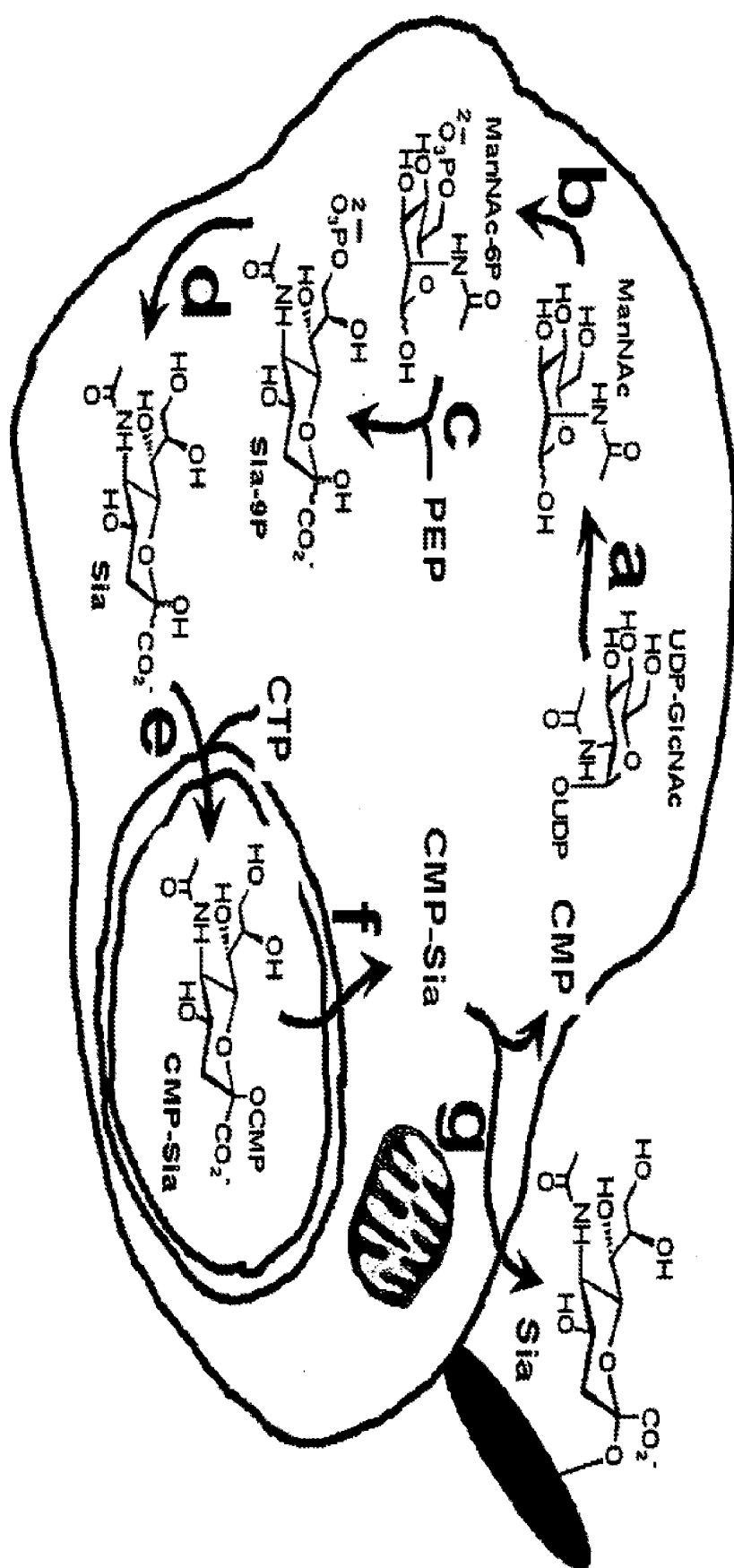


Figura 6

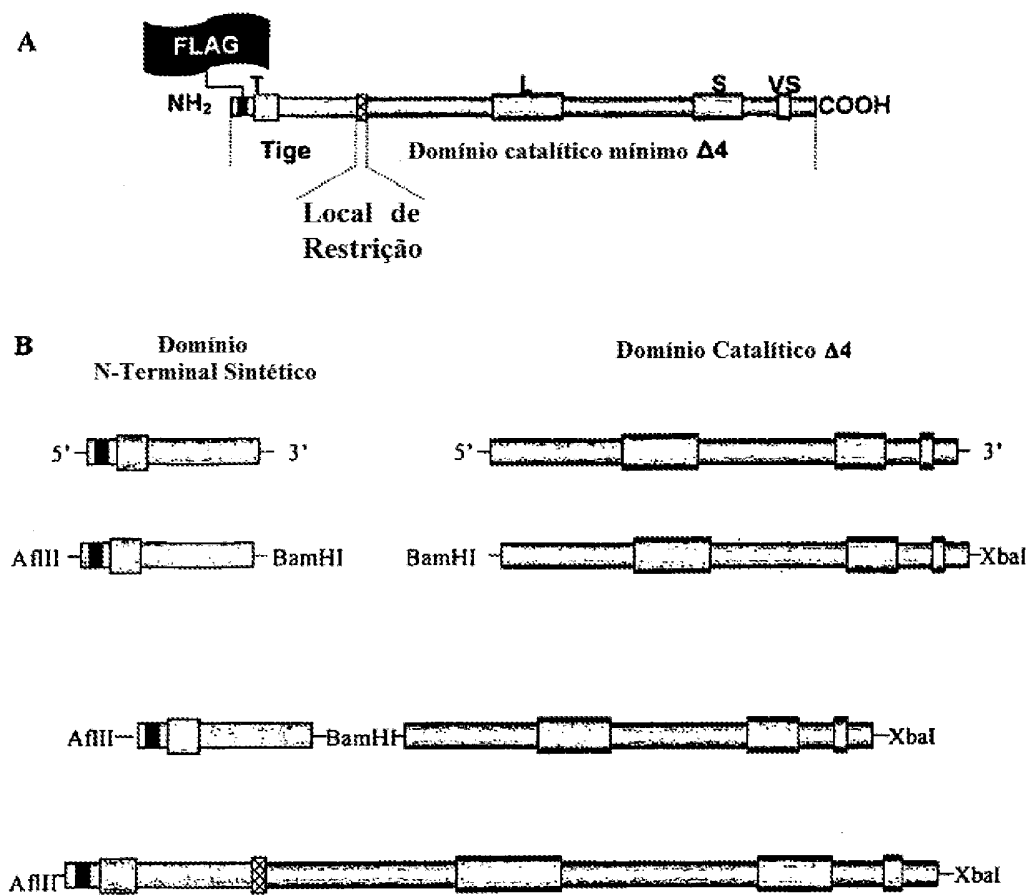


Figura 7

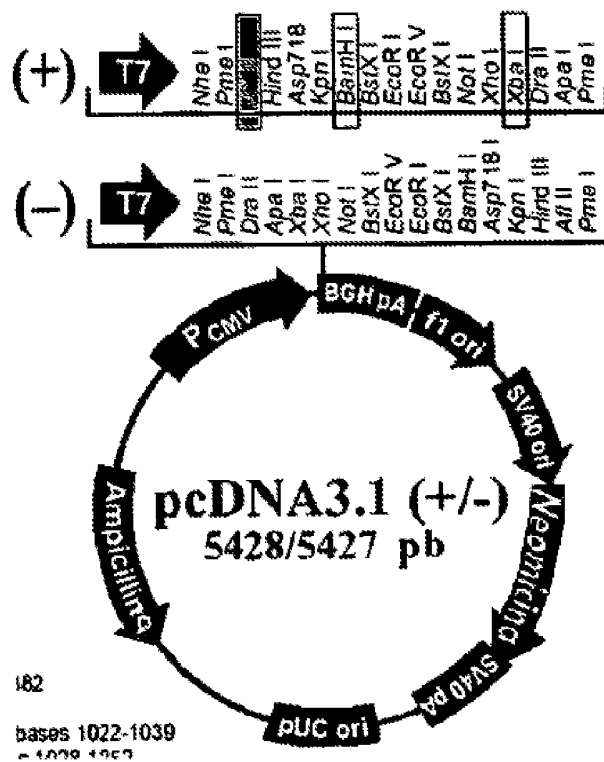
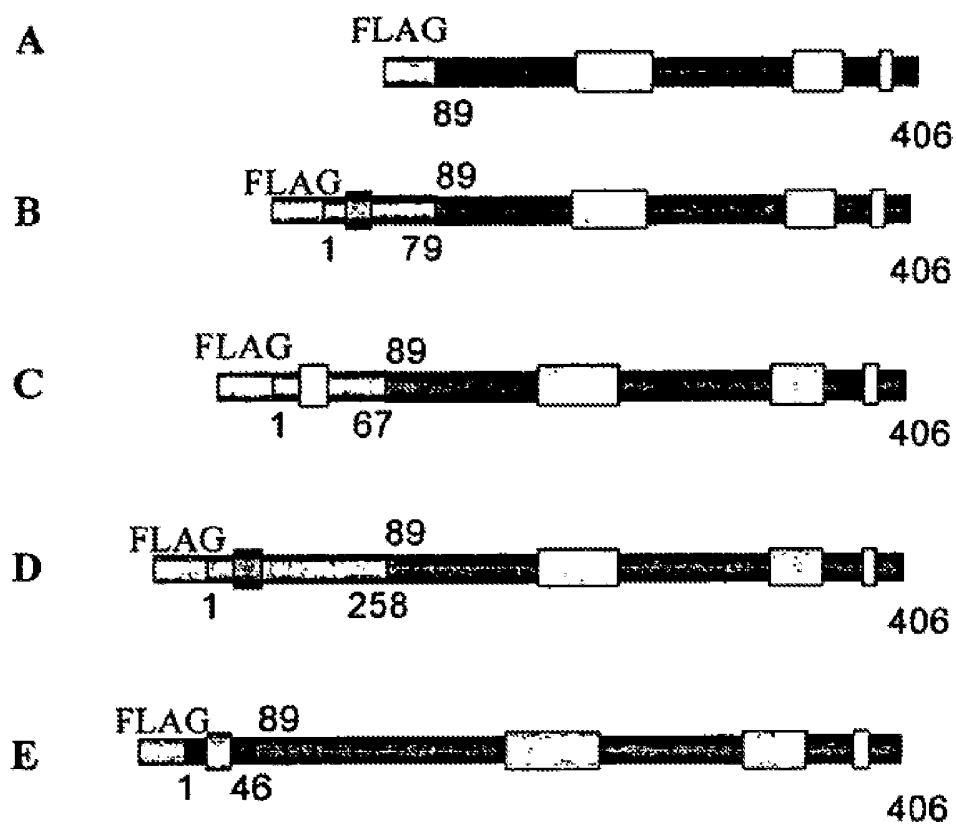


Figura 8

**Figura 9**

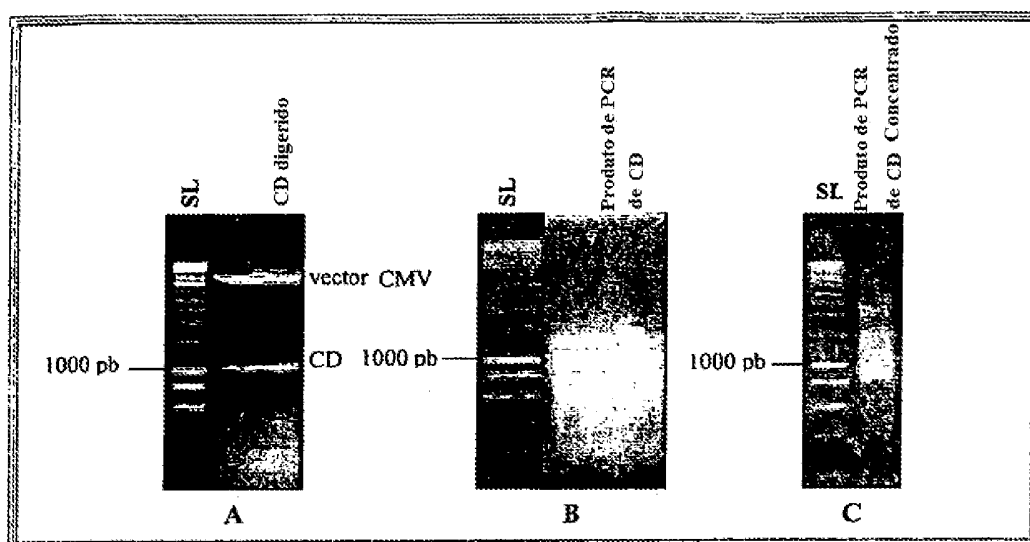
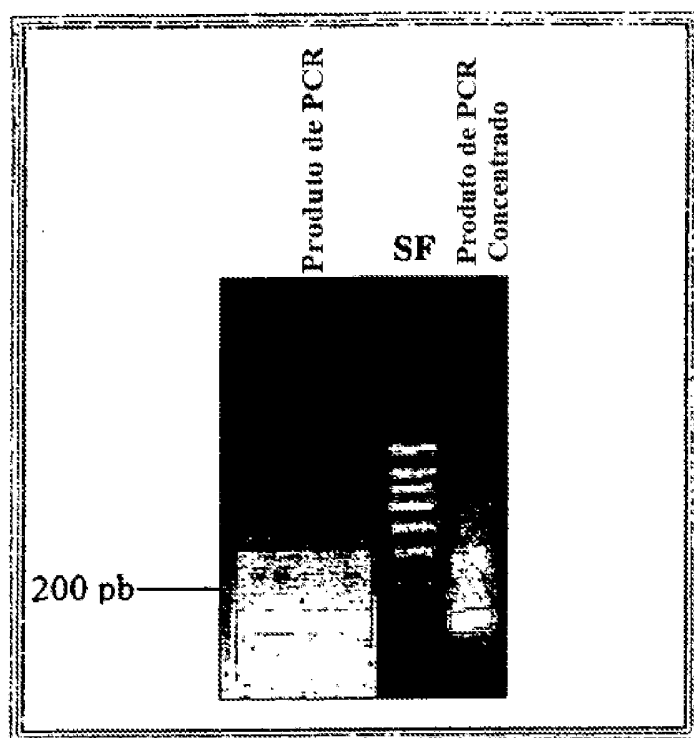


Figura 10

**Figura 11**

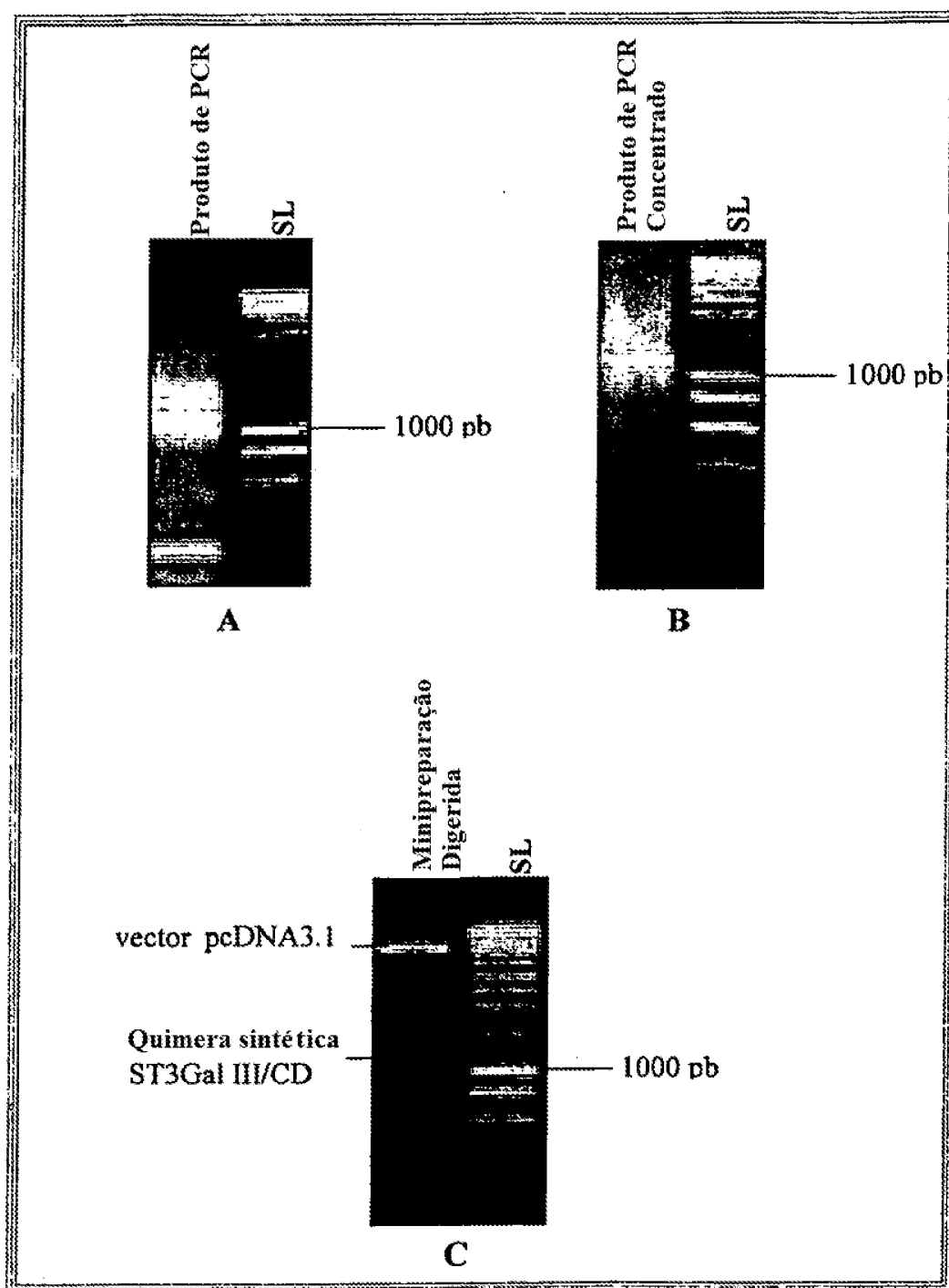
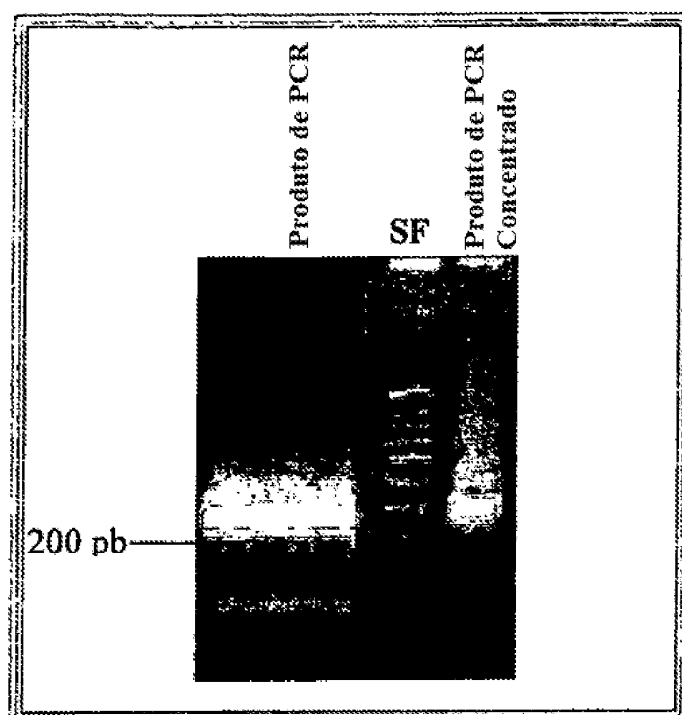


Figura 12

**Figura 13**

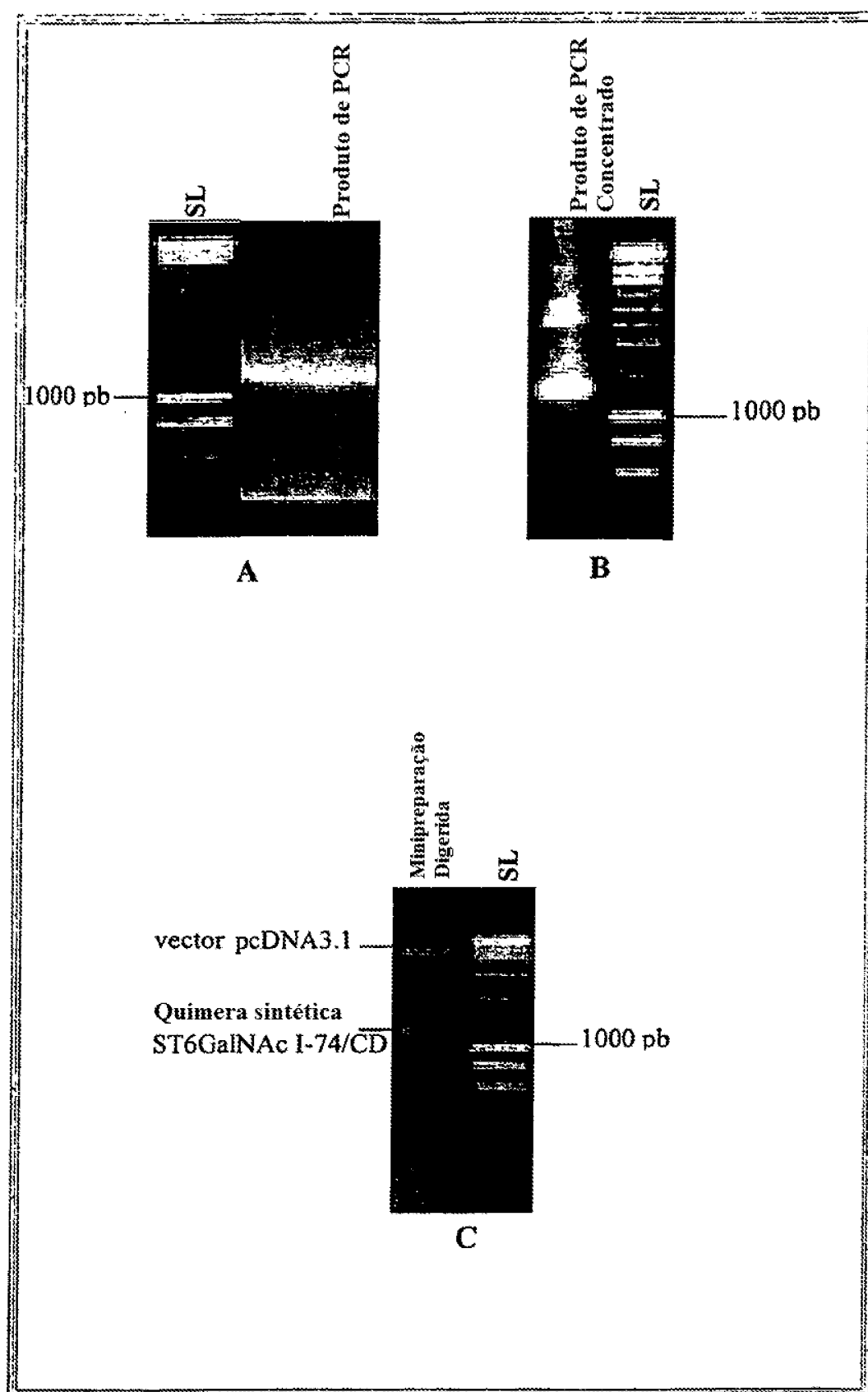


Figura 14

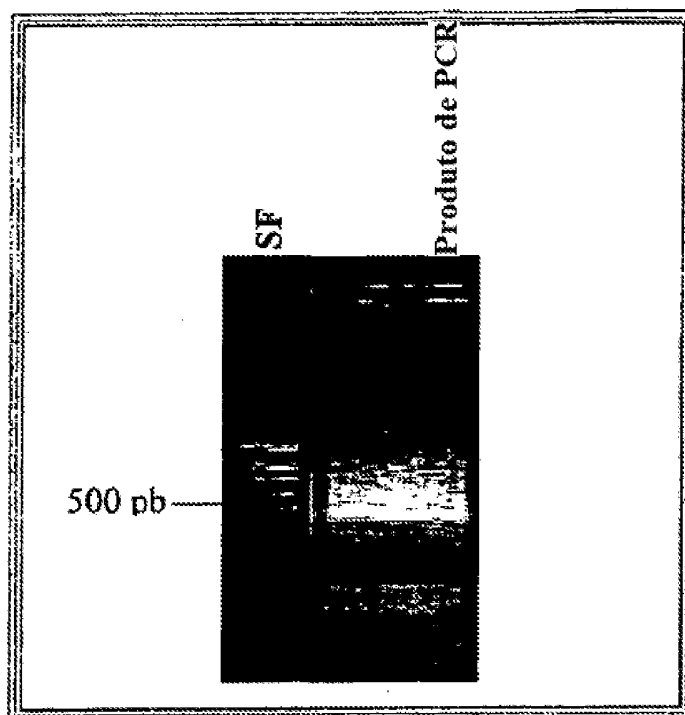


Figura 15

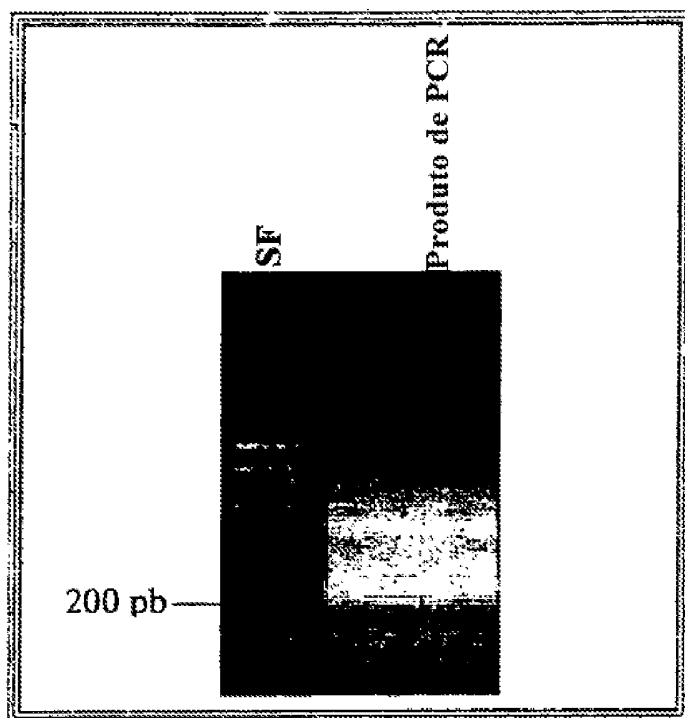
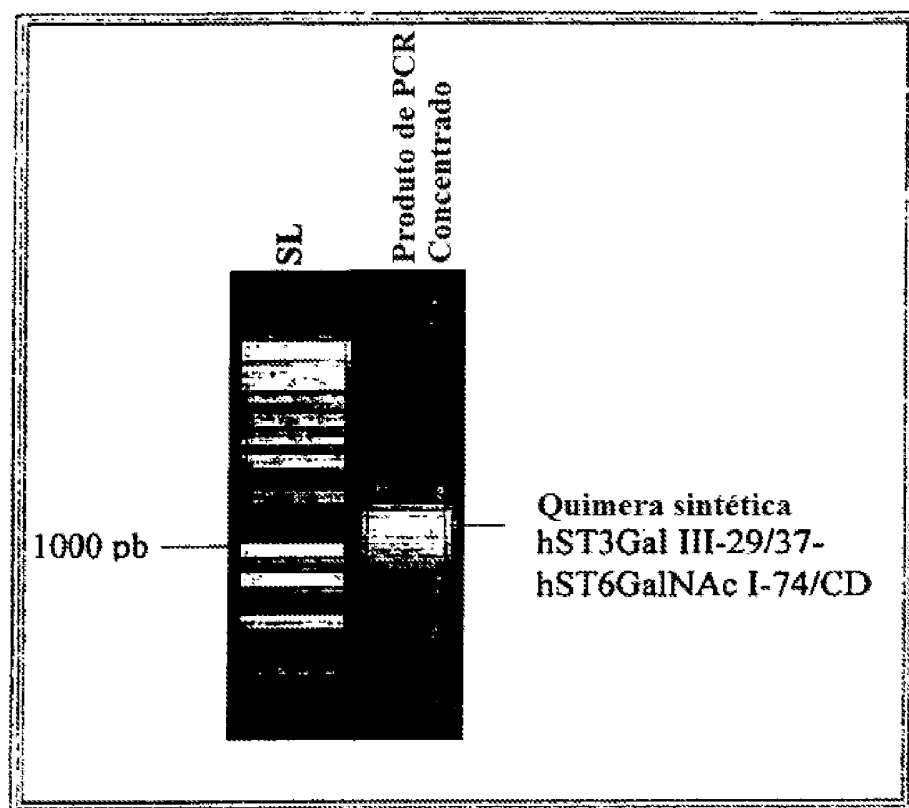


Figura 16

**Figura 17**

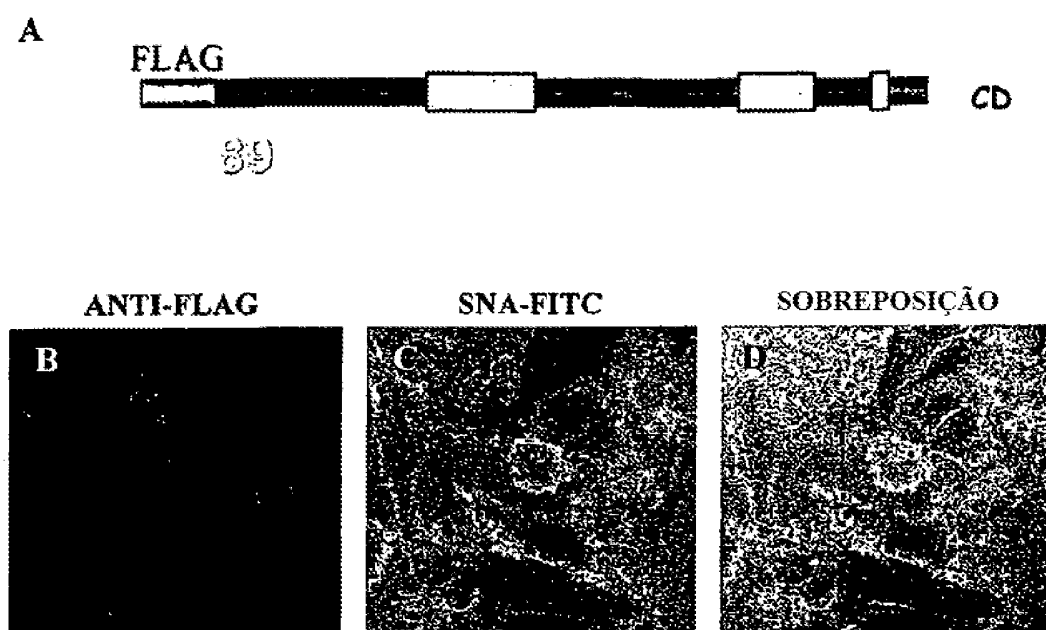


Figura 18

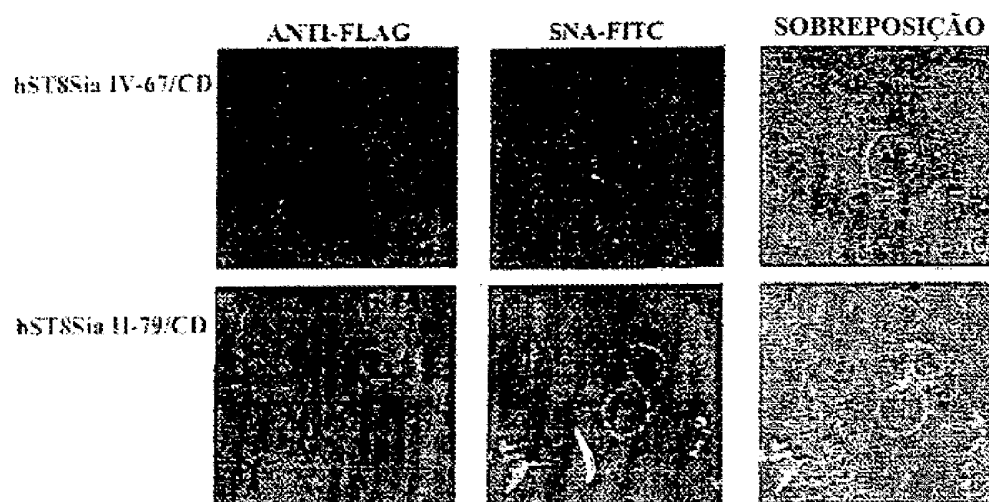


Figura 19

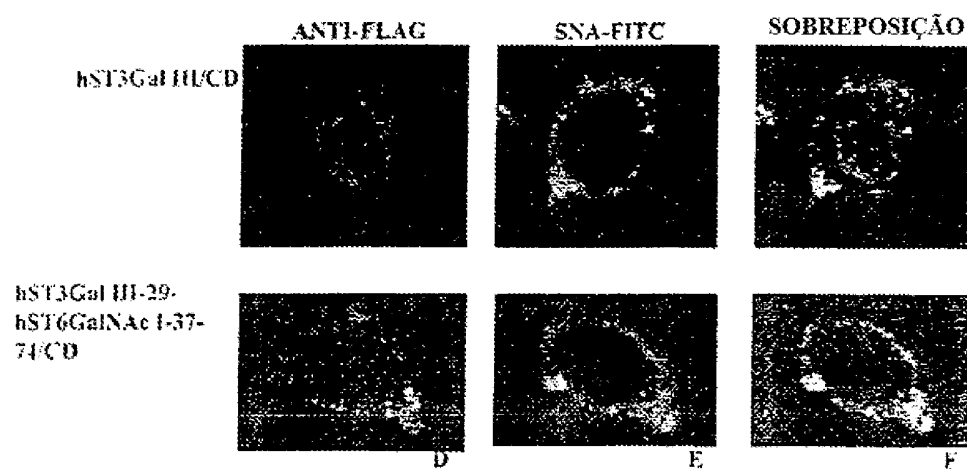


Figura 20

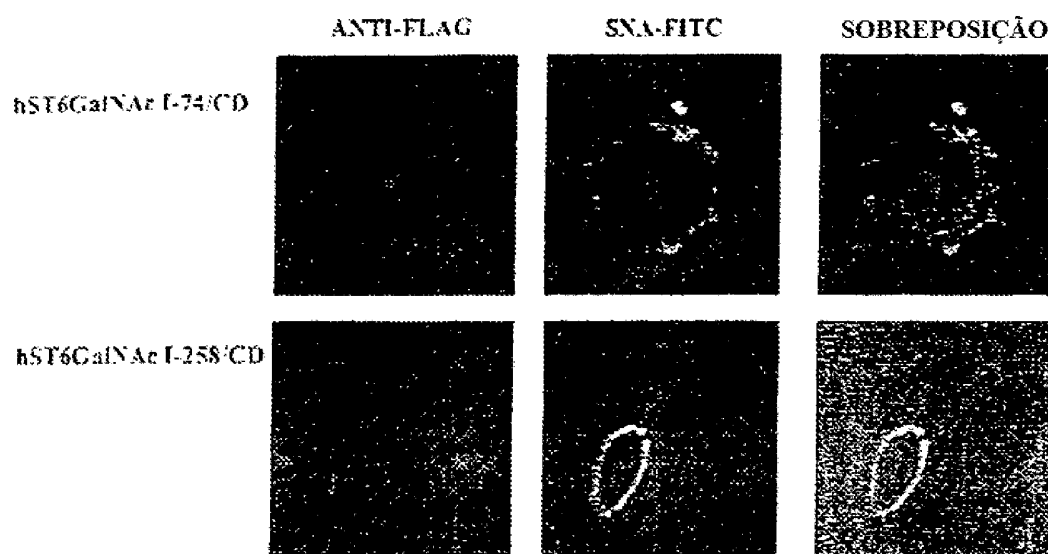
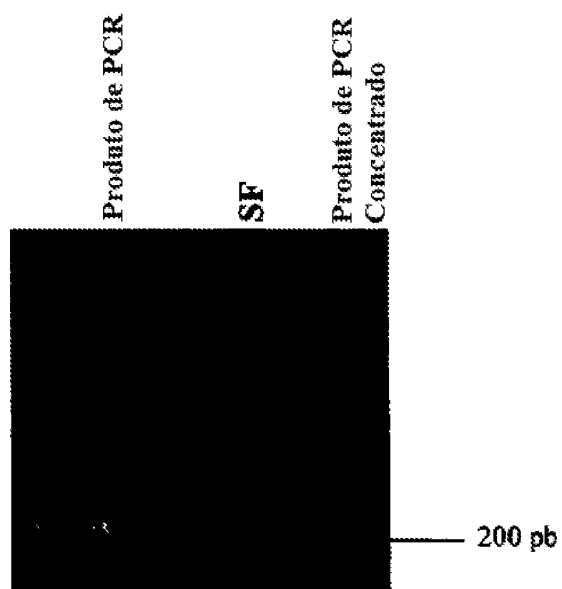
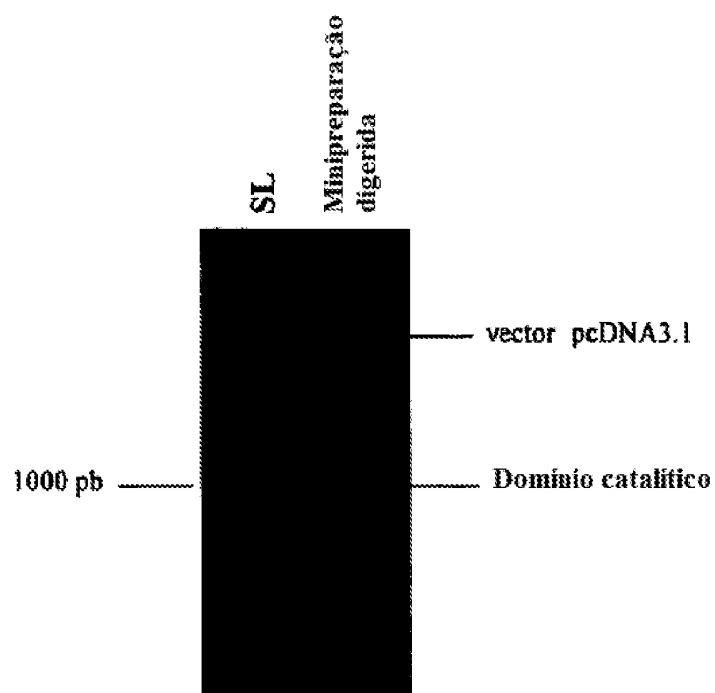


Figura 21

**Figura 22**

**Figura 23**

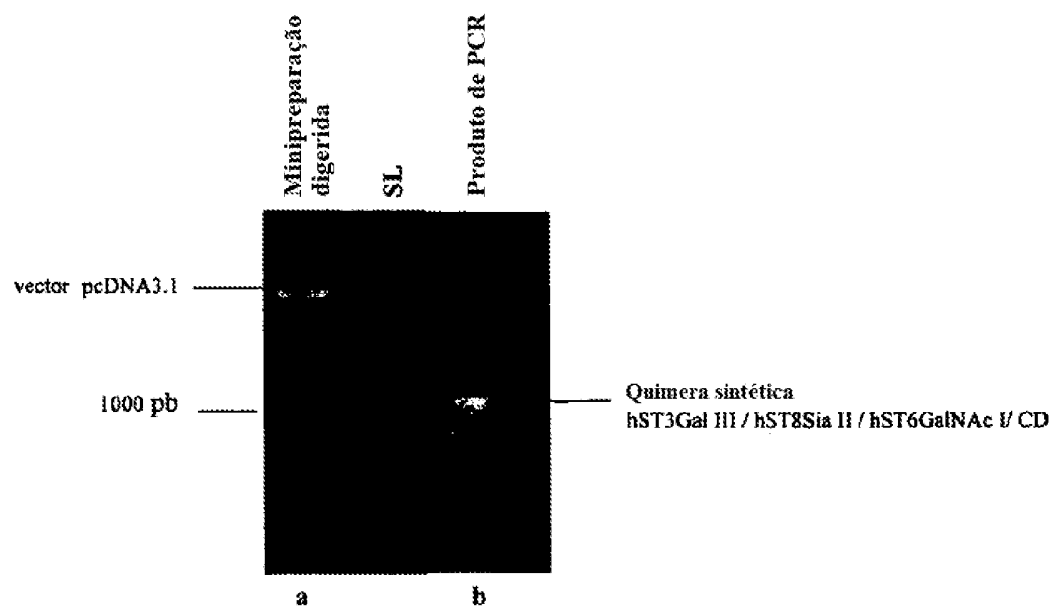
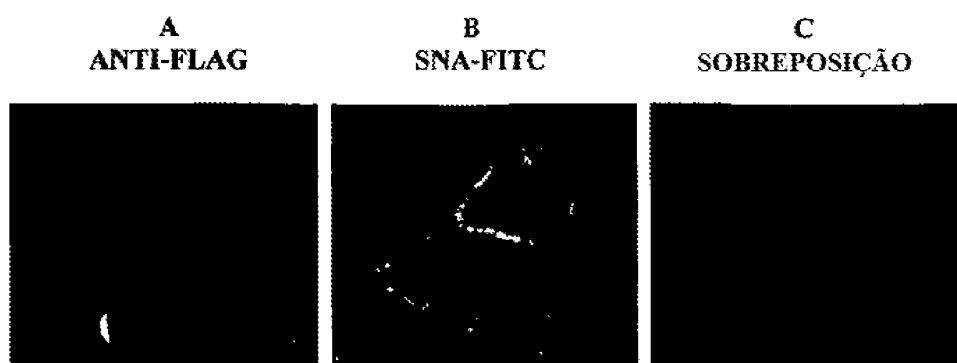


Figura 24

**Figura 25**