

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

(22) Data de pedido: 2004.06.17	(73) Titular(es): SEMBIOSYS GENETICS INC. BAY 110, 2985-23RD AVENUE N.E. CALGARY, ALBERTA T1Y 7L3 CA
(30) Prioridade(s): 2003.06.17 US 478818 P 2004.03.04 US 549539 P	
(43) Data de publicação do pedido: 2006.03.15	(72) Inventor(es): GIJS VAN ROOIJEN CA MAURICE M. MOLONEY CA JOSEPH BOOTHE CA RICHARD KEON CA CORY NYKIFORUK CA
(45) Data e BPI da concessão: 2008.08.20 215/2008	(74) Mandatário: PEDRO DA SILVA ALVES MOREIRA RUA DO PATROCÍNIO, N.º 94 1399-019 LISBOA PT

(54) Epígrafe: **MÉTODOS PARA A PRODUÇÃO DE INSULINA EM PLANTAS**

(57) Resumo:

DESCRIÇÃO

"MÉTODOS PARA A PRODUÇÃO DE INSULINA EM PLANTAS"

CAMPO DA INVENÇÃO

A presente invenção refere-se a métodos de engenharia genética de plantas e à produção de insulina. Mais especificamente, a presente invenção refere-se a métodos para a produção de insulina nas sementes das plantas.

ANTECEDENTES DA INVENÇÃO

A insulina é uma hormona peptídica importante necessária para manter a homeostase da glucose no sangue em mamíferos, incluindo humanos e outros vertebrados. Em indivíduos saudáveis, um aumento do nível de glucose no sangue estimula a secreção de insulina pelas células β do pâncreas. O polipéptido da insulina liga-se então a receptores específicos no músculo, fígado e tecido adiposo, conduzindo a um aumento da incorporação de glucose por estes tecidos alvo, um aumento no metabolismo e uma diminuição na produção de glucose hepática. Os efeitos cumulativos destas respostas servem para manter as concentrações de glucose no sangue a um nível constante.

Em indivíduos que sofrem de diabetes mellitus, uma concentração de insulina anormalmente baixa, apresenta-se por si só própria, como hiperglicemia crónica. As manifestações

clínicas de hiperglicemia crónica são várias e incluem a cegueira, insuficiência renal e, se não for tratada, em última instância resultará na morte. Estimativas colocam a diabetes mellitus como a terceira maior causa de morte em países industrializados, depois das doenças cardiovasculares e do cancro (Barfoed, H.C., 1987, *Chem. Eng. Prog.* 83:49-54). De modo a permitir a incorporação e o metabolismo, eficientes, de glucose no sangue pelas células, os indivíduos diabéticos podem ser tratado através da administração de rotina de insulina. Aproximadamente 0,7% da população mundial sofre de diabetes dependente de insulina (diabetes mellitus Tipo I) (Winter, J. et al., 2000, *J. of Biotechnol.* 84:175-185). Adicionalmente, estima-se que o número de indivíduos diagnosticados com diabetes duplicará para aproximadamente 300 milhões, nos próximos 25 anos (Kjeldsen, T. et al., 2001, *Biotechnol. Gen. Eng. Rev.* 18:89-121). Consequentemente, a capacidade para o fabrico rentável de insulina humana em quantidades para satisfazer a procura mundial que se antecipa crescente para a insulina é altamente desejável.

In vivo, o polipéptido da insulina humana é produzido pelas células β pancreáticas como um precursor de cadeia polipeptídica única de 110 aminoácidos, pré-pró-insulina que inclui uma pré-sequência de 24 aminoácidos localizada no terminal N que é clivada imediatamente após a finalização da biossíntese da cadeia (Steiner, D.F. 2000. *J. Ped. Endocrinol. Metab.* 13:229-239). A pró-insulina consiste numa cadeia B e A, ligada por um péptido de ligação (péptido C). Durante o empacotamento da hormona para secreção, o péptido C é clivado e removido por convertases de pró-hormona, PC2 e PC1/PC3 (Steiner, D.F. 2000. *J. Ped. Endocrinol. Metab.* 13:229-239). O que permanece é a insulina humana madura, uma proteína de 51 aminoácidos

consistindo em duas cadeias de polipéptido, A (21 aminoácidos de comprimento) e B (30 aminoácidos de comprimento), ligadas por duas ligações persulfureto intercadeia. Adicionalmente, a cadeia A compreende uma ligação persulfureto intracadeia.

A insulina humana foi preparada utilizando uma variedade de metodologias diferentes. Microrganismos, tais como *Escherichia coli* (Frank *et al.*, 1981, em *Peptides: Proceedings of the 7th American Peptide Chemistry Symposium* (Rich & Gross, eds.), Pierce Chemical Co., Rockford. Ill pp 729-739; Chan *et al.*, 1981, *Proc Natl. Acad. Sci. USA* 78: 5401-5404), *Saccharomyces cerevisiae* (Thim *et al.*, 1986, *Proc. Natl. Acad. Sci. USA* 83: 6766-6770) são habitualmente empregues para produzir insulina de forma recombinante. Wang *et al.* (*Biotechnol. Bioeng.*, 2001, 73:74-79) demonstraram que fungos, tal como *Pichia pastoris*, também são adequados para a produção de insulina. Opções de fabrico alternativas incluem a produção em linhas celulares de mamíferos não humanos (Yanagita, M., *et al.*, 1992, *FEBS Lett* 311:55-59), isolamento a partir de pâncreas humano, síntese de péptidos ou a conversão semi-sintética para insulina humana a partir de insulina porcina e bovina. Todavia, todos estes métodos sofrem de rendimentos mais baixos e custos mais elevados do que o desejado.

A utilização de plantas como biorreactores para a produção em larga escala de proteínas recombinantes é bem conhecida e foram produzidas numerosas proteínas, incluindo proteínas terapêuticas humanas. Por exemplo, As Patentes US 4956282, 5550038 e 5629175 revelam a produção de interferão γ em plantas; as patentes US 5650307, 5716802 e 5763748 apresentam em detalhe a produção de albumina de soro bovino em plantas e as Patentes US 5202422, 5639947 e 5959177 referem-se à produção de

anticorpos em plantas. Uma das vantagens significativas, oferecidas pelos sistemas de produção de proteína recombinante à base de plantas, é que aumentando a superfície de plantas cultivadas, a produção de proteína pode ser aumentada de escala de forma não dispendiosa para proporcionar grandes quantidades de proteína. Por contraste, os sistemas de fermentação e de cultura de células possuem requisitos de grandes espaços, de equipamento e de energia, tornando o aumento de escala da produção dispendioso. Todavia, apesar do facto de a utilização de plantas como biorreactores estar amplamente documentada e, apesar do aumento prodigioso antecipado acima mencionado, da necessidade de grandes quantidades de insulina, a técnica anterior proporciona apenas um número limitado de métodos que resultam de forma demonstrável na produção de insulina em plantas (ver: Arakawa *et al.* *Nature Biotech.*, 1998, 16: 934 - 938; documento PCT 01/72959).

Arakawa *et al.* revelam a produção de uma fusão de proteína compreendendo insulina nos tubérculos de plantas de batata transgénicas. Todavia, a insulina representa apenas até 0,05% do conteúdo em proteína solúvel total presente nos tubérculos transgénicos. Ao nível de 0,05% da proteína solúvel total, grandes quantidades de biomassa devem ser submetidas à extracção de proteína tornando desfavorável a economia da produção associada à utilização de tubérculos de batata. Além disso, Arakawa *et al.* não estão preocupados com o isolamento de insulina a partir do tecido de tubérculo de batata, mas em vez disso sugerem uma abordagem que previne o despoletar de diabetes de Tipo I, induzindo imunotolerância que envolve administração oral da insulina através da alimentação de tubérculos de batatas transgénicas.

O Pedido de Patente PCT WO 01/72959 revela a produção de uma proteína de fusão compreendendo insulina em cloroplastos de tabaco transgénico. Todavia, embora pretendendo abordar as dificuldades em relação aos níveis de acumulação de proteínas humanas em tecido de plantas, a invenção à qual o documento WO 01/72959 pretende estar limitado nessa produção em cloroplastos resulta na acumulação de insulina em tecido verde, principalmente nas folhas do tabaco. Devido ao conteúdo relativamente elevado em água do tecido verde, deve ser processada uma grande quantidade de biomassa. Além disso, a produção de insulina deve requerer a extracção imediata da biomassa após a colheita, na medida em que o material das folhas se irá deteriorar rapidamente quando armazenado.

Assim, à luz das dificuldades associadas aos métodos para a produção recombinante de insulina em plantas proporcionados pela técnica anterior, presentemente não é claro se e como, a capacidade sintética das plantas pode ser aproveitada para atingir a produção comercial de insulina em plantas. Existe uma necessidade na técnica anterior para melhorar métodos para a produção comercial de insulina em plantas.

SUMÁRIO DA INVENÇÃO

A presente invenção refere-se a métodos melhorados para a produção de insulina em plantas. Em particular, a presente invenção refere-se a métodos para a produção de insulina em sementes.

Consequentemente, a presente invenção proporciona um método para a expressão de insulina em sementes de plantas

compreendendo:

(a) proporcionar uma construção de ácido nucleico quimérico compreendendo na direcção 5' para 3' da transcrição como componentes ligados operacionalmente:

(i) uma sequência de ácido nucleico que compreende um promotor preferido pelas sementes;

(ii) uma sequência de ácido nucleico codificando o polipéptido da insulina; e

(iii) uma sequência de ácido nucleico codificando um polipéptido que é capaz de reter o polipéptido da insulina no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada do ER.

(b) introduzir uma construção de ácido nucleico quimérica numa célula vegetal; e

(c) cultivar a célula vegetal até ficar uma planta madura capaz de produzir sementes em que a semente expressa insulina; em que um polipéptido da insulina se acumula no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER e em que, pelo menos, 0,1% da proteína total da semente na semente é insulina; e

(d) obter insulina substancialmente pura a partir das referidas sementes.

Numa forma de realização preferida, a construção de ácido nucleico quimérico é introduzida na célula vegetal em condições

de integração genómica nuclear. Nessas condições, a sequência de ácido nucleico quimérica é integrada estavelmente no genoma da planta.

Ainda noutra forma de realização preferida, a sequência de ácido nucleico codificando insulina é otimizada para a utilização de códons de plantas e a sequência de ácido nucleico codificando o péptido de ligação (péptido C) é encurtada. As sequências de ácido nucleico preferidas, utilizadas de acordo com a presente invenção, codificam a insulina humana, bovina ou porcina. De acordo com a presente invenção, é utilizada a sequência de ácido nucleico codificando a sequência de pró-insulina em que a pró-insulina é modificada em que o péptido C é encurtado no comprimento.

Noutro aspecto, a presente invenção proporciona um método de recuperação de sementes de plantas compreendendo insulina. Consequentemente, de acordo com a presente invenção, é proporcionado um método para obter sementes de plantas compreendendo insulina, compreendendo:

(a) proporcionar uma construção de ácido nucleico quimérica compreendendo na direcção 5' para 3' da transcrição como componentes ligados operacionalmente:

(i) uma sequência de ácido nucleico que compreende um promotor preferido das sementes;

(ii) uma sequência de ácido nucleico codificando um polipéptido da insulina; e

(iii) uma sequência de ácido nucleico codificando um polipéptido que é capaz de reter o polipéptido da insulina no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER.

(b) introduzir a construção de ácido nucleico quimérica numa célula vegetal;

(c) cultivar a célula vegetal até ficar uma planta madura capaz de produzir sementes; e

(d) obter sementes da referida planta em que a semente compreende insulina, em que o referido polipéptido da insulina se acumula no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER e em que pelo menos 0,1% da proteína total de semente presente na semente é insulina.

As sementes podem ser utilizadas para obter uma população de progenia de plantas compreendendo cada uma delas uma pluralidade de sementes que expressam insulina.

A presente invenção também proporciona plantas capazes de produzir sementes que expressam insulina. Numa forma de realização preferida da invenção, é proporcionada uma planta capaz de produzir semente compreendendo uma sequência de ácido nucleico quimérica, compreendendo na direcção 5' para 3' da transcrição:

(a) uma primeira sequência de ácido nucleico que compreende um promotor preferido das sementes ligadas operacionalmente a;

(b) uma segunda sequência de ácido nucleico codificando um polipéptido da insulina; e

(c) uma sequência de ácido nucleico codificando um polipéptido que é capaz de reter um polipéptido da insulina no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER, em que a semente contém insulina, em que o referido polipéptido da insulina se acumula no retículo endoplásmico (ER) ou vesícula de armazenamento derivada de ER e em que, pelo menos, 0,1% da proteína total solúvel presente na semente é insulina.

Numa forma de realização preferida, a sequência de ácido nucleico quimérica é integrada no genoma nuclear da planta.

Noutra forma de realização preferida da presente invenção, a planta que é utilizada é uma açafrão, uma planta de linho ou uma planta de *Arabidopsis*.

Ainda noutro aspecto, a presente invenção proporciona sementes de plantas que expressam insulina. Numa forma de realização preferida da presente invenção, é proporcionada uma semente de planta compreendendo uma sequência de ácido nucleico quimérica, compreendendo na direcção 5' para 3' de transcrição:

(a) uma primeira sequência de ácido nucleico que compreende um promotor preferido das sementes ligada operacionalmente a;

(b) uma segunda sequência de ácido nucleico codificando um polipéptido da insulina; e

(c) uma sequência de ácido nucleico codificando um polipéptido que é capaz de reter um polipéptido da insulina no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER.

De um modo preferido, pelo menos 0,1% da proteína total de semente presente na semente é insulina. As sementes são uma fonte a partir da qual o polipéptido da insulina desejado que é sintetizado pelas células da semente, pode ser extraído e a insulina pode ser utilizada para tratar doentes diabéticos.

Outras características e vantagens da presente invenção serão rapidamente óbvias a partir da seguinte descrição detalhada. Todavia deve ser entendido que a descrição detalhada e os exemplos específicos, embora indicando formas de realização preferidas da invenção, são apresentados apenas a título de ilustração.

BREVE DESCRIÇÃO DOS DESENHOS

A invenção será agora descrita em relação aos desenhos, nos quais:

a Figura 1 apresenta a sequência de nucleótidos (SEQ ID N° 1) e a sequência de aminoácidos deduzida (SEQ ID N° 2) de uma proteína de fusão de insulina (PRS-D9scFv-Klip27-MI-KDEL) de pSBS4404. A sequência de aminoácidos prevista é apresentada no código de letra única. A sequência de aminoácidos deduzida do péptido de sinal PRS está em itálico, a sequência de aminoácidos deduzida do D9 scFv está a cheio, a sequência de

aminoácidos deduzida da sequência de KLIP 27 está sublinhada, a sequência de aminoácidos deduzida da sequência de mini-insulina está em itálico e cheio e finalmente, a sequência KDEL está a cheio e sublinhada.

A Figura 2 apresenta uma sequência de nucleótidos (SEQ ID N° 3) e sequência de aminoácidos deduzida (SEQ ID N° 4) de uma proteína de fusão de insulina (OLEO-KLIP8-KLIP27-MI) de pSBS4405. As sequências de aminoácidos previstas são apresentadas no código de uma letra. A sequência de aminoácidos deduzida da oleosina de *Arabidopsis thaliana* de 18 kDa está em itálico, a sequência de aminoácidos deduzida da sequência de KLIP 8 está a cheio, a sequência de aminoácidos deduzida da sequência de KLIP27 está a sublinhado e a sequência de aminoácidos deduzida da mini-insulina está em itálico e a cheio.

A Figura 3 apresenta a sequência de ácido nucleico completa (SEQ ID N°:5) e a sequência de aminoácidos (SEQ ID N°:6) da proteína de fusão de insulina 4414 (PRS-MI-ligando tetrabásico-D9Scfv-KDEL). As sequências de aminoácidos previstas estão apresentadas no código de uma letra. A sequência de aminoácidos deduzida do péptido de sinal PRS está em itálico, a sequência de aminoácidos deduzida da mini-insulina (B30 tetrabásico) está em cheio, a sequência de aminoácidos deduzida da sequência do ligando tetrabásico está sublinhada, a sequência de aminoácidos deduzida de D9 scFv está em itálico e a cheio e finalmente, a sequência KDEL está a cheio e sublinhada.

As Figuras 4(A-D) apresentam expressão recombinante de proteínas de fusão de insulina em linhas transformadas de

Arabidopsis thaliana (4404-2, -17, -20 e 4405-4) com base na análise de SDS-PAGE com coloração de Coomassie e Transferência de Western. As setas apresentam a posição dos polipéptidos de fusão de migração de 38,5 kDa e 34,2 kDa, PRS-D9(scfv)-KLIP27-MIw/KDEL e OLEO-KLIPB-KLIP27-MI respectivamente, em condições redutoras.

As Figuras 4A (gel corado com Coomassie) e 4B (Transferência de Western correspondente rastreada com sonda de E2E3 anti-insulina) apresentam a proteína total de semente de linhas de semente tipo selvagem (wt) e transgênicas que expressam as construções 4404 e 4405.

As Figuras 4B (gel corado com Coomassie) e 4D (Transferência de Western correspondente rastreada com sonda de E2E3 anti-insulina) apresentam a proteína de óleo corporal preparada a partir de sementes selvagens e transgênicas que expressam as mesmas construções 4404 e 4405.

As Figuras 4 (E-F) apresentam a expressão recombinante de proteínas de fusão de insulina em linhas transformadas de *Arabidopsis thaliana* (4419-9 e 4414-20) com base na análise de SDS-PAGE corado com Coomassie e Transferência de Western. Os marcadores de peso molecular (M) são 10, 15, 25, 37, 50, 75, 100, 150 kDa. Os controlos incluem, hIN (insulina humana recombinante convencional) e hProIN (pró-insulina humana recombinantes convencional), separados em condições não redutoras.

A Figura 5 apresenta níveis de expressão determinados nas linhas de semente T3 disponíveis (4404-2, -17, -20, 4405-4,

-13, -19) e linhas de semente T2 (4414-9 e-20). Os níveis de transgene e % molar de expressão MI foram determinados com base na densitometria.

A Figura 6 apresenta a análise de SDS-PAGE corado com Coomassie (15%) das preps de corpo oleoso antes da eluição (-OB), prep de OB após eluição com ácido fórmico (-OB') e o material eluído concentrado (-E). A seta representa a posição do polipéptido de fusão de migração. O controlo de tipo selvagem está, essencialmente, isento de quaisquer proteínas principais, após eluição, enquanto que o material de 4404 concentrado contém a proteína de fusão, alguns produtos truncados (possível proteína de fusão hidrolisada) e possivelmente algumas albuminas que co-eluíram.

A Figura 7 apresenta os cromatogramas apresentando os tempos de retenção característicos da insulina humana padrão em comparação com a tripsina clivada eluída proteína de fusão 4404 na coluna C18. A hIN convencional é insulina humana recombinante convencional (0,5 µg).

A Figura 8 apresenta a análise de espectros de massa de padrão de insulina humana (A), em comparação com tripsina clivada e fracções de 4404 (B) de purificação por HPLC, recolhidas a partir dos 17,0-17,5 minutos.

A Figura 9 apresenta a análise de SDS-PAGE corado com Coomassie (15%) da proteína total extraível de semente e proteína preparada com corpo oleoso (OB) a partir das linhas que expressam 4405 em comparação com sementes do tipo selvagem (não recombinante). A seta apresenta a posição do polipéptido de fusão de migração.

A Figura 10 apresenta os cromatogramas apresentando os tempos de retenção característicos do padrão da insulina humana, em comparação com as preparações de 4405 OB por RP-HPLC em coluna C18 clivadas com tripsina. O padrão de hIN é padrão de insulina humana recombinante (0,5 µg).

A Figura 11 apresenta a análise de espectros de massa de padrão de insulina humana (A) em comparação com fracções de 4405 (B) clivados com tripsina e purificadas por HPLC recolhidas a partir dos 17,0-17,5 minutos.

A Figura 12 apresenta o cromatograma das preparações de corpo oleoso de 4405 clivadas com tripsina (linha a tracejado) em comparação com padrão de insulina humana (linha normal). As fracções da insulina clivada eluída foram recolhidas entre 7-35 mS/cm e concentradas por liofilização para bioensaio de insulina.

A Figura 13 apresenta as alterações nos níveis de glucose do soro em murganhos machos B6, após injeção de controlos negativos (círculos a branco = placebo de soro fisiológico, círculos a cheio = corpos oleosos de tipo selvagem clivados com tripsina), controlos positivos (quadrados a branco = Humulina R, triângulos a branco = hIN da Roche) em comparação com insulina derivada de plantas preparada a partir de corpos oleosos de 4405 (losangos a cheio = SBS hIN DesB₃₀).

A Figura 14 apresenta um gel corado com Coomassie de proteínas de corpo oleoso de duas linhas representativas (4409-6 e 4409-8) comparando a migração de proteína de fusão Oleosina-hPIN (como apresentado pela seta preta) com

Arabidopsis não transformada (wt). O nível de expressão foi determinado por densitometria para medir em média cerca de 0,10% da proteína total da semente. Este nível foi calculado acima e para além da co-migração de uma proteína endógena do mesmo peso molecular na semente não transformada (wt) que constituiu aproximadamente 0,04% da proteína total da semente.

DESCRIÇÃO DETALHADA DA INVENÇÃO

Como aqui mencionado anteriormente, a presente invenção refere-se a métodos melhorados para a produção de insulina em plantas transgênicas. A presente requerente verificou surpreendentemente que podem ser alcançados em plantas níveis de acumulação de insulina que excedem 0,1% da proteína total celular através da produção de forma recombinante de insulina nas sementes das plantas. Estes níveis de expressão que são, pelo menos, dez vezes superiores aqueles até agora alcançados, tornam a produção comercial de insulina em plantas viável. A produção em sementes oferece flexibilidade no armazenamento e transporte de insulina como uma matéria bruta, uma vez que a insulina retém a sua actividade após extracção a partir das sementes armazenadas. Além disso, a quantidade de biomassa que precisa de ser sujeita a extracção, é limitada devido ao teor em água, relativamente, baixo presente nas sementes de plantas.

Consequentemente, de acordo com a presente invenção, é proporcionado um método para a expressão de insulina em sementes de plantas, compreendendo:

(a) proporcionar uma construção de ácido nucleico quimérica compreendendo, na direcção 5' para 3' da transcrição, como componentes ligados operacionalmente:

(i) uma sequência de ácido nucleico que compreende um promotor preferido das sementes;

(ii) uma sequência de ácido nucleico codificando um polipéptido da insulina; e

(iii) uma sequência de ácido nucleico codificando um polipéptido que é capaz de reter um polipéptido da insulina no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER;

(b) introduzir a construção de ácido nucleico quimérica numa célula vegetal; e

(c) cultivar a célula vegetal até ficar uma planta madura capaz de produzir sementes. em que a semente expressa insulina; em que um polipéptido da insulina se acumula no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER e em que, pelo menos, 0,1 % da proteína total de semente na semente é insulina; e

(d) obter insulina substancialmente pura a partir das referidas sementes.

De acordo com a presente invenção, verificou-se surpreendentemente que a insulina se acumula nas sementes de plantas em níveis até agora nunca alcançados se a insulina for expressa em semente de um modo que permita o sequestro de um

polipéptido da insulina nas células da semente sem retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER.

Termos e Definições

Salvo definido em contrário, todos os termos técnicos e científicos aqui utilizados devem possuir o mesmo significado que é normalmente atribuído por um especialista na técnica à qual a presente invenção pertende. Quando permitido, todas as patentes, pedidos, pedidos publicados e outras publicações, incluindo sequências de ácido nucleico e polipéptido de GenBank, SwissPro e outras bases de dados a que é feita referência na descrição são incorporadas por referência na sua totalidade.

O termo "sequência de ácido nucleico", como aqui utilizado, refere-se a uma sequência de monómeros de nucleósido ou nucleótido consistindo em bases de ocorrência naturalmente, açúcares e ligações interaçúcar (esqueleto). O termo também inclui sequências modificadas ou substituídas compreendendo monómeros que não ocorrem naturalmente ou as suas porções. As sequências de ácido nucleico da presente invenção podem ser sequências de ácido desoxirribonucleico (ADN) ou sequências de ácido ribonucleico (ARN) e podem incluir bases que ocorrem naturalmente, incluindo adenina, guanina, citosina, timidina e uracilo. As sequências também podem conter bases modificadas. Exemplos dessas bases modificadas incluem aza e desaza adenina, guanina, citosina, timidina e uracilo; e xantina e hipoxantina.

Os termos “sequência de ácido nucleico codificando insulina” e “sequência de ácido nucleico codificando um polipéptido da insulina” que podem ser aqui utilizados indistintamente, referem-se a qualquer e a todas as sequências de ácido nucleico codificando um polipéptido da insulina, incluindo os polipéptidos da insulina listados na Tabela 1 (SEQ ID N°: 7 até 145), assim como a qualquer polipéptido da insulina de mamífero e quaisquer sequências de ácido nucleico que codificam pró-insulina e pré-pró-insulina. Como aqui utilizado “pró-insulina” refere-se a um polipéptido da insulina que inclui o péptido de ligação ou “péptido C” ligando as cadeias de polipéptido B e A da insulina. Na insulina humana nativa, o péptido C é a cadeia de polipéptido de 31 resíduos de aminoácido ligando o resíduo B30 ao resíduo A1. O termo “pré-pró-insulina” refere-se a uma molécula de pró-insulina compreendendo adicionalmente uma sequência sinal N-terminal que coordena a tradução para ocorrer nos ribossomas do ER. Sequências de ácido nucleico codificando um polipéptido da insulina incluem ainda qualquer e todas as sequências de ácido nucleico que (i) codificam polipéptidos que são substancialmente idênticos a um polipéptido das sequências de insulina aqui apresentadas; ou (ii) hibridam com quaisquer sequências de ácido nucleico aqui apresentadas em condições de hibridação, pelo menos, moderadamente restritivas ou que podem hibridar com estas em condições pelo menos moderadamente restritivas, mas para a utilização de códons sinónimos.

Pelo termo “substancialmente idêntico” entende-se que duas sequências de polipéptido são, de um modo preferido, pelo menos, 75% idênticas e, de um modo mais preferido, são, pelo menos, 85% idênticas e de um modo muito preferido pelo menos 95% idênticas, por exemplo, 96%, 97%, 98% ou 99% idênticas. De modo a

determinar a percentagem de identidade entre duas sequências de polipéptido, as sequências de aminoácido dessas duas sequências são alinhadas, de um modo preferido utilizando o algoritmo Clustal W (Thompson, JD, Higgins DG, Gibson TJ, 1994, *Nucleic Acids Res.* 22 (22): 4673-4680, juntamente com a matriz de classificação BLOSUM 62 (Henikoff S. e Henikoff J.G., 1992, *Proc. Natl. Acad. Sci. USA* 89: 10915-10919) e uma penalização de abertura de intervalo de 10 e penalização de extensão de intervalo de 0,1, de modo a que seja obtido o emparelhamento de ordem mais elevada entre duas sequências em que, pelo menos, 50% do comprimento total de uma das sequências está envolvido no alinhamento. Outros métodos que podem ser utilizados para alinhar sequências são o método de alinhamento de Needleman e Wunsch (*J. Mol. Biol.*, 1970,48: 443), como revisto por Smith e Waterman (*Adv. Appl. Math.*, 1981, 2: 482) de modo a que seja obtido o emparelhamento de ordem mais elevada entre as duas sequências e o número de aminoácidos idênticos é determinado entre as duas sequências. Outros métodos para calcular a percentagem de identidade entre duas sequências de aminoácidos são geralmente reconhecidos na técnica e incluem, por exemplo, os descritos por Carillo e Lipton (*SIAM J. Applied Math.*, 1988, 48:1073) e os descritos em *Computational Molecular Biology*, Lesk, e.d. Oxford University Press, Nova Iorque, 1988, *Biocomputing: Informatics and Genomics Projects*. Geralmente, serão utilizados programas de computador para esses cálculos. Programas de computador que podem ser utilizados a este respeito incluem, mas não estão limitados a, GCG (Devereux et al., *Nucleic Acids Res.*, 1984, 12: 387) BLASTP, BLASTN e FASTA (Altschul et al., *J. Molec. Biol.*, 1990: 215: 403).

Por "condições de hibridação pelo menos moderadamente restritivas" entende-se que são seleccionadas condições que

promovem hibridação selectiva entre duas moléculas de ácido nucleico complementares em solução. A hibridação pode ocorrer com a totalidade ou uma porção de uma sequência de molécula de ácido nucleico. A porção de hibridação tem tipicamente pelo menos 15 (e. g. 20, 25, 30, 40 ou 50) nucleótidos de comprimento. Os especialistas na técnica reconhecem que a estabilidade de um duplex de ácido nucleico ou híbridos, é determinada pela T_m que em tampões que contêm sódio é uma função da concentração de ião sódio e temperatura ($T_m = 81,5^\circ\text{C} - 16,6 (\text{Log}_{10} [\text{Na}^+]) + 0,41(\%(G+C) - 600/1)$ ou equação semelhante). Conseqüentemente, os parâmetros nas condições de lavagem que determinam a estabilidade do híbrido são a concentração do ião sódio e a temperatura. De modo a identificar moléculas que são semelhantes, mas não idênticas, a uma molécula de ácido nucleico conhecido, pode ser assumido um desemparelhamento de 1% como resultando numa diminuição de cerca de 1 °C em T_m , por exemplo se as moléculas de ácido nucleico forem consideradas como tendo uma identidade >95%, a temperatura de lavagem final será reduzida em cerca de 5 °C. Com base nestas considerações, os especialistas na técnica são capazes de seleccionar rapidamente condições de hibridação apropriadas. Em formas de realização preferidas, são seleccionadas condições de hibridação restritivas. A título de exemplo, podem ser empregues as seguintes condições para atingir hibridação restritiva: hibridação a 5 x cloreto de sódio/citrato de sódio (SSC)/5 x solução de Denhardt/1,0% SDS a $T_m - 5$ °C com base na equação acima, seguido por uma lavagem de 0,2 x SSC/0,1% SDS a 60 °C. Condições de hibridação moderadamente restritivas, incluem um passo de lavagem em 3 x SSC a 42 °C. Todavia, deve ser entendido que podem ser alcançadas restrições equivalentes utilizando tampões, sais e temperaturas alternativas. Instruções adicionais em relação às condições de

hibridação podem ser encontradas em: *Current Protocols in Molecular Biology*, John Wiley & Sons, N.I., 1989, 6.3.1.-6.3.6 e em: Sambrook *et al.*, *Molecular Cloning, a Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1989, Vol.3.

Como aqui utilizado, os termos “insulina” e “polipéptido da insulina” que podem ser aqui utilizados indistintamente, referem-se a qualquer e a todos os polipéptidos da insulina incluindo polipéptidos da insulina listados na Tabela 1 (SEQ ID N°: 7 até 145) assim como uma molécula de polipéptido compreendendo uma sequência de resíduos de aminoácidos que é (i) substancialmente idêntica às sequências de aminoácidos que constituem quaisquer polipéptidos da insulina aqui apresentados ou (ii) codificada por uma sequência de ácido nucleico capaz de hibridar em condições, pelo menos, moderadamente restritivas com qualquer sequência de ácido nucleico codificando insulina aqui apresentada ou capaz de hibridar em condições pelo menos moderadamente restritivas com qualquer sequência de ácido nucleico codificando insulina aqui apresentada, mas para a utilização de códons sinónimos. Os termos insulina e polipéptido da insulina incluem pró-polipéptidos da insulina e mini-polipéptidos da insulina. Um polipéptido da insulina é, de um modo preferido, de origem humana, porcina ou bovina.

O termo “polipéptido capaz de reter um polipéptido da insulina num compartimento intracelular numa membrana”, como aqui utilizado, refere-se a qualquer polipéptido que quando ligado a um polipéptido da insulina, é capaz de sequestrar um polipéptido da insulina numa estrutura subcelular rodeado por uma membrana e localizado no espaço intracelular da célula vegetal, como definido pela membrana plasmática da célula vegetal.

O termo “um polipéptido capaz de reter um polipéptido da insulina no ER ou numa vesícula de armazenamento derivada de ER” como aqui utilizado refere-se a qualquer polipéptido que quando ligado a um polipéptido da insulina é capaz de sequestrar um polipéptido da insulina no retículo endoplásmico ou num compartimento de armazenamento que é derivado de um retículo endoplásmico, tal como por exemplo um corpo oleoso, numa célula vegetal.

O termo “corpo oleoso” ou “corpos oleosos”, como aqui utilizado refere-se a qualquer organito de armazenamento de óleo ou gordura numa célula de semente vegetal (descrito em por exemplo: Huang (1992) *Ann. Rev. Plant Mol. Biol.* 43: 177-200).

O termo “quimérico”, como aqui utilizado no contexto de sequências de ácido nucleico, refere-se a, pelo menos, duas sequências de ácido nucleico ligadas, que são estão ligadas naturalmente. Sequências de ácido nucleico quiméricas incluem sequências de ácido nucleico ligadas de diferentes origens naturais. Por exemplo, uma sequência de ácido nucleico constituindo um promotor vegetal ligado a uma sequência de ácido nucleico codificando insulina humana é considerada quimérica. Sequências de ácido nucleico quiméricas também podem compreender sequências de ácido nucleico da mesma origem natural, desde que não estejam ligadas naturalmente. Por exemplo, a sequência de ácido nucleico constituindo um promotor obtido a partir de um tipo de célula particular pode estar ligado a uma sequência de ácido nucleico codificando um polipéptido obtida a partir do mesmo tipo de célula, mas não ligada normalmente à sequência de ácido nucleico constituindo o promotor. Sequências de ácido nucleico quiméricas também incluem sequências de ácido nucleico compreendendo qualquer sequência de ácido nucleico que ocorre

naturalmente, ligada a qualquer sequência de ácido nucleico que não ocorre naturalmente.

Preparação de vetores de expressão recombinante compreendendo sequências de ácido nucleico quiméricas codificando insulina e um promotor capaz de controlar a expressão numa célula de semente de planta

As sequências de ácido nucleico codificando insulina que podem ser utilizadas de acordo com os métodos e composições aqui proporcionadas podem ser qualquer sequência de ácido nucleico codificando um polipéptido da insulina, incluindo qualquer pró-insulina e pré-pró-insulina.

Sequências de ácido nucleico exemplares codificando insulina são bem conhecidas na técnica e estão geralmente prontamente disponíveis a partir de uma variedade diversa de fontes de mamíferos, incluindo humanas (Bell, G.I. *et al.*, 1980, *Nature* 284:26-32), porcinas (Chance, R.E. *et al.*, 1968, *Science* 161:165-167), bovinas (D'Agostino, J. *et al.*, 1987, *Mol. Endocrinol.* 1:327-331), ovinas (Peterson, J.D. *et al.*, 1972, *Biol. Chem.* 247:4866-4871) e semelhantes, assim como a partir de fontes vegetais (Oliveira, A.E.A. *et al.*, 1999, *Protein Pept. Lett.* 6:15-21). As sequências codificando insulina que podem ser utilizadas incluem as que codificam cadeias de polipéptido apresentadas como SEQ ID N°:7 até SEQ ID N°:145. As sequências de ácido nucleico correspondentes respectivas codificando cadeias de polipéptido da insulina podem ser rapidamente identificadas através dos números identificadores de Swiss Protein proporcionados na Tabela 1. Utilizando as sequências de ácido nucleico, sequências de ácido nucleico novas codificando

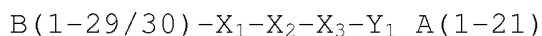
insulina adicionais podem ser rapidamente identificadas utilizando técnicas conhecidas dos especialistas na técnica. Por exemplo, bibliotecas, tais como bibliotecas de expressão, bibliotecas de ADNc e bibliotecas genómicas, podem ser rastreadas e as bases de dados contendo a informação de sequência de projectos de sequenciação pode ser pesquisadas para sequências semelhantes. Podem ser utilizados métodos alternativos para isolar sequências de ácido nucleico adicionais codificando polipéptidos de insulina e podem ser verificadas novas sequências e utilizadas de acordo com a presente invenção. Em formas de realização preferidas, as sequências de ácido nucleico codificando insulina são de insulina humana, porcina e bovina.

São conhecidos da técnica anterior, numerosos análogos da insulina (ver por exemplo Patentes US 5461031; 5474978; 5164366 e 5008241) e podem ser utilizados de acordo com a presente invenção. Os análogos que podem ser aqui utilizados incluem moléculas de insulina humana na qual o resíduo de aminoácido 28 da cadeia B (B28) foi substituído do seu resíduo de prolina natural por aspartato, lisina ou isoleucina. Noutra forma de realização, o resíduo de lisina em B29 é modificado para uma prolina. Além disso, a aspargina em A21 pode ser alterada para alanina, glutamina, glutamato, glicina, histidina, isoleucina, leucina, metionina, serina, treonina, triptofano, tirosina ou valina. Também, a aspargina em B3 pode ser modificada para lisina. Outros exemplos de análogos de insulina que podem ser aqui utilizados incluem: insulina humana sem o resíduo B30, também, frequentemente referido como o "desB30" ou "B(1-29)"; aqueles que não têm os últimos 3 resíduos de aminoácidos de insulina "B(1-27)"; moléculas de insulina sem o resíduo de fenilalanina em B1; e análogos em que a cadeia A ou a cadeia B

possuem uma extensão N-terminal ou C-terminal, por exemplo a cadeia B pode ser estendida no N-terminal pela adição de dois resíduos de arginina.

Em formas de realização preferidas, a sequência de ácido nucleico codificando insulina que é utilizada, é pró-insulina. Em formas de realização preferidas adicionais, são utilizadas moléculas de sequência de ácido nucleico codificando insulina nas quais o péptido C foi modificado em relação à sua forma nativa. Resíduos de aminoácidos no péptido C podem ser substituídos e o péptido C pode ser alongado ou encurtado. A este respeito, como aqui utilizado, o termo "mini-insulina" refere-se a um polipéptido da insulina que pode ser modificado de modo a que o péptido C possa ser encurtado no seu comprimento em relação à sua forma nativa. Em formas de realização preferidas é utilizada uma mini-insulina. De um modo preferido, o péptido C de uma molécula de mini-insulina será mais curta do que 20 resíduos de aminoácidos, de um modo mais preferido mais curto do que 15 resíduos de aminoácidos e, de um modo muito preferido, mais curto do que 9 resíduos de aminoácidos, por exemplo, 7, 5, ou 3 resíduos. De um modo preferido, como é o caso com moléculas de insulina natural, o péptido C da mini-insulina compreenderá um sítio de clivagem nos seus terminais C e N. Esses sítios de clivagem podem ser quaisquer sítios convenientes conhecidos na técnica, por exemplo metionina, clivagem por brometo de cianogénio, um resíduo básico único ou um par de resíduos básicos, clivagem por tripsina ou proteases do tipo tripsina ou através de uma carboxi-peptidase. Por exemplo, o péptido C pode compreender uma lisina C-terminal, por exemplo Ala-Ala-Lys (SEQ ID N°:146) ou um sítio de processamento dibásico imediatamente antes do resíduo GlyA1, por exemplo Asn-Lys-Arg (SEQ ID N°:147), ou Arg-Arg-Lys-Gln-Lys-Arg.

(SEQ ID N°:148) ou um sítio de processamento tetrabásico imediatamente antes do resíduo Gly A1, por exemplo Arg-Arg-Lys-Arg (SEQ ID N°:149). Consequentemente, as moléculas de mini-insulina que podem ser utilizadas de acordo com a presente invenção incluem:



em que

X_1 é qualquer aminoácido;

X_2 é qualquer aminoácido;

X_3 é Lys ou Arg;

Y_1 é uma ligação peptídica ou 1-17 resíduos de aminoácidos;

B(1-29/30) é a cadeia B da insulina humana contendo os resíduos de aminoácidos 1-29 ou 1-30; e

A(1-21) é a cadeia A da insulina humana contendo os resíduos de aminoácidos 1-21.

Numa forma de realização preferida, X_1 é um resíduo de aminoácido básico (Lys ou Arg) e Y_1 é ou uma ligação peptídica ou 1-17 resíduos de aminoácidos, em que o resíduo C-terminal é um resíduo de aminoácido básico (Lys ou Arg).

Além disso, moléculas de mini-insulina que podem ser aqui utilizadas, compreendem aquelas que podem ser representadas pela fórmula:

B(1-27)-X₂-X₃-X₁-Y-A(1-21)

em que

X₁ é um péptido de 1-8 resíduos de aminoácidos compreendendo, pelo menos, um resíduo de aminoácido aromático;

X₂ é um de Pro, Asp, Lys ou Ile na posição 28 da cadeia B;

X₃ é um de Pro, Lys, Ala, Arg ou Pro-Thr na posição 29 da cadeia B;

Y é Lys ou Arg;

B(1-27) é a cadeia B, da cadeia B da insulina humana contendo os resíduos de aminoácidos 1-27; e

A(1-21) é a cadeia A, da cadeia A da insulina humana contendo os resíduos de aminoácidos 1-21.

Exemplos adicionais de moléculas de ácido nucleico codificando mini-polipéptidos de uma insulina que podem ser utilizados de acordo com a presente invenção incluem aqueles descritos em: Markussen *et al.*, Walter de Gruyter & Co. 1987, em: *Peptides* pp 189-194; Thim *et al.*, 1989, em: *Genetics and molecular biology of industrial microorganisms*, American Society for Microbiology pp 322-328; e aqueles apresentados nas Patentes US 4916212; 5324641 e 6521738. Alterações à sequência de ácido nucleico codificando insulina para preparar análogos da insulina podem ser realizados utilizando uma variedade de técnicas de modificação de ácido nucleico conhecidas dos especialistas na técnica, incluindo, por exemplo, mutagênese dirigida a um sítio,

mutagênese direccionada, mutagênese aleatória, a adição de solventes orgânicos, mistura de genes ou uma combinação destas e outras técnicas conhecidas dos especialistas na técnica (Shraishi *et al.*, 1988, *Arch. Biochem. Biophys*, 358: 104-115; Galkin *et al.*, 1997, *Protein Eng.* 10: 687-690; Carugo *et al.*, 1997, *Proteins* 28: 10-28; Hurley *et al.*, 1996, *Biochem*, 35 : 5670-5678; Holmberg *et al.*, 1999, *Protein Eng.* 12: 851-856).

De acordo com a presente invenção foi surpreendentemente verificado que a insulina se acumula nas sementes de plantas até níveis até agora não alcançados, se a insulina for expressa em sementes, de um modo que um polipéptido da insulina nas células da semente é sequestrada no ER ou numa vesícula de armazenamento derivada de ER. De modo a alcançar esse sequestro da insulina no ER ou numa vesícula de armazenamento derivada de ER, de acordo com a presente invenção, o polipéptido codificando insulina está ligado a um polipéptido que provoca a retenção de um polipéptido da insulina no ER ou numa vesícula de armazenamento derivada de ER em vez de ser transportado para fora do ER até, por exemplo, o apoplasto. Polipéptidos que podem ser utilizados de acordo com a presente invenção para reter um polipéptido da insulina no ER, incluem qualquer polipéptido capaz de sequestrar uma insulina no ER. Esses polipéptidos podem ser sintetizados ou obtidos a partir de uma fonte biológica. Numa forma de realização preferida da presente invenção, o polipéptido que é capaz de reter uma insulina é um polipéptido compreendendo um motivo C-terminal de retenção no ER. Exemplos desses motivos de retenção no ER C-terminais incluem as sequências KDEL, HDEL, DDEL, ADEL e SDEL (SEQ ID N°: 150 até 154 respectivamente). Outros exemplos incluem HDEF (SEQ ID N°: 155) (Lehmann *et al.*, 2001. *Plant Physiol* 127(2): 436-49.), ou dois resíduos de arginina próximo do terminal N localizados nas posições 2 e 3, 3 e 4, ou 4 e 5

(Resumo da *Plant Biology* 2001 Program, ASPB, Junho de 2001, Providence, Rhode Island, EUA). Sequências de ácido nucleico codificando um motivo C-terminal de retenção no ER estão, de um modo preferido, ligadas à sequência de ácido nucleico codificando um polipéptido da insulina de modo a que o polipéptido capaz de reter uma insulina no ER está ligado à extremidade do terminal C de um polipéptido da insulina.

De modo a alcançar o sequestro de um polipéptido da insulina numa vesícula de armazenamento derivada de ER, um polipéptido da insulina está ligado a um polipéptido que é capaz de reter um polipéptido da insulina numa vesícula de armazenamento derivada de ER. O polipéptido capaz de reter um polipéptido da insulina numa vesícula de armazenamento derivada de ER que pode ser utilizada de acordo com os métodos aqui apresentados pode ser qualquer polipéptido capaz de sequestrar um polipéptido da insulina numa vesícula de armazenamento derivada de ER. Polipéptidos capazes de reter insulina numa vesícula de armazenamento derivada de ER podem ser sintetizados ou obtidos a partir de qualquer fonte biológica. Numa forma de realização preferida, a vesícula de armazenamento derivada de ER é um corpo oleoso e um polipéptido da insulina é ligada a uma proteína de corpo oleoso ou uma porção suficiente desta capaz de reter um polipéptido da insulina na vesícula de armazenamento derivada de ER. Proteínas de corpo oleoso que podem ser utilizadas a este respeito incluem qualquer proteína que se associa naturalmente a um corpo oleoso. Proteínas de corpo oleoso que são particularmente preferidas são oleosinas, por exemplo uma oleosina de *Arabidopsis* (van Rooijen *et al.* (1991) *Plant Mol Biol.* 18: 1177-1179) oleosina de milho (Bowman-Vance *et al.*, 1987, *J. Biol. Chem.* 262: 11275-11279; Qu *et al.*, 1990, *J. Biol. Chem.* 265: 2238-2243), oleosina de cenoura (Hatzopoulos

et al. (1990) *Plant Cell* 2: 457-457) ou oleosina de *Brassica* (Lee *et al.*, 1991, *Plant Physiol.* 96: 1395-1397), caleosinas, ver por o número de acesso de Genbank AF067857) e estereoleosinas (Lin *et al.*, 2002 *Plant Physiol.* 128(4):1200-11). Numa outra forma de realização preferida, a proteína do corpo oleoso é uma oleosina vegetal e partilha uma semelhança de sequência com outras oleosinas vegetais, tais como a oleosina isolada de *Arabidopsis thaliana* (SEQ ID N°:156) ou *Brassica napus* (SEQ ID N°:157). Noutra forma de realização, a proteína do corpo oleoso é uma caleosina ou proteína de ligação ao cálcio de plantas, fungos ou outras fontes e partilha homologia de sequência com caleosinas de plantas, tais como a caleosina isolada de *Arabidopsis thaliana* (SEQ ID N°:158 e SEQ ID N°:159) Noutra forma de realização, a proteína do corpo oleoso é uma estereoleosina (SEQ ID N°:160), uma desidrogenase de ligação ao esterol (Lin L-J *et al.*, (2002) *Plant Physiol* 128: 1200-1211). O polipéptido codificando insulina pode ser ligado à proteína do corpo oleoso ao terminal N, assim como ao terminal C e aos fragmentos de uma proteína de corpo oleoso, tais como por exemplo o domínio central de uma oleosina. Podem ser descobertas novas proteínas de corpo oleoso, por exemplo preparando corpos oleosos (para metodologias para preparar corpos oleosos, ver por exemplo, Patente US 6650554) e identificar proteínas nas preparações do corpo oleoso através de, por exemplo, electroforese em gel na presença de SDS. Podem ser criados anticorpos policlonais contra as proteínas e utilizados para rastrear bibliotecas de ADNc de modo a identificar sequências de ácido nucleico codificando proteínas de corpo oleoso. Podem ser descobertas ainda outras detectadas novas proteínas de corpo oleoso utilizando sequências de ácido nucleico conhecidas codificando proteínas de corpo oleoso utilizando, por exemplo, as sequências de proteína de corpo oleoso aqui mencionadas

anteriormente codificando proteínas de corpo oleoso, para rastrear por exemplo bibliotecas de ADNc ou genómicas para a presença de proteínas de corpo oleoso.

Polipéptidos capazes de reter a insulina no ER ou num organito de armazenamento derivado de ER são tipicamente não clivados e a insulina pode acumular-se na forma de uma proteína de fusão que é, por exemplo, tipicamente o caso quando um sinal de retenção KDEL é utilizado para reter o polipéptido no ER ou quando uma proteína de corpo oleoso é utilizado para reter o polipéptido no organito de armazenamento derivado de ER.

A sequência de ácido nucleico quimérica adicionalmente pode conter uma sequência nucleica que tem como alvo a sequência de ácido nucleico para o sistema de endomembrana ("péptico sinal"). Nas formas de realização da presente invenção nas quais o polipéptido da insulina é retido no ER utilizando uma sequência capaz de reter o polipéptido no ER, tais como polipéptido KDEL, HDEL ou SDEL, é particularmente desejável incluir uma sequência de ácido nucleico codificando um péptido sinal. Péptidos sinal exemplares que podem ser aqui utilizados incluem a sequência sinal da proteína relacionada com a patogénese do tabaco (PR-S) (SEQ ID N°:161) (Sijmons *et al.*, 1990, *Bio/technology*, 8:217-221), sequência sinal da lectina (Boehn *et al.*, 2000, *Transgenic Res*, 9(6):477-86), sequência sinal de glicoproteína rica em hidroxiprolina de *Phaseolus vulgaris* (Yan *et al.*, 1997, *Plant Physiol.* 115(3):915-24 e Corbin *et al.*, 1987, *Mol Cell Biol* 7(12):4337-44), sequência sinal da patatina de batata (Iturriaga, G *et al.*, 1989, *Plant Cells* 1:381-390 e Bevan *et al.*, 1986, *Nuc. Acids Res.* 41:4625-4638.) e a sequência sinal da amilase alfa de cevada (Rasmussen e Johansson, 1992, *Plant Mol. Biol.* 18(2):423-7). Esses sinais de direccionamento podem ser

removidos por clivagem *in vivo* de uma sequência de insulina que, por exemplo, é tipicamente o caso quando é utilizado um sinal de direccionamento de apoplasto, tal como a sequência sinal da proteína S relacionada com a patogénese do tabaco (PR-S) (Sijmons *et al.*, 1990, *Bio/technology*, 8:217-221). Podem ser previstos outros péptidos sinal utilizando o servidor da World Wide Web SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) que prevê a presença e a localização de sítios de clivagem de péptido sinal em sequências de aminoácidos de diferentes organismos. Em geral, há pouca conservação da sequência de aminoácidos primários, embora as propriedades físico-químicas gerais sejam conservadas até certo ponto. A estrutura genérica de péptidos sinal possui 3 regiões, a “região n” amino-terminal curta contém resíduos carregados positivamente, a “região h” central hidrofóbica varia de tamanho desde 7 até 15 aminoácidos e a “região c” carboxi-terminal contém aminoácidos polares e um sítio de clivagem que é reconhecido por enzimas peptidase de sinal ligadas à membrana (Nakai K., 2000, *Advances in Protein Chem* 54: 277-344). Um sinal de direccionamento que também pode ser utilizado de acordo com este, inclui a sequência sinal de insulina nativa (24 aminoácidos de comprimento no caso da sequência humana). Em formas de realização preferidas aqui, uma sequência de direccionamento de apoplasto localizada no terminal N, tal como a sequência aqui mencionada anteriormente de PR-S de tabaco é utilizada combinada com uma sequência de retenção em ER localizada no terminal C, tal como a sequência KDEL.

Numa outra forma de realização preferida, a sequência de ácido nucleico codificando uma sequência líder de factor α de levedura é ligada à extremidade N-terminal da sequência de ácido nucleico codificando insulina. As sequências líder de levedura ou sequências derivadas das sequências líder de levedura que

podem ser utilizadas de acordo com o exposto incluem as listadas na SEQ ID N°:162 até SEQ ID N°:171 (Kjeldsen *et al.*, 2001, *Biotechnology and Genetic Engineering Reviews* 18: 89-121). Essas sequências líder podem ainda compreender um péptido espaçador localizado no terminal C do ácido nucleico que codifica a sequência líder e N-terminalmente da sequência que codifica insulina. De acordo com o exposto, essas sequências espaçadoras possuem tipicamente entre 2 e 20 aminoácidos de comprimento. Assim, por exemplo, podem ser utilizadas as sequências espaçadoras SEQ ID N°:172 e SEQ ID N°:173 (Kjeldsen *et al.*, 2001, *Biotechnology and Genetic Engineering Reviews* 18: 89-121). Em formas de realização da presente invenção nas quais é utilizada uma sequência líder de levedura, a sequência de ácido nucleico codificando um polipéptido da insulina é, de um modo preferido, um mini-polipéptido da insulina. De acordo com o exposto, numa forma de realização particularmente preferida, é utilizada a sequência de ácido nucleico codificando um anticorpo de cadeia única ligado a uma sequência de ácido nucleico codificando um péptido líder de secreção de levedura, como aqui descrito adiante no Exemplo 1.

A sequência de ácido nucleico quimérica também pode compreender polipéptidos resultando em extensões da proteína de estabilização N- e/ou C-terminal. Essas extensões podem ser utilizadas para estabilizar e/ou assistir na dobragem de uma cadeia de polipéptido de insulina e adicionalmente podem ser utilizadas para facilitar a purificação da insulina. Extensões de polipéptido que podem ser utilizadas a este respeito incluem, por exemplo, uma sequência de ácido nucleico codificando um anticorpo de cadeia simples, um ácido nucleico codificando uma molécula de Affibody® (Affibody AB), uma sequência de ácido nucleico codificando a subunidade não-tóxica B da toxina da

cólera (CTB) (Arakawa, T. *et al.*, 1998, *Nat. Biotechnol.* 16:938) ou combinações desses polipéptidos. Numa forma de realização particularmente preferida, um polipéptido da insulina é retido num compartimento rodeado por uma membrana, tal como o ER, utilizando, por exemplo, uma sequência KDEL como aqui descrito anteriormente, combinada com um polipéptido estabilizador que permite a associação de um polipéptido da insulina com um corpo oleoso após quebra da integridade da célula vegetal, tal como irá ocorrer quando um polipéptido da insulina é recuperado da célula vegetal. Um exemplo de um tal polipéptido estabilizador é um anticorpo de cadeia simples com especificidade para um corpo oleoso. Sequências de ácido nucleico codificando anticorpos de cadeia simples com especificidade para um corpo oleoso podem ser preparadas a partir de linhas celulares de hibridoma expressando anticorpos monoclonais criados contra uma proteína de corpo oleoso. Numa forma de realização, o anticorpo de cadeia simples liga-se especificamente a uma oleosina, como descrito por Alting-Mees *et al.* (2000) IBC's International Conference on Antibody Engineering, Poster #1. Esta forma de realização da presente invenção é aqui ainda mais detalhada no Exemplo 1.

Noutra forma de realização, um sítio de clivagem pode estar localizado a montante do terminal N e a jusante do terminal C de uma insulina permitindo que um polipéptido da insulina seja clivado a partir do parceiro de fusão, obtendo desse modo insulina isolada. Exemplos desses sítios de clivagem podem ser encontrados no documento WO 98/49326 (Método para a clivagem de proteínas de fusão) e aplicações relacionadas e LaVallie *et al.* (1994) *Enzymatic and chemical cleavage of fusion proteins* In *Current Protocols in Molecular Biology* pp 16.4.5-16.4.17, John Wiley and Sons, Inc., Nova Iorque NY. Numa forma de realização preferida, o sítio de clivagem é um ligando tetrabásico (por

exemplo Arg-Arg-Lys-Arg- SEQ ID N°: 149) que é clivável pela tripsina. Numa outra forma de realização preferida, o sítio de clivagem é KLLP 8 (SEQ ID N°: 174) que é clivável por proteases aspárticas, incluindo quimosinas.

A invenção proporciona ainda métodos para a separação de proteínas heterólogas de componentes da célula hospedeira repartindo a fracção do corpo oleoso e subsequente libertação da proteína heteróloga através da clivagem específica da proteína heteróloga - proteína de fusão do corpo oleoso. Opcionalmente, um sítio de clivagem pode estar localizado a montante do terminal N e a jusante do terminal C do polipéptido heterólogo, permitindo que um polipéptido de fusão seja clivado e separado por separação de fase nos seus péptidos componentes.

A sequência de ácido nucleico codificando insulina pode ser alterada, para melhorar ainda mais os níveis de expressão por exemplo, otimizando a sequência de ácido nucleico de acordo com a utilização de codões preferida para um tipo de célula vegetal particular que é seleccionado para a expressão de um polipéptido da insulina ou alterando os motivos conhecidos para desestabilizar o ARNm (ver por exemplo: Pedido de Patente PCT 97/02352). A comparação entre a utilização de codões da sequência de ácido nucleico codificando um polipéptido da insulina e a utilização de codões do tipo de célula vegetal irá permitir a identificação dos codões que pode ser alterada. A construção de genes sintéticos alterando a utilização de codões é descrita, por exemplo, no Pedido de Patente PCT 93/07278.

Numa forma de realização preferida, a sequência de ácido nucleico codificando insulina que é utilizada é representada por SEQ ID N°:1, SEQ ID N°:3, SEQ ID N°:5, ou SEQ ID N°: 195.

São aqui ensinados promotores derivados de plantas capazes de controlar a expressão de polipéptidos de plantas. Geralmente, serão utilizados promotores obtidos a partir de espécies de plantas dicotiledóneas quando é seleccionada uma planta dicotiledónea, de acordo com o exposto, enquanto que será utilizado um promotor de planta monocotiledónea quando é seleccionada uma espécie de planta monocotiledónea. Promotores constitutivos que incluem, por exemplo, o promotor 35S do vírus do mosaico da couve-flor (CaMV) (Rothstein *et al.*, 1987, *Gene* 53: 153-161), o promotor da actina do arroz (McElroy *et al.*, 1990, *Plant Cell* 2: 163-171; Patente US 6429357), um promotor da ubiquitina, tal como o promotor da ubiquitina de milho (Patentes US 5879903; 5273894) e o promotor da ubiquitina da salsa (Kawalleck, P. *et al.*, 1993, *Plant Mol. Biol.* 21:673-684 são aqui descritos.

O promotor que é utilizado no método da presente invenção é um promotor que resulta na expressão preferencial de um polipéptido da insulina em tecido de semente. "Promotores preferido das sementes", neste contexto, são promotores que controlam a expressão de uma proteína recombinante (*i. e.*, insulina) de modo a que, de um modo preferido, pelo menos, 80% da quantidade total de proteína recombinante presente na planta madura esteja presente na semente. De um modo mais preferido, pelo menos, 90% da quantidade total de proteína recombinante presente na planta madura está presente na semente. De um modo muito preferido, pelo menos, 95% da quantidade total de proteína recombinante presente na planta madura está presente na semente. Promotores preferidos das sementes que podem ser utilizados neste contexto incluem, por exemplo, o promotor da faseolina de feijão (Sengupta-Gopalan *et al.*, 1985, *Proc. Natl. Acad. Sci. USA* 82: 3320-3324); o promotor da oleosina de *Arabidopsis* de 18

kDa (Patente US 5 792922) ou o promotor da oleosina de linho (documento WO 01/16340); o promotor da proteína de armazenamento da semente do tipo linho (linina) (documento WO 01/16340); o promotor da proteína de armazenamento de 2S de linho (documento WO 01/16340); um promotor de endosperma preferido, tal como o promotor Amy32b (Rogers e Milliman, *J. Biol. Chem.*, 1984, 259: 12234-12240, o promotor Amy6-4 (Kursheed e Rogers, *J. Biol. Chem.*, 1988, 263: 18953-18960 ou o promotor Aleurain (Whittier *et al.*, 1987, *Nucleic Acids Res.*, 15: 2515-2535) ou o promotor arcelino de feijão (Jaeger GD, *et al.*, 2002, *Nat. Biotechnol.* Dez; 20:1265-8). São constantemente detectados novos promotores úteis em várias plantas. Numerosos exemplos de promotores de planta podem ser encontrados em Ohamuro *et al.* (*Biochem. of Plnts.*, 1989, 15: 1-82).

Podem ser aqui utilizados determinados elementos genéticos capazes de aumentar a expressão de um polipéptido da insulina. Estes elementos incluem as sequências líder não traduzidas de certos vírus, tais como a sequência líder AMV (Jobling e Gehrke, 1987, *Nature*, 325: 622-625) e o intrão associado ao promotor da ubiquitina de milho (Patente US 5504200). Geralmente, a sequência de ácido nucleico quimérica será preparada de modo a que elementos genéticos capazes de aumentar a expressão sejam localizados a 5' da sequência de ácido nucleico codificando um polipéptido da insulina.

De acordo com a presente invenção, as sequências de ácido nucleico quiméricas compreendendo um promotor capaz de controlar a expressão em sementes de plantas ligadas a uma sequência de ácido nucleico codificando um polipéptido da insulina podem ser integradas num vector de expressão recombinante que assegura a boa expressão no óleo de semente. Um vector de expressão

recombinante compreendendo na direcção 5' para 3' da transcrição como componentes ligados operacionalmente:

(i) uma sequência de ácido nucleico capaz de controlar a expressão em células vegetais da semente; e

(ii) uma sequência de ácido nucleico codificando um polipéptido da insulina; em que o vector de expressão é adequado para a expressão numa célula de semente é aqui abordada.

O termo "adequado para expressão numa célula de semente" significa que o vector de expressão recombinante compreende a sequência de ácido nucleico quimérica da presente invenção ligada a elementos genéticos necessários para alcançar a expressão numa célula de semente. Elementos genéticos que podem ser incluídos no vector de expressão neste contexto, incluem uma região de terminação de transcrição, uma ou mais sequências de ácido nucleico codificando genes marcadores, uma ou mais origens de replicação e semelhantes. Em formas de realização preferidas, o vector de expressão compreende ainda elementos genéticos necessários para a integração do vector ou uma porção deste no genoma nuclear da célula vegetal, por exemplo as sequências de fronteira à esquerda e direita de ADN-T que facilitam a integração no genoma nuclear da planta em formas de realização da invenção nas quais as células vegetais são transformadas utilizando *Agrobacterium*.

Como aqui mencionado anteriormente, o vector de expressão recombinante compreende geralmente um terminador de transcrição que, para além de servir como um sinal para a terminação da transcrição ainda pode servir como um elemento protector capaz

de prolongar a semi-vida do ARNm (Guarneros *et al.*, 1982, *Proc. Natl. Acad. Sci. USA*, 79: 238-242). O terminador de transcrição tem geralmente desde cerca de 200 nucleótidos até cerca de 1000 nucleótidos e o vector de expressão é preparado de modo a que o terminador de transcrição esteja localizado a 3' da sequência de ácido nucleico codificando insulina. As sequências de terminação que podem ser aqui utilizadas incluem, por exemplo, a região de terminação da nopalina (Bevan *et al.*, 1983, *Nucl. Acids. Res.*, 11: 369-385), o terminador da faseolina (van der Geest *et al.*, 1994, *Plant J.* 6: 413-423), o terminador da arcelina ((Jaeger GD, *et al.*, 2002, *Nat. Biotechnol.* Dez; 20:1265-8)), o terminador para os genes da sintase da octopina de *Agrobacterium tumefaciens* ou outros elementos de funcionamento semelhante. Os terminadores de transcrição podem ser obtidos como descrito por An (An, 1987, *Methods in Enzym.* 153: 292).

De acordo com a presente invenção, o vector de expressão pode conter ainda um gene marcador. Genes marcadores que podem ser utilizados de acordo com a presente invenção incluem todos os genes que permitem a distinção de células transformadas de células não transformadas, incluindo todos os genes marcadores seleccionáveis e rastreáveis. Um gene marcador pode ser um marcador de resistência, tal como um marcador de resistência a antibiótico contra, por exemplo, canamicina (Patente US 6174724), ampicilina, G418, bleomicina, higromicina que permite a selecção de uma característica através de meios químicos ou um marcador de tolerância contra um agente químico, tal como a manose de açúcar normalmente fitotóxico (Negrotto *et al.*, 2000, *Plant Cell Rep.* 19: 798-803). Outros marcadores convenientes que podem ser aqui utilizados incluem marcadores capazes de conferir resistência contra herbicidas, tais como

glifosato (Patentes US 4940935; 5188642), fosfinotricina (Patente US 5879903) ou sulfonilureias (Patente US 633437). os marcadores de resistência, quando ligados em íntima proximidade com a sequência de ácido nucleico codificando um polipéptido da insulina, podem ser utilizados para manter a pressão de selecção numa população de células vegetais ou plantas que não perderam a sequência de ácido nucleico codificando um polipéptido da insulina. Marcadores rastreáveis que podem ser empregues para identificar transformantes através de inspecção visual incluem β -glucuronidase (GUS) (Patentes US 5268463 e 5599670) e proteína verde fluorescente (GFP) (Niedz *et al.*, 1995, *Plant Cell Rep.*, 14: 403).

Vectores recombinantes adequados para a introdução de sequências de ácido nucleico em plantas incluem vectores baseados em *Agrobacterium* e *Rhizobium*, tais como os plasmídeos Ti e Ri, incluindo por exemplo pBIN19 (Bevan, *Nucl. Acid. Res.*, 1984, 22: 8711-8721), pGKB5 (Bouchez *et al.*, 1993, *C R Acad. Sci. Paris, Life Sciences*, 316:1188-1193), a série pCGN de vectores binários (McBride e Summerfelt, 1990, *Plant Mol. Biol.*, 14:269-276) e outros vectores binários (e. g., Patente US 4940838).

Os vectores de expressão recombinante, sequências de ácido nucleico e sequências de ácido nucleico quiméricas da presente invenção podem ser preparadas de acordo com metodologias bem conhecidas dos especialistas na técnica da biologia molecular. Essa preparação envolverá tipicamente a espécie bacteriana *Escherichia coli* como um hospedeiro de clonagem intermédio. A preparação dos vectores de *E. coli*, bem como os vectores de transformação de plantas pode ser realizada utilizando técnicas comuns conhecidas, tais como digestão de restrição, ligação,

electroforese em gel, sequenciação de ADN, Reacção da Polimerase em Cadeia (PCR) e outras metodologias. Estas metodologias permitem a ligação de sequências de ácido nucleico e polipéptidos às quais a presente invenção se refere. Está disponível uma vasta variedade de vectores de clonagem para realizar os passos necessários para preparar um vector de expressão recombinante. Entre os vectores com um sistema funcional de replicação em *E. coli*, estão vectores, tais como pBR322, os vectores da série pUC, os vectores da série M13mp, pBluescript etc. Tipicamente, estes vectores de clonagem contêm um marcador que permite a selecção das células transformadas. Podem ser introduzidas sequências de ácido nucleico nestes vectores e os vectores podem ser introduzidos em *E. coli* cultivada num meio apropriado. Os vectores de expressão recombinante podem ser rapidamente recuperados das células após recolha e lise das células. Além disso, podem ser encontradas informações gerais em relação à preparação de vectores recombinantes em, por exemplo: Sambrook *et al.*, *Molecular Cloning, a Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1989, Vol.3.

Preparação de plantas compreendendo sementes capazes de expressar insulina

De acordo com a presente invenção, a sequência de ácido nucleico quimérica é introduzida numa célula vegetal e as células são cultivadas até ficarem plantas maduras capazes de produzir semente, em que a semente expressa um polipéptido da insulina.

De acordo com o exposto, qualquer espécie de planta ou célula vegetal pode ser seleccionada. Células particulares aqui utilizada incluem células que se podem obter a partir de *Arabidopsis thaliana*, castanha do Pará (*Betholettia excelsa*); carrapateira (*Ricinus communis*); coco (*Cocus nucifera*); coentros (*Coriandrum sativum*); algodão (*Gossypium* spp.); de amendoim (*Arachis Hypogaea*); jojoba (*Simmondsia chinensis*); linho/linhaça (*Linum usitatissimum*); milho (*Zea mays*); mostarda (*Brassica* spp. e *Sinapis alba*); palmeira dendem (*Elaeis guineensis*); oliveira (*Olea eurpaea*); colza (*Brassica* spp.); arroz (*Oryza sativa*); açafroa (*Carthamus tinctorius*); soja (*Glycine max*); abóbora-menina (*Cucurbita maxima*); cevada (*Hordeum vulgare*); trigo (*Traeticum aestivum*); e girassol (*Helianthus annuus*).

De acordo com o exposto numa forma de realização preferida são utilizadas espécies de plantas ou células vegetais de plantas oleaginosas. As plantas oleaginosas que podem ser aqui utilizadas incluem amendoim (*Arachis hypogaea*); mostarda (*Brassica* spp. e *Sinapis alba*); colza (*Brassica* spp.); grão-de-bico (*Cicer arietinum*); soja (*Glycine max*); algodão (*Gossypium hirsutum*); girassol (*Helianthus annuus*); lentilha (*Lens culinaris*); linho/linhaça (*Linum usitatissimum*); trevo branco (*Trifolium repens*); oliveira (*Olea eurpaea*); palmeira dendem (*Elaeis guineensis*); açafroa (*Carthamus tinctorius*) e faveta de Beja (*Vicia narbonensis*).

De acordo com o exposto, numa forma de realização particularmente preferida é utilizada açafroa, *Arabidopsis* ou linhaça.

Metodologias para introduzir vectores de expressão recombinante vegetais numa célula vegetal, também aqui referidas como "transformação", são bem conhecidas na técnica e tipicamente variam dependendo da célula vegetal que é seleccionada. Técnicas gerais para introduzir vectores de expressão recombinante em células incluem, electroporação; técnicas mediadas quimicamente, por exemplo, incorporação de ácido nucleico mediado por CaCl_2 ; bombardeamento de partículas (biolística); a utilização de sequências de ácido nucleico naturalmente infecciosas, por exemplo, sequências de ácido nucleico derivadas de vírus ou sequências derivadas de *Agrobacterium* ou *Rhizobium*, incorporação de ácido nucleico mediada por polietilenoglicol (PEG), microinjecção e a utilização de palhetas de carboneto de silício.

Em formas de realização preferidas, é seleccionada uma metodologia de transformação que irá permitir a integração da sequência de ácido nucleico quimérica no genoma da célula vegetal e de um modo preferido o genoma nuclear da célula vegetal. De acordo com o exposto, isto é considerado particularmente desejável na medida em que a utilização dessa metodologia resultará na transferência da sequência de ácido nucleico quimérica para as plantas de progenia após reprodução sexual. Métodos de transformação que podem ser utilizados a este respeito, incluem biolística e métodos mediados por *Agrobacterium*.

Metodologias de transformação para a espécie de plantas dicotiledóneas são bem conhecidas. Geralmente, a transformação mediada por *Agrobacterium* é utilizada devido à sua eficiência, assim como à sensibilidade geral por parte de muitas, senão todas, as espécies de plantas dicotiledóneas. A transformação

com *Agrobacterium* envolve geralmente a transferência de um vector binário, tais como um dos vectores binários aqui mencionados anteriormente, compreendendo a sequência de ácido nucleico quimérica da presente invenção a partir de *E. coli* para uma estirpe de *Agrobacterium* adequada (e. g. EHA101 e LBA4404) através de, por exemplo, emparelhamento triparental com uma estirpe de *E. coli* comportando o vector binário recombinante e uma estirpe de *E. coli* comportando um plasmídeo auxiliar capaz de mobilizar o vector binário para a estirpe de *Agrobacterium* alvo, ou por transformação de ADN da estirpe de *Agrobacterium* (Hofgen et al., *Nucl. Acids Res.*, 1988, 16:9877). Outras técnicas que podem ser utilizadas para transformar células vegetais de dicotiledónea, incluem biolística (Sanford, 1988, *Trends in Biotechn.* 6:299-302); electroporação (Fromm et al., 1985, *Proc. Natl. Acad. Sci. USA.*, 82:5824-5828); incorporação de ADN mediada por PEG (Potrykus et al., 1985, *Mol. Gen. Genetics*, 199:169-177); microinjecção (Reich et al., *Bio/Techn.*, 1986, 4:1001-1004); e palhetas de carboneto de silício (Kaepler et al., 1990, *Plant Cell Rep.*, 9:415-418) ou em transformação de plantas utilizando, por exemplo, uma metodologia de gotejamento em flores (Clough e Bent, 1998, *Plant J.*, 16:735-743).

A espécie de plantas monocotiledóneas pode ser transformada utilizando uma variedade de metodologias incluindo bombardeamento de partículas (Christou et al., 1991, *Biotechn.* 9:957-962; Weeks et al., *Plant Physiol.*, 1993, 102:1077-1084; Gordon-Kamm et al., *Plant Cell*, 1990, 2:5603-618); incorporação de AND mediada por PEG (Patentes Europeias 0292435; 0392225) ou transformação mediada por *Agrobacterium* (Goto-Fumiyuki et al., 1999, *Nature-Biotech.* 17:282-286).

A metodologia de transformação da planta exacta pode variar em certa medida dependendo da espécie vegetal e o tipo de célula vegetal (e. g., tipos celulares derivados de plantas derivadas de sementes, tais como hipocótilos e cotilédones ou tecido embrionário) que é seleccionado como a célula alvo para transformação. Como aqui mencionado anteriormente numa forma de realização particularmente preferida é utilizada açafroa, *Arabidopsis* ou linho. Uma metodologia para obter transformantes de açafroa está disponível em Baker e Dyer (*Plant Cell Rep.*, 1996, 16:106-110). Podem ser encontrados protocolos de transformação específica de espécies de plantas adicionais em: *Biotechnology in Agriculture and Forestry 46: Transgenic Crops I* (Y.P.S. Bajaj ed.), Springer-Verlag, Nova Iorque (1999) e *Biotechnology in Agriculture and Forestry 47: Transgenic Crops II* (Y.P.S. Bajaj ed.), Springer-Verlag, Nova Iorque (2001).

Após transformação, as células vegetais são cultivadas e após a emergência de tecido em diferenciação, tais como rebentos e raízes, são regeneradas plantas maduras. Tipicamente, é regenerada uma pluralidade de plantas. Metodologias para regenerar plantas são geralmente dependentes da espécie de planta e do tipo de célula e serão bem conhecidos dos especialistas na técnica. Informação adicional em relação a cultura de tecidos vegetais pode ser encontrada em, por exemplo: *Plant Cell and Tissue Culture*, 1994, Vasil e Thorpe Eds., Kluwer Academic Publishers; e em: *Plant Cell Culture Protocols (Methods in Molecular Biology 111)*, 1999, Hall Eds, Humana Press.

Num aspecto, a presente invenção proporciona um método de recuperação de sementes de plantas compreendendo insulina. Consequentemente, a presente invenção proporciona um método para obter sementes de plantas compreendendo insulina, compreendendo:

(a) proporcionar uma construção de ácido nucleico quimérica compreendendo na direcção 5' para 3' da transcrição como componentes ligados operacionalmente:

(i) uma sequência de ácido nucleico que compreende um promotor preferido das sementes;

(ii) uma sequência de ácido nucleico codificando um polipéptido da insulina; e

(iii) uma sequência de ácido nucleico codificando um polipéptido que é capaz de reter um polipéptido da insulina no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER.

(b) introduzir a construção de ácido nucleico quimérica numa célula vegetal;

(c) cultivar a célula vegetal até ficar uma planta madura capaz de produzir semente; e

(d) obter sementes da referida planta em que a semente compreende insulina, em que o referido polipéptido da insulina se acumula no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER e em que, pelo menos, 0,1% da proteína total de semente presente na semente é insulina.

Em formas de realização preferidas, é obtida uma pluralidade de plantas transformadas, cultivadas e rastreadas, para a presença da sequência de ácido nucleico quimérica

desejada, cuja presença em transformantes putativos podem ser testados por, por exemplo, cultivo num meio selectivo, quando são utilizados marcadores de resistência a herbicida, por aplicação directa do herbicida na planta ou por transferência de Southern. Se a presença da sequência de ácido nucleico quimérica é detectada, as plantas transformadas podem ser seleccionadas para criar progenia e por último plantas maduras compreendendo uma pluralidade de sementes compreendendo a sequência de ácido nucleico quimérica desejada. Essas sementes podem ser utilizadas para isolar insulina ou podem ser plantadas para criar duas ou mais gerações subsequentes. Será geralmente desejável plantar uma pluralidade de sementes transgênicas para obter uma população de plantas transgênicas, cada uma compreendendo sementes compreendendo uma sequência de ácido nucleico quimérica codificando insulina. Além disso, será geralmente desejável assegurar a homozigotidade nas plantas para assegurar a continuação da herança do polipéptido recombinante. Métodos para seleccionar plantas homozigóticas são bem conhecidos dos especialistas na técnica. Métodos para obter plantas homozigóticas que podem ser utilizadas incluem a preparação e transformação de células haplóides ou tecidos seguidos pela regeneração de plântulas haplóides e a conversão subsequente em plantas diplóides, por exemplo, pelo tratamento com colchicina ou outros agentes de rebentamento de microtúbulos. As plantas podem ser cultivadas de acordo com práticas agrícolas em todo o resto convencionais.

Noutro aspecto, a presente invenção também proporciona plantas capazes de estabelecer sementes que expressam insulina. Numa forma de realização preferida da invenção, é proporcionada uma planta capaz de produzir sementes compreendendo uma sequência de ácido nucleico quimérica compreendendo na direcção

5' para 3' da transcrição:

(a) uma primeira sequência de ácido nucleico que compreende um promotor preferido das sementes ligado operacionalmente a;

(b) uma segunda sequência de ácido nucleico codificando um polipéptido da insulina; e

(c) uma sequência de ácido nucleico codificando um polipéptido que é capaz de reter um polipéptido da insulina no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER, em que a semente contém insulina, em que o referido polipéptido da insulina se acumula no retículo endoplásmico (ER) ou na vesícula de armazenamento derivada de ER e em que, pelo menos, 0,1% da proteína solúvel total presente na semente é insulina.

Numa forma de realização preferida a sequência de ácido nucleico quimérica é integrada estavelmente no genoma nuclear da planta.

Ainda noutro aspecto, a presente invenção proporciona sementes de plantas expressando insulina. Numa forma de realização preferida da invenção, é proporcionada uma semente de planta compreendendo uma sequência de ácido nucleico quimérica, compreendendo na direcção 5' para 3' da transcrição:

(a) uma primeira sequência de ácido nucleico que compreende um promotor preferido das sementes ligadas operacionalmente a;

(b) uma segunda sequência de ácido nucleico codificando um polipéptido da insulina; e

(c) uma sequência de ácido nucleico codificando um polipéptido que é capaz de reter um polipéptido da insulina no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER.

De acordo com a presente invenção, é obtida uma semente na qual, de um modo preferido, pelo menos, 0,1% da proteína solúvel total presente na semente é insulina. Noutras formas de realização preferidas da presente invenção, é obtida uma semente na qual, pelo menos, 0,2%, 0,3%, 0,5%, ou 1,0% da proteína solúvel total presente na semente é insulina. Um polipéptido da insulina pode estar presente numa variedade de diferentes tipos de células da semente incluindo, por exemplo, os hipocótilos e o eixo embrionário, incluindo nas raízes embrionárias e nas folhas embrionárias e no caso de espécies de plantas monocotiledóneas, incluindo cereais e milho, são utilizados no tecido do endosperma.

Preparação de insulina das sementes das plantas

Uma vez obtidas as sementes de plantas, a proteína da insulina pode ser purificada a partir da semente utilizando quaisquer metodologias de purificação de proteína conhecidas na técnica. Consequentemente, um método de purificação de insulina a partir das sementes de plantas, em que o método compreende:

(a) proporcionar uma construção de ácido nucleico quimérica compreendendo na direcção 5' para 3' da transcrição como

componentes ligados operacionalmente:

(i) uma sequência de ácido nucleico capaz de controlar a expressão em células vegetais da semente; e

(ii) uma sequência de ácido nucleico codificando um polipéptido da insulina;

(b) introduzir a construção de ácido nucleico quimérica numa célula vegetal;

(c) cultivar a célula vegetal até ficar uma planta madura capaz de produzir semente em que a semente expressa insulina;

(d) obter semente expressando insulina; e

(e) purificar a referida insulina a partir da semente, é aqui ensinado.

As sementes de plantas podem ser moídas utilizando qualquer processo de trituração resultando num rebentamento substancial da membrana celular da semente e das paredes celulares. Podem ser utilizadas quer condições secas ou húmidas para a moagem (Patente US3 971856; Lawhon *et al.*, 1977, *J. Am. Oil Chem. Soc.*, 63:533-534). Equipamento de moagem adequado neste contexto inclui moinhos colóides, moinhos de disco, moinhos IKA, homogeneizadores de escala industrial e semelhantes. A selecção do equipamento de moagem dependerá do tipo de semente e das necessidades de processamento. Contaminantes de semente sólidos, tais como cascas de sementes, materiais fibrosos, hidratos de carbono não dissolvidos, proteínas e outros contaminantes

insolúveis em água podem ser removidos da fracção da semente utilizando por exemplo metodologias com base na exclusão por tamanho, tal como filtração ou processos à base de gravitacional, tal como centrifugação. Em formas de realização preferidas, é evitada a utilização de solventes orgânicos normalmente utilizados na extracção oleosa, tais como hexano, é evitada porque esses solventes podem destruir um polipéptido da insulina. A insulina substancialmente pura pode ser recuperada da semente utilizando uma variedade de metodologias de purificação adicionais, tal como técnicas à base de centrifugação; metodologias à base de exclusão por tamanho, incluindo, por exemplo, ultrafiltração por membrana e ultrafiltração de fluxo cruzado; e técnicas cromatográficas, incluindo, por exemplo, cromatografia de permuta iónica, cromatografia de exclusão de tamanho, cromatografia de afinidade, cromatografia líquida de elevado desempenho (HPLC), cromatografia líquida de proteína rápida (FPLC), cromatografia de interacção hidrofóbica e semelhantes. Geralmente, uma combinação dessas técnicas será utilizada para obter insulina substancialmente pura.

Numa forma de realização particularmente preferida da presente invenção, um polipéptido da insulina é isolado dos contaminantes da semente colocando em contacto um polipéptido da insulina com corpos oleosos. Este método é considerado como sendo particularmente vantajoso, na medida em que permite a remoção de contaminantes da semente incluindo proteínas da semente de um modo particularmente eficaz e não dispendioso. Como aqui mencionado anteriormente, esse contacto de um polipéptido da insulina com os corpos oleosos pode ser alcançado ligando um polipéptido da insulina com uma proteína de corpo oleoso ou ligando um polipéptido da insulina a um polipéptido

com afinidade para um corpo oleoso, tal como um anticorpo de cadeia simples com afinidade para um corpo oleoso. Nesta última forma de realização, o polipéptido da insulina será sequestrado na célula nos corpos oleosos e, desse modo, co-purificado com os corpos oleosos. Na forma de realização anterior, após ser expresso num compartimento intracelular encerrado numa membrana, tal como o ER, um polipéptido da insulina estará associado com o corpo oleoso após rebentamento das células da semente durante o processo de trituração. Um processo para isolar os corpos oleosos é descrito na Patente US 650554.

Podem ser preparadas formulações farmacêuticas de insulina a partir da insulina purificada e essas formulações podem ser utilizadas para tratar a diabetes. Geralmente, a insulina purificada será misturada com um veículo farmacêuticamente aceitável ou diluente em quantidades suficientes para exercer um efeito terapêuticamente útil na ausência de efeitos colaterais indesejáveis no doente tratado. Para formular uma composição de insulina, a fracção em peso da insulina é dissolvida, suspensa, dispersa ou, de outra forma, misturada num veículo ou diluente seleccionado a uma concentração eficaz, tal que o estado a ser tratado seja melhorado. As formulações de insulina farmacêuticas são, de um modo preferido, formuladas para administração de dosagem única. Doses terapêuticamente eficazes para a distribuição parentérica de insulina humana são bem conhecidas na técnica. Quando são utilizados análogos da insulina ou são utilizados outros modos de distribuição, podem ser rapidamente determinadas empiricamente doses terapêuticamente eficazes pelos especialistas na técnica utilizando protocolos de teste conhecidos ou por extrapolação de resultados dos testes *in vivo* ou *in vitro*. Deve ser entendido, todavia, que as concentrações e dosagens podem variar de acordo com a gravidade do estado a ser

aliviado. É ainda entendido que, para um dado indivíduo particular, podem ser ajustados regimes de dosagem específicos ao longo do tempo, de acordo com a avaliação individual da pessoa que administra ou supervisiona a administração das formulações.

Soluções ou suspensões farmacêuticas podem incluir, por exemplo, um diluente estéril, tais como, por exemplo, água, lactose, sacarose, fosfato de dicálcio ou carboximetilcelulose. Veículos que podem ser utilizados incluem água, solução de soro fisiológico, dextrose aquosa, glicerol, glicóis, etanol e semelhantes, para desse modo formar uma solução ou suspensão. Se desejado, as composições farmacêuticas podem também conter substâncias auxiliares não tóxicas, tal como agentes humidificantes; agentes emulsificantes; agentes solubilizantes; agentes antimicrobianos, tal como álcool benzílico e metilparabenos; antioxidantes, tais como ácido ascórbico e bissulfito de sódio; agentes quelantes, tais como ácido etilenodiaminotetra-acético (EDTA); agentes de tamponamento de pH, tais como tampões de acetato, citrato ou fosfato; e as suas combinações.

A formulação final de uma preparação de insulina irá geralmente depender do modo de distribuição da insulina. A insulina preparada de acordo com a presente invenção pode ser distribuída de qualquer modo desejado; todavia as formas de distribuição parentérica, oral, pulmonar, bucal e nasal são considerados os modos de distribuição de mais provável utilização. As preparações parentéricas podem ser inseridas em ampolas, seringas descartáveis ou frasquinhos de dose única ou múltipla, feitos de vidro, plástico ou outros materiais igualmente adequados.

Exemplos

Os exemplos seguintes são apresentados a título ilustrativo e não para limitação.

Exemplo 1

Preparação de uma proteína de insulina expressa como uma proteína de fusão de Mini-Insulina (MI) com um pró-péptido de tripsina clivável.

Construção de pSBS4404: proteína de fusão PRS-D9scFv-Klip27-MI-KDEL

Uma das proteínas de fusão estudadas começa com a sequência relacionada com o patogénio do tabaco (PRS) (Sijmons *et al.*, 1990, *Bio/technology*, 8:217-221) que serviu como o péptido sinal para direccionar a expressão do ER de um modo co-tradução. Imediatamente a jusante estava uma sequência codificando para um anticorpo Fv de cadeia única (scFv) com afinidade específica para a espécie contra a oleosina de 18 kDa de *Arabidopsis thaliana* denominado D9scFv, seguido por um pró-péptido de tripsina clivável (KLIP27) derivado do pró-péptido TA57 de levedura (Kjeldsen *et al.*, 2001, *Biotechnology and Genetic Engineering Reviews* 18:89-121). Isto foi seguido por mini insulina (MI) descrita por Kjeldsen *et al.* (2001) com a adição de um sinal de retenção em ER KDEL na extremidade do terminal C do polipéptido.

A estrutura deste plasmídeo pSBS4055 foi baseada no vector binário da planta, pPZP200, descrito por Hajdukiewicz *et al.*

(*Plant Molecular Biology*, 1994, 25:989-994). No lugar do sítio de clonagem múltiplo descrito, foi inserido um gene *pat* que confere resistência à fosfinotricina à planta hospedeira (Wohlleben et al., 1988, *Gene* 70:25-37)) controlado pelo promotor/terminador da ubiquitina de *Petroselinum crispum* (Kawalleck et al., 1993, *Plant. Mol. Bio.*, 21:673-684), entre as sequências de fronteira esquerda e direita. Adicionalmente a esta cassete, foi sub-clonado o promotor/terminador da β -faseolina de *Phaseolus vulgaris* (Slightom et al., 1983, *Proc. Natl. Acad. Sc. USA* 80:1897-1901) controlando o PRS. Foi utilizada PCR convencional (Horton et al., 1989, *Gene* 77:61-68) para fundir a sequência codificando PRS sintética com sítios de endonuclease de restrição *SphI/HindIII* ligados à extremidade 3' do promotor da faseolina para produzir pSBS4011. Foi criada uma sequência de inserção *SphI-D9scFv-XhoI, SwaI, HindIII* por amplificação por PCR de um clone de ADNc D9scFv (Sean Hemmingsen lab, não publicado) com os iniciadores 1325 (GCATGCTGACATTG TGATGACACAGTC) - SEQ ID N°:175 e 1326 (AAGCTTGCATTTAAATACTCGAGA CTGTGAGAGTGGTGCCTTG) - SEQ ID N°:176. A ligação subsequente deste fragmento nos sítios *SphI/HindIII* de pSBS4011 resultou no plasmídeo pSBS4055.

A sequência Klip27-MI foi sintetizada a partir de quatro oligonucleótidos parcialmente sobreponíveis que incorporaram a utilização de códons de *Arabidopsis thaliana* para aumentar o sucesso da tradução eficiente num sistema de expressão à base de plantas. Os oligonucleótidos 1324 (GAAGAAGGAGAGCCTAAGTTTGTTAATCA ACATCTTTGTGGATCTCATCTTGTTGAGGCTCTCTACCTTG) - SEQ ID N°:177 e 1323 (CCTTAGGAGTGTAGAAAAATCCTCTTTCTCCACACACAAGGTAGAGAGCCTCAACA) - SEQ ID N°:178 foram emparelhadas na sua sobreposição complementar de 20 nucleótidos e estendidas para formar a extremidade 5' da fusão Klip27-MI enquanto que o mesmo foi

realizado com os oligonucleótidos 1322 (CTAAGGCTGCTAAGGGAATTG) - SEQ ID N°:179 e 1321 (AAGCTTCAGTTGCAATAGTTCTCCAATTGGTAAAGTGAGCA AATAGAAGTGCAACATTGTTCAACAATCCCTTAGCAGCCTT) - SEQ ID N°:180 para formar a extremidade 3'. As duas metades foram ligadas após digestão de restrição com *Bsu36I*, para produzir a sequência codificante total Klip27-MI. A PCR desta fusão de genes utilizando os iniciadores 1364 (CTCGAGTCAACCAATTGATGACACTGAATC) - SEQ ID N°:181 e 1334 (AAGCTTCAAAGTTCATCCTTGTTGCAATAGTTCTCCAATT G) - SEQ ID N°:182 ligadas ao sítio de clivagem da endonuclease de restrição 5' *XhoI* e a sequência de ADN de KDEL em 3' mais o sítio de clivagem de *HindIII* para ligação subsequente em pSBS4055 cortado com *XhoI/HindIII*. O resultado foi o plasmídeo pSBS4404: uma sequência de ADN codificando a proteína de fusão PRS-D9scFv-Klip27-MI-KDEL sendo colocada num vector binário sob o controlo de expressão do promotor/terminador da faseolina. O promotor da faseolina controla a expressão específica temporal e específica para o tecido do transgene durante o desenvolvimento da semente. A sequência de ácido nucleico completa (SEQ ID N°:1) e a sequência de aminoácidos (SEQ ID N°:2) da proteína de fusão de insulina 4404 (PRS-D9ScFv-Klip27-MIKDEL) são apresentadas na Figura 1.

Construção de pSBS4405: proteína de fusão OLE0-Klip8-Klip27-MI

A segunda proteína de fusão estudada começou com a oleosina de 18 kDa de *Arabidopsis thaliana* seguida em grelha por um pró-péptido de quimosina clivável (Klip8) - SEQ ID N°:175. Imediatamente a jusante estava uma sequência codificando para o pró-péptido da tripsina clivável (Klip27) derivado do pró-péptido TA57 de levedura como descrito acima (Kjeldsen

et al., 2001, *Biotechnology and Genetic Engeneering Reviews* 18:89-121). Este foi fundido com mini-insulina (MI) descrita acima (Kjeldsen *et al.*, 2001). A expressão desta proteína de fusão foi direccionada para os corpos oleosos nascentes formados durante o desenvolvimento do embrião.

A estrutura deste plasmídeo pSBS4055 foi baseada no vector binário de plantas, pPZP200, descrito por Hajdukiewicz *et al.* (*Plant Molecular Biology*, 1994, 25:989-994). Em vez do sítio de clonagem múltipla descrito, foi inserido um gene *pat* conferindo resistência à fosfinotricina à planta hospedeira (Wohlleben *et al.*, 1988, *Gene* 70:25-37)) controlada pelo promotor/terminador da ubiquitina de *Petroselinum crispum* (Kawalleck *et al.*, 1993, *Plant. Mol. Bio.*, 21:673-684), entre as sequências de fronteira da esquerda e da direita. Adicionalmente a esta cassette, foi subclonado o promotor/terminador da β -faseolina de *Phaseolus vulgaris* (Slightom *et al.*, 1983, *Proc. Natl. Acad. Sc. USA* 80:1897-1901) controlando a fusão da sequência genómica de oleosina de *Arabidopsis* de 18 kDa-Klip8. Foi utilizada PCR convencional (Horton *et al.*, 1989, *Gene* 77:61-68) para fundir a sequência do gene da oleosina-Klip8 com sítios de endonuclease de restrição *XhoI/HindIII* ligadas à extremidade 3' do promotor da faseolina para produzir pSBS4010.

A sequência Klip27-MI foi sintetizada a partir de quatro oligonucleótidos parcialmente sobreponíveis que incorporaram a utilização de codões de *Arabidopsis thaliana* para aumentar o sucesso de tradução eficiente num sistema de expressão à base de plantas. Os oligonucleótidos 1324 (GAAGAAGGAGAGCCTAAGTTTGTTAATAC ATCTTTGTGGATCTCATCTTGTTGAGGCTCTCTACCTTG) - SEQ ID N°:177 e 1323 (CCTTAGGAGTGTAGAAAAATCCTCTTTCTCCACACACAAGGTAGAGAGCCTCAACA) - SEQ ID N°:178 foram emparelhados na sua sobreposição de

20 nucleótidos complementar e estendidos para formar a extremidade 5' da fusão Klip27-MI enquanto que o mesmo foi realizado com os oligonucleótidos 1322 (CTAAGGCTGCTAAGGGAATTG) - SEQ ID N°:179 e 1321 (AAGCTTCAGTTGCAATAGTTCTCCAATTGGTAAAGTGAGCAAATAGAAGTGCAACAT TGTTCACAATTCCCTTAGCAGCCTT) - SEQ ID N°:180 para formar a extremidade 3'. As duas metades foram ligadas após digestão de restrição com *Bsu36I*, para produzir a sequência codificante completa Klip27-MI. A PCR desta fusão de genes utilizando os iniciadores 1364 (CTCGAGTCAACCAATTGATGACACTGAATC) - SEQ ID N°:181 e 1329 (AAGCTTCAGTTGCAATAGTTC) - SEQ ID N°:183 ligados ao sítio de clivagem da endonuclease de restrição *XhoI* a 5' e um sítio de clivagem *HindIII* a 3', respectivamente, para ligação subsequente em pSBS4010 cortado com *XhoI/HindIII*. O resultado foi o plasmídeo pSBS4405: uma sequência de ADN codificando a proteína de fusão Oleosina-Klip8-Klip27-MI estando ligada num vector binário sob o controlo de expressão do promotor/terminador da faseolina. O promotor da faseolina controla a expressão específica temporal e específica para o tecido do transgene durante o desenvolvimento da semente. A sequência de ácido nucleico completa SEQ ID N°:3 e a sequência de aminoácido SEQ ID N°:4 da proteína de fusão de insulina 4405 (OLE0-Klip8-Klip27-MI) são apresentadas na Figura 2.

Construção de pSBS4414: proteína de fusão PRS-MI-ligando tetrabásico-D9Scfv-KDEL

Outra proteína de fusão estudada começou com a sequência relacionada com o patogénio do tabaco (PRS) (Sijmons *et al.*, 1990, *Bio/technology*, 8:217-221) que serviu como o péptido sinal para direccionar a expressão para o ER de um modo co-tradução.

Imediatamente a jusante estava a sequência codificando mini-insulina (MI) descrita por Kjeldsen *et al.* (2001) com a exceção de a região de mini-pró-peptídeo C (AAK - SEQ ID N°:146) ter sido substituída com uma sequência do sítio B₃₀ treoninatetrabásica (B30T-RRKR) (SEQ ID N°:149) de intervalo entre as cadeias B₍₁₋₂₉₎- e A₍₁₋₂₁₎ da insulina humana. Isto foi, imediatamente, seguido por uma sequência codificando para um segundo ligando tetrabásico seguido por um anticorpo Fv de cadeia simples (scFv) com afinidade específica para a espécie contra a oleosina de 18 kDa de *Arabidopsis thaliana* denominada D9scFv. Na extremidade do terminal C do polipeptídeo estava a adição de um sinal de retenção em ER de KDEL.

A estrutura do plasmídeo, pSBS4055, foi baseada no vector binário da planta, pPZP200, descrito por Hajdukiewicz *et al.* (*Plant Molecular Biology*, 1994, 25:989-994). Em vez do sítio de clonagem múltipla descrito, foi inserido um gene *pat* gene conferindo resistência à fosfinotricina à planta hospedeira (Wohlleben *et al.*, 1988, *Gene* 70:25-37) controlado pelo promotor/terminador da ubiquitina de *Petroselinum crispum* (Kawalleck *et al.*, 1993, *Plant. Mol. Bio.*, 21:673-684), entre as sequências de fronteira da esquerda e da direita. Adicionalmente a esta cassette, foi sub-clonado o promotor/terminador da β -faseolina de *Phaseolus vulgaris* (Slightom *et al.*, 1983, *Proc. Natl. Acad. Sc. USA* 80:1897-1901) controlando a PRS. Foi utilizada PCR convencional (Horton *et al.*, 1989, *Gene* 77:61-68) para fundir a sequência codificando PRS sintética com sítios de endonuclease de restrição *SphI/HindIII* com a extremidade 3' do promotor da faseolina para produzir pSBS4011.

A sequência Klip27-MI foi sintetizada a partir de quatro oligonucleótidos parcialmente sobreponíveis que incorporaram a

utilização de codões de *Arabidopsis thaliana* para aumentar o sucesso da tradução eficiente num sistema de expressão à base de plantas. Os oligonucleótidos 1324 (GAAGAAGGAGAGCCTAAGTTTGTAAATC AACATCTTTGTGGATCTCATCTTGTGAGGCTCTCTACCTTG) SEQ ID N°:177 e 1323 (CCTTAGGAGTGTAGAAAAATCCTCTTTCTCCACACACAAGGTAGAGAGCCTCAACA) SEQ ID N°:178 foram emparelhados na sua sobreposição de 20 nucleótidos complementar e estendidos para formar a extremidade 5' da fusão Klip27-MI enquanto que o mesmo foi realizado com os oligonucleótidos 1322 (CTAAGGCTGCTAAGGGAATTG) SEQ ID N°:179 e 1321 (AAGCTTCAGTTGCAATAGTTCTCCAATTGGTAAAGTGAGCAAATAGAAGTGCAACAT TGTTCACAATTCCCTTAGCAGCCTT) SEQ ID N°:180 para formar a extremidade 3'. As duas metades foram ligadas após a digestão de restrição com *Bsu36I*, para produzir a sequência codificante Klip27-MI completa. A PCR desta fusão de genes utilizando os iniciadores 1363 (GCATGCCCAACCAATTGATGACACTG) SEQ ID N°:184 e 1334 (AAGCTTCAAAGTTCATCCTTGTGCAATAGTTCTCCAATTG) SEQ ID N°:182 ligou-se a um sítio de clivagem de endonuclease de restrição *SphI* em 5' e a sequência de ADN de KDEL a 3' mais o sítio de clivagem *HindIII* para subsequente ligação em pSBS4011 cortado com *SphI/HindIII*. O resultado foi o plasmídeo pSBS4402: uma sequência de ADN codificando a proteína PRS-Klip27-MI-KDEL de fusão sendo colocado num vector binário sob controlo de expressão do promotor/terminador da faseolina. O promotor da faseolina controla a expressão específica temporal e específica para o tecido do transgene durante o desenvolvimento da semente. O vector de expressão de plantas pSBS4402, serviu como o molde para introduzir sítios tetrabásicos entre as cadeias B e A de insulina e entre MI e D9 Scfv.

Um sítio tetrabásico (B30T-RRKR) de intervalo foi colocado entre as cadeias autênticas B₍₁₋₂₉₎- e A₍₁₋₂₁₎ de insulina humana

por PCR utilizando os iniciadores 1515 (GCATGCATGCCTTTGTTAATCAAC ATCTTTGTGG) SEQ ID N°:185 e 1518 (ACATTGTTCAACAATTCTCTCTTTCTTCT AGTCTTAGGAGTGTAGAAAAATCC) SEQ ID N°:186 utilizando pSBS4402 como o molde. O fragmento de 124 pb resultante foi utilizado em combinação com o iniciador 1517 (GCATAAGCTTCAAAGCTCATCCTTTGAGC) SEQ ID N°:187 utilizando pSBS3400 como um molde. Notar que o pSBS3400 é um plasmídeo que contém o fragmento D9 scFv-KDEL com um sítio de restrição *HindIII*. Esta reacção de PCR produziu um produto de 955 pb que introduziu um (RRKR)-D9 Scfv-KDEL-*HindIII* tetrabásico no fragmento de 124 pb *SphI*-MI. O fragmento de 955 pb foi então ligado e subclonado em pGEM-T (Promega) para resultar em pSBS3403. O fragmento *SphI*-MI (com pró-peptídeo C modificado com B₃₀T-RRKR)-RRKR-D9 Scfv-KDEL-*HindIII* inteiro foi inserido em pSBS4402 pré-cortado (*SphI/HindIII*) para criar pSBS4414. A sequência de ácido nucleico SEQ ID N°:5 completa e a sequência de aminoácidos SEQ ID N°:6 da proteína de fusão de insulina 4414 (PRS-MI-ligando tetrabásico-D9Scfv-KDEL) é apresentada na Figura 3.

Transformação e Cultura de *E. coli* recombinante e *Agrobacterium* com pSBS4404, pSBS4405 ou pSBS4414.

Após confirmar a integridade do ADNc codificando para a proteína de fusão por análise de sequências, os plasmídeos pSBS4404, pSBS4405 e pSBS4414 foram transformados na estirpe de *E. coli* DH5 α para permitir a expressão de nível elevado. O ADN do plasmídeo isolado (100 ng) foi misturado em gelo com 100 μ L de células competentes DH5 α durante 20 min. As células foram então submetidas a choque térmico a 42 °C durante 45 segundos e recolocadas em gelo durante 2 min. Depois foi adicionado 1 mL de meio SOC e as células foram incubadas a 37 °C num agitador

orbital a 225 rpm durante 1 h antes de plaquear as células transformadas em placas de LB-espectinomicina (10 g/L triptona, 5 g/L extracto de levedura, 5 g/L NaCl, 15 g/L agar) e incubando durante a noite a 37 °C. Foi utilizada uma única colónia para inocular 5 mL de meio LB-espectinomicina. Estas culturas foram cultivadas durante a noite a 37 °C. O plasmídeo recombinante foi isolado a partir de 1 mL da cultura durante a noite de acordo com o QIAprep®Spin Miniprep Kit (Qiagen). O plasmídeo isolado foi então utilizado para transformar a estirpe de *Agrobacterium* EH101 competente (Hood et al., 1986; *J. Bacteriol.* 144: 732-743) por electroporação (25 µF, 2,5 kV, 200 Ω). Foram plaqueadas *Agrobacterium* recombinante em AB-espectinomicina/canamicina (20x sais AB, glucose a 2 M, 0,25 mg/mL de FeSO₄ e 7H₂O, 1 M MgSO₄, CaCl₂ a 1 M) e foi utilizada uma colónia isolada para inocular 5 mL de meio AB-espectinomicina/canamicina. Estas culturas foram cultivadas durante a noite a 28 °C. Os *Agrobacterium* recombinante foram então utilizados para transformar plantas de *Arabidopsis thaliana* pelo método de imersão da flor (Clough et al., 1998, *Plant J.*, 16:735-743) *Arabidopsis thaliana* cv. (C24) é utilizada para todas as experiências. As sementes são plantadas na superfície de uma mistura de solo (dois terços de terra Redi e um terço de perlite com um pH = 6,7) ou uma mistura de solo de *Arabidopsis* fornecida por Lehle Seeds (perlite, vermiculite, turfa, terra-verde, com um pH = 5,5) em vasos de 4 polegadas. As plantas originadas de sementes foram deixadas crescer até um estágio de roseta de 6-8 folhas até um diâmetro de aproximadamente 2,5 cm. Os vasos são colocados dentro de uma cúpula a 4 °C durante quatro dias para um tratamento a frio e subsequentemente transferidos para uma sala de cultura a 24 °C com luz constante a cerca de 150 µE e 60-70% de humidade relativa. As plantas são irrigadas a 2-3 dias de intervalo e fertilizadas semanalmente com 1% de Peters 20-19-18. Cada vaso

contém cinco a seis plantas. Quando as plantas atingem cerca de 2 cm em altura, os botões primários são cortados para encorajar o crescimento de botões secundários e terciários. 4 a 5 dias após cortar os botões primários, as plantas estão prontas para serem infectadas com *Agrobacterium*. Os vasos com plantas de *Arabidopsis* são invertidos para permitir que as plantas de *Arabidopsis* sejam infectadas com 500 mL de uma ressuspensão numa cultura de *Agrobacterium* durante a noite contendo o vector de transformação da planta de interesse durante 20 segundos. É importante que a cultura de *Agrobacterium* contenha sacarose a 5% e 0,05% do tensoactivo Silwet L-77 (Lehle Seeds). Os vasos são subsequentemente cobertos com uma cúpula de plástico transparente durante 24 horas para manter a humidade mais elevada. As plantas são deixadas a cultivar até à maturidade e é recolhida uma mistura de sementes, não transformadas e transformadas. Para selecção de linhas transgénicas, as sementes transformadas putativas são esterilizadas numa lavagem rápida de 70% etanol, depois uma lixívia comercial a 20% durante 15 min e depois passadas pelo menos quatro vezes por ddH₂O. Cerca de 1000 sementes esterilizadas são misturadas com 0,6% de top agar fundido e espalhadas uniformemente numa placa MS com metade da concentração (Murashige e Skoog, 1962, *Physiologia Plantarum* 15: 473-497) contendo 0,3% de sacarose e 80 µM do herbicida fosfinotricina (PPT) DL. As placas foram então colocadas numa sala de cultura com um regime de luz de 8 h no escuro e 16 h de luz a 24 °C. Após 7 a 10 dias, as plantas derivadas de sementes transgénicas putativas são verdes e em crescimento, enquanto que as plantas derivadas de sementes não transformadas plantas derivadas de sementes são lixiviadas. Após o estabelecimento de raízes, as plantas derivadas de sementes transgénicas putativas são transferidas individualmente para vasos (as plantas individualmente são irrigadas em 3 dias de intervalo e

fertilizadas com 1% Peters 20-19-18 em intervalos de 7 dias) e deixadas crescer até à maturidade. Os vasos são cobertos com uma cúpula de plástico transparente durante três dias para proteger as plantas derivadas de sementes sensíveis. Após 7 dias, as plantas derivadas de sementes são cobertas com um sistema de recolha de sementes de Lehle Seeds para evitar a perda de semente devido a espalhamento. Sementes destas plantas transgênicas são recolhidas individualmente e preparadas para análise.

Exemplo 2

Níveis de expressão de insulina em *Arabidopsis thaliana*

No segundo exemplo, os níveis de expressão da proteína de fusão D9scfv-KLIP27-MI-KDEL (4404), OLEO-KLIP8-KLIP27-MI (4405) e PRS-MI-RRKR-D9Scfv-KDEL (4414) foram determinados em sementes de *Arabidopsis thaliana* transgênica madura. O produto de transgene demonstrou estar presente nos extractos celulares da semente madura. Aproximadamente 40 sementes de *Arabidopsis thaliana* transgênicas foram moídas com um almofariz e pilão em 50 µL de Tris-HCl a 50 mM pH 8,0. Depois, foi adicionado um tampão de amostra de SDS-PAGE de redução (amostra de tampão 6x SDS, 0,35 M Tris-HCl pH 6,8, 30% glicerol, 10% SDS, 0,012% azul de bromofenol, 5% β-mercaptoetanol) à mistura densa e misturados por breve agitação em vórtice. A amostra foi então rapidamente centrifugada e colocada a 99 °C durante 10 minutos. Após arrefecer em gelo durante 2 minutos a amostra foi centrifugada brevemente. As amostras foram aplicadas (10 µL - equivalente a aproximadamente 7 sementes) em condições de redução.

Para as amostras preparadas a partir de corpo oleoso as sementes transgênicas e de tipo selvagem (20 mg) foram moídas em 250 µL de tampão de extração de corpo oleoso (0,4 M de sacarose, 0,5 M NaCl, Tris-HCl a 50 mM pH 8,0). As amostras foram microcentrifugadas a 10000 g durante 10 min. A fracção solúvel aquosa foi removida com uma seringa 26 G 5/8 de 1 mL e a pilha de gordura foi ressuspensa em 100 µL de tampão fosfato suplementado com sal (Na_2HPO_4 a 20 mM pH 8,0, 0,5 M NaCl). A pilha de gordura ressuspensa foi transferida para um tubo de microcentrífuga e centrifugado a 10000 g durante 10 min. O processo foi repetido mais 3 vezes com uma ressuspensão final da pilha de gordura em 100 µL de tampão fosfato sem sal (20 mM Na_2HPO_4 pH 8,0). Foram realizadas mais duas lavagens em tampão fosfato sem tampão sem sal foram realizados com passos de centrifugação intermitente como delineado acima. O sedimento de gordura final foi ressuspensa em tampão fosfato a 10 µL (20 mM Na_2HPO_4 pH 8,0). Foi recolhida uma fracção de 5 µL e a proteína do corpo oleoso foi solubilizada fervendo em 1/10 (v/v) Tris-HCl a 50 mM pH 8,0 com 2% SDS. A amostra foi arrefecida em gelo durante 2 min e centrifugada a 10000 g durante 5 min. O teor de proteína do subnadante foi determinado por ensaio de proteína BCA (Pierce, Rockford, IL). Para géis corados com coomassie e análise de transferência de Western, 20 µg da proteína total foi separada em géis de 15% SDS-PAGE sob as condições de redução utilizando o tampão de amostra de SDS-PAGE.

A(s) amostra(s) foi(foram) então aplicada(s) em géis de SDS-PAGE a 15% descontínuos e separados a 150 volts durante aproximadamente 1,5 horas. Os géis foram então corados com Coomassie ou transferidos para membrana de PVDF (Immobilon-P, Millipore Corporation, Bedford, MA) para análise de transferência de Western. As amostras transferidas foram

rastreadas com anticorpo monoclonal dirigido contra insulina (Clone E2-E3; Roth et al., 1992) adquirido a Abcam (Cambridge, Reino Unido). As bandas de insulina foram detectadas utilizando um conjugado secundário de IgG F(ab')₂ AP de Ovelha X murganho (Chemicon International, Temecula, CA) e desenvolvidos utilizando NBT-BCIP em tampão GARAP (Tris-HCl pH 9,5, 100 mM NaCl, 5 mM Mg Cl₂). A banda imunorreactiva correspondeu a uma banda de polipéptido, migrando ao peso molecular previsto da proteína de fusão, como apresentado nas Figuras 4A-4F. As Figuras 4(A- F) apresentam expressão recombinante de proteínas de fusão de insulina em linhas transformadas de *Arabidopsis thaliana* (4404-2, -17, -20, 4405-4, 4414-19 e 4414-20) com base em SDS-PAGE corado com Coomassie e análise de Transferência de Western. As setas apresentam a posição dos polipéptidos de fusão que migram a 38,5 kDa, 34,2 kDa e 34,2 kDa, PRS-D9(scfv)-KLIP27-MIw/KDEL (4404), OLEO-KLIP8-KLIP27-MI (4405) e PRS-MI-RRKR-D9Scfv-KDEL (4414) respectivamente, em condições de redução. Deve ser notado que a proteína de fusão 4414 possui um peso molecular esperado de 34,2 kDa mas possui um peso molecular aparente elevado num gel de SDS-PAGE). As Figuras 4A (gel corado com Coomassie) e 4B (correspondendo à Transferência de Western rastreada com sonda com anti-insulina E2E3) apresenta proteína de semente total de tipo selvagem (wt) e linhas de semente transgênicas expressando as construções 4404 e 4405. As Figuras 4C (gel corado com Coomassie) e 4D (correspondendo à Transferência de Western rastreada com sonda com E2E3 anti-insulina) apresentam proteína de corpo oleoso preparada a partir de sementes de tipo selvagem e transgênicas expressando as mesmas construções 4404 e 4405. As Figuras 4D (gel corado com Coomassie) e 4E (correspondendo à Transferência de Western rastreada com sonda com E2E3 anti-insulina) apresentam proteína de corpo oleoso preparado a partir de semente de tipo selvagem e

transgénica expressando as mesmas construções 4414. Os marcadores de peso molecular (M) são 10, 15, 20, 25, 37, 50, 75, 100, 150, 250 kDa. Os controlos incluem, hIN (padrão de insulina humana recombinante) e hProIN (padrão de pró-insulina humana recombinante), separados sob condições de não redução. As diferenças nos níveis de expressão são o resultado de variação clonal entre os transformantes. Os níveis aproximados de expressão de proteína de transgene e MI são apresentados na Figura 5. Os níveis de expressão foram determinados utilizando a banda de oleosina de 18 kDa como um padrão interno (equivalente a 1,5% da proteína total da semente) por densitometria da banda do transgene. O nível médio de expressão para as construções PRS-D9(scfv)-KLIP27-MIw/KDEL (4404), OLEO-KLIPB-KLIP27-MI (4405) e PRS-MI-RRKRD9Scfv- KDEL (4414) foram 0,21% da proteína total da semente, 0,12% da proteína total da semente e 0,79% da proteína total da semente, respectivamente.

Exemplo 3

Clivagem de pSBS4404 e purificação de HPLC

Eluição do corpo oleoso

No terceiro exemplo, 1 g de semente transgénica foi homogeneizada em 12 mL de tampão de extracção (0,4 M de sacarose, 0,5 M NaCl, Tris-HCl a 50 mM pH 8,0) e centrifugada a 10000 g durante 10 min, as pilhas de gordura foram removidas e colocadas em 1 mL de 20 mM Na₂HPO₄, 0,5 M NaCl e re-centrifugadas como acima. Isto foi repetido duas vezes, antes da lavagem e centrifugação da pilha de gordura duas vezes em 750 µL de Na₂HPO₄ a 20 mM. A proteína de fusão 4404 foi eluída do corpo oleoso

para o subnadante por lavagem da pilha de gordura final 5 vezes em 750 μ L de ácido fórmico a 20 mM pH 4,1, com passos de centrifugação de 10000 g entre cada lavagem. As fracções de eluição recolhidas (subnadantes) foram reunidas e neutralizadas com NaOH 2 N a pH 8,0. A solução inteira foi então colocada a -80 °C para congelar e liofilizada durante a noite para concentrar a proteína de fusão. A amostra liofilizada foi ressuspensa em 500 μ L de Tris-HCl a 50 mM pH 8,0. A proteína de fusão 4404 ressuspensa foi então dessalinizada numa coluna NAP-5 (Amersham Pharmacia Biotech Ab, Uppsala, Suécia) e re-substituída com tampão (Tris-HCl a 50 mM pH 8,0). A fracção dessalinizada foi então novamente congelada e liofilizada durante a noite para concentrar. A amostra final concentrada foi ressuspensa num volume final de 105 μ L de H₂O destilada duas vezes. Os resultados da eluição são apresentados na Figura 6. A Figura 6 é uma análise de SDS-PAGE corado com Coomassie (15%) de preparações de corpo oleoso antes da eluição (-OB), prep. De OB após eluição com ácido fórmico (-OB') e o material eluído concentrado (-E). A seta apresenta a posição do polipéptido de fusão em migração. O controlo de tipo selvagem está essencialmente isento de quaisquer proteínas principais após eluição, enquanto que o material 4404 concentrado contém a proteína de fusão, alguns produtos truncados (possível proteína de fusão hidrolizada) e possivelmente algumas albuminas que co-eluíram.

Clivagem e análise de HPLC de Semente de *Arabidopsis* expressando 4404

A amostra concentrada foi ressuspensa em 105 μ L de água duplamente destilada e o teor de proteína foi avaliado por

ensaio de proteína BCA de acordo com o fabricante (Pierce, Rockford, IL, EUA). As amostras foram então clivadas com tripsina (proporção de 1:300 tripsina:proteína total, em Tris-HCl a 50 mM pH 8,0, em gelo durante 90 min). As reacções foram paradas com um excesso molar de 10 vezes de TLCK (N-p-tosil-L-lisina clorometil cetona). As reacções inteiras foram então filtradas através de filtros de 0,2 µm (filtro de seringa Aerodisc® de 13 mm com membrana de 0,2 µm Supof®, Pall Corporation, Ann Arbor, MI, EUA) e analisadas por (RP)-HPLC de fase reversa utilizando um coluna C18 (Zorbax 300SB-C18, Agilent Technologies, Waldbronn, Alemanha). As amostras foram aplicadas na coluna e eliminadas por eluição a 1,0 mL/min utilizando um gradiente linear de 19-min de 5-50% (v/v) acetonitrilo em 0,1% (v/v) TFA. O cromatograma resultante desta análise é observado na Figura 7. O vestígio revela um produto de tripsina clivado de proteína de fusão 4404, com propriedades quase idênticas na coluna ao padrão de insulina humana (tempos de retenção de 17,011 min e 17,179 min, respectivamente). A fracção de HPLC foi recolhida a partir de 17,0-17,5 min e analisada por espectrometria de massa PSD MALDI/TOF utilizando um espectrómetro de massa Voyager-DE STR (Applied Biosystems). A análise de MS foi realizada pelo serviço de Espectroscopia BioAnalítica fornecido por NRC-Plant Biotechnology Institute, Saskatoon, Saskatchewan, Canadá. A resolução do produto 4404 clivado purificado por HPLC como descrito acima é apresentado na Figura 8B em comparação com o padrão de insulina humana apresentado na Figura 8A. A massa observada de proteína de fusão 4404 clivada com tripsina foi de 6191,51 Da. A discrepância entre padrão de insulina humana (Figura 8A) e o padrão do produto 4404 clivado (Figura 8B) corresponde a uma Insulina Des-B30 com o sinal KDEL retido na cadeia A do produto clivado (Des-B₃₀ Insulina-KDEL).

Exemplo 4

Clivagem de pSBS4405 e purificação por HPLC

Preparação de Corpo Oleoso

A proteína de fusão (OLEO-KLIP8-KLIP27-MI) pode ser parcialmente purificada realizando preparações de corpo oleoso como descrito abaixo. Aproximadamente, 1 g de semente transgénica foi homogeneizada em 12 mL de tampão de extracção (0,4 M de sacarose, 0,5 M de NaCl, Tris-HCl a 50 mM pH 8,0) e centrifugada a 10000 g durante 10 min, as pilhas de gordura foram removidas e colocadas em 1 mL de Tris-HCl de 50 mM pH 8,0, NaCl a 0,5 M e re-centrifugadas como acima. Isto foi repetido duas vezes, antes de lavar e centrifugar a pilha de gordura duas vezes em 750 µL de Tris-HCl a 50 mM pH 8,0. A preparação do corpo oleoso resulta na remoção da maioria das proteínas de fundo. O perfil de proteína típico de uma preparação do corpo oleoso de semente de *Arabidopsis* transgénica expressando a construção de 4405 é demonstrada na Figura 9.

Clivagem e análise de HPLC de Semente de *Arabidopsis* expressando 4405

O conteúdo em proteína total dos corpos oleosos ressuspensos foi avaliado solubilizando uma fracção da preparação (5 µL) diluída 10 vezes em SDS a 2%, Tris-HCl a 50 mM pH 8,0, fervida durante 5 minutos e centrifugada durante 3 min a 10000 g. Posteriormente, o conteúdo de proteína foi determinado por ensaio de proteína BCA de acordo com o fabricante (Pierce, Rockford, IL, EUA). As amostras foram então clivadas com

tripsina (proporção de 1:300 tripsina:proteína total, em Tris-HCl a 50 mM pH 8,0, em gelo durante 90 min) para libertar o fragmento Klip27-MI da proteína de fusão. As reacções foram paradas com um excesso molar de 10 vezes de TLCK (N-p-tosil-L-lisina clorometil cetona). As amostras foram centrifugadas a 10000 g durante 10 min e os subnádantes das reacções inteiras foram então filtradas através de filtros de 0,2 µm (filtro de seringa Aerodisc® 13 mm com membrana de 0,2 µm Supof®, Pall Corporation, Ann Arbor, MI, EUA). A Figura 9 apresenta análise de SDS-PAGE corado com Coomassie (15%) de proteína de semente total extraível e proteína de corpo oleoso (OB) preparada a partir de linhas que expressam 4405 em comparação com semente de tipo selvagem (não recombinante). A seta apresenta a posição da migração do polipéptido de fusão. Os subnádantes foram ainda analisados por (RP)-HPLC de fase reversa utilizando uma coluna C18 (Zorbax 300SB-C18, Agilent Technologies, Waldbronn, Alemanha). As amostras foram aplicadas à coluna e removidas por eluição a 1,0 mL/min utilizando um gradiente linear de 19 min⁻ de 5-50% (v/v) acetonitrilo em 0,1% (v/v) TFA. O cromatógrafo resultante desta análise é observado na Figura 10. O vestígio revela um produto de tripsina clivado a partir da proteína de fusão 4405, com propriedades quase idênticas na coluna como o padrão de insulina humana (tempos de retenção de 17,220 min e 17,179 min, respectivamente). A fracção de HPLC foi recolhida a partir de 17,0-17,5 min e analisada por espectrometria de massa PSD MALDI/TOF utilizando um espectrómetro de massa Voyager-DE STR (Applied Biosystems). A análise de MS foi realizada pelo serviço de Espectroscopia BioAnalítica fornecido por NRC-Plant Biotechnolgy Institute, Saskatoon, Saskatchewan, Canadá. Como apresentado na Figura 11, a massa observada de proteína de fusão 4405 clivada com tripsina foi de 5706,30 Da. Uma discrepância entre padrão de insulina humana (Figura 8A) e o produto clivado

4405 (Figura 11) corresponde a um produto de Des-B₃₀ Insulina (Des-B₃₀ Insulin). A Des-B₃₀ Insulina é o produto esperado a partir da maturação de tripsina correcta da fusão 4405.

Exemplo 5

Purificação de Tripsina Clivada MI utilizando AKTA Explorer (FPLC)

A purificação de MI clivada de 4405 foi também parcialmente purificada das reacções de clivagem de escala aumentada por permuta aniónica (Mono Q FF 1 mL, Amersham Pharmacia) num AKTA explorer (Amersham Pharmacia). As reacções de clivagem foram realizadas em corpo oleoso 4405 como descrito acima preparado a partir de até 30 g de semente transgénica. Os subnadaentes das reacções de clivagem foram filtrados através de filtros de 0,2 µm ou concentrados por liofilização num Savant Speed Vac. As reacções de amostra filtradas podem ser aplicadas a uma coluna directamente, mas as amostras concentradas necessitaram da remoção de sais para se ligarem eficazmente à coluna. As amostras concentradas podem ser dessalinizadas passando o material clivado através de uma coluna PD-10 (Amersham Pharmacia), por diálise ou diluição para uma concentração de sais equivalente ou inferior a 5 mS/cm. As amostras dessalinizadas foram equilibradas com Tris-HCl a 20 mM pH 6,5. As amostras foram separadas utilizando um gradiente descontínuo com NaCl de 0-40% NaCl com um caudal de 1 mL/min. A detecção foi realizada a 214 nm (a detecção a 280 nm é relativamente fraca devido ao baixo conteúdo de aminoácidos aromáticos na insulina). O Solvente A foi Tris-HCl a 20 mM pH 6,5 enquanto que o solvente B foi de Tris-HCl a 20 mM pH 6,5, NaCl a 1,0 M. As fracções

(1 mL) que eluíram à mesma condutividade que a insulina padrão da Roche, entre 7-35 mS/cm, foram recolhidas (referência à Figura 12). A Figura 12 apresenta o cromatograma das preparações de tripsina clivada do corpo oleoso 4405 (linha a tracejado) em comparação com a insulina humana padrão (linha a cheio). A presença de Insulina foi verificada nas fracções recolhidas por HPLC, ELISA ou análise de Western (resultados não apresentados). As amostras recolhidas foram então concentradas por liofilização e utilizadas num bioensaio de insulina descrito no exemplo 6.

Exemplo 6

Teste de Tolerância a Insulina: Bioensaio em Murganhos Machos C57BI/6 (B6)

Este bioensaio foi realizado para determinar o efeito *in vivo* de (DesB₃₀ IN) recombinante derivado de planta de tripsina clivada 4405 em comparação com a insulina humana. Os níveis de glucose no plasma em murganhos B6 foram determinados antes de e após a injeção intraperitoneal de padrões de insulina, controlos negativos e insulina SBS. Quinze murganhos machos C57BI/6 (B6) de aproximadamente 2 meses de idade foram adquiridos de Jackson Laboratories (Bar Harbor, ME). Os níveis de glucose no plasma foram determinados com um glucómetro automático (OneTouch Ultra, Lifescan, Johnson e Johnson). Controlos positivos incluíram HumulinR® (Eli Lilly) e padrão de insulina humana recombinante de levedura da Roche. Uma solução de soro fisiológico serviu como o placebo. Foi incluído um controlo negativo que representou corpos oleosos clivados com tripsina purificada a partir de semente de *Arabidopsis* tipo selvagem (não-recombinante) que foi processado identicamente a

preparações de corpo oleoso recombinante 4405 clivado com tripsina.

Foram mantidos em biotério murganhos B6 e alimentados *ad libitum*, num ciclo de 12 horas escuro-luz.

Para os testes de tolerância à insulina foram injectados murganhos intraperitonealmente (IP) com insulina (1 U/kg de peso corporal) e os níveis de glucose foram determinados a 0, 15, 30 e 60 minutos utilizando o glucómetro automático. Todos os testes de tolerância à insulina foram realizados no mesmo intervalo a cada dia (9:00 da manhã). Os testes de tolerância a insulina foram realizados com pelo menos 2 dias de intervalo até à administração do teste seguinte. Os resultados para os testes de tolerância à insulina são apresentados na Figura 13. A Insulina SBS DesB₃₀ derivada de sementes 4405 (losangos a cheio) teve um comportamento quase idêntico (estatisticamente não foi diferente, $p < 0,05$) a Humulin R® (quadrados a branco) e Insulina da Roche (padrões com triângulos a branco após injeção durante o decurso do estudo. Todas as insulina(s) testadas reduziram significativamente os níveis de glucose no plasma ($p < 0,05$) em comparação com placebo de soro fisiológico (círculos a branco) e corpos oleosos de *Arabidopsis* de tipo selvagem clivados com tripsina (círculos a cheio) (controlo negativo).

Exemplo 7

Construção de pSBS4401: proteína de fusão PRS-Klip27-MI

Uma das proteínas de fusão estudada começou com a sequência relacionada com o patogénio do tabaco (PRS) (Sijmons *et al.*,

1990, *Bio/technology*, 8:217-221) que serviu como o péptido sinal para controlar a expressão no ER de um modo de co-tradução. Imediatamente a jusante estava um pró-péptido clivável com tripsina (KLIP27) derivado do pró-péptido TA57 de levedura (Kjeldsen et al., 2001, *Biotechnology and Genetic Engineering Reviews* 18:89-121). Isto foi seguido por mini-insulina (MI) descrito por Kjeldsen et al. (2001).

A estrutura deste plasmídeo pSBS4055, foi baseada no vector binário de planta, pPZP200, descrito por Hajdukiewicz et al. (*Plant Molecular Biology*, 1994, 25:989-994). Em vez do sítio de clonagem múltipla descrito, foi inserido um gene *pat* conferindo resistência à fosfinotricina à planta hospedeira (Wohlleben et al., 1988, *Gene* 70:25-37)) controlado pelo promotor/terminador da ubiquitina de *Petroselinum crispum* (Kawalleck et al., 1993, *Plant. Mol. Bio.*, 21:673-684), entre as sequências de fronteira esquerda e direita. Adicionalmente a esta cassete, foi subclonado o promotor/terminador da β -faseolina de *Phaseolus vulgaris* (Slightom et al., 1983, *Proc. Natl. Acad. Sc. USA* 80:1897-1901) controlado PRS. Foi utilizada PCR convencional (Horton et al., 1989, *Gene* 77:61-68) para fundir a sequência sintética codificando PRS com sítios de endonuclease de restrição *SphI/HindIII* ligados à extremidade 3' do promotor da faseolina para produzir pSBS4011.

A sequência de Klip27-MI foi sintetizada a partir de quatro oligonucleótidos parcialmente sobreponíveis que incorporaram a utilização de códons de *Arabidopsis thaliana* para aumentar o sucesso da tradução eficiente num sistema de expressão à base de plantas. Os oligonucleótidos 1324 (GAAGAAGGAGAGCCTAAGTTTGTTAATCAACATCTTTGTGGATCTCATCTTGTTGAGGCTCTCTACCTTG) - SEQ ID N°:177 e 1323 (CCTTAGGAGTGTAGAAAAATCCTCTTTCTCCACACACAAGGTAGAGAGCCTCAACA)

- SEQ ID N°:178 foram emparelhados na sua sobreposição de 20 nucleótidos complementares e estendidos para formar a extremidade 5' da fusão Klip27-MI enquanto que o mesmo foi realizado com os oligonucleótidos 1322 (CTAAGGCTGCTAAGGGAATTG) - SEQ ID N°:179 e 1321 (AAGCTTCAGTTGCAATAGTTCTCCAATTGGTAAAGTGAGCA AATAGAAGTGCAACATTGTTCAACAATTCCTTAGCAGCCTT) - SEQ ID N°:180 para formar a extremidade 3'. As duas metades foram ligadas após a digestão de restrição com *Bsu36I*, para produzir a sequência codificante de Klip27-MI completa. A PCR deste gene de fusão utilizando os iniciadores 1363 (GCATGCCCAACCAATTGATGACACTG) - SEQ ID N°:184 e iniciador 1329 (AAGCTTCAGTTGCAATAGTTC) - SEQ ID N°:183 ligados a sítio de clivagem por endonuclease de restrição 5' *SphI* e 3' *HindIII* para subsequente ligação em pSBS4011 cortado com *SphI/HindIII* (como mencionado acima). O resultado foi o plasmídeo pSBS4401: uma sequência de ADN (SEQ ID N°:188) codificando a proteína de fusão PRS-Klip27-MI (SEQ ID N°:189) sendo colocado num vector binário sob o controlo de expressão do promotor/terminador de faseolina. O promotor da faseolina controla a expressão específica temporal e específica para o tecido do transgene durante o desenvolvimento da semente.

Transformação e Cultura de *E. coli* recombinante e *Agrobacterium* com pSBS4401.

Após confirmar a integridade do ADNc codificando para a proteína de fusão por análise de sequência, o plasmídeo pSBS4401 foi transformado na estirpe de *E. coli* DH5 α para permitir o nível de expressão elevado. O ADN de plasmídeo isolado (100 ng) foi misturado em gelo com 100 μ L de células competentes de DH5 α durante 20 min. As células foram então sujeitas a choque térmico a 42 °C durante 45 segundos e recolocadas no gelo durante 2 min.

Depois, foi adicionado 1 mL de meio SOC e as células foram incubadas a 37 °C num agitador orbital a 225 rpm durante 1 h antes de plaquear células transformadas em placas de LB-espectinomicina (10 g/L triptona, 5 g/L de extracto de levedura, 5 g/L NaCl, 15 g/L agar) e incubando durante a noite a 37 °C. Foi utilizada uma colónia isolada para inocular 5 mL de meio de LB espectinomicina. Estas culturas foram cultivadas durante a noite a 37 °C. O plasmídeo recombinante foi isolado a partir de 1 mL da cultura durante a noite de acordo com a mini prep Qiagen. O plasmídeo isolado foi então utilizado para transformar a estirpe de *Agrobacterium* EH101 competente (Hood et al., 1986; *J. Bacteriol.*: 144: 732-743) por electroporação (25 µF, 2,5 kV, 200 Ω). O *Agrobacterium* recombinante foram plaqueados em AB-espectinomicina/canamomicina (20x sais AB, glucose a 2 M, 0,25 mg/mL de FeSO₄ e 7H₂O, MgSO₄ a 1 M, CaCl₂ a 1 M) e foi utilizada uma colónia isolada para inocular 5 mL de meio AB-espectinomicina/canamomicina. Estas culturas foram cultivadas durante a noite a 28 °C. Os *Agrobacterium* recombinantes foram então utilizados para transformar plantas de *Arabidopsis thaliana* pelo método de imersão da flor (Clough et al., 1998, *Plant J.*, 16:735-743) como descrito no Exemplo 1.

Níveis de expressão de insulina em *Arabidopsis thaliana*

Os níveis de expressão da proteína de fusão KLIP27-MI (4401) foram determinados em sementes maduras de *Arabidopsis thaliana* transgénicas utilizando o processo delineado no Exemplo 2 acima. O produto do transgene não estava presente nos extractos celulares da semente madura.

Exemplo 8

Construção de pSBS4409: proteína de fusão OLEO-pró-insulina humana (OLEO-hPIN)

Esta proteína de fusão começa com a oleosina de 18 kDa de *Arabidopsis thaliana* seguida em grelha pelo gene codificando para a pró-insulina humana (hPIN). A expressão desta proteína de fusão foi direcionada para os corpos oleosos nascentes formados durante o desenvolvimento do embrião.

A estrutura deste plasmídeo pSBS4008, foi baseada no vector binário de plantas, pPZP200, descrito por Hajdukiewicz *et al.* (*Plant Molecular Biology*, 1994, 25:989-994). Em vez do sítio de clonagem múltipla descrito, foi inserido um gene *pat* conferindo resistência à fosfinotricina à planta hospedeira (Wohlleben *et al.*, 1988, *Gene* 70:25-37)) controlado pelo promotor/terminador da ubiquitina de *Petroselinum crispum* (Kawalleck *et al.*, 1993, *Plant. Mol. Bio.*, 21:673-684), entre as sequências de fronteira da esquerda e da direita. Adicionalmente a esta cassete, foi subclonado o promotor/terminador da β -faseolina de *Phaseolus vulgaris* (Slightom *et al.*, 1983, *Proc. Natl. Acad. Sc. USA* 80:1897-1901) controlando a sequência genómica da oleosina de 18 kDa de *Arabidopsis*. Foi utilizada PCR convencional (Horton *et al.*, 1989, *Gene* 77:61-68) para fundir a sequência do gene da oleosina (menos o codão de terminação) com os sítios de endonuclease de restrição *NcoI* e *HindIII* ligados à extremidade 3' do promotor da faseolina para produzir pSBS4008.

Foi sintetizado um gene de pre-pró-insulina *NcoI*-humana - *HindIII* como um fragmento de 335 pb por Aptagen utilizando a utilização de codões preferida de plantas. A

ligação subsequente em pSBS4008 *NcoI/HindIII* resultou no plasmídeo pSBS4400: uma sequência de ADN codificando a proteína de fusão Oleosina-fiuman pre-pró-insulina sendo colocada num vector binário sob controlo de expressão do promotor/terminador da faseolina. O plasmídeo pSBS4400 serviu como o molde para criar pró-insulina humana (hPIN) através de PCR convencional utilizando polimerase de ADN pfu com os iniciadores dirigidos contra a extremidade 5' (1457 oligo TTCGTGAACCAACTTG - SEQ ID N°:190) e extremidade 3' (1458 oligo AAGCTTTCAGTTACAGTAGT - SEQ ID N°:191) incluindo o sítio *HindIII* da região pró-insulina existente do vector. Foi amplificado um segundo fragmento utilizando polimerase de ADN pfu com um iniciador dirigido para o sítio *SphI* disponível (oligo 1455 GCATGCATGTGTTGAGC - SEQ ID N°:192) para a extremidade 3' do gene da oleosina de *Arabidopsis* (oligo 1456 GGTAGTGTGCTGGCCA - SEQ ID N°:193) no vector pSBS4400. Após a PCR, os produtos foram separados num gel de agarose e bandas correspondentes a um fragmento de 267 pb (hPIN-*HindIII*) e um fragmento de 360 pb (*SphI*-OLEO(extremidade 3')) foram purificadas do gel utilizando um kit de extracção do gel (Qiagen). Os dois fragmentos foram fundidos por uma segunda ronda de amplificação por PCR utilizando polimerase de ADN *Taq* com os iniciadores 1455 (SEQ ID N°:192) e 1458 (SEQ ID N°:193) em combinação com 0,001 µM de um iniciador de PCR de ponte sobreponível (oligo 1459 GGTGGCCAGCACACTACCTTCGTGAACCAACTTGTG - SEQ ID N°:194) durante dois ciclos com uma temperatura de emparelhamento de 58 °C seguido por 31 ciclos a 52 °C de modo a amplificar um fragmento de 627 pb *SphI*-OLEO(extremidade 3')-hPIN-*HindIII*. O fragmento de 627 pb *SphI*-OLEO(extremidade 3')-hPIN-*HindIII* foi então ligado à extremidade pendente T/A de pGEMT Easy Vector System™ (Promega) e utilizado para transformar bactérias DH5α para resultar em pSBS3409 (pGEMT-*SphI*-OLEO(extremidade 3')-hPIN-

HindIII).

O fragmento *SphI/HindIII* de pSBS3409 foi substituído com o fragmento *SphI/HindIII* de pSBS4400. Foram realizadas digestões de restrição convencionais em ambos pSBS3409 e pSBS4404 utilizando *SphI/HindIII* (New England Biolabs). Os fragmentos foram separados em géis de agarose a 1,5% e purificados utilizando um kit de extracção de gel (Qiagen). O fragmento de 617 pb *SphI/HindIII* libertado de pSBS3409 foi então ligado ao sítio aceitador *SphI/HindIII* na estrutura do vector pSBS4400 pré-cortado (fragmento interno *SphI/HindIII* removido) utilizando ligase de ADN de T4 (NEB) durante a noite a 15 °C.

O resultado foi o plasmídeo pSBS4409: uma sequência de ADN (SEQ ID N°:195) codificando a proteína de fusão OLEOSIN-hPIN (SEQ ID N°:196 estando colocada num vector binário sob o controlo de expressão do promotor/terminador da faseolina. O promotor da faseolina controla a expressão específica temporal e específica para o tecido do transgene durante o desenvolvimento da semente.

Transformação e Cultura de *E. coli* recombinante e *Agrobacterium* com pSBS4409.

Após confirmar a integridade do ADNc codificando para a proteína de fusão através da análise de sequência, o plasmídeo pSBS4409 foi transformado na estirpe de *E. coli* DH5 α para permitir a expressão de nível elevado. O ADN de plasmídeo isolado (100 ng) foi misturado em gelo com 100 μ L de células competentes de DH5 α durante 20 min. As células foram então sujeitas a choque térmico a 42 °C durante 45 segundos e

recolocadas no gelo durante 2 min. Depois, foi adicionado 1 mL de meio SOC e as células foram incubadas a 37 °C num agitador orbital a 225 rpm durante 1 h antes de plaquear células transformadas em placas de LB-espectinomicina (10 g/L triptona, 5 g/L extracto de levedura, 5 g/L NaCl, 15 g/L agar) e incubadas durante a noite a 37 °C. Foi utilizada uma colónia única para inocular 5 mL de meio de LB espectinomicina. Estas culturas foram cultivadas durante a noite a 37 °C. O plasmídeo recombinante foi isolado a partir de 1 mL da cultura durante a noite de acordo com mini prep da Qiagen. O plasmídeo isolado foi então utilizado para transformar a estirpe de *Agrobacterium* EH101 competente (Hood *et al.*, 1986; *J. Bacteriol.* 144: 732-743) por electroporação (25 µF, 2,5 kV, 200 Ω). Foram plaqueados *Agrobacterium* recombinante em AB-espectinomicina/canamicina (sais AB a 20x, glucose a 2 M, 0,25 mg/mL de FeSO₄ e 7H₂O, MgSO₄ a 1 M, CaCl₂ a 1 M) e foi utilizada uma colónia isolada para inocular 5 mL de meio AB-espectinomicina/canamicina. Estas culturas foram cultivadas durante a noite a 28 °C. Os *Agrobacterium* recombinante foram então utilizados para transformar plantas de *Arabidopsis thaliana* pelo método de imersão da flor (Clough *et al.*, 1998, *Plant J.*, 16:735-743) como descrito no Exemplo 1.

Níveis de expressão de insulina em *Arabidopsis thaliana*

Os níveis de expressão da proteína de fusão OLEO-hPIN (4409) foram determinados em semente madura de *Arabidopsis thaliana* transgénica utilizando o processo apresentado no Exemplo 2 acima. Gel corado com Coomassie de proteínas de corpo oleoso de duas linhas representativas (4409-6 e 4409-8) comparando a migração da proteína de fusão Oleosina-hPIN (como

apresentado pela seta preta) com *Arabidopsis* não transformada (de tipo selvagem) (Figura 14). O nível de expressão foi determinado por densitometria para medir em média cerca de 0,10% da proteína da semente total. Este nível foi calculado acima e para além da co-migração de uma proteína endógena do mesmo peso molecular na semente não transformada (de tipo selvagem) que constituiu aproximadamente 0,04% da proteína total da semente.

Exemplo 9

Transformação de açafroa

Este protocolo de transformação é semelhante ao delineado por Orlilcowska T. K. *et al.* ((1995) *Plant Cell, Tissue and Organ Culture* 40: 85-91), mas com modificações e melhoramentos, ambos para transformar S-317 e para utilizar fosfinotricina como o marcador de selecção. Sementes descontaminadas de variedade de açafroa S-317 da Califórnia, que não estão danificadas, quebradas ou doentes, em HCl₂ a 0,1% durante 12 minutos, seguido por 4-5 passagens por água destilada estéril. Sementes estéreis germinadas na escuridão em meio MS (Murashige T. & Skoog F (1962) *Physiol. Plant.* 15: 473-497) com sacarose a 1% e 0,25% ed Gelrite. Culturas de *Agrobacterium* iniciais de stocks de glicerol congelados em 5 mL de meio líquido mínimo AB com selecção de antibióticos e cultivado durante 48 horas a 28 °C. Cultivado numa fracção desta cultura cultivada durante a noite em 5 mL de meio Luria com selecção para transformação. Lavar com 6-8 mL de células bacterianas duas vezes com meio AB e ressuspensas para uma densidade celular final de 0,4-0,5 (OD₆₀₀).

Remover cotilédones de dois dias de idade de plantas

germinadas derivadas de sementes, imersas nas células de *Agrobacterium* preparadas e as placas em meio MS com sacarose a 3%, 4 µM de N6-benziladenina (BA) e 0,8 µM de ácido naftalenoacético (NAA). Placas incubadas, a 21° C, em condições de escuridão. Após 3 dias, a transferência para o mesmo meio com 300 mg/L de timentina. Após mais 4 dias, mudar todas as culturas para a luz. Após 3 dias, colocar os explantes no meio de selecção com fosfinotricina adicionado a 0,5 mg/L. Para alongação do botão continuada, transferir os explantes semanalmente para meio MS sem fito-hormonas, mas com duas vezes a quantidade basal de KNO₃. Excisar os rebentos que tinham alongado até serem maiores do que 10 mm a partir do explante inicial e individualmente crescer em selecção. Para o enraizamento, colocar rebentos verdes, representando tecido transgénico putativo, em meio MS com sacarose a 2%, ácido indolebutírico a 10 µM e NAA a 0,5 µM. Transferir os rebentos enraizados para uma mistura sem solo bem drenada e cultivar sob humidade elevada e 12 horas de luz.

Exemplo 10

Protocolo de transformação de linho

Este processo de transformação é semelhante ao apresentado por Dong J. e McHughen A. (*Plant Cell Reports* (1991) 10:555-560), Dong J. e McHughen A. (*Plant Sciences* (1993) 88:61-71) e Mlynarova et al. (*Plant Cell Reports* (1994) 13: 282-285). Sementes de linho descontaminadas que não estão danificadas, quebradas ou doentes, numa solução de etanol a 70% durante 5 a 7 minutos, seguido por 25 minutos numa solução de lixívia a 50% com Tween 20 (3-4 gotas por 100 mL) com agitação

contínua. Lavar as sementes 5 a 7 vezes com água estéril destilada. Germinar sementes descontaminadas à luz em meio MS (Murashige T. & Skoog F (1962) *Physiol. Plant.* 15: 473-497) com sacarose a 2% e Gelrite® a 0,3% em frascos Magenta. Para transformação, cultivar culturas de *Agrobacterium* durante a noite em meio AB mais o antibiótico apropriado para selecção. Lavar 6 a 8 mL de células durante a noite duas vezes e ressuspender em 5 mL de meio AB; adicionar 2 mL deste stock a 98 mL de meio de indução (meio basal MS com sacarose a 3%, 6-benzilaminopurina a 5 µM (BA) e ácido alfa-naftalenoacético a 0,25 µM (NAA) e ajustar para uma OD₆₀₀ final de 1,0.

Seleccionar explantes de hipocótilo e inocular na solução de células de *Agrobacterium* preparadas durante cerca de 4 h (agitar placas cuidadosamente 1-2 vezes durante este período). Após o período de infecção, remover os explantes do meio de inoculação líquido e transferência para papel de filtro estéril. Plaquear 15-20 explantes em 0,7% de meio de indução solidificada com agar em placas de cultura de tecidos. Selar as placas com película aderente e co-cultivar os explantes durante 48 h sob condições de luz (23-24 °C). Após 2 dias, transferir os explantes verdes, meristemáticos para o mesmo meio contendo 300 mg/L de Timentina (meio de pré-selecção) e embrulhar com película aderente. Após 3 dias, transferir as culturas para o meio acima contendo 10 mg/L de DL PPT (Seleccção 1). Embrulhar as placas com Parafilm® e incubar a 24 °C em condições de luz. Transferir as culturas a cada duas semanas e manter neste meio durante um mês. Para o alongamento dos rebentos, transferir as culturas a cada duas semanas em meio de selecção II (meio basal MS contendo sacarose a 2%, 500 mg/L de tampão MES, 300 mg/L de Timentina e 10 mg/L de DL PPT) em frascos Magenta. Os rebentos transformados putativos, que sobreviveram à selecção, são

verde-escuro e formam rebentos vigorosos em 7-10 dias, quando plantados individualmente em meio de selecção II. Transferir os rebentos enraizados para uma mistura de solo de estufa esterilizado em vasos pequenos e cobrir as plântulas com copos de plástico transparentes para aclimatização. Para a maturação, transferir plantas de crescerem activamente para vasos de um galão com uma mistura de solo bem drenada e cultivar em condições de estufa.

Tabela 1 Exemplos de Sequências de Insulina Conhecidas

SEQ ID Nº	Motivo de Insulina (Identificador da Sequência de Aminoácidos) {Identificador da Sequência de Ácido Nucleico}
Insulina Nativa Humana	
7	(P01308)Pré-pró-insulina humana {Compreende os genes V00565, M10039, J00265, X70508, L15440, BC005255 e AJ009655}
Insulina Nativa Não Humana	
Mamíferos	
8	(AAB25818) Péptido C de Pró-insulina de <i>Equus przewalskii</i> (cavalos, zebra, rinoceronte e tapir (Perissodactyla))
9	(P01310)Pré-pró-insulina <i>Equus caballus</i> (cavalo)
10	(P01311)Pré-pró-insulina {Inclui os genes U03610 e M61153} <i>Oryctolagus cuniculus</i> (coelho doméstico)
11	(P01312) insulina <i>Balaenoptera physalus</i> (baleia comum)

(continuação)

SEQ ID Nº	Motivo de Insulina (Identificador da Sequência de Aminoácidos) {Identificador da Sequência de Ácido Nucleico}
12	(P01314) Insulina <i>Balaenoptera borealis</i> (baleia sardinheira)
13	(P01315) Pré-pró-insulina {Compreende os genes AF064555 e AY044828} <i>Sus scrofa</i> (porco) suíno
14	(P01316) Insulina <i>Elephas maximus</i> (elefante asiático)
15	(P01317) Pré-pró-insulina {Gene M54979} <i>Bos taurus</i> (vaca)
16	(P01318) Pré-pró-insulina {Gene U00659} <i>Ovis aries</i> (ovelha)
17	(P01320) Insulina <i>Camelus dromedarius</i> (camelo árabe)
18	(P01321) Pré-pró-insulina {Gene V00179} <i>Canis sp.</i> (caninos)
19	(P01328) Insulina <i>Hystrix cristata</i> (porco-espinho)
20	(P10604) Pré-pró-insulina {Gene J02989} <i>Aotus trivirgatus</i> (douroucouli) macaco noturno
21	(P30406) Pré-pró-insulina { Gene J00336} <i>Macaca fascicularis</i> (macaca maura)
22	(P30407) Pré-pró-insulina {Gene X61092} <i>Cercopithecus aethiops</i> (macacos vervet)
23	(P30410) Pré-pró-insulina {Gene X61089} <i>Pan troglodytes</i> (chimpanzé)
24	(Q9TQY7) Insulina <i>Ornithorhynchus anatinus</i> (platypus)
25	(AAM76641) Insulina {Gene AY092024} <i>Pongo pygmaeus</i> (orangotango)

(continuação)

SEQ ID Nº	Motivo de Insulina (Identificador da Sequência de Aminoácidos) {Identificador da Sequência de Ácido Nucleico}
26	(AAN06935) Pré-pró-insulina {Genes: AH011815 (contendo AY137498, AY137499 e AY137500)} <i>Gorilla gorilla</i> (gorila)
27	(INMKSQ) Insulina <i>Saimiri sciureus</i> (macaco esquilo comum)
28	(P01313) Pré-pró-insulina {Gene M26328} <i>Cricetulus longicaudatus</i> (hamster de cauda longa)
29	(P01322) precursor de Insulina 1 {inclui os genes V01242, V01242 e M25584} <i>Rattus norvegicus</i> (ratazana castanha)
30	(P01323) precursor da Insulina 2 {inclui os genes V01243, J00748, M25583 e M25585} <i>Rattus norvegicus</i> (ratazana castanha)
31	(P01324) Insulina <i>Acomys cahirinus</i> (ratinho Egípcio)
32	(P01325) precursor da Insulina 1 (inclui os genes X04725 e AK007482) <i>Mus musculus</i> (ratinho doméstico)
33	(P01326) precursor da Insulina 2 {Gene X04724} <i>Mus musculus</i> (ratinho doméstico)
34	(P01327) Insulina <i>Chinchilla brevicaudata</i> (chinchila)
35	(P01329) Pré-pró-insulina {inclui os genes K02233 e M11713} <i>Cavia porcellus</i> (porquinho da Índia)
36	(P17715) Pré-pró-insulina {Gene M57671} <i>Octodon degus</i> (degu)
37	(P18109) Insulina <i>Didelphis virginiana</i> (gambá da Virgínia)
38	(P21563) Pré-pró-insulina [<i>Rodentia</i> sp.]

(continuação)

SEQ ID Nº	Motivo de Insulina (Identificador da Sequência de Aminoácidos) {Identificador da Sequência de Ácido Nucleico}
39	(Q62587) Pré-pró-insulina {Gene X98241} <i>Psammomys obesus</i> (rato da areia)
40	(Q91XI3) Pré-pró-insulina {Gene AY038604} <i>Spermophilus tridecemlineatus</i> (Esquilo de treze linhas)
41	(740063A) Péptido C da Insulina <i>Cavia porcellus</i> (porquinho da Índia)
Aves	
42	(P01332) Pré-pró-insulina {inclui os genes AH002454 (contendo, J00872, J00873 e J00874), V00416, V00418 e X58993} <i>Gallus gallus</i> (galinha)
43	(P01333) Pré-pró-insulina <i>Anas platyrhynchos</i> (pato real)
44	(P07454) Insulina <i>Anser anser anser</i> (ganso bravo)
45	(P51463) Pré-pró-insulina {Genes: AH006925 (contendo S66611 e S66612)} <i>Selasphorus rufus</i> (colibri)
Peixes	
46	(073727) Pré-pró-insulina {Gene: AF036326} <i>Danio rerio</i> (peixe zebra)
47	(P01335) Pré-pró-insulina {Gene: X00989} <i>Cyprinus carpio</i> (caspa comum)
48	(P01337) Insulina <i>Batrachoididae</i> gen. sp. (peixe sapo)
49	(P01339) Insulina <i>Thunnus thynnus</i> (albacora)
50	(P01340) Insulina <i>Katsuwonus pelamis</i> (gaiado)
51	(P01341) Pré-pró-insulina {Gene: V00634} <i>Lophius piscatorius</i> (tamboril)

(continuação)

SEQ ID Nº	Motivo de Insulina (Identificador da Sequência de Aminoácidos) {Identificador da Sequência de Ácido Nucleico}
52	(P01342)Pré-pró-insulina {Gene: V00649} <i>Myxine glutinosa</i> (enguia de casulo)
53	(P04667)Pré-pró-insulina {Inclui os genes: X00148, J00936, K01655 e X13559} <i>Oncorhynchus keta</i> (salmão keta)
54	(P07453) <i>Myoxocephalus scorpius</i> (peixe escorpião)
55	(P09476)Insulina <i>Lepisosteus spatula</i> (peixe agulha)
56	(P09477)Insulina <i>Platichthys flesus</i> (patruça)
57	(P09536)Insulina <i>Hydrolagus colliei</i> (quimera)
58	(P12704)Insulina <i>Squalus acanthias</i> (cação)
59	(P12705)Pré-pró-insulina <i>Torpedo marmorata</i> (raia eléctrica)
60	(P13190)Pró-insulina {Gene: U82395} <i>Callorhinchus milii</i> (peixe elefante)
61	(P14806)Insulina <i>Petromyzon marinus</i> (lampreia do mar)
62	(P23187)Insulina <i>Oncorhynchus gorbuscha</i> (salmão cor-de-rosa)
63	(P29335)Insulina <i>Amia calva</i> (âmia)
64	(P42633)Insulina <i>Anguilla rostrata</i> (enguia americana)
65	(P81025)Pré-pró-insulina {Gene: AF038123} <i>Oreochromis niloticus</i> (tilápia do Nilo)
66	(P81423)Insulina <i>Acipenser gueldenstaedtii</i> (esturjão russo)
67	(P81881)Insulina <i>Piaractus mesopotamicus</i> (Pacu)
68	(Q9W7R2)Pré-pró-insulina {Gene AB029318} <i>Verasper moseri</i> (solha)

(continuação)

SEQ ID Nº	Motivo de Insulina (Identificador da Sequência de Aminoácidos) {Identificador da Sequência de Ácido Nucleico}
69	(1603264A) Péptido C de Insulina <i>Anguilla anguilla</i> (enguia europeia)
Anfíbios	
70	(P12706) precursor da Insulina 1 {Gene: M24443} <i>Xenopus laevis</i> (rã Africana)
71	(P12707) precursor da Insulina 2 {Gene: M24442} <i>Xenopus laevis</i> (rã Africana)
Répteis	
72	(P31887) Insulina <i>Trachemys scripta</i> (tartaruga da Flórida)
73	(P12703) Insulina <i>Alligator mississippiensis</i> (alligator do Mississippi)
74	(P12708) Insulina <i>Zaocys dhumnades</i> (cobra)
75	(P01334) Insulina <i>Crotalus atrox</i> (cascavel)
Insulina Modificada	
Humana	
76	(AAA72172) Pré-pró-insulina Sintética {Gene: J02547}
77	(AAA72916) extremidade 3' da cadeia alfa da insulina sintética {Genes: AH003171 ou M38610}
78	(AAA72917) extremidade 3' da cadeia beta da insulina sintética {Genes: AH003171 ou M38611}
79	(CAA00712) Insulina Sintética {Gene: A07755}
80	(CAA00713) Insulina Sintética {Gene: A07758}
81	(CAA00714) Insulina Sintética {Gene: A07761}
82	(CAA00715) Produto de proteína sem nome {Gene: A07764}

(continuação)

SEQ ID Nº	Motivo de Insulina (Identificador da Sequência de Aminoácidos) {Identificador da Sequência de Ácido Nucleico}
83	(CAA00736) pró-Insulina Sintética {Gene: A08012} (EP 0367163-A)
84	(CAA00783) Insulina Sintética {Gene: A08468} (EP 0376156-A)
85	(CAA01581) precursor da Insulina Modificada {Gene: A21951} (documento WO 9011299)
86	(CAA01254) Insulina Sintética {Gene: A15938} (EP 0214826-A)
87	(CAA01799) Asp(B1), Asp(B4), Asp(B10), Asp(B16), Glu(B27)A26317} construção da Insulina Sintética {Gene: A26317}
88	(CAA01798) Glu(B9), Glu(A12) construção do precursor de Insulina Sintética {Gene: A26314}
89	(CAA23424) pró-Insulina Sintética {Gene: V00082}
90	(CAA24707) Cadeia C da Insulina Sintética {Gene: V01461}
91	(CAA25151) Cadeia B da Insulina Sintética {Gene: X00462}
92	(CAD60056) Produto de proteína sintética não denominada {Gene: AX573757} (Pat. WO 02/079250)
93	(Gene: M31026) Cadeias de Insulina Sintética humana B e mini-C utilizando cromatografia de sílica gel desactivada
94	(1BZVA) Cadeia A [d-Alab26]-Des(B27-B30)-insulina-B26-Amida A Análogo da Insulina de Substituição Supervasonte Única

(continuação)

SEQ ID Nº	Motivo de Insulina (Identificador da Sequência de Aminoácidos) {Identificador da Sequência de Ácido Nucleico}
95	(1BZVB) Cadeia B [d-Alab26]-Des(B27-B30)-Insulina-B26-Amida A Análogo da Insulina de Substituição Supervasonte Única
96	(1HUIA) Mutante de Insulina Cadeia A (B1, B10, B16, B27)glu, Des-B30, Nmr
97	(1HUIB) Mutante de Insulina Cadeia B (B1, B10, B16, B27)glu, Des-B30, Nmr
98	(1HLSA) Mutante de Insulina Humana Cadeia A -His(B16)
99	(1HLSB) Mutante de Insulina Humana Cadeia B-His(B16)
100	(1JCAA) Concepção não convencional de cadeia A de Análogos de Insulina Instáveis com Actividade Aumentada
101	(1JCAB) Concepção não convencional de cadeia B de Análogos de Insulina Instáveis com Actividade Aumentada
102	(1JCAC) Concepção não convencional de cadeia C de Análogos de Insulina Instáveis com Actividade Aumentada
103	(1JCAD) Concepção não convencional de cadeia D de Análogos de Insulina Instáveis com Actividade Aumentada
104	(1J73A) Cadeia A de um Análogo de Insulina Instável com Actividade Nativa
105	(1J73B) Cadeia B de um Análogo de Insulina Instável com Actividade Nativa

(continuação)

SEQ ID Nº	Motivo de Insulina (Identificador da Sequência de Aminoácidos) {Identificador da Sequência de Ácido Nucleico}
106	(1J73C) Cadeia C de um Análogo de Insulina Instável com Actividade Nativa
107	(1J73D) Cadeia D de um Análogo de Insulina Instável com Actividade Nativa
108	(1KMFA) Cadeia A de Mutante de Insulina Humana Ile-A2-Allo-Ile, His-B10-Asp, Pro-B28-Lys, Lys-B29-Pro
109	(1KMFB) Cadeia B de Mutante de Insulina Humana Ile-A2-Allo-ile, His-B10-Asp, Pro-B28-Lys, Lys-B29-Pro
110	(1K3MA) Cadeia A de Mutante de Insulina Humana Ile-A2-Ala, His-B10-Asp, Pro-B28-Lys, Lys-B29-Pro
111	(1K3MB) Cadeia B de Mutante de Insulina Humana Ile-A2-Ala, His-B10-Asp, Pro-B28-Lys, Lys-B29-Pro
112	(1LW8A) Cadeia A Allo-ilea2-insulina, Um Análogo Quiral Inactivo
113	(1LW8B) Cadeia B Allo-ilea2-Insulina, Um Análogo Quiral Inactivo
114	(1LW8C) Cadeia C Allo-Ilea2-Insulina, Um Análogo Quiral Inactivo
115	(1LW8D) Cadeia D Allo-ilea2-Insulina, Um Análogo Quiral Inactivo
116	(1LKQA) Cadeia A de Mutante de Insulina Humana Ile-A2-Gly, Val-A3-Gly, His-B10-Asp, Pro-B28-Lys, Lys-B29-Pro
117	(1LKQB) Cadeia B de Mutante de Insulina Humana Ile-A2-Gly, Val-A3-Gly, His-B10-Asp, Pro-B28-Lys, Lys-B29-Pro

(continuação)

SEQ ID Nº	Motivo de Insulina (Identificador da Sequência de Aminoácidos) {Identificador da Sequência de Ácido Nucleico}
118	(1MHIA) Cadeia A de mutante B9 (Asp)
119	(1MHIB) Cadeia B de mutante B9 (Asp)
120	(1MHJA) Cadeia A de mutante B25 (phe)
121	(1MHJB) Cadeia B de mutante B25 (phe)
122	(1VKTA) Cadeia A, Modelo de dois Persulfuretos de Insulina Humana
123	(1VKTB) Cadeia B, Modelo de dois Persulfuretos de Insulina Humana
Não-Humana	
124	(AAG59607) Insulina albebetina Sintética {Gene: AY017185}
125	(AAG59606) Insulina albeferão Sintética {Gene: AY017184}
Proteínas de Fusão de Insulina	
Humana	
126	(AAB27046) N-terminal de uma proteína de fusão interlequina 2-Insulina
127	(AAB27047) proteína de fusão Beta-galactosidase-Insulina N-terminal
128	(PC7082) proteína de fusão factor de crescimento da epiderme/precursor da Insulina humana de cadeia simples Des-B30
Não-humana	
129	(AAA72177) proteína de fusão penicilinase de <i>E.coli</i> /extremidade 5' da insulina I de rato {G AH003149 ou J02553}

(continuação)

SEQ ID Nº	Motivo de Insulina (Identificador da Sequência de Aminoácidos) {Identificador da Sequência de Ácido Nucleico}
130	(AAA72178) proteína de fusão penicilinase de <i>E.coli</i> /extremidade 3' da insulina I de rato {Genes: AH003149 ou J02554}
131	(AAA72179) proteína de fusão sequência sinal de insulina de rato/beta-galactosidase de <i>E.coli</i> {Gene: J02555}
132	(AAA72181) proteína de fusão vírus de símio 40 (SV40)/Pré-pró-insulina I de rato {Gene: J02559}
Mini Insulina	
Humana	
133	(1EFEA) Cadeia A, Uma Mini-Pró-insulina Activa, M2pi
134	(1JCAA) Cadeia A, concepção Não Convencional de Análogos da Insulina Instáveis Com Actividade Aumentada
135	(1JCAB) Cadeia B, concepção Não Convencional de Análogos da Insulina Instáveis Com Actividade Aumentada
136	(1JCAC) Cadeia C, concepção Não Convencional de Análogos da Insulina Instáveis Com Actividade Aumentada
137	(1JCAD) Cadeia D, concepção Não Convencional de Análogos da Insulina Instáveis Com Actividade Aumentada
138	(1JK8C) Cadeia C, Um Complexo de Péptido de Insulina Humana-HIa-Dq8

(continuação)

SEQ ID Nº	Motivo de Insulina (Identificador da Sequência de Aminoácidos) {Identificador da Sequência de Ácido Nucleico}
139	(1J73A) Cadeia A, Um Análogo da Insulina Instável Com Actividade Nativa
140	(1J73B) Cadeia B, Um Análogo da Insulina Instável Com Actividade Nativa
141	(1J73C) Cadeia C, Um Análogo da Insulina Instável Com Actividade Nativa
142	(1J73D) Cadeia D, Um Análogo da Insulina Instável Com Actividade Nativa
143	(1SJTA) Cadeia A, Mini-Pró-insulina, Mutante Análogo da Insulina com duas cadeias: Des B30,His(B 10)asp, Pro(B 28)asp
144	(1SJTB) Cadeia B, , Mini-Pró-insulina, Mutante Análogo da Insulina com duas cadeias: Des B30,His(B 10)asp, Pro(B 28)asp
145	(1SJU) Mini-Pró-insulina, , Mini-Pró-insulina, Mutante Análogo da Insulina com cadeia única: Des B30, His(B10)asp, Pro(B28)asp e ligação peptídica entre Lys B29 e Gly A1

SUMÁRIO DAS SEQUÊNCIAS

SEQ ID N°:1 e 2 apresentam a sequência de nucleótidos e a sequência de aminoácidos deduzida, respectivamente, da proteína de fusão PRS-D9scFv-KLIP27-MI-KDEL no plasmídeo pSBS4404.

SEQ ID N°:3 e 4 apresenta a sequência de nucleótidos e a sequência de aminoácidos deduzida, respectivamente, da proteína de fusão Oleo-KLIP8-KLIP27-MI no plasmídeo pSBS4405.

SEQ ID N°:5 e 6 apresentam a sequência de nucleótidos e a sequência de aminoácidos deduzida, respectivamente, da proteína de fusão PRS-MI-ligando tetrabásico-D9Scfv-KDEL no plasmídeo pSBS4414.

SEQ ID N°:7 até 145 apresentam as sequências de insulina conhecidas que são descritas na Tabela 1.

SEQ ID N°: 146 até 148 apresentam as sequências de aminoácidos ou fragmentos do péptido C da insulina.

SEQ ID N°:149 apresentam a sequência de aminoácidos do péptido de processamento tetrabásico.

SEQ ID N°:150 até 155 apresentam as sequências de aminoácidos dos polipéptidos capazes de reter o polipéptido da insulina no ER.

SEQ ID N°:156 até 160 apresentam as sequências de aminoácidos dos polipéptidos capazes de reter o polipéptido da insulina num organito de armazenamento derivado de ER.

SEQ ID N°:161 apresenta a sequência de aminoácidos de uma sequência sinal PRS.

SEQ ID N°:162 até 171 apresentam as sequências de aminoácidos de sequências líder de levedura e as suas sequências derivadas.

SEQ ID N°:172 até 173 apresentam as sequências de aminoácidos dos péptidos espaçadores.

SEQ ID N°:174 apresenta a sequência de aminoácidos da sequência KLIP8

SEQ ID N°:175 apresenta a sequência de nucleótidos do iniciador de sentido directo 1325 que é complementar à região 5' do clone de ADNc D9ScFv e é concebido para adicionar um sítio *SphI* à região 5' para facilitar a ligação subsequente.

SEQ ID N°:176 apresenta a sequência de nucleótidos do iniciador reverso 1326 que é complementar à região 3' do clone de ADNc D9ScFv e é concebido para adicionar um sítio *XhoI* site à região 3' para facilitar a ligação subsequente.

SEQ ID N°:177 apresenta a sequência de nucleótidos do iniciador de sentido directo 1324 que é complementar à região de 20 nucleótidos do iniciador reverso 1323 e é concebido para formar a extremidade 5' da fusão Klip27-MI.

SEQ ID N°:178 apresenta a sequência de nucleótidos do iniciador reverso 1323 que é complementar à região de 20 nucleótidos do iniciador de sentido directo 1324 e é

concebido para formar a extremidade 5' da proteína de fusão Klip27-MI.

SEQ ID N°:179 apresenta a sequência de nucleótidos do iniciador em sentido directo 1322 que é complementar à região de 19 nucleótidos do iniciador reverso 1321 e é concebida para formar a extremidade 3' da fusão Klip27-MI.

SEQ ID N°:180 apresenta a sequência de nucleótidos do iniciador reverso 1321 que é complementar à região de 19 nucleótidos do iniciador em sentido directo 1322 e é concebido para formar a extremidade 3' da fusão Klip27-MI.

SEQ ID N°:181 apresenta a sequência de nucleótidos do iniciador em sentido directo 1364 que é complementar à região 5' da sequência de Klip27-MI e é concebida para adicionar um sítio *XhoI* à região 5' para facilitar a ligação subsequente.

SEQ ID N°:182 apresenta a sequência de nucleótidos do iniciador reverso 1334 que é complementar à região 3' da sequência Klip27-MI e é concebida para adicionar um sítio *HindIII* à região 3' para facilitar a ligação subsequente e uma sequência 3' KDEL.

SEQ ID N°:183 apresenta a sequência de nucleótidos do iniciador reverso 1329 que é concebida para adicionar um sítio *HindIII* à região 3' para facilitar a ligação subsequente.

SEQ ID N°:184 apresenta a sequência de nucleótidos do iniciador em sentido directo 1363 que é complementar à

região 5' da sequência Klip27-MI e é concebida para adicionar um sítio *SphI* à região 5' para facilitar a ligação subsequente.

SEQ ID N°:185 apresenta a sequência de nucleótidos do iniciador em sentido directo 1515 que é complementar à região 5' da sequência da cadeia B da Insulina e é concebida para inserir o sítio tetrabásico intercalar entre as autênticas cadeias A e B da insulina humana em conjugação com o iniciador reverso 1518.

SEQ ID N°:186 apresenta a sequência de nucleótidos do iniciador reverso 1518 que é complementar à região 3' da cadeia B da insulina e a região 5' da cadeia A da insulina com a sequência de mini-péptido C tetrabásico intercalar e é concebida para inserir o sítio tetrabásico intercalar entre as cadeias A e B autênticas de insulina humana.

SEQ ID N°:187 apresenta a sequência de nucleótidos do iniciador reverso 1517 que é complementar à região 3' da sequência D9 scFV/KDEL e é concebida para amplificar o MI-ligando tetrabásico-D9Scfv-KDEL inteiro para criar a inserção pSBS4414.

SEQ ID N°:188 e 189 apresentam a sequência de nucleótidos e a sequência de aminoácidos deduzida, respectivamente, da proteína de fusão PRS-Klip27-MI no plasmídeo pSBS4401.

SEQ ID N°:190 apresenta a sequência de nucleótidos do iniciador em sentido directo 1457 que é complementar à região 5' da sequência da cadeia B da insulina e é concebida

para criar o fragmento da pró-insulina humana (hRIN) em conjugação com o iniciador reverso 1591.

SEQ ID N°:191 apresenta a sequência de nucleótidos do iniciador reverso 1458 que é complementar à região 3' da pró-insulina humana (hPIN) é concebida para criar a pro(hPIN) humana e adiciona um sítio de clonagem 3' *HindIII*.

SEQ ID N°:192 apresenta a sequência de nucleótidos do iniciador de sentido directo 1455 que é complementar à região 5' do sítio *SphI* de pSBS4404 e é concebida para amplificar o gene da oleosina de *Arabidopsis* em conjugação com o iniciador reverso 1456.

SEQ ID N°:193 apresenta a sequência de nucleótidos do iniciador reverso 1456 que é complementar à região 3' do gene da oleosina de *Arabidopsis* e é concebido para amplificar o gene da oleosina de *Arabidopsis* em conjugação com o iniciador de sentido directo 1455.

SEQ ID N°:194 apresenta a sequência de nucleótidos do iniciador de PCR que faz a ponte sobreponível que é complementar à região 3' do gene da oleosina de *Arabidopsis* e a extremidade 5' do gene da pró-insulina humana gene e é concebida para criar a inserção pSBS4409 em conjugação com o iniciador em sentido directo 1455 e o iniciador reverso 1456.

SEQ ID N°:195 e 196 apresentam a sequência de nucleótidos e a sequência de aminoácidos deduzida, respectivamente, da proteína de fusão OLEO-hPIN no plasmídeo pSBS4409.

SEQ ID N°:1

atgaacttccttaagtctttccctttctacgctttcctttgtttcgggtcaatacttc
gttgctgttacgcatgctgacattgtgatgacacagtctccatcctccctggctatg
tcagtgggacagcgggtcactatgcgctgcaagtccagtcagagccttttaaaaagt
accaatcaaaagaactatattggcctgggtaccagcagaaaccaggacagtctcctaaa
cttctgggtatactttgcatccactagggaaatctggggcctgatcgcttcataggc
agtggatctgggacagatttcaactcttaccatcagcagtggtgcaggctgaagacctg
gcagattacttctgtcagcaacattataaactcctcccacggtcgggtgctgggacc
aagctggagcttaagcgggtctccgaacgggtgcttctcatagcgggtctgcaccaggc
actagctctgcatctggatctcaggtgcacctgcagcagctcggagctgagctgatg
aagcctggggcctcaatgaagatacctgcaaggctactggctacacattcagtagc
tactggatagagtgggtaaagcagaggcctggacatggccttgagtgattggagag
atthtacctggcagtggttagtactacctaataatgagaagttcaagggcaaggccaca
ttcactgcagatacatcctccaacacagcctacatgcaactcagcagcctgacatct
gaggactctgccgtctattactgtgcaagattggatggtgactcctggggccaaggc
accactctcacagtctcgagtcaaccaattgatgacactgaatcccagaccacgtca
gtgaacctcatggccgatgatactgagagcgcgtttgctacacaaacaaattcggga
ggtcttgacgttgtcggattgatctccatggetaagagagaagaaggagagcctaag
tttgtaatacaacatctttgtggatctcatcttggaggctctctaccttggtgtg
ggagaaagaggatthttctacactcctaaggctgctaagggaaattggtgaacaatgt
tgcacttctatthgctcactttaccaattggagaactattgcaacaaggatgaactt
tga

SEQ ID N°:2

MNFLKSFPHYAFLCFGQYFVAVTHADIVMTQSPSSLAMSVGQRVTMRCKSSQSLLKS
TNQKNYLAWYQKPGQSPKLLVYFASTRESGVPDRFIGSGSGTDFTLTISSVQAEDL
ADYFCQOHYNTPTFGAGTKLELKRSPNGASHSGSAPGTSSASGSQVHLQQSGAELM
KPGASMKISCKATGYTFSSYWIEWVKQRPHGLEWIGEILPGSGSTTYNEKFKGKAT
FTADTSSNTAYMQLSSLTSEDSAVYYCARLDVDSWGQGTTLTVSSQPIDDTESQTT
VNLMAADTESAFATQTNSGGLDVVGLISMAREEGEPKFVNQHLGSHLVEALYLVC
GERGFFYTPKAAKGIVEQCCTSI CSLYQLENYCNKDEL

SEQ ID N°:3

atggcggatacagctagaggaacccatcacgatatcatcggcagagaccagtaaccgatg
atgggccgagaccgagaccagtagcagatgtccggacgaggatctgactactccaagtct
aggcagattgctaaagctgcaactgctgtcacagctgggtggttccctccttggttctctcc
agccttacccttggttggaactgtcatagctttgactggtgcaacacctctgctcgttatc
ttcagcccaatccttgtccggctctcatcacagttgcaactcctcatcacgggttttctt
tcctctggagggtttggcattgccgctataaccggttttctcttggatttacgcaacggga
gagcaccacagggatcagacaagttggacagtgaaggatgaagttgggaagcaagct
caggatctgaaagacagagctcagtagcagcaacatactgggggaaacatgac
cgtgaccgtactcgtgggtggccagcactaccatggctgagatcacccgattcctctc
tacaaggttaagctctcctcgaaggcgtgaaggacaatggacttctagaagacttcttg
cagaacaacagatggcatctcgagcaagttccaaccaattgatgacactgaatcccag
accacgtcagtgaaacctcatggccgatgatactgagagcgcgtttgctacacaaacaaat
tcgggagggtcttgacgttgtcggattgatctccatggcetaagagagaagaaggagagcct
aagtttgtaatacaacatctttgtggatctcatcttggaggctctctaccttggtgtg
ggagaaagaggatthttctacactcctaaggctgctaagggaaattggtgaacaatggtg
acttctatthgctcactttaccaattggagaactattgcaactga

SEQ ID N°:4

MADTARGTHHDIIGRDQYPMGRDRDQYQMSGRGSDYKSRQIAKAATAVTTAGGSLLVLSLTLVGTVI
ALTVATPLLVI FSPILVPALITVALLITGFLSSGGFGIAAITVFSWIYATG
EHPQGSCLKDSARMKLGSKAQDLKDRAQYYGQOHTGGEHDRDRTRGGQHTTMAEITRIPLYK GKSLRKA
LKEHGLLEDFLQKQOYGISSKFQPIDDTESQTTSVNLMADDTESAFATQTN
SGGLDVVGLISMAREEGEPK FVNQHLGSHLVEALYLVCGERGFYTPKAAKGIVEQCCTSICSLYQL
ENYCN

SEQ ID N°:5

atgaacttccttaagtctttccctttctacgctttcctttgtttcgggtcaatacttcggtgctggttacg
catgcctttgtaatacaacatctttgtggatctcatcttggtaggctctctaccttgggtgaggagaa
agaggatctttctacactcctaagactagaagaaagagaggaattggtgaacaatggtgcaacttctatt
tgctcactttaccaattggagaactattgcaacagaagaaagagagacattggtgatgacacagctcca
tcctccctggctatgtcagtgaggacagcgggtcactatgctgctgcaagtccagtcagagccttttaaaa
agtaccaatcaaaagaactatttggcctggtaccagcagaaaccaggacagctctcctaaacttctggta
tactttgcatccactagggaaatctggggctccctgctcagcttcataggcagtggtctgggacagatttc
actcttaccatcagcagtggtgcaggctgaagacctggcagattacttctgtcagcaacattataact
cctcccacgcttcggtgctgggaccaagttggagcttaagcggctctccgaacggtgcttctcatagcgg
tctgcaccaggcactagctctgcatctggatctcaggtgcacctgcagcagctctggagctgagctgatg
aagcctggggcctcaatgaagatatcctgcaaggctactggctacacattcagtagctactggatagag
tgggtaaaagcagagggcctggacatggccttgagtggttggagagatcttacctggcagtggttagtact
acctacaatgagaagttcaagggcaaggccacattcactgcagatacatcctccaacacagcctacatg
caactcagcagcctgacatctgaggactctgccgtctattactgtgcaagattggatggtgactcctgg
ggccaaggcaccactctcacagtgagctcaaaggatgagctttga

SEQ ID N°:6

MNFLKSFYAFYFLCFGQYFVAVTHAFVNQHLGSHLVEALYLVCGERGFYTPKTRRKR GIVEQCCTSI
CSLYQLENYCNRRKRDI VMTQSPSSLAMSVGQRV TMRCKSSQSLLKSTNQKNYLAWYQKPGQSPKLLV
YFASTRESGVPDRFIGSGSGTDFTLTISVQAEDLADYFCQOQHYNT PPTFGAGTKLELKRSPNGASHG
SAPGTSSASGSQVHLQOQSGAELMKPGASKIKSCKATGYTFSSYWIEWVKQRP HGLEWIGELLPGSGST
TYNEKFKGKATFTADTSSNTAYMQLSSLTSEDSAVYYCARLDVDSWGQGTTLTVSSKDEL

SEQ ID N°:7

MALWMRLLPLLALLALWGPDPAAAFVNQHLGSHLVEALYLVCGERGFYTPKTRRE
AEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

SEQ ID N°:8

EAEDPQVGEVELGGGPGGLGGLQPLALAGPQQ

SEQ ID N°:9

FVNQHLGSHLVEALYLVCGERGFYTPKAXXEAEDPQVGEVELGGGPGGLGGLQPLALAGPQQXXGIVE
QCCTGICSLYQLENYCN

SEQ ID N°:10

MASLAALLPLLALLVLCRLDPAQAFVNQHLGSHLVEALYLVCGERGFYTPKSRREVEELQVGQ AELG
GGPGAGGLQPSALELALQKRGIVEQCCTSICSLYQLENYCN

SEQ ID N°:11

GIVEQCCTSICSLYQLENYCN

SEQ ID N°:12

GIVEQCCASTCSLYQLENYCN

SEQ ID N°:13
MALWTRLPLLLALLALWAPAPAQAFVNQHLGSHLVEALYLVCGERGFFYTPKARRE
AENPQAGAVELGGGLGGLQALALEGPPQKRGIVEQCCTSICSLYQLENYCN

SEQ ID N°:14
GIVEQCCTGVCSLYQLENYCN

SEQ ID N°:15
MALWTRLRPLLLALLALWPPPPARAFVNQHLGSHLVEALYLVCGERGFFYTPKARRE
VEGPQVGALELAGGPGAGGLEGPPQKRGIVEQCCASVCSLYQLENYCN

SEQ ID N°:16
MALWTRLVPLLLALLALWAPAPAHAFVNQHLGSHLVEALYLVCGERGFFYTPKARRE
VEGPQVGALELAGGPGAGGLEGPPQKRGIVEQCCAGVCSLYQLENYCN

SEQ ID N°:17
GIVEQCCASVCSLYQLENYCN

SEQ ID N°:18
MALWMRLLPLLLALLALWAPAPTRAFVNQHLGSHLVEALYLVCGERGFFYTPKARRE
VEDLQVRDVELAGAPGEGGLQPLALEGALQKRGIVEQCCTSICSLYQLENYCN

SEQ ID N°:19
GIVDQCCTGVCSLYQLQNYCN

SEQ ID N°:20
MALWMHLLPLLLALLALWGPEPAPAFVNQHLGPHLVEALYLVCGERGFFYAPKTRRE
AEDLQVGQVELGGGSITGSLPPELEGPMQKRGVVDQCCTSICSLYQLQNYCN

SEQ ID N°:21
MALWMRLLPLLLALLALWGPDPAFAFVNQHLGSHLVEALYLVCGERGFFYTPKTRRE
AEDPQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

SEQ ID N°:22
MALWMRLLPLLLALLALWGPDVPFAFVNQHLGSHLVEALYLVCGERGFFYTPKTRRE
AEDPQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

SEQ ID N°:23
MALWMRLLPLLVLLALWGPDPAFAFVNQHLGSHLVEALYLVCGERGFFYTPKTRRE
AEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

SEQ ID N°:24
GIVEECCKGVCSMYQLENYCN

SEQ ID N°:25

CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQ
KRGIVEQC

SEQ ID N°:26

MALWMRLLPFLALALWGFDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTEKTRRE
AEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

SEQ ID N°:27

FVNQHLCGPHLVEALYLVCGERGFFYAPKTVVDQCCTSICSLYQLENYCN

SEQ ID N°:28

MTLWMRLLPFLTLLVWEPNPAQAFVNQHLCGSHLVEALYLVCGERGFFYTPKSRRG
VEDPQVAQLELGGGPGADDLQTLALEVAQKRGIVDQCCTSICSLYQLENYCN

SEQ ID N°:29

MALWMRFLPFLALLVWEPKPAQAFVKQHLCGPHLVEALYLVCGERGFFYTPKSRRR
VEDPQVPQLELGGGPEAGDLQTLALEVARQKRGIVDQCCTSICSLYQLENYCN

SEQ ID N°:30

MALWIRFLPFLALLLWEPKPAQAFVKQHLCGSHLVEALYLVCGERGFFYTPMSRRR
VEDPQVAQLELGGGPGAGDLQTLALEVARQKRGIVDQCCTSICSLYQLENYCN

SEQ ID N°:31

GIVDQCCTSICSLYQLENYCN

SEQ ID N°:32

MALIVHFLPFLALLLWEPKPTQAFVKQHLCGPHLVEALYLVCGERGFFYTPKSRRR
VEDPQVEQLELGGSPGDLQTLALEVARQKRGIVDQCCTSICSLYQLENYCN

SEQ ID N°:33

MALWMRFLPFLALLFLWESHPTQAFVKQHLCGSHLVEALYLVCGERGFFYTPMSRRR
VEDPQVAQLELGGGPGAGDLQTLALEVAQKRGIVDQCCTSICSLYQLENYCN

SEQ ID N°:34

GIVDQCCTSICTLYQLENYCN

SEQ ID N°:35

MALWMHLLTVLALLALWGPNTGQAFVSRHLCGSNLVETLYSVCQDDGFFYIPKDRRE
LEDPQVEQTELGMLGAGGLQPLALEMALQKRGIVDQCCTGTCTRHLQSYCN

SEQ ID N°:36

MAPWMHLLTVLALLALWGFNSVQAYSSQHLCGSNLVEALYMTCGRSGFYRPHDRREL
EDLQVEQAEGLGAGGLQPSALEMILQKRGIVDQCCNICTFNQLQNYCNVP

SEQ ID N°:37

GIVEQCCNSICSLYQLETYCN

SEQ ID N°:38

MALWILLPLLALLILWGPDPAQAFVNQHLCGSHLVEALYLVCGERGFFYTPMSRRE
VEDPQVGVQVELGAGPGAGSEQTLALEVARQARIVQQCTSGICSLYQENYCN

SEQ ID N°:39

MALWMRLLPLLAFLLWEPSPAHAFVNQHLCGSHLVEALYLVCGERGFFYTPKFRRG
VDDPQMPQLELGGSPGAGDLRALALEVARQKRGIVEQCCTGICSLYQLENYCN

SEQ ID N°:40

MALWTRLLPLLALLLGGPDPAQAFVNQHLCGSHLVEALYLVCGERGFFYTPKSRRE
VEEQQGGQVELGGGPGAGLPQPLALEMALQKRGIVEQCCTSICSLYQLENYCN

SEQ ID N°:41

ELEDPQVEQTELMGLGAGGLQFLQALQ

SEQ ID N°:42

MALWIRSLPLLALLVFSGPGTSYAAANQHLCGSHLVEALYLVCGERGFFYSPKARRD
VEQPLVSSPLRGEAGVLPFQQEYKVKRGIVEQCCHNTCSLYQLENYCN

SEQ ID N°:43

AANQHLCGSHLVEALYLVCGERGFFYSPKTXDVEQPLVNGPLHGEVGELEPFQHEEY
QXXGIVEQCCENPCSLYQLENYCN

SEQ ID N°:44

GIVEQCCENPCSLYQLENYCN

SEQ ID N°:45

IQSLPLLALLALSFGPGTSHAAVNQHLCGSHLVEALYLVCGERGFFYSPKARRDAEHP
LVNGPLHGEVGDLPFQQEYKVKRGIVEQCCHNTCSLYQLENYCN

SEQ ID N°:46

MAVWLQAGALLVLLVSSVSTNPGTFQHLGSHLVDALYLVCGPTGFFYNPKRDVEP
LLGFLPPKSAQETEVAADFADKHAELIRKRGIVEQCCHKPCSI FELQNYCN

SEQ ID N°:47

MAVWIQAGALLFLLAVSSVNANAGAPQHLCGSHLVDALYLVCGPTGFFYNPKRDVDP
PLGFLPPKSAQETEVAADFADKHAEVIRKRGIVEQCCHKPCSI FELQNYCN

SEQ ID N°:48

GIVEQCCHRPCDIFDLQSYCN

SEQ ID N°:49

GIVEQCCHKPCNIFDLQNYCN

SEQ ID N°:50

GIHZZCCHKPCBIFZLZBYCN

SEQ ID N°:51
MAALWLQSFSLVLLVSVWPGSQAVAPAQHLCGSHLVDALYLVCGDRGFFYNPKRDV
DQLLGFLPPKSGGAAAAGADNEVAEFAFKDQMEMMVKRGIVEQCCHRPCNIFDLQNY
CN

SEQ ID N°:52
MALSPFLAAVIPLVLLSRAPPSADTRTTGHLGSKDLVNALYIACGVRGFFYDPTKM
KRDTGALAAFLPLAYAEDNESQDDESIGINEVLKSKRGIVEQCCHKRCSIYDLENYC
N

SEQ ID N°:53
MAFWLQAASLLVLLALS PGVDAAAQHLCGSHLVDALYLVCGEKGFFYTPKRDVDPL
IGFLSPKSAKENEY PFKDQTEMMVKRGIVEQCCHKPCNIFDLQNYCN

SEQ ID N°:54
GIVEQCCHRPCNIRVLENYCN

SEQ ID N°:55
GIVEQCCHKPCTIYELENYCN

SEQ ID N°:56
GIVEQCCHKPCNIFDLQNYCN

SEQ ID N°:57
GIVEQCCHNTCSLANLEGYCN

SEQ ID N°:58
GIVEHCCHNTCSLYDLEGYCNQ

SEQ ID N°:59
GIVEHCCHNTCSLFDLEGYCN

SEQ ID N°:60
VPTQRLCGSHLVDALYFVCGERGFYSPKQIRDVGPLSAFRDLEPPLDTEMEDRFPY
RQQLAGSKMKRGIVEQCCHNTCSLVNLEGYCN

SEQ ID N°:61
GIVEQCCHRKCSIYDMENYCN

SEQ ID N°:62
GIVEQCCHKPCNIFDLQNYCN

SEQ ID N°:63
GIVEQCCLKPCTIYEMEKYCN

SEQ ID N°:64
GIVEQCCHKPCSI FDLQNYCN

SEQ ID N°:65
MAALWLOAFSLLVLMVSWPGSQAVGGPQHLCGSHLVDALYLVCGRGFFYNPRRDV
DFLLGFLPPKAGGAVVQGGENEVTFKDQMEMMVKRGIVEECCHKPCTI FDLQNYCN

SEQ ID N°:66
GIVEQCCHSPCSLYDLENYCN

SEQ ID N°:67
GIVEQCCHKPCSI FDLQNYCN

SEQ ID N°:68
MAALWLOQSVSLLVLMVSWSGSQAVLPPQHLCGAHLVDALYLVCGERGFFYTPKRDRV
DFLLGFLPAKSGGAAAGGENEVAEFAFKDQMEMMVKRGIVEQCCHKPCNI FDLQNYC
N

SEQ ID N°:69
DVEPLLGFSPKSGQENEVD DFPYKGGQEL

SEQ ID N°:70
MALWMQCLPLVLVLLFFSTPNTEALVNQHLCGSHLVEALYLVCGRGFFYYPKVKRDM
EQALVSGPQDNELDGMQLQPQEYQKMKRGIVEQCCHSTCSLFQLESYCN

SEQ ID N°:71
MALWMQCLPLVLVLLFFSTPNTEALANQHLCGSHLVEALYLVCGRGFFYYPKIKRDI
EQAQVNGPQDNELDGMQFQPQEYQKMKRGIVEQCCHSTCSLFQLENYCN

SEQ ID N°:72
GIVEQCCHNTCSLYQLENYCN

SEQ ID N°:73
GIVEQCCHNTCSLYQLENYCN

SEQ ID N°:74
GIVEQCCENTCSLYELENYCN

SEQ ID N°:75
GIVEQCCENTCSLYQLENYCN

SEQ ID N°:76
MGLWIRLLPLIALLLILWGPDPAAAEFRMFVNQHLCGSHLVEALYLVCGERGFFYTPK
TRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

SEQ ID N°:77
GIVEQCCTSICSLYQLENYCN

SEQ ID N°:78
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID N°:79
FVDQHLCGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSICSLYELEDYCN

SEQ ID N°:80
FVEQHLCGSDLVEALYLVCGERGFFYTPKAAKGIVEQCCTSICSLYQLEEYCN

SEQ ID N°:81
FVQQHLCGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSICSLYQLENYCG

SEQ ID N°:82
FVTQHLCGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSICSLYQLEHYCS

SEQ ID N°:83
NSNGKFVNQHLCGSHLVEALYLVCGERGFFYTPKTKGIVEQCCTSICSLYQLENYCN

SEQ ID N°:84
NSNGKFVNQHLCGSHLVEALYLVCGERGFFYTPKTKRGIVEQCCTSICSLYQLENYC
N

SEQ ID N°:85
FVNQHLCGSHLVEALYLVCGERGFFYTPKGIVEQCCTSICSLYQLENYCN

SEQ ID N°:86
RFVNQHLCGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSICSLYQLENYCN

SEQ ID N°:87
KETLTITCAVPTWLKLTWFAVKEVSSTNLRLRLVLSNNAVPPSAPCTNWKTTATTR
SPQA

SEQ ID N°:88
KDSLNTNTCAVSTWLKLTWFAVKEVSSTLLRLRLVLSNNAVPPSANYTNWKTTATTR
SPQA

SEQ ID N°:89
MFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPL
ALEGSLQKRGIVEQCCTSICSLYQLENYCN

SEQ ID N°:90
RREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKR

SEQ ID N°:91
GPETLCGAELVDALQFVCGDRGF

SEQ ID N°:92
MKLKTVRSVAVLSSLFASQVLGQPIDDTESQTTSVNLMADDTESAFATQTNSSGGLDVV
GLISMAKREEGEPK FVNQHLCGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSIC
SLYQLENYCN

SEQ ID N°:93
tttgtcaatcagcacctttgtggttctcacctggaggctctgtacctgggtgtgt
ggggaacgtggtttcttctacacaccaagaccgctcgtaagcttaagcgtggcatt
gtggagcagtgctgcaccagcatctgctccctctaccaactggagaactactgcaac

SEQ ID N°:94
GIVEQCCTSICSLYQLENYCN

SEQ ID N°:95
FVNQHLCGSHLVEALYLVCGERGFFX

SEQ ID N°:96
GIVEQCCTSICSLYQLENYCN

SEQ ID N°:97
EVNQHLCGSELVEALELVCGERGFFYEPK

SEQ ID N°:98
GIVEQCCTSICSLYQLENYCN

SEQ ID N°:99
FVNQHLCGSHLVEALHLVCGERGFFYTPKT

SEQ ID N°:100
GIVEQCCKSICSLYQLENYCN

SEQ ID N°:101
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID N°:102
GIVEQCCKSICSLYQLENYCN

SEQ ID N°:103
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID N°:104
GIVEQCCXSICSLYQLENYCN

SEQ ID N°:105
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID N°:106
GIVEQCXSICSLYQLENYCN

SEQ ID N°:107
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID N°:108
GXVEQCCTSICSLYQLENYCN

SEQ ID N°:109
FVNQHLCGSDLVEALYLVCGERGFFYTKPT

SEQ ID N°:110
GAVEQCCTSICSLYQLENYCN

SEQ ID N°:111
FVNQHLCGSDLVEALYLVCGERGFFYTKPT

SEQ ID N°:112
GXVEQCCTSICSLYQLENYCN

SEQ ID N°:113
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID N°:114
GXVEQCCTSICSLYQLENYCN

SEQ ID N°:115
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID N°:116
GGGEQCCTSICSLYQLENYCN

SEQ ID N°:117
FVNQHLCGSDLVEALYLVCGERGFFYTKPT

SEQ ID N°:118
XIVEQCCTSICSLYQLENYCN

SEQ ID N°:119
XVNQHLCGDHLVEALYLVCGERGFFYTPKT

SEQ ID N°:120
GIVEQCCTSICSLYQLENYCN

SEQ ID N°:121
FVNQHLCGSHLVEALYLVCGERGFYTPKT

SEQ ID N°:122
GIVEQSCSISSELYQLENYCN

SEQ ID N°:123
FVNQHLCGSDLVEALYLVCGERGFFYTKPT

SEQ ID N°:124
MDPGDPECLEQLLRRLGGSVEVEVTGGTVHVEVSPEDPGDPECLEQLLRRLGGSVEV
EVTGGTVHVEVSPGERGFFYCN

SEQ ID N°:125
MLKEKKYSPDPGDPECLEQLLRRLGGSVEVEVTGGTVHVEVSPEDPGDPECLEQLLR
RLGGSVEVEVTGGTVHVEVSPGERGFFYCN

SEQ ID N°:126
MATXSSTKKTQLQLEHLXLDLQM

SEQ ID N°:127
TMITDSLAVVLQRXDWXPGVTQL

SEQ ID N°:128
NSVLASALALTVAPMAFANSDESPLSHDGYSLHDGVSMYIEALDKFVNQHLCGSHL
VEALYLVCGERGFFYTPKGIVEQCCTSICSLYQLENYCN

SEQ ID N°:129
DTMPAGGGGGGQHLCGPHLVEALY

SEQ ID N°:130
LENYCN

SEQ ID N°:131
MTMITDSLEFQAWGGGGGWMRF

SEQ ID N°:132
MVLRFLLALLVLWEPKPAQA

SEQ ID N°:133
FVNQHLCGSHLVEALYLVCGERGFFYTPKTRRYPGDVKRGIVEQCCTSICSLYQLEN
YCN

SEQ ID N°:134
GIVEQCCKSICSLYQLENYCN

SEQ ID N°:135
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID N°:136
GIVEQCCKSICSLYQLENYCN

SEQ ID N°:137
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID N°:138
LVEALYLVCGERGG

SEQ ID N°:139
GIVEQCCKSICSLYQLENYCN

SEQ ID N°:140
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID N°:141
GIVEQCCKSICSLYQLENYCN

SEQ ID N°:142
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

SEQ ID N°:143
GIVEQCCTSICSLYQLENYCN

SEQ ID N°:144
FVNQHLCGSDLVEALYLVCGERGFFYTDK

SEQ ID N°:145
FVNQHLCGSDLVEALYLVCGERGFFYTDKGVIVEQCCTSICSLYQLENYCN

SEQ ID N°:146
AAK

SEQ ID N°:147
NKR

SEQ ID N°:148
RRKQKR

SEQ ID N°:149
RRKR

SEQ ID N°:150
KDEL

SEQ ID N°:151
HDEL

SEQ ID N°:152

DDEL

SEQ ID N°:153

ADEL

SEQ ID N°:154

SDEL

SEQ ID N°:155

HDEF

SEQ ID N°:156

Met Ala Asp Thr Ala Arg Gly Thr His His Asp Ile Ile Gly Arg Asp
1 5 10 15

Gln Tyr Pro Met Met Gly Arg Asp Arg Asp Gln Tyr Gln Met Ser Gly
20 25 30

Arg Gly Ser Asp Tyr Ser Lys Ser Arg Gln Ile Ala Lys Ala Ala Thr
35 40 45

Ala Val Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Ser Leu Thr Leu
50 55 60

Val Gly Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile
65 70 75 80

Phe Ser Pro Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Leu Ile
85 90 95

Thr Gly Phe Leu Ser Ser Gly Gly Phe Gly Ile Ala Ala Ile Thr Val
100 105 110

Phe Ser Trp Ile Tyr Lys
115

SEQ ID N°:157

Met Ala Asp Thr Ala Arg Thr His His Asp Val Thr Ser Arg Asp Gln
1 5 10 15

Tyr Pro Arg Asp Arg Asp Gln Tyr Ser Met Ile Gly Arg Asp Arg Asp
20 25 30

Gln Tyr Ser Met Met Gly Arg Asp Arg Asp Gln Tyr Asn Met Tyr Gly
35 40 45

Arg Asp Tyr Ser Lys Ser Arg Gln Ile Ala Lys Ala Val Thr Ala Val
50 55 60

Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Ser Leu Thr Leu Val Gly
65 70 75 80

Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile Phe Ser
85 90 95

Pro Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Leu Ile Thr Gly
100 105 110

Phe Leu Ser Ser Gly Gly Phe Ala Ile Ala Ala Ile Thr Val Phe Ser
115 120 125

Trp Ile Tyr Lys Tyr Ala Thr Gly Glu His Pro Gln Gly Ser Asp Lys
130 135 140

Leu Asp Ser Ala Arg Met Lys Leu Gly Thr Lys Ala Gln Asp Ile Lys
145 150 155 160

Asp Arg Ala Gln Tyr Tyr Gly Gln Gln His Thr Gly Gly Glu His Asp
165 170 175

Arg Asp Arg Thr Arg Gly Gly Gln His Thr Thr
180 185

SEQ ID N°:158

taccatgggg tcaaagacgg agatgatgga gagagacgca atggctacgg tggctcccta 60
tgcgccggtc acttaccatc gccgtgctcg tgttgacttg gatgatagac ttcctaaacc 120
ttatatgcca agagcattgc aagcaccaga cagagaacac ccgtacggaa ctccaggcca 180
taagaattac ggacttagtg ttcttcaaca gcatgtctcc ttcttcgata tcgatgataa 240
tggcatcatt tacccttggg agacctactc tggactgcca atgcttggtt tcaatatcat 300
tgggtcgctt ataatagccg ctgttatcaa cctgaccctt agctatgcca ctcttccggg 360
gtggttacct tcacctttct tccctatata catacacaac atacacaagt caaagcatgg 420
aagtgattca aaaacatatg acaatgaagg aaggtttatg ccggtgaatc ttgagttgat 480
atttagcaaa tatgcgaaaa ccttgccaga caagttgagt cttggagaac tatgggagat 540
gacagaagga aaccgtgacg cttgggacat ttttgatgg atcgcaggca aaatagagtg 600
gggactgttg tacttgctag caagggatga agaagggtt ttgtcaaaag aagctattag 660
gcgggtgttc gatggaagct tgttcgagta ctgtgccaaa atctacgctg gtatcagtga 720
agacaagaca gcatactacg ccatggat 748

SEQ ID N°:159

atgggggtcaa agacggagat gatggagaga gacgcaatgg ctacggtggc tccctatgcg 60
ccggtcactt accaccgccg tgctcgtggt gacttggatg atagacttcc taaaccttat 120
atgccaagag cattgcaagc accagacaga gaacaccogt acggaactcc aggcataag 180
aattacggac ttagtgttct tcaacagcat gtctccttct tcgatatcga tgataatggc 240
atcatttacc cttgggagac ctactctgga ctgccaatgc ttggtttcaa tatcattggg 300
tcgcttataa tagccgctgt tatcaacctg acccttagct atgccactct tccgggggtg 360
ttaccttac ctttcttccc tatatacata cacaacatac acaagtcaaa gcatggaagt 420
gattcaaaaa catatgacaa tgaaggaag tttatgccg tgaatcttga gttgatattt 480
agcaaatatg cgaaaacctt gccagacaag ttgagtcttg gagaactatg ggagatgaca 540
gaaggaaacc gtgacgcttg ggacatthttt ggatggatcg caggcaaat agagtgggga 600
ctgttgact tgctagcaag gtagaagaa gggttttgt caaagaagc tattaggcgg 660
tgtttcgatg gaagcttgtt cgagtactgt gccaaatct acgctggtat cagtgaagac 720
aagacagcat actactaa 738

SEQ ID N°:160

ATGGATCTAATCCACACTTTCTCAACTTAATAGCTCCCCCTTTCACCTTCTTCTTC
CTTCTCTTTTTCTTGCCACCCTTCCAGATTTTCAAGTTCTTCTTTCAATCTTGGGC
ACCCTTTTCAGCGAGGATGTCGCTGGAAAAGTCGTCGTCATCACCGGCGCCTCCTCC
GGCATCGGCGAAAGTCTTGCTTACGAGTATGCTAAGAGAGGGGCGTGCTTGGTGCTT
GCTGCAAGAAGGGAAAGGAGTCTTCAAGAAGTGGCCGAAAGGGCGCGGATTTGGGG
TCGCCGACGTCGTGGTGGTCCGGGCCGATGTTTCAAGGCGGAGGACTGCAGGAAG
GTTGTTGATCAGACTATGAATCGCTTTGGAAGATTGGATCACCTGGTCAATAACGCT
GGAATTATGTCAGTTTCAATGCTGGAAGAAGTTGAAGATATTACTGGTTACAGAGAA
ACTATGGATATCAACTTCTGGGGCTATGTGTATATGACCCGATTTGCCGCCCCATAC
CTTAGGAATAGCAGAGGCCGAATTGTTGTACTTTCTTCATCCAGTTCTTGGATGCCT
ACTCCGAGGATGAGTTTTTACAATGCAAGCAAAGCGGCGATTTCACAATTTTTTGAG
ACACTGCGGGTGAATTCGGCCCCGATATAGGCATAACCCTTGTGACTCCAGGATTC
ATAGAATCTGAACTTACCCAAGGCAAATCTACAATGCTGGCGAACGTGTAATTGAT
CAGGACATGAGAGATGTACAAGTGAGCACGACTCCAATCCTGAGGGTGGAAAGTGCG
GCAAGGTCAATCGTGAGGAGCGCGATCCGTGGAGAAAGATACGTGACAGAGCCGGCC
TGGTTTAGGGTTACTTATTGGTGGAAAGCTATTCTGCCCTGAGGTGATGGAGTGGTA
TTTAGACTGATGTACTTGGCCAGCCGGGTGAGCCGGAGAAGGAAACGTTTGGCAAG
AAGTTTTTGGATTACACAGGAGTGAAGTCCTTGCTTTACCCGAAACCGTGCAAGTT
CCGGAGCCCAAGAATGATTAA

SEQ ID N°:161

MNFLKSFPHYAFLCFGQYFVAVTHA

SEQ ID N°:162

APVNTTEDETAQAEAVIGYSDLEGDFDVAVL PFSNSTNGLLFIBTTIASIAAKEEG
VSLMAKR

SEQ ID N°:163
APVNTTEDETAQAEAVIGYSOLEGDFDVAVLPFSNSTNGLLFIBTTIASIAAKEEG
VSMAKR

SEQ ID N°:164
QPIDEDNDTSSMAKR

SEQ ID N°:165
QPIDDTESNTTSVNLMADDTEDRFATNTTLALDVVNLI SMAKR

SEQ ID N°:166
QPIDDTESQTTSVNLMADDTEDRFATQTTLALDVVNLI SMAKR

SEQ ID N°:167
QPIDDTESQTTSVNLMADDTEDRFATQTTLALDVVNLI SMAAA

SEQ ID N°:168
QPIDDTESNTTSVNLMADDTEDRFATNTTLALDVVNLI SMAAA

SEQ ID N°:169
QPIDDTESNTTSVNLMADDTEDRFATNTTLAGGLDVVNLI SMAKR

SEQ ID N°:170
QPIDDTESQTTSVNLMADDTESAFATQTNSSGLDVVGLI SMAKR

SEQ ID N°:171
QPIDDTESQTTSVNLMADDTESAFATQTNSSGLDVVGLI SMAAA

SEQ ID N°:172
EEAEAEAEPK

SEQ ID N°:173
EEGEPK

SEQ ID N°:174
MAEITRIPLYKGKSLRKALKEHGLEDFLQKQYGISSKF

SEQ ID N°:175
GCATGCTGACATTGTGATGACACAGTC

SEQ ID N°:176
AAGCTTGCAFTTAAATACTCGAGACTGTGAGAGTGGTGCCTTG

SEQ ID N°:177
GAAGAAGGAGAGCCCTAAGTTTGTTAATCAACATCTTTGTGGATCTCATCTTGTTGAG
GCTCTCTACCTTG

SEQ ID N°:178
CCTTAGGAGTGTAGAAAAATCCTCTTTCTCCACACACAAGGTAGAGAGCCTCAACA

SEQ ID N°:179
CTAAGGCTGCTAAGGGAATTG

SEQ ID N°:180
AAGCTTCAGTTGCAATAGTTCTCCAATTGGTAAAGTGAGCAAATAGAAGTGCAACAT
TGTTCAACAATTCCTTAGCAGCCTT

SEQ ID N°:181
CTCGAGTCAACCAATTGATGACACTGAATC

SEQ ID N°:182
AAGCTTCAAAGTTCATCCTTGTTGCAATAGTTCTCCAATTG

SEQ ID N°:183
AAGCTTCAGTTGCAATAGTTC

SEQ ID N°:184
GCATGCCCAACCAATTGATGACACTG

SEQ ID N°:185
GCATGCATGCCTTTGTTAATCAACATCTTTGTGG

SEQ ID N°:186
ACATFGTTCAACAATTCCTCTCTTTCTTCTAGTCTTAGGAGTGTAGAAAAATCC

SEQ ID N°:187
GCATAAGCTTCAAAGCTCATCCTTTGAGC

SEQ ID N°:188
ATGAACCTCCTFAAGTCTTTCCCTTTCTACGCTTTCTTTGTTTCGGTCAATACTTCGTTGCT
GTTACGCATGCCCAACCAATTGATGACACTGAATCCCAGACCACGTCAGTGAACCTCATG
GCCGATGATACTGAGAGCGCGTTTGCTACACAAACAAATTCGGGAGGTCCTTGACGTTGTC
GGATTGATCTCCATGGCTAAGAGAGAAGAAGGAGAGCCTAAGTTTGTTAATCAACATCTT
TGTGGATCTCATCTTGTTGAGGCTCTCTACCTTGTGTGGAGAAAAGAGGATTTTTCTACA
CTCCTAAGGCTGCTAAGGGAATTGTTGAACAATGTTGCACTTCTATTTGCTCACTTTACCA
ATTGGAGAACTATTGCAACTGA

SEQ ID N°:189
MNFLKSFFFYAFLCFGQYFVAVTHAQPIDDTESQTTSVNLMADDTESAFATQTNSSGLDVVGL
ISMAREEGEPKFVNQHLGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSICSLYQLENY
CN

SEQ ID N°:190
TTCGTGAACCAACACTTG

SEQ ID N°:191
AAGCTTTCAGTTACAGTAGT

SEQ ID N°:192
GCATGCATGTGTTGAGC

SEQ ID N°:193
GGTAGTGTGCTGGCCA

SEQ ID N°:194
GGTGGCCAGCACACTACCTTCGTGAACCAACACTTGTG

SEQ ID N°:195
ATGGCGGATACAGCTAGAGGAACCCATCACGATATCATCGGCAGAGACCAGTACCCGATG
ATGGGCCGAGACCGAGACCAGTACCAGATGTCCGGACGAGGATCTGACTACTCCAAGTCT
AGGCAGATTGCTAAAGCTGCAACTGCTGTACAGCTGGTGGTTCCCTCCTTGTCTCTCCA
GCCTTACCCTTGTGGAACTGTACATAGCTTTGACTGTTGCAACACCTCTGCTCGTTATCTTC
AGCCCAATCCTTGTCCCGCTCTCATCACAGTTGCACTCCTCATCAACCGTTTTCTTTCCTC
TGGAGGGTTTGGCATTGCCGCIATAACCGTTTTCTCTTGGATTTACAAGTAAGCACACATT
TATCATCTTACTTCATAATTTTTGTGCAATATGTGCATGCATGTGTTGAGCCAGTAGCTTTGG
ATCAATTTTTTTGGTAGAATAACAAATGTAACAATAAGAAATGCAAATTCTAGGGAACA
TTTGGTFAACTAAATACGAAATTTGACCTAGCTAGCTTGAATGTGTCTGTATATCATCT
ATATAGGTAAAATGCTTGGTATGATACCTATTGATTGTGAATAGGTACGCAACGGGAGAG
CACCCACAGGGATCAGACAAGTTGGACAGTGCAAGGATGAAGTTGGGAAGCAAAGCTCA
GGATCTGAAAGACAGAGCTCAGTACTACGGACAGCAACATACTGGTGGGGAACATGACC
GTGACCGTACTCGTGGTGGCCAGCACACTACCTTCGTGAACCAACACTTGTGTGGATCTCA
TCTCGTTGAAGCTCTCTACTTGGTTTGTGGTGAGAGAGGATTCTTCTACTCCTAAGACC
AGAAGGGAAGCTGAGGACTTGCAGGTGGGACAAGTTGAGTTGGGTGGAGGTCTGGAGC
AGGATCTTTGCAACCTCTCGCTTTGGAAGGTTCTTTGCAGAAGAGAGGAATCGTTGAACA
ATGTTGCACTTCAATCTGTTCTTGTATCAGTTGGAGA ACTACTGTA ACTGA

SEQ ID N°:196
MADTARGTHHDIIGRDQYPMGRDRDQYQMSGRGSDYSKSRQLAKAATAVTAAGSLLVLSS
LTLVGTVIALTVATPLLVI FSPILVPALITVALLITGFLSSGGFGIAAITVFSWIYATGEHPQSDK
LDSARMKLGSKAQDLKDRAQYYGQHTGGEHDRDRTRGGQHTTFVNQHLCGSHLVEALYL
VCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKR GIVEQCCTSICSLYQ
LENYCN

LISTAGEM DE SEQUÊNCIAS

<110> SemBioSys Genetics Inc.

Moloney, Maurice M.

Boothe, Joseph

Keon, Richard

Nykiforuk, Cory

Van Rooijen, Gijs

<120> Métodos para a Produção de Insulina em Plantas

<130> 9369-296

<150> Documento 60/478818

<151> 2003-06-17

<150> Documento 60/549539

<151> 2004-03-04

<160> 196

<170> PatentIn versão 1.0

<210> 1

<211> 1143

<212> ADN

<213> Sequência Artificial

<220>

<223> Sequência de ácido nucleico da proteína de fusão da
insulina

<400> 1

```
atgaacttcc ttaagtcttt ccctttctac gctttccttt gtttcggtca atacttcgtt      60
gctgttacgc atgtgacat tgtgatgaca cagtctccat cctccctggc tatgtcagtg      120
ggacagcggg tcactatgcg ctgcaagtcc agtcagagcc ttttaaaaag taccaatcaa      180
aagaactatt tggcctggta ccagcagaaa ccaggacagt ctcctaaact tctggtatac      240
tttgcataca ctagggaatc tggggtcctt gatcgcttca taggcagtgg atctgggaca      300
gatttcactc ttaccatcag cagtgtgcag gctgaagacc tggcagatta cttctgtcag      360
caacattata aactcctcc cacgttcggt gctgggacca agctggagct taagcggctc      420
ccgaacggtg cttctcctag cggttctgca ccaggcacta gctctgcatc tggatctcag      480
gtgcacctgc agcagtctgg agctgagctg atgaagcctg gggcctcaat gaagatatcc      540
tgcaaggcta ctggctacac attcagtagc tactggatag agtgggtaaa gcagaggcct      600
ggacatggcc ttgagtggat tggagagatt ttacctggca gtggtagtac tacctacaat      660
gagaagttca agggcaaggc cacattcact gcagatacat cctccaacac agcctacatg      720
caactcagca gcctgacatc tgaggactct gccgtctatt actgtgcaag attggatggt      780
gactcctggg gcccaaggcac cactctcaca gtctogagtc aaccaattga tgacactgaa      840
tcccagacca cgtcagtgaa cctcatggcc gatgatactg agagcgcggt tgctacacaa      900
acaaattcgg gaggtcttga cgttgtcggg ttgatctcca tggctaagag agaagaagga      960
gagcctaagt ttgttaatca acatctttgt ggatctcatc ttgttgagge tctctacctt     1020
gtgtgtggag aaagaggatt tttctacact cctaaggctg ctaagggaat tgttgaacaa     1080
tgttgcactt ctatttgctc actttaccaa ttggagaact attgcaacaa ggatgaactt     1140
tga                                                                                   1143
```

<210> 2

<211> 380

<212> PRT

<213> Sequência Artificial

<220>

<223> Proteína de fusão da insulina

<400> 2

```
Met Asn Phe Leu Lys Ser Phe Pro Phe Tyr Ala Phe Leu Cys Phe Gly
 1          5          10          15

Gln Tyr Phe Val Ala Val Thr His Ala Asp Ile Val Met Thr Gln Ser
      20          25          30

Pro Ser Ser Leu Ala Met Ser Val Gly Gln Arg Val Thr Met Arg Cys
      35          40          45

Lys Ser Ser Gln Ser Leu Leu Lys Ser Thr Asn Gln Lys Asn Tyr Leu
 50          55          60

Ala Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Lys Leu Leu Val Tyr
 65          70          75          80

Phe Ala Ser Thr Arg Glu Ser Gly Val Pro Asp Arg Phe Ile Gly Ser
      85          90          95

Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Val Gln Ala Glu
      100          105          110

Asp Leu Ala Asp Tyr Phe Cys Gln Gln His Tyr Asn Thr Pro Pro Thr
      115          120          125

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

Ser His Ser Gly Ser Ala Pro Gly Thr Ser Ser Ala Ser Gly Ser Gln
 145 150 155 160

Val His Leu Gln Gln Ser Gly Ala Glu Leu Met Lys Pro Gly Ala Ser
 165 170 175

Met Lys Ile Ser Cys Lys Ala Thr Gly Tyr Thr Phe Ser Ser Tyr Trp
 180 185 190

Ile Glu Trp Val Lys Gln Arg Pro Gly His Gly Leu Glu Trp Ile Gly
 195 200 205

Glu Ile Leu Pro Gly Ser Gly Ser Thr Thr Tyr Asn Glu Lys Phe Lys
 210 215 220

Gly Lys Ala Thr Phe Thr Ala Asp Thr Ser Ser Asn Thr Ala Tyr Met
 225 230 235 240

Gln Leu Ser Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Tyr Cys Ala
 245 250 255

Arg Leu Asp Val Asp Ser Trp Gly Gln Gly Thr Thr Leu Thr Val Ser
 260 265 270

Ser Gln Pro Ile Asp Asp Thr Glu Ser Gln Thr Thr Ser Val Asn Leu
 275 280 285

Met Ala Asp Asp Thr Glu Ser Ala Phe Ala Thr Gln Thr Asn Ser Gly
 290 295 300

Gly Leu Asp Val Val Gly Leu Ile Ser Met Ala Lys Arg Glu Glu Gly
 305 310 315 320

Glu Pro Lys Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu
 325 330 335

Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys
 340 345 350

Ala Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu
 355 360 365

Tyr Gln Leu Glu Asn Tyr Cys Asn Lys Asp Glu Leu
 370 375 380

<210> 3

<211> 945

<212> ADN

<213> Sequência Artificial

<220>

<223> Sequência de ácido nucleico de proteína de fusão da
 insulina

<400> 3

```

atggcggata cagctagagg aaccatcac gatatcatcg gcagagacca gtacccgatg      60
atgggccgag accgagacca gtaccagatg tccggacgag gatctgacta ctccaagtct      120
aggcagattg ctaaagctgc aactgctgtc acagctgggtg gttccctcct tgttctctcc      180
agccttacc ttgttgaac tgtcatagct ttgactgttg caacacctct gctcgttatc      240
ttcagcccaa tccttgtccc ggctctcatc acagttgcac tcctcatcac cggttttctt      300
tcctctggag ggtttggcat tgccgtata accgttttct cttggattta cgcaacggga      360
gagcaccac agggatcaga caagttggac agtgcaagga tgaagttggg aagcaaagct      420
caggatctga aagacagagc tcagtactac ggacagcaac atactgggtg ggaacatgac      480
cgtgaccgta ctcgtgggtg ccagcacact accatggctg agatcaccgg cattcctctc      540
taciaaggta agtctctccg taaggcgctg aaggaacatg gaottctaga agacttcttg      600
cagaaacaac agtatggcat ctcgagcaag ttccaaccaa ttgatgacac tgaatcccag      660
accacgtcag tgaacctcat ggccgatgat actgagagcg cgtttgctac acaaacaat      720
tcgggaggtc ttgacgttgt cggattgatc tccatggcta agagagaaga aggagagcct      780
aagtttgta atcaacatct ttgtggatct catcttggtg aggctctcta ccttgtgtgt      840
ggagaaagag gatTTTTtcta cactcctaag gctgctaagg gaattgttga acaatgttgc      900
acttctattt gctcacttta ccaattggag aactattgca actga                        945

```

<210> 4

<211> 314

<212> PRT

<213> Sequência Artificial

<220>

<223> Proteína de fusão da insulina

<400> 4

```
Met Ala Asp Thr Ala Arg Gly Thr His His Asp Ile Ile Gly Arg Asp
1           5           10           15

Gln Tyr Pro Met Met Gly Arg Asp Arg Asp Gln Tyr Gln Met Ser Gly
20           25           30

Arg Gly Ser Asp Tyr Ser Lys Ser Arg Gln Ile Ala Lys Ala Ala Thr
35           40           45

Ala Val Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Ser Leu Thr Leu
50           55           60

Val Gly Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile
65           70           75           80

Phe Ser Pro Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Leu Ile
85           90           95

Thr Gly Phe Leu Ser Ser Gly Gly Phe Gly Ile Ala Ala Ile Thr Val
100          105          110

Phe Ser Trp Ile Tyr Ala Thr Gly Glu His Pro Gln Gly Ser Asp Lys
115          120          125

Leu Asp Ser Ala Arg Met Lys Leu Gly Ser Lys Ala Gln Asp Leu Lys
130          135          140
```

Asp Arg Ala Gln Tyr Tyr Gly Gln Gln His Thr Gly Gly Glu His Asp
145 150 155 160

Arg Asp Arg Thr Arg Gly Gly Gln His Thr Thr Met Ala Glu Ile Thr
165 170 175

Arg Ile Pro Leu Tyr Lys Gly Lys Ser Leu Arg Lys Ala Leu Lys Glu
180 185 190

His Gly Leu Leu Glu Asp Phe Leu Gln Lys Gln Gln Tyr Gly Ile Ser
195 200 205

Ser Lys Phe Gln Pro Ile Asp Asp Thr Glu Ser Gln Thr Thr Ser Val
210 215 220

Asn Leu Met Ala Asp Asp Thr Glu Ser Ala Phe Ala Thr Gln Thr Asn
225 230 235 240

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

Glu Gly Glu Pro Lys Phe Val Asn Gln His Leu Cys Gly Ser His Leu
260 265 270

Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr
275 280 285

Pro Lys Ala Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys
290 295 300

Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
305 310

<210> 5

<211> 1011

<212> ADN

<213> Sequência Artificial

<220>

<223> Sequência de ácido nucleico de proteína de fusão da
insulina

<400> 5

```
atgaacttcc ttaagtcttt ccctttctac gctttccttt gtttcggtca atacttcggt      60
gctgttacgc atgcctttgt taatcaacat ctttgtggat ctcatcttgt tgaggctctc      120
taccttgtgt gtggagaaag aggatthttc tacactccta agactagaag aaagagagga      180
attggtgaac aatggtgcac ttctatttgc tcactttacc aattggagaa ctattgcaac      240
agaagaaaga gagacattgt gatgacacag tctccatcct ccctggctat gtcagtggga      300
cagcgggtca ctatgcgctg caagtccagt cagagccttt taaaaagtac caatcaaaag      360
aactatthgg cctggtacca gcagaaacca ggacagtctc ctaaacttct ggtatacttt      420
gcatccacta gggaatctgg ggtccctgat cgcttcatag gcagtggatc tgggacagat      480
ttcactctta ccatcagcag tgtgcaggct gaagacctgg cagattactt ctgtcagcaa      540
cattataaca ctccctccac gttcgggtgct gggaccaagt tggagcttaa gcggtctccg      600
aacggtgctt ctcatagcgg ttctgcacca ggcactagct ctgcatctgg atctcaggtg      660
cacctgcagc agtctggagc tgagctgatg aagcctgggg cctcaatgaa gatatcctgc      720
aaggctactg gctacacatt cagtagctac tggatagagt gggtaaagca gaggcctgga      780
catggccttg agtggattgg agagatthta cctggcagtg gtagtactac ctacaatgag      840
aagttcaagg gcaaggccac attcactgca gatacatcct ccaacacagc ctacatgcaa      900
ctcagcagcc tgacatctga ggactctgcc gtctattact gtgcaagatt ggatggtgac      960
tcctggggcc aaggcaccac tctcactgtg agctcaaagg atgagctttg a      1011
```

<210> 6

<211> 366

<212> PRT

<213> Sequência Artificial

<220>

<223> Proteína de fusão da insulina

<400>

Met Asn Phe Leu Lys Ser Phe Pro Phe Tyr Ala Phe Leu Cys Phe Gly
1 5 10 15

Gln Tyr Phe Val Ala Val Thr His Ala Phe Val Asn Gln His Leu Cys
20 25 30

Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly
35 40 45

Phe Phe Tyr Thr Pro Lys Thr Arg Arg Lys Arg Gly Ile Val Glu Gln
50 55 60

Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
65 70 75 80

Arg Arg Lys Arg Asp Ile Val Met Thr Gln Ser Pro Ser Ser Leu Ala
85 90 95

Met Ser Val Gly Gln Arg Val Thr Met Arg Cys Lys Ser Ser Gln Ser
100 105 110

Leu Leu Lys Ser Thr Asn Gln Lys Asn Tyr Leu Ala Trp Tyr Gln Gln
115 120 125

Lys Pro Gly Gln Ser Pro Lys Leu Leu Val Tyr Phe Ala Ser Thr Arg
130 135 140

Glu Ser Gly Val Pro Asp Arg Phe Ile Gly Ser Gly Ser Gly Thr Asp
145 150 155 160

Phe Thr Leu Thr Ile Ser Ser Val Gln Ala Glu Asp Leu Ala Asp Tyr
165 170 175

Phe Cys Gln Gln His Tyr Asn Thr Pro Pro Thr Phe Gly Ala Gly Thr
180 185 190

Lys Leu Glu Leu Lys Arg Ser Pro Asn Gly Ala Ser His Ser Gly Ser
195 200 205

Ala Pro Gly Thr Ser Ser Ala Ser Gly Ser Gln Val His Leu Gln Gln
210 215 220

Ser Gly Ala Glu Leu Met Lys Pro Gly Ala Ser Met Lys Ile Ser Cys
225 230 235 240

Lys Ala Thr Gly Tyr Thr Phe Ser Ser Tyr Trp Ile Glu Trp Val Lys
245 250 255

Gln Arg Pro Gly His Gly Leu Glu Trp Ile Gly Glu Ile Leu Pro Gly
260 265 270

Ser Gly Ser Thr Thr Tyr Asn Glu Lys Phe Lys Gly Lys Ala Thr Phe
275 285

Thr Ala Asp Thr Ser Ser Asn Thr Ala Tyr Met Gln Leu Ser Ser Leu
290 295 300

Thr Ser Glu Asp Ser Ala Val Tyr Tyr Cys Ala Arg Leu Asp Val Asp
305 310 315 320

Ser Trp Gly Gln Gly Thr Thr Leu Thr Val Ser Ser Lys Asp Glu Leu
325 330 335

<210> 7

<211> 110

<212> PRT

<213> *Homo sapiens*

<400> 7

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Gly Pro Asp Pro Ala Ala Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln Val Gly
50 55 60

Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu
65 70 75 80

Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 8

<211> 31

<212> PRT

<213> *Equus przewalskii*

<400> 8

Glu Ala Glu Asp Pro Gln Val Gly Glu Val Glu Leu Gly Gly Gly Pro
1 5 10 15

Gly Leu Gly Gly Leu Gln Pro Leu Ala Leu Ala Gly Pro Gln Gln
20 25 30

<210> 9

<211> 86

<212> PRT

<213> *Equus caballus*

<220>

<221> CARACTERÍSTICA_MISC

<222> (31)..(32)

<223> X = qualquer aminoácido

<220>

<221> CARACTERÍSTICA_MISC

<222> (64)..(65)

<223> X = qualquer aminoácido

<400> 9

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Ala Xaa Xaa
20 25 30

Glu Ala Glu Asp Pro Gln Val Gly Glu Val Glu Leu Gly Gly Gly Pro
35 40 45

Gly Leu Gly Gly Leu Gln Pro Leu Ala Leu Ala Gly Pro Gln Gln Xaa
50 55 60

Xaa Gly Ile Val Glu Gln Cys Cys Thr Gly Ile Cys Ser Leu Tyr Gln
65 70 75 80

Leu Glu Asn Tyr Cys Asn
85

<210> 10

<211> 110

<212> PRT

<213> *Oryctolagus cuniculus*

<400> 10

Met Ala Ser Leu Ala Ala Leu Leu Pro Leu Leu Ala Leu Leu Val Leu
1 5 10 15

Cys Arg Leu Asp Pro Ala Gln Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ser Arg Arg Glu Val Glu Glu Leu Gln Val Gly
50 55 60

Gln Ala Glu Leu Gly Gly Gly Pro Gly Ala Gly Gly Leu Gln Pro Ser
65 70 75 80

Ala Leu Glu Leu Ala Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 11

<211> 21

<212> PRT

<213> *Balaenoptera physalus*

<400> 11

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 12

<211> 21

<212> PRT

<213> *Balaenoptera borealis*

<400> 12

Gly Ile Val Glu Gln Cys Cys Ala Ser Thr Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 13

<211> 108

<212> PRT

<213> *Sus scrofa*

<400> 13

Met Ala Leu Trp Thr Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Ala Pro Ala Pro Ala Gln Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ala Arg Arg Glu Ala Glu Asn Pro Gln Ala Gly
50 55 60

Ala Val Glu Leu Gly Gly Gly Leu Gly Gly Leu Gln Ala Leu Ala Leu
65 70 75 80

Glu Gly Pro Pro Gln Lys Arg Gly Ile Val Glu Gln Cys Cys Thr Ser
85 90 95

Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105

<210> 14

<211> 21

<212> PRT

<213> *Elephas maximus*

<400> 14

Gly Ile Val Glu Gln Cys Cys Thr Gly Val Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 15

<211> 105

<212> PRT

<213> *Bos taurus*

<400> 15

Met Ala Leu Trp Thr Arg Leu Arg Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Pro Pro Pro Pro Ala Arg Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ala Arg Arg Glu Val Glu Gly Pro Gln Val Gly
50 55 60

Ala Leu Glu Leu Ala Gly Gly Pro Gly Ala Gly Gly Leu Glu Gly Pro
65 70 75 80

Pro Gln Lys Arg Gly Ile Val Glu Gln Cys Cys Ala Ser Val Cys Ser
85 90 95

Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105

<210> 16

<211> 105

<212> PRT

<213> *Ovis aries*

<400> 16

Met Ala Leu Trp Thr Arg Leu Val Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Ala Pro Ala Pro Ala His Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ala Arg Arg Glu Val Glu Gly Pro Gln Val Gly
50 55 60

Ala Leu Glu Leu Ala Gly Gly Pro Gly Ala Gly Gly Leu Glu Gly Pro
65 70 75 80

Pro Gln Lys Arg Gly Ile Val Glu Gln Cys Cys Ala Gly Val Cys Ser
85 90 95

Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105

<210> 17

<211> 21

<212> PRT

<213> *Camelus dromedaries*

<400> 17

Gly Ile Val Glu Gln Cys Cys Ala Ser Val Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 18

<211> 110

<212> PRT

<213> *Canis sp.*

<400> 18

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Ala Pro Ala Pro Thr Arg Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ala Arg Arg Glu Val Glu Asp Leu Gln Val Arg
50 55 60

Asp Val Glu Leu Ala Gly Ala Pro Gly Glu Gly Gly Leu Gln Pro Leu
65 70 75 80

Ala Leu Glu Gly Ala Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 19

<211> 21

<212> PRT

<213> *Histrix cristata*

<400> 19

Gly Ile Val Asp Gln Cys Cys Thr Gly Val Cys Ser Leu Tyr Gln Leu
1 5 10 15

Gln Asn Tyr Cys Asn
20

<210> 20

<211> 110

<212> PRT

<213> *Aotus trivirgatus*

<400> 20

Met Ala Leu Trp Met His Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Gly Pro Glu Pro Ala Pro Ala Phe Val Asn Gln His Leu Cys Gly
 20 25 30

Pro His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
 35 40 45

Phe Tyr Ala Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln Val Gly
 50 55 60

Gln Val Glu Leu Gly Gly Gly Ser Ile Thr Gly Ser Leu Pro Pro Leu
65 70 75 80

Glu Gly Pro Met Gln Lys Arg Gly Val Val Asp Gln Cys Cys Thr Ser
 85 90 95

Ile Cys Ser Leu Tyr Gln Leu Gln Asn Tyr Cys Asn
 100 105

<210> 21

<211> 110

<212> PRT

<213> *Macaca fascicularis*

<400> 21

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Gly Pro Asp Pro Ala Pro Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Pro Gln Val Gly
50 55 60

Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu
65 70 75 80

Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 22

<211> 110

<212> PRT

<213> *Oryctolagus cuniculus*

<400> 22

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Gly Pro Asp Pro Val Pro Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Pro Gln Val Gly
50 55 60

Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu
65 70 75 80

Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 23

<211> 110

<212> PRT

<213> *Pan troglodytes*

<400> 23

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Val Leu Leu Ala Leu
1 5 10 15

Trp Gly Pro Asp Pro Ala Ser Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln Val Gly
50 55 60

Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu
65 70 75 80

Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 24

<211> 21

<212> PRT

<213> *Ornytorhynchus anatinus*

<400> 24

Gly Ile Val Glu Glu Cys Cys Lys Gly Val Cys Ser Met Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 25

<211> 65

<212> PRT

<213> *Pongo pygmaeus*

<400> 25

Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg
1 5 10 15

Gly Phe Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln
20 25 30

Val Gly Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln
35 40 45

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

Cys
65

<210> 26

<211> 110

<212> PRT

<213> *Gorilla gorilla*

<400> 26

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Gly Pro Asp Pro Ala Ala Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln Val Gly
50 55 60

Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu
65 70 75 80

Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 27

<211> 51

<212> PRT

<213> *Saimiri sciureus*

<400> 27

Phe Val Asn Gln His Leu Cys Gly Pro His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Ala Pro Lys Thr Gly Val
20 25 30

Val Asp Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Gln Asn
35 40 45

Tyr Cys Asn
50

<210> 28

<211> 110

<212> PRT

<213> *Cricetulus longicaudatus*

<400> 28

Met Thr Leu Trp Met Arg Leu Leu Pro Leu Leu Thr Leu Leu Val Leu
1 5 10 15

Trp Glu Pro Asn Pro Ala Gln Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ser Arg Arg Gly Val Glu Asp Pro Gln Val Ala
50 55 60

Gln Leu Glu Leu Gly Gly Gly Pro Gly Ala Asp Asp Leu Gln Thr Leu
65 70 75 80

Ala Leu Glu Val Ala Gln Gln Lys Arg Gly Ile Val Asp Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 29

<211> 110

<212> PRT

<213> *Rattus norvegicus*

<400> 29

Met Ala Leu Trp Met Arg Phe Leu Pro Leu Leu Ala Leu Leu Val Leu
1 5 10 15

Trp Glu Pro Lys Pro Ala Gln Ala Phe Val Lys Gln His Leu Cys Gly
20 25 30

Pro His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ser Arg Arg Glu Val Glu Asp Pro Gln Val Pro
50 55 60

Gln Leu Glu Leu Gly Gly Gly Pro Glu Ala Gly Asp Leu Gln Thr Leu
65 70 75 80

Ala Leu Glu Val Ala Arg Gln Lys Arg Gly Ile Val Asp Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 30

<211> 110

<212> PRT

<213> *Rattus norvegicus*

<400> 30

Met Ala Leu Trp Ile Arg Phe Leu Pro Leu Leu Ala Leu Leu Ile Leu
1 5 10 15

Trp Glu Pro Arg Pro Ala Gln Ala Phe Val Lys Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Met Ser Arg Arg Glu Val Glu Asp Pro Gln Val Ala
50 55 60

Gln Leu Glu Leu Gly Gly Gly Pro Gly Ala Gly Asp Leu Gln Thr Leu
65 70 75 80

Ala Leu Glu Val Ala Arg Gln Lys Arg Gly Ile Val Asp Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 31

<211> 21

<212> PRT

<213> *Acomys cahirinus*

<400> 31

Gly Ile Val Asp Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 32

<211> 108

<212> PRT

<213> *Mus musculus*

<400> 32

Met Ala Leu Leu Val His Phe Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Glu Pro Lys Pro Thr Gln Ala Phe Val Lys Gln His Leu Cys Gly
20 25 30

Pro His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ser Arg Arg Glu Val Glu Asp Pro Gln Val Glu
50 55 60

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

Glu Val Ala Arg Gln Lys Arg Gly Ile Val Asp Gln Cys Cys Thr Ser
85 90 95

Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105

<210> 33

<211> 110

<212> PRT

<213> *Mus musculus*

<400> 33

Met Ala Leu Trp Met Arg Phe Leu Pro Leu Leu Ala Leu Leu Phe Leu
1 5 10 15

Trp Glu Ser His Pro Thr Gln Ala Phe Val Lys Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Met Ser Arg Arg Glu Val Glu Asp Pro Gln Val Ala
50 55 60

Gln Leu Glu Leu Gly Gly Gly Pro Gly Ala Gly Asp Leu Gln Thr Leu
65 70 75 80

Ala Leu Glu Val Ala Gln Gln Lys Arg Gly Ile Val Asp Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 34

<211> 21

<212> PRT

<213> *Chinchila brevicaudata*

<400> 34

Gly Ile Val Asp Gln Cys Cys Thr Ser Ile Cys Thr Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
 20

<210> 35

<211> 110

<212> PRT

<213> *Cavia porcellus*

<400> 35

Met Ala Leu Trp Met His Leu Leu Thr Val Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Gly Pro Asn Thr Gly Gln Ala Phe Val Ser Arg His Leu Cys Gly
 20 25 30

Ser Asn Leu Val Glu Thr Leu Tyr Ser Val Cys Gln Asp Asp Gly Phe
 35 40 45

Phe Tyr Ile Pro Lys Asp Arg Arg Glu Leu Glu Asp Pro Gln Val Glu
 50 55 60

Gln Thr Glu Leu Gly Met Gly Leu Gly Ala Gly Gly Leu Gln Pro Leu
65 70 75 80

Ala Leu Glu Met Ala Leu Gln Lys Arg Gly Ile Val Asp Gln Cys Cys
 85 90 95

Thr Gly Thr Cys Thr Arg His Gln Leu Gln Ser Tyr Cys Asn
 100 105 110

<210> 36

<211> 109

<212> PRT

<213> *Octodon degus*

<400> 36

Met Ala Pro Trp Met His Leu Leu Thr Val Leu Ala Leu Leu Ala Leu
1 5 10 15

Trp Gly Pro Asn Ser Val Gln Ala Tyr Ser Ser Gln His Leu Cys Gly
20 25 30

Ser Asn Leu Val Glu Ala Leu Tyr Met Thr Cys Gly Arg Ser Gly Phe
35 40 45

Tyr Arg Pro His Asp Arg Arg Glu Leu Glu Asp Leu Gln Val Glu Gln
50 55 60

Ala Glu Leu Gly Leu Glu Ala Gly Gly Leu Gln Pro Ser Ala Leu Glu
65 70 75 80

Met Ile Leu Gln Lys Arg Gly Ile Val Asp Gln Cys Cys Asn Asn Ile
85 90 95

Cys Thr Phe Asn Gln Leu Gln Asn Tyr Cys Asn Val Pro
100 105

<210> 37

<211> 21

<212> PRT

<213> *Didelphis virginiana*

<400> 37

Gly Ile Val Glu Gln Cys Cys Asn Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Thr Tyr Cys Asn
20

<210> 38

<211> 108

<212> PRT

<213> *Rodentia sp.*

<400> 38

Met Ala Leu Trp Ile Leu Leu Pro Leu Leu Ala Leu Leu Ile Leu Trp
1 5 10 15

Gly Pro Asp Pro Ala Gln Ala Phe Val Asn Gln His Leu Cys Gly Ser
20 25 30

His Leu Val Glu Ala Leu Tyr Ile Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Met Ser Arg Arg Glu Val Glu Asp Pro Gln Val Gly
50 55 60

Gln Val Glu Leu Gly Ala Gly Pro Gly Ala Gly Ser Glu Gln Thr Leu
65 70 75 80

Ala Leu Glu Val Ala Arg Gln Ala Arg Ile Val Gln Gln Cys Thr Ser
85 90 95

Gly Ile Cys Ser Leu Tyr Gln Glu Asn Tyr Cys Asn
100 105

<210> 39

<211> 110

<212> PRT

<213> *Psammomys obesus*

<400> 39

Met Ala Leu Trp Met Arg Leu Leu Pro Leu Leu Ala Phe Leu Ile Leu
1 5 10 15

Trp Glu Pro Ser Pro Ala His Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Phe Arg Arg Gly Val Asp Asp Pro Gln Met Pro
50 55 60

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

Ala Leu Glu Val Ala Arg Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95

Thr Gly Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 40

<211> 110

<212> PRT

<213> *Spermophilus tridecemlineatus*

<400> 40

Met Ala Leu Trp Thr Arg Leu Leu Pro Leu Leu Ala Leu Leu Ala Leu
1 5 10 15

Leu Gly Pro Asp Pro Ala Gln Ala Phe Val Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Thr Pro Lys Ser Arg Arg Glu Val Glu Glu Gln Gln Gly Gly
50 55 60

Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Leu Pro Gln Pro Leu
65 70 75 80

Ala Leu Glu Met Ala Leu Gln Lys Arg Gly Ile Val Glu Gln Cys Cys
85 90 95

Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105 110

<210> 41

<211> 29

<212> PRT

<213> *Cavia porcellus*

<400> 41

Glu Leu Glu Asp Pro Gln Val Glu Gln Thr Glu Leu Gly Met Gly Leu
1 5 10 15

Gly Ala Gly Gly Leu Gln Pro Leu Gln Gly Ala Leu Gln
20 25

<210> 42

<211> 107

<212> PRT

<213> *Ballus gallus*

<400> 42

Met Ala Leu Trp Ile Arg Ser Leu Pro Leu Leu Ala Leu Leu Val Phe
1 5 10 15

Ser Gly Pro Gly Thr Ser Tyr Ala Ala Ala Asn Gln His Leu Cys Gly
20 25 30

Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe
35 40 45

Phe Tyr Ser Pro Lys Ala Arg Arg Asp Val Glu Gln Pro Leu Val Ser
50 55 60

Ser Pro Leu Arg Gly Glu Ala Gly Val Leu Pro Phe Gln Gln Glu Glu
65 70 75 80

Tyr Glu Lys Val Lys Arg Gly Ile Val Glu Gln Cys Cys His Asn Thr
85 90 95

Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
100 105

<210> 43

<211> 81

<212> PRT

<213> *Anas platyrhynchos*

<220>

<221> CARACTERÍSTICA_MISC

<222> (31)..(32)

<223> X = qualquer aminoácido

<220>

<221> CARACTERÍSTICA_MISC

<222> (59)..(60)

<223> X = qualquer aminoácido

<400> 43

Ala Ala Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Ser Pro Lys Thr Xaa Xaa
20 25 30

Asp Val Glu Gln Pro Leu Val Asn Gly Pro Leu His Gly Glu Val Gly
35 40 45

Glu Leu Pro Phe Gln His Glu Glu Tyr Gln Xaa Xaa Gly Ile Val Glu
50 55 60

Gln Cys Cys Glu Asn Pro Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys
65 70 75 80

Asn

<210> 44

<211> 21

<212> PRT

<213> *Anser anser*

<400> 44

Gly Ile Val Glu Gln Cys Cys Glu Asn Pro Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 45

<211> 103

<212> PRT

<213> *Selasphorus rufus*

<400> 45

Ile Gln Ser Leu Pro Leu Leu Ala Leu Leu Ala Leu Ser Gly Pro Gly
1 5 10 15

Thr Ser His Ala Ala Val Asn Gln His Leu Cys Gly Ser His Leu Val
20 25 30

Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Ser Pro
35 40 45

Lys Ala Arg Arg Asp Ala Glu His Pro Leu Val Asn Gly Pro Leu His
50 55 60

Gly Glu Val Gly Asp Leu Pro Phe Gln Gln Glu Glu Phe Glu Lys Val
65 70 75 80

Lys Arg Gly Ile Val Glu Gln Cys Cys His Asn Thr Cys Ser Leu Tyr
85 90 95

Gln Leu Glu Asn Tyr Cys Asn
100

<210> 46

<211> 48

<212> PRT

<213> *Danio rerio*

<400> 46

Met Ala Val Trp Leu Gln Ala Gly Ala Leu Leu Val Leu Leu Val Val
1 5 10 15

Ser Ser Val Ser Thr Asn Pro Gly Thr Pro Gln His Leu Cys Gly Ser
20 25 30

His Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Pro Thr Gly Phe Phe
35 40 45

Tyr Asn Pro Lys Arg Asp Val Glu Pro Leu Leu Gly Phe Leu Pro Pro
50 55 60

Lys Ser Ala Gln Glu Thr Glu Val Ala Asp Phe Ala Phe Lys Asp His
65 70 75 80

Ala Glu Leu Ile Arg Lys Arg Gly Ile Val Glu Gln Cys Cys His Lys
85 90 95

Pro Cys Ser Ile Phe Glu Leu Gln Asn Tyr Cys Asn
100 105

<210> 47

<211> 108

<212> PRT

<213> *Cyprinus caprio*

<400> 47

Met Ala Val Trp Ile Gln Ala Gly Ala Leu Leu Phe Leu Leu Ala Val
1 5 10 15

Ser Ser Val Asn Ala Asn Ala Gly Ala Pro Gln His Leu Cys Gly Ser
20 25 30

His Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Pro Thr Gly Phe Phe
35 40 45

Tyr Asn Pro Lys Arg Asp Val Asp Pro Pro Leu Gly Phe Leu Pro Pro
50 55 60

Lys Ser Ala Gln Glu Thr Glu Val Ala Asp Phe Ala Phe Lys Asp His
65 70 75 80

Ala Glu Val Ile Arg Lys Arg Gly Ile Val Glu Gln Cys Cys His Lys
85 90 95

Pro Cys Ser Ile Phe Glu Leu Gln Asn Tyr Cys Asn
100 105

<210> 48

<211> 21

<212> PRT

<213> *Batrachoididae gen. sp.*

<400> 48

Gly Ile Val Glu Gln Cys Cys His Arg Pro Cys Asp Ile Phe Asp Leu
1 5 10 15

Gln Ser Tyr Cys Asn
20

<210> 49

<211> 21

<212> PRT

<213> *Thunnus thynnus*

<400> 49

Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Asn Ile Phe Asp Leu
1 5 10 15

Gln Asn Tyr Cys Asn
20

<210> 50

<211> 21

<212> PRT

<213> *Katsuwonus pelamis*

<400> 50

Gly Ile His Glx Glx Cys Cys His Lys Pro Cys Asx Ile Phe Glx Leu
1 5 10 15

Glx Asx Tyr Cys Asn
20

<210> 51

<211> 116

<212> PRT

<213> *Lophius piscatorius*

<400> 51

Met Ala Ala Leu Trp Leu Gln Ser Phe Ser Leu Leu Val Leu Leu Val
1 5 10 15

Val Ser Trp Pro Gly Ser Gln Ala Val Ala Pro Ala Gln His Leu Cys
20 25 30

Gly Ser His Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Asp Arg Gly
35 40 45

Phe Phe Tyr Asn Pro Lys Arg Asp Val Asp Gln Leu Leu Gly Phe Leu
50 55 60

Pro Pro Lys Ser Gly Gly Ala Ala Ala Ala Gly Ala Asp Asn Glu Val
65 70 75 80

Ala Glu Phe Ala Phe Lys Asp Gln Met Glu Met Met Val Lys Arg Gly
85 90 95

Ile Val Glu Gln Cys Cys His Arg Pro Cys Asn Ile Phe Asp Leu Gln
100 105 110

Asn Tyr Cys Asn
115

<210> 52

<211> 115

<212> PRT

<213> *Myxine glutinosa*

<400> 52

Met Ala Leu Ser Pro Phe Leu Ala Ala Val Ile Pro Leu Val Leu Leu
1 5 10 15

Leu Ser Arg Ala Pro Pro Ser Ala Asp Thr Arg Thr Thr Gly His Leu
20 25 30

Cys Gly Lys Asp Leu Val Asn Ala Leu Tyr Ile Ala Cys Gly Val Arg
35 40 45

Gly Phe Phe Tyr Asp Pro Thr Lys Met Lys Arg Asp Thr Gly Ala Leu
50 55 60

Ala Ala Phe Leu Pro Leu Ala Tyr Ala Glu Asp Asn Glu Ser Gln Asp
65 70 75 80

Asp Glu Ser Ile Gly Ile Asn Glu Val Leu Lys Ser Lys Arg Gly Ile
85 90 95

Val Glu Gln Cys Cys His Lys Arg Cys Ser Ile Tyr Asp Leu Glu Asn
100 105 110

Tyr Cys Asn

115

<210> 53

<211> 105

<212> PRT

<213> *Oncorhynchus keta*

<400> 53

Met Ala Phe Trp Leu Gln Ala Ala Ser Leu Leu Val Leu Leu Ala Leu
1 5 10 15

Ser Pro Gly Val Asp Ala Ala Ala Ala Gln His Leu Cys Gly Ser His
20 25 30

Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Glu Lys Gly Phe Phe Tyr
35 40 45

Thr Pro Lys Arg Asp Val Asp Pro Leu Ile Gly Phe Leu Ser Pro Lys
50 55 60

Ser Ala Lys Glu Asn Glu Glu Tyr Pro Phe Lys Asp Gln Thr Glu Met
65 70 75 80

Met Val Lys Arg Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Asn
85 90 95

Ile Phe Asp Leu Gln Asn Tyr Cys Asn
100 105

<210> 54

<211> 21

<212> PRT

<213> *Myoxocephalus scorpius*

<400> 54

Gly Ile Val Glu Gln Cys Cys His Arg Pro Cys Asn Ile Arg Val Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 55

<211> 21

<212> PRT

<213> *Lepisosteus spatula*

<400> 55

Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Thr Ile Tyr Glu Leu
1 5 10 15

Glu Asn Tyr Cys Asn
 20

<210> 56

<211> 21

<212> PRT

<213> *Platichthys flesus*

<400> 56

Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Asn Ile Phe Asp Leu
1 5 10 15

Gln Asn Tyr Cys Asn
 20

<210> 57

<211> 21

<212> PRT

<213> *Hydrolagus colliei*

<400> 57

Gly Ile Val Glu Gln Cys Cys His Asn Thr Cys Ser Leu Ala Asn Leu
1 5 10 15

Glu Gly Tyr Cys Asn
20

<210> 58

<211> 22

<212> PRT

<213> *Squalus acanthias*

<400> 58

Gly Ile Val Glu, His Cys Cys His Asn Thr Cys Ser Leu Tyr Asp Leu
1 5 10 15

Glu Gly Tyr Cys Asn Gln
20

<210> 59

<211> 21

<212> PRT

<213> *Torpedo marmorata*

<400> 59

Gly Ile Val Glu His Cys Cys His Asn Thr Cys Ser Leu Phe Asp Leu
1 5 10 15

Glu Gly Tyr Cys Asn
20

<210> 60

<211> 89

<212> PRT

<213> *Callorhinchus milii*

<400> 60

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

<210> 61

<211> 21

<212> PRT

<213> *Petromyzon marinus*

<400> 61

Gly Ile Val Glu Gln Cys Cys His Arg Lys Cys Ser Ile Tyr Asp Met
1 5 10 15
Glu Asn Tyr Cys Asn
 20

<210> 62

<211> 21

<212> PRT

<213> *Oncorhynchus gorbuscha*

<400> 62

Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Asn Ile Phe Asp Leu
1 5 10 15

Gln Asn Tyr Cys Asn
20

<210> 63

<211> 21

<212> PRT

<213> *Amia calva*

<400> 63

Gly Ile Val Glu Gln Cys Cys Leu Lys Pro Cys Thr Ile Tyr Glu Met
1 5 10 15

Glu Lys Tyr Cys Asn
20

<210> 64

<211> 21

<212> PRT

<213> *Anguilla rostrata*

<400> 64

Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Ser Ile Phe Asp Leu
1 5 10 15

Gln Asn Tyr Cys Asn
20

<210> 65

<211> 113

<212> PRT

<213> *Oreochromis niloticus*

<400> 65

Met Ala Ala Leu Trp Leu Gln Ala Phe Ser Leu Leu Val Leu Met Met
1 5 10 15

Val Ser Trp Pro Gly Ser Gln Ala Val Gly Gly Pro Gln His Leu Cys
20 25 30

Gly Ser His Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Asp Arg Gly
35 40 45

Phe Phe Tyr Asn Pro Arg Arg Asp Val Asp Pro Leu Leu Gly Phe Leu
50 55 60

Pro Pro Lys Ala Gly Gly Ala Val Val Gln Gly Gly Glu Asn Glu Val
65 70 75 80

Thr Phe Lys Asp Gln Met Glu Met Met Val Lys Arg Gly Ile Val Glu
85 90 95

Glu Cys Cys His Lys Pro Cys Thr Ile Phe Asp Leu Gln Asn Tyr Cys
100 105 110

Asn

<210> 66

<211> 21

<212> PRT

<213> *Acipenser gueldenstaedti*

<400> 66

Gly Ile Val Glu Gln Cys Cys His Ser Pro Cys Ser Leu Tyr Asp Leu
1 5 10 15

Glu Asn Tyr Cys Asn
 20

<210> 67

<211> 21

<212> PRT

<213> *Piaractus mesopotamicus*

<400> 67

Gly Ile Val Glu Gln Cys Cys His Lys Pro Cys Ser Ile Phe Asp Leu
1 5 10 15

Gln Asn Tyr Cys Asn
 20

<210> 68

<211> 115

<212> PRT

<213> *Verasper moseri*

<400> 68

Met Ala Ala Leu Trp Leu Gln Ser Val Ser Leu Leu Val Leu Met Leu
1 5 10 15

Val Ser Trp Ser Gly Ser Gln Ala Val Leu Pro Pro Gln His Leu Cys
20 25 30

Gly Ala His Leu Val Asp Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly
35 40 45

Phe Phe Tyr Thr Pro Lys Arg Asp Val Asp Pro Leu Leu Gly Phe Leu
50 55 60

Pro Ala Lys Ser Gly Gly Ala Ala Ala Gly Gly Glu Asn Glu Val Ala
65 70 75 80

Glu Phe Ala Phe Lys Asp Gln Met Glu Met Met Val Lys Arg Gly Ile
85 90 95

Val Glu Gln Cys Cys His Lys Pro Cys Asn Ile Phe Asp Leu Gln Asn
100 105 110

Tyr Cys Asn
115

<210> 69

<211> 30

<212> PRT

<213> *Anguilla anguilla*

<400> 69

Asp Val Glu Pro Leu Leu Gly Phe Leu Ser Pro Lys Ser Gly Gln Glu
1 5 10 15

Asn Glu Val Asp Asp Phe Pro Tyr Lys Gly Gln Gly Glu Leu
20 25 30

<210> 70

<211> 106

<212> PRT

<213> *Xenopus laevis*

<400> 70

Met Ala Leu Trp Met Gln Cys Leu Pro Leu Val Leu Val Leu Phe Phe
1 5 10 15

Ser Thr Pro Asn Thr Glu Ala Leu Val Asn Gln His Leu Cys Gly Ser
20 25 30

His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Asp Arg Gly Phe Phe
35 40 45

Tyr Tyr Pro Lys Val Lys Arg Asp Met Glu Gln Ala Leu Val Ser Gly
50 55 60

Pro Gln Asp Asn Glu Leu Asp Gly Met Gln Leu Gln Pro Gln Glu Tyr
65 70 75 80

Gln Lys Met Lys Arg Gly Ile Val Glu Gln Cys Cys His Ser Thr Cys
85 90 95

Ser Leu Phe Gln Leu Glu Ser Tyr Cys Asn
100 105

<210> 71

<211> 106

<212> PRT

<213> *Xenopus laevis*

<400> 71

Met Ala Leu Trp Met Gln Cys Leu Pro Leu Val Leu Val Leu Leu Phe
1 5 10 15

Ser Thr Pro Asn Thr Glu Ala Leu Ala Asn Gln His Leu Cys Gly Ser
20 25 30

His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Asp Arg Gly Phe Phe
35 40 45

Tyr Tyr Pro Lys Ile Lys Arg Asp Ile Glu Gln Ala Gln Val Asn Gly
50 55 60

Pro Gln Asp Asn Glu Leu Asp Gly Met Gln Phe Gln Pro Gln Glu Tyr
65 70 75 80

Gln Lys Met Lys Arg Gly Ile Val Glu Gln Cys Cys His Ser Thr Cys
85 90 95

Ser Leu Phe Gln Leu Glu Asn Tyr Cys Asn
100 105

<210> 72

<211> 21

<212> PRT

<213> *Trachemys scripta*

<400> 72

Gly Ile Val Glu Gln Cys Cys His Asn Thr Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 73

<211> 21

<212> PRT

<213> *Alligator mississippiensis*

<400> 73

Gly Ile Val Glu Gln Cys Cys His Asn Thr Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 74

<211> 21

<212> PRT

<213> *Zaocys dhumnades*

<400> 74

Gly Ile Val Glu Gln Cys Cys Glu Asn Thr Cys Ser Leu Tyr Glu Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 75

<211> 21

<212> PRT

<213> *Crotalus atrox*

<400> 75

Gly Ile Val Glu Gln Cys Cys Glu Asn Thr Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 76

<211> 114

<212> PRT

<213> Sequência Artificial

<220>

<223> Pré-pro-insulina

<400> 76

Met Gly Leu Trp Ile Arg Leu Leu Pro Leu Ile Ala Leu Leu Ile Leu
1 5 10 15

Trp Gly Pro Asp Pro Ala Ala Ala Glu Phe Arg Met Phe Val Asn Gln
20 25 30

His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly
35 40 45

Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp
50 55 60

Leu Gln Val Gly Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser
65 70 75 80

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

Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr
100 105 110

Cys Asn

<210> 79

<211> 53

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina

<400> 79

Phe Val Asp Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Ala Ala Lys
20 25 30

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Glu Leu
35 40 45

Glu Asp Tyr Cys Asn
50

<210> 80

<211> 53

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina

<400> 80

Phe Val Glu Gln His Leu Cys Gly Ser Asp Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Ala Ala Lys
20 25 30

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
35 40 45

Glu Glu Tyr Cys Asn
50

<210> 81

<211> 53

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina

<400> 81

Phe Val Gln Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Ala Ala Lys
20 25 30

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
35 40 45

Glu Asn Tyr Cys Gly
50

<210> 82

<211> 53

<212> PRT

<213> Sequência Artificial

<220>

<223> Produto proteico sem nome com homologia com insulina

<400> 82

Phe Val Thr Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Ala Ala Lys
 20 25 30

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
 35 40 45

Glu His Tyr Cys Ser
 50

<210> 83

<211> 57

<212> PRT

<213> Sequência Artificial

<220>

<223> Pró-insulina

<400> 83

Asn Ser Asn Gly Lys Phe Val Asn Gln His Leu Cys Gly Ser His Leu
1 5 10 15

Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr
20 25 30

Pro Lys Thr Lys Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser
35 40 45

Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
50 55

<210> 84

<211> 58

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina

<400> 84

Asn Ser Asn Gly Lys Phe Val Asn Gln His Leu Cys Gly Ser His Leu
1 5 10 15

Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr
20 25 30

Pro Lys Thr Lys Arg Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys
35 40 45

Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
50 55

<400> 86

Arg Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu
1 5 10 15

Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Ala Ala
20 25 30

Lys Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln
35 40 45

Leu Glu Asn Tyr Cys Asn
50

<210> 87

<211> 61

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina

<400> 87

Lys Glu Thr Leu Thr Ile Thr Cys Ala Val Pro Thr Trp Leu Lys Leu
1 5 10 15

Trp Thr Trp Phe Ala Val Lys Glu Val Ser Ser Thr Asn Leu Arg Leu
20 25 30

Leu Arg Val Leu Ser Asn Asn Ala Val Pro Pro Ser Ala Pro Cys Thr
35 40 45

Asn Trp Lys Thr Thr Ala Thr Arg Arg Ser Pro Gln Ala
50 55 60

<210> 88

<211> 61

<212> PRT

<213> Sequência Artificial

<220>

<223> Pré-pro-insulina

<400> 88

Lys Asp Ser Leu Thr Asn Thr Cys Ala Val Ser Thr Trp Leu Lys Leu
1 5 10 15

Cys Thr Trp Phe Ala Val Lys Glu Val Ser Ser Thr Leu Leu Arg Leu
20 25 30

Leu Arg Val Leu Ser Asn Asn Ala Val Pro Pro Ser Ala Asn Tyr Thr
35 40 45

Asn Trp Lys Thr Thr Ala Thr Arg Arg Ser Pro Gln Ala
50 55 60

<210> 89

<211> 87

<212> PRT

<213> Sequência Artificial

<220>

<223> Pró-insulina

<400> 89

Met Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu
1 5 10 15

Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr Arg
20 25 30

Arg Glu Ala Glu Asp Leu Gln Val Gly Gln Val Glu Leu Gly Gly Gly
35 40 45

Pro Gly Ala Gly Ser Leu Gln Pro Leu Ala Leu Glu Gly Ser Leu Gln
50 55 60

Lys Arg Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr
65 70 75 80

Gln Leu Glu Asn Tyr Cys Asn
85

<210> 90

<211> 35

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina

<400> 90

Arg Arg Glu Ala Glu Asp Leu Gln Val Gly Gln Val Glu Leu Gly Gly
1 5 10 15

Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu Ala Leu Glu Gly Ser Leu
20 25 30

Gln Lys Arg
35

Thr Asn Ser Gly Gly Leu Asp Val Val Gly Leu Ile Ser Met Ala Lys
50 55 60

Arg Glu Glu Gly Glu Pro Lys Phe Val Asn Gln His Leu Cys Gly Ser
65 70 75 80

His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe
85 90 95

Tyr Thr Pro Lys Ala Ala Lys Gly Ile Val Glu Gln Cys Cys Thr Ser
100 105 110

Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
115 120

<210> 93

<211> 171

<212> ADN

<213> Sequência Artificial

<220>

<223> Insulina

<400> 93

tttgtcaatc agcacctttg tggttctcac ctggtggagg ctctgtacct ggtgtgtggg 60
gaacgtggtt tcttctacac acccaagacc cgctgtaagc ttaagcgtgg cattgtggag 120
cagtgtgca ccagcatctg ctccctctac caactggaga actactgcaa c 171

<210> 94

<211> 21

<212> PRT

<213> Sequência Artificial

<220>

<223> Análogo da insulina

<400> 94

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 95

<211> 26

<212> PRT

<213> Sequência Artificial

<220>

<223> Análogo da insulina

<220>

<221> CARACTERÍSTICA_MISC

<222> (26)..(26)

<223> X = qualquer aminoácido

<400> 95

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Xaa
20 25

<210> 96

<211> 21

<212> PRT

<213> *Homo sapiens*

<400> 96

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 97

<211> 29

<212> PRT

<213> *Homo sapiens*

<400> 97

Glu Val Asn Gln His Leu Cys Gly Ser Glu Leu Val Glu Ala Leu Glu
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Glu Pro Lys
20 25

<210> 98

<211> 21

<212> PRT

<213> *Homo sapiens*

<400> 98

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 99

<211> 30

<212> PRT

<213> *Homo sapiens*

<400> 99

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu His
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
20 25 30

<210> 100

<211> 21

<212> PRT

<213> *Homo sapiens*

<400> 100

Gly Ile Val Glu Gln Cys Cys Lys Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 101

<211> 30

<212> PRT

<213> *Homo sapiens*

<400> 101

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
 20 25 30

<210> 102

<211> 21

<212> PRT

<213> *Homo sapiens*

<400> 102

Gly Ile Val Glu Gln Cys Cys Lys Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
 20

<210> 103

<211> 30

<212> PRT

<213> *Homo sapiens*

<400> 103

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
 20 25 30

<210> 104

<211> 21

<212> PRT

<213> *Homo sapiens*

<220>

<221> CARACTERÍSTICA_MISC

<222> (8)..(8)

<223> X = qualquer aminoácido

<400> 104

Gly Ile Val Glu Gln Cys Cys Xaa Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 105

<211> 30

<212> PRT

<213> *Homo sapiens*

<400> 105

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
20 25 30

<210> 106

<211> 21

<212> PRT

<213> *Homo sapiens*

<220>

<221> CARACTERÍSTICA_MISC

<222> (8)..(8)

<223> X = qualquer aminoácido

<400> 106

Gly Ile Val Glu Gln Cys Cys Xaa Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 107

<211> 30

<212> PRT

<213> *Homo sapiens*

<400> 107

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
20 25 30

<210> 108

<211> 30

<212> PRT

<213> *Homo sapiens*

<220>

<221> CARACTERÍSTICA_MISC

<222> (2)..(2)

<223> X = quaisquer aminoácidos

<400> 108

Gly Xaa Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 109

<211> 30

<212> PRT

<213> *Homo sapiens*

<400> 109

Phe Val Asn Gln His Leu Cys Gly Ser Asp Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Lys Pro Thr
20 25 30

<210> 110

<211> 21

<212> PRT

<213> *Homo sapiens*

<400> 110

Gly Ala Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 111

<211> 30

<212> PRT

<213> *Homo sapiens*

<400> 111

Phe Val Asn Gln His Leu Cys Gly Ser Asp Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Lys Pro Thr
20 25 30

<210> 112

<211> 21

<212> PRT

<213> Sequência Artificial

<220>

<223> Análogo da insulina

<220>

<221> CARACTERÍSTICA_MISC

<222> (2)..(2)

<223> X = qualquer aminoácido

<400> 112

Gly Xaa Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 113
<211> 30
<212> PRT
<213> Sequência Artificial

<220>
<223> Análogo da insulina

<400> 113

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
 20 25 30

<210> 114
<211> 21
<212> PRT
<213> Sequência Artificial

<220>
<223> Análogo da insulina

<220>
<221> CARACTERÍSTICA_MISC
<222> (2)..(2)
<223> X = qualquer aminoácido

<400> 114

Gly Xaa Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
 20

<210> 115

<211> 30

<212> PRT

<213> Sequência Artificial

<220>

<223> Análogo da insulina

<400> 115

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
 20 25 30

<210> 116

<211> 21

<212> PRT

<213> *Homo sapiens*

<400> 116

Gly Gly Gly Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
 20

<210> 117

<211> 30

<212> PRT

<213> *Homo sapiens*

<400> 117

Phe Val Asn Gln His Leu Cys Gly Ser Asp Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Lys Pro Thr
20 25 30

<210> 118

<211> 21

<212> PRT

<213> Sequência Artificial

<220>

<223> Mutante da insulina de *Homo sapiens*

<220>

<221> CARACTERÍSTICA_MISC

<222> (1)..(1)

<223> X = qualquer aminoácido

<400> 118

Xaa Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 119

<211> 30

<212> PRT

<213> Sequência Artificial

<220>

<223> Mutante da insulina de *Homo sapiens*

<220>

<221> CARACTERÍSTICA_MISC

<222> (1)..(1)

<223> X = qualquer aminoácido

<400> 119

Xaa	Val	Asn	Gln	His	Leu	Cys	Gly	Asp	His	Leu	Val	Glu	Ala	Leu	Tyr
1				5					10					15	

Leu	Val	Cys	Gly	Glu	Arg	Gly	Phe	Phe	Tyr	Thr	Pro	Lys	Thr
			20					25					30

<210> 120

<211> 30

<212> PRT

<213> Sequência Artificial

<220>

<223> Mutante da insulina de *Homo sapiens*

<400> 120

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 121

<211> 29

<212> PRT

<213> Sequência Artificial

<220>

<223> Mutante da insulina de *Homo sapiens*

<400> 121

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Tyr Thr Pro Lys Thr
20 25

<210> 122

<211> 21

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina de *Homo sapiens*

<400> 122

Gly Ile Val Glu Gln Ser Cys Thr Ser Ile Ser Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 123

<211> 30

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina de *Homo sapiens*

<400> 123

Phe Val Asn Gln His Leu Cys Gly Ser Asp Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Lys Pro Thr
20 25 30

<210> 124

<211> 79

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina

<400> 124

Met Asp Pro Gly Asp Pro Glu Cys Leu Glu Gln Leu Leu Arg Arg Leu
1 5 10 15

Gly Gly Ser Val Glu Val Glu Val Thr Gly Gly Thr Val His Val Glu
20 25 30

Val Ser Pro Glu Asp Pro Gly Asp Pro Glu Cys Leu Glu Gln Leu Leu
35 40 45

Arg Arg Leu Gly Gly Ser Val Glu Val Glu Val Thr Gly Gly Thr Val
50 55 60

His Val Glu Val Ser Pro Gly Glu Arg Gly Phe Phe Tyr Cys Asn
65 70 75

<210> 125

<211> 87

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina

<400> 125

Met Leu Lys Glu Lys Lys Tyr Ser Pro Asp Pro Gly Asp Pro Glu Cys
1 5 10 15

Leu Glu Gln Leu Leu Arg Arg Leu Gly Gly Ser Val Glu Val Glu Val
20 25 30

Thr Gly Gly Thr Val His Val Glu Val Ser Pro Glu Asp Pro Gly Asp
35 40 45

Pro Glu Cys Leu Glu Gln Leu Leu Arg Arg Leu Gly Gly Ser Val Glu
50 55 60

Val Glu Val Thr Gly Gly Thr Val His Val Glu Val Ser Pro Gly Glu
65 70 75 80

Arg Gly Phe Phe Tyr Cys Asn
85

<210> 126

<211> 23

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina proteína de fusão

<220>

<221> CARACTERÍSTICA_MISC

<222> (5)..(5)

<223> X = qualquer aminoácido

<220>

<221> CARACTERÍSTICA_MISC

<222> (18)..(18)

<223> X = qualquer aminoácido

<400> 126

Met Ala Thr Ser Xaa Ser Thr Lys Lys Thr Gln Leu Gln Leu Glu His
1 5 10 15

Leu Xaa Leu Asp Leu Gln Met
20

<210> 127

<211> 23

<212> PRT

<213> *Escherichia coli*

<220>

<221> CARACTERÍSTICA_MISC

<222> (14)..(14)

<223> X = qualquer aminoácido

<220>

<221> CARACTERÍSTICA_MISC

<222> (17)..(17)

<223> X = qualquer aminoácido

<400> 127

Thr Met Ile Thr Asp Ser Leu Ala Val Val Leu Gln Arg Xaa Asp Trp
1 5 10 15

Xaa Pro Gly Val Thr Gln Leu
20

<210> 128

<211> 96

<212> PRT

<213> *Brevibacillus brevis*

<400> 128

Asn Ser Val Leu Ala Ser Ala Leu Ala Leu Thr Val Ala Pro Met Ala
1 5 10 15

Phe Ala Asn Ser Asp Ser Glu Ser Pro Leu Ser His Asp Gly Tyr Ser
20 25 30

Leu His Asp Gly Val Ser Met Tyr Ile Glu Ala Leu Asp Lys Phe Val
35 40 45

Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val
50 55 60

Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Gly Ile Val Glu Gln
65 70 75 80

Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
85 90 95

<210> 129

<211> 30

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina proteína de fusão

<400> 129

Asp Thr Thr Met Pro Ala Gly Gly Gly Gly Gly Gly Gln His Leu Cys
1 5 10 15

Gly Pro His Leu Val Glu Ala Leu Tyr
20 25

<210> 130

<211> 6

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina proteína de fusão

<400> 130

Leu Glu Asn Tyr Cys Asn
1 5

<210> 131

<211> 22

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina proteína de fusão

<400> 131

Met Thr Met Ile Thr Asp Ser Leu Glu Phe Gln Ala Trp Gly Gly Gly
1 5 10 15

Gly Gly Trp Met Arg Phe
 20

<210> 132

<211> 22

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina proteína de fusão

<400> 132

Met Val Leu Arg Phe Leu Pro Leu Leu Ala Leu Leu Val Leu Trp Glu
1 5 10 15

Pro Lys Pro Ala Gln Ala
 20

<210> 133

<211> 60

<212> PRT

<213> Sequência Artificial

<220>

<223> Mini pró-insulina

<400> 133

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr Arg Arg
20 25 30

Tyr Pro Gly Asp Val Lys Arg Gly Ile Val Glu Gln Cys Cys Thr Ser
35 40 45

Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
50 55 60

<210> 134

<211> 21

<212> PRT

<213> *Homo sapiens*

<400> 134

Gly Ile Val Glu Gln Cys Cys Lys Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 135

<211> 30

<212> PRT

<213> *Homo sapiens*

<400> 135

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
20 25 30

<210> 136

<211> 21

<212> PRT

<213> *Homo sapiens*

<400> 136

Gly Ile Val Glu Gln Cys Cys Lys Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 137

<211> 30

<212> PRT

<213> *Homo sapiens*

<400> 137

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
20 25 30

<210> 138

<211> 14

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina de *Homo sapiens*

<400> 138

Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Gly
1 5 10

<210> 139

<211> 21

<212> PRT

<213> *Homo sapiens*

<220>

<221> CARACTERÍSTICA_MISC

<222> (8)..(8)

<223> X = qualquer aminoácido

<400> 139

Gly Ile Val Glu Gln Cys Cys Xaa Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

<210> 140

<211> 30

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina de *Homo sapiens*

<400> 140

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
 20 25 30

<210> 141

<211> 21

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina de *Homo sapiens*

<220>

<221> CARACTERÍSTICA_MISC

<222> (8)..(8)

<223> X = qualquer aminoácido

<400> 141

Gly Ile Val Glu Gln Cys Cys Xaa Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
 20

<210> 142

<211> 30

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina de *Homo sapiens*

<400> 142

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

<210> 143

<211> 21

<212> PRT

<213> Sequência Artificial

<220>

<223> Mutante da mini pró-insulina

<400> 143

Gly	Ile	Val	Glu	Gln	Cys	Cys	Thr	Ser	Ile	Cys	Ser	Leu	Tyr	Gln	Leu
1			5					10						15	
Glu	Asn	Tyr	Cys	Asn											
			20												

<210> 144

<211> 29

<212> PRT

<213> Sequência Artificial

<220>

<223> Mutante da mini pró-insulina

<400> 144

Phe Val Asn Gln His Leu Cys Gly Ser Asp Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Asp Lys
 20 25

<210> 145

<211> 50

<212> PRT

<213> Sequência Artificial

<220>

<223> Mutante da mini pró-insulina

<400> 145

Phe Val Asn Gln His Leu Cys Gly Ser Asp Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Asp Lys Gly Ile Val
 20 25 30

Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr
 35 40 45

Cys Asn
 50

<210> 146

<211> 3

<212> PRT

<213> Sequência Artificial

<220>

<223> Péptido C da insulina

<400> 146

Ala Ala Lys
1

<210> 147

<211> 3

<212> PRT

<213> Sequência Artificial

<220>

<223> Péptido C da insulina

<400> 147

Asn Lys Arg
1

<210> 148

<211> 6

<212> PRT

<213> Sequência Artificial

<220>

<223> Péptido C da insulina

<400> 148

Arg Arg Lys Gln Lys Arg
1 5

<210> 149

<211> 4

<212> PRT

<213> Sequência Artificial

<220>

<223> Sítio de Clivagem

<400> 149

Arg Arg Lys Arg
i

<210> 150

<211> 4

<212> PRT

<213> Sequência Artificial

<220>

<223> Sequência de retenção de ER

<400> 150

Lys Asp Glu Leu
i

<210> 151

<211> 4

<212> PRT

<213> Sequência Artificial

<220>

<223> Sequência de retenção de ER

<400> 151

His Asp Glu Lsu
1

<210> 152

<211> 4

<212> PRT

<213> Sequência Artificial

<220>

<223> Sequência de retenção de ER

<400> 152

Asp Asp Glu Leu
1

<210> 153

<211> 4

<212> PRT

<213> Sequência Artificial

<220>

<223> Sequência de retenção de ER

<400> 153

Ala Asp Glu Leu
1

<210> 154

<211> 4

<212> PRT

<213> Sequência Artificial

<220>

<223> Sequência de retenção de ER

<400> 154

Ser Asp Glu Leu
|

<210> 155

<211> 4

<212> PRT

<213> *Lycopersicon esculentum Mill.*

<400> 155

His Asp Glu Phe
|

<210> 156

<211> 118

<212> PRT

<213> *Arabidopsis thaliana*

<400> 156

Met Ala Asp Thr Ala Arg Gly Thr His His Asp Ile Ile Gly Arg Asp
1 5 10 15

Gln Tyr Pro Met Met Gly Arg Asp Arg Asp Gln Tyr Gln Met Ser Gly
20 25 30

Arg Gly Ser Asp Tyr Ser Lys Ser Arg Gln Ile Ala Lys Ala Ala Thr
35 40 45

Ala Val Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Ser Leu Thr Leu
50 55 60

Val Gly Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile
65 70 75 80

Phe Ser Pro Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Leu Ile
85 90 95

Thr Gly Phe Leu Ser Ser Gly Gly Phe Gly Ile Ala Ala Ile Thr Val
100 105 110

Phe Ser Trp Ile Tyr Lys
115

<210> 157

<211> 187

<212> PRT

<213> *Brassica napus*

<400> 157

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

<210> 158

<211> 748

<212> ADN

<213> *Arabidopsis thaliana*

<400> 158

```
taocatgggg tcaaagacgg agatgatgga gagagacgca atggctacgg tggctcccta    60
tgcgcgggtc acttaccato gccgtgetog tgttgacttg gatgatagac ttectaaacc    120
ttatatgcca agagcattgc aagcaccaga cagagaacac cggtaaggaa ctccaggcca    180
taagaattac ggacttagtg ttcttcaaca gcatgtctoc ttottogata togatgataa    240
tggcatoatt taccottggg agacotactc tggactgoga atgottgggt tcaatatcat    300
tgggtcgctt ataatagccg ctgttatcaa cctgacocctt agctatgcca ctcttccggg    360
gtggttacct tcaactttct tccctatata cctacacaac atacacaagt caaagcatgg    420
aagtgattca aaaacatctg acaatgaagg aaggtttatg ccggtgaate ttgagttgat    480
atttagcaaa tatgogaaaa ccttgccaga caagttgagt ctgggagaac tatgggagat    540
gacagaagga aaccgtgacg cttgggacat ttttggatgg atcgcaggca aaatagagtg    600
gggactggtg tacttgctag caagggatga agaagggttt ttgtcaaaaag aagctattag    660
gcgggtgttc gatggaagct tgttcgagta ctgtgcaaaa atctaogctg gtatcagtgga    720
agacaagaca gcatactacg ccatggat                                     748
```

<210> 159

<211> 738

<212> ADN

<213> *Arabidopsis thaliana*

<400> 159

```
atgggggtcaa agacggagat gatggagaga gacgcaatgg ctacgggtggc tccctatgcg    60
ccggtcactt accaccgocg tgctcgtggt gacttggatg atagacttcc taaaccttat    120
atgccaagag cattgcaagc accagacaga gaacaccctg acggaactcc aggccataag    180
aattacggac ttagtgttct tcaacagcat gtctccttct tcgatatcga tgataatggc    240
atcatttacc cttgggagac ctactctgga ctgcgaatgc ttggtttcaa tatcattggg    300
tcgcttataa tagccgctgt tatcaacctg accottagct atgccactct tccgggggtg    360
ttaccttcac ctttcttccc tatatacata cacaacatac acaagtcaaa gcatggaagt    420
gattcaaaaa catatgacaa tgaaggaagg tttatgccgg tgaatcttga gttgatattt    480
```

agcaaatatg cgaaaacctt gccagacaag ttgagtcttg gagaactatg ggagatgaca 540
gaaggaaacc gtgacgcttg ggacatTTTT ggatggatcg caggcaaaat agagtgggga 600
ctgttgtaact tgctagcaag ggatgaagaa gggTTTTTgt caaagaagc tattaggcgg 660
tgTTTTgatg gaagcttggt cgagtactgt gccaaaatct acgctggtat cagtgaagac 720
aagacagcat actactaa 738

<210> 160

<211> 1047

<212> ADN

<213> *Sesamum indicum*

<400> 160

atggatctaa tccacacttt cctcaactta atagetcccc ctttcacott ctctctcett 60
ctctTTTTot tgcacacctt ccagattttc aagttottcc ttcaaatett gggcaacctt 120
ttcagcgggg atgtccttgg aaaagtctgc gtcactcccc gcgcctcttc cggcatcggc 180
gaaagtcttg cttacagagta tgctaagaga ggggctgtct tgggtcttgc tgcaagaagg 240
gaaaggagtc ttcaagaagt ggcgaaagg gcgcgcgatt tggggtctgc ggaactcttg 300
gtggtccggg ccgatgtttc gaaggcggag gactgcagga aggttgttga tcagaactatg 360
aatcgtcttg gaagattgga tcacctggtc aataacgctg gaattatgtc agtttcaatg 420
ctggaagaag ttgaagatat tactggttac agagaaacta tggatatcaa cttctggggc 480
tatgtgtata tgacctgatt tgccgcccca tcccttagga atagcagagg ccgaattggt 540
gtactttctt cctccagttc ttggatgctt actccgagga tgagtTTTTa caatgcaagc 600
aaagcggcga tttcacaatt ttttgagaca ctgcgggtgg aattcggccc cgatataggc 660
ataaccttg tgactccagg attcatagaa tctgaaacta cccaaggcaa attctacaat 720
gctggcgaa cgtgtaattga tcaggacatg agagatgtac aagtgagcac gactccaatc 780
ctgaggggtg aaagtgcggc aaggtaactc gtgaggagcg cgatccgtgg agaaagatac 840
gtgacagagc cggccttggt tagggttact tattgggtga agctattctg cactgaggtg 900
atggagtggg tatttagact gatgtacttg gccagccccg gtgagccgga gaaagaaacg 960
tttgcaaga aggttttga ttacacagga gtgaagtctt tgccttcccc ggaacctgtg 1020
caagttccgg agcccaagaa tgattaa 1047

<210> 161

<211> 25

<212> PRT

<213> Tabaco, sequência sinal relacionada com proteína patogénica (PR-S)

<400> 161

Met Asn Phe Leu Lys Ser Phe Pro Phe Tyr Ala Phe Leu Cys Phe Gly
1 5 10 15

Gln Tyr Phe Val Ala Val Thr His Ala
20 25

<210> 162

<211> 64

<212> PRT

<213> Sequência Artificial

<220>

<223> Sequência líder do factor alfa

<400> 162

Ala Pro Val Asn Thr Thr Glu Asp Glu Thr Ala Gln Ala Glu Ala Val
1 5 10 15

Ile Gly Tyr Ser Asp Leu Glu Gly Asp Phe Asp Val Ala Val Leu Pro
20 25 30

Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu Phe Ile Asx Thr Thr Ile
35 40 45

Ala Ser Ile Ala Ala Lys Glu Glu Gly Val Ser Leu Met Ala Lys Arg
50 55 60

<210> 163

<211> 63

<212> PRT

<213> Sequência Artificial

<220>

<223> Sequência líder do factor alfa

<400> 163

Ala Pro Val Asn Thr Thr Glu Asp Glu Thr Ala Gln Ala Glu Ala Val
1 5 10 15

Ile Gly Tyr Ser Asp Leu Glu Gly Asp Phe Asp Val Ala Val Leu Pro
20 25 30

Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu Phe Ile Asx Thr Thr Ile
35 40 45

Ala Ser Ile Ala Ala Lys Glu Glu Gly Val Ser Met Ala Lys Arg
50 55 60

<210> 164

<211> 15

<212> PRT

<213> Sequência Artificial

<220>

<223> Sequência líder do factor alfa

<400> 164

Gln Pro Ile Asp Glu Asp Asn Asp Thr Ser Ser Met Ala Lys Arg
1 5 10 15

<210> 165

<211> 43

<212> PRT

<213> Sequência Artificial

<220>

<223> Sequência líder do factor alfa

<400> 165

Gln Pro Ile Asp Asp Thr Glu Ser Asn Thr Thr Ser Val Asn Leu Met
1 5 10 15

Ala Asp Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Thr Leu Ala Leu
20 25 30

Asp Val Val Asn Leu Ile Ser Met Ala Lys Arg
35 40

<210> 166

<211> 43

<212> PRT

<213> Sequência Artificial

<220>

<223> Sequência líder do factor alfa

<400> 166

Gln Pro Ile Asp Asp Thr Glu Ser Gln Thr Thr Ser Val Asn Leu Met
1 5 10 15

Ala Asp Asp Thr Glu Asp Arg Phe Ala Thr Gln Thr Thr Leu Ala Leu
20 25 30

Asp Val Val Asn Leu Ile Ser Met Ala Lys Arg
35 40

<210> 167

<211> 43

<212> PRT

<213> Sequência Artificial

<220>

<223> Sequência líder do factor alfa

<400> 167

Gln Pro Ile Asp Asp Thr Glu Ser Gln Thr Thr Ser Val Asn Leu Met
1 5 10 15

Ala Asp Asp Thr Glu Asp Arg Phe Ala Thr Gln Thr Thr Leu Ala Leu
20 25 30

Asp Val Val Asn Leu Ile Ser Met Ala Ala Ala
35 40

<210> 168

<211> 43

<212> PRT

<213> Sequência Artificial

<220>

<223> Sequência líder do factor alfa

<400> 168

Gln Pro Ile Asp Asp Thr Glu Ser Asn Thr Thr Ser Val Asn Leu Met
1 5 10 15

Ala Asp Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Thr Leu Ala Leu
20 25 30

Asp Val Val Asn Leu Ile Ser Met Ala Ala Ala
35 40

<210> 169

<211> 45

<212> PRT

<213> Sequência Artificial

<220>

<223> Sequência líder do factor alfa

<400> 169

Gln Pro Ile Asp Asp Thr Glu Ser Asn Thr Thr Ser Val Asn Leu Met
1 5 10 15

Ala Asp Asp Thr Glu Asp Arg Phe Ala Thr Asn Thr Thr Leu Ala Gly
20 25 30

Gly Leu Asp Val Val Asn Leu Ile Ser Met Ala Lys Arg
35 40 45

<210> 170

<211> 44

<212> PRT

<213> Sequência Artificial

<220>

<223> Sequência líder do factor alfa

<400> 170

Gln Pro Ile Asp Asp Thr Glu Ser Gln Thr Thr Ser Val Asn Leu Met
1 5 10 15

Ala Asp Asp Thr Glu Ser Ala Phe Ala Thr Gln Thr Asn Ser Gly Gly
20 25 30

Leu Asp Val Val Gly Leu Ile Ser Met Ala Lys Arg
35 40

<210> 171

<211> 44

<212> PRT

<213> Sequência Artificial

<220>

<223> Sequência líder do factor alfa

<400> 171

Gln Pro Ile Asp Asp Thr Glu Ser Gln Thr Thr Ser Val Asn Leu Met
1 5 10 15

Ala Asp Asp Thr Glu Ser Ala Phe Ala Thr Gln Thr Asn Ser Gly Gly
20 25 30

Leu Asp Val Val Gly Leu Ile Ser Met Ala Ala Ala
35 40

<210> 172

<211> 10

<212> PRT

<213> Péptido espaçador

<400> 172

Glu Glu Ala Glu Ala Glu Ala Glu Pro Lys
1 5 10

<210> 173

<211> 6

<212> PRT

<213> Sequência Artificial

<220>

<223> Péptido espaçador

<400> 173

Glu Glu Gly Glu Pro Lys
1 5

<210> 174

<211> 40

<212> PRT

<213> Sequência Artificial

<220>

<223> Sítio de clivagem

<400> 174

Met Ala Glu Ile Thr Arg Ile Pro Leu Tyr Lys Gly Lys Ser Leu Arg
1 5 10 15

Lys Ala Leu Lys Glu His Gly Leu Leu Glu Asp Phe Leu Gln Lys Gln
20 25 30

Gln Tyr Gly Ile Ser Ser Lys Phe
35 40

<210> 175

<211> 27

<212> ADN

<213> Sequência Artificial

<220>

<223> Iniciador

<400> 175

gcctgctgac attgtgatga cacagtc

27

<210> 176

<211> 43

<212> ADN

<213> Sequência Artificial

<220>

<223> Iniciador

<400> 176

aagcttgcat ttaaatactc gagactgtga gactggtgcc ttg

43

<210> 177

<211> 70

<212> ADN

<213> Sequência Artificial

<220>

<223> Iniciador

<400> 177

gaagaaggag agocaaagt tgtaaatcaa catctttgtg gatctcatct tgttgaggct 60

ctctaccttg 70

<210> 178

<211> 56

<212> ADN

<213> Sequência Artificial

<220>

<223> Iniciador

<400> 178

ccttaggagt gtagaaaaat cctctttctc cacacacaag gtagagagcc tcaaca 56

<210> 179

<211> 21

<212> ADN

<213> Sequência Artificial

<220>

<223> Iniciador

<400> 179

ctaaggctgc taaggaatt g 21

<210> 180

<211> 83

<212> ADN

<213> Sequência Artificial

<220>

<223> Iniciador

<400> 180

aagcttcagt tgcaatagtt ctccaattgg taaagtgcgc aaatagaagt gcaacattgt 60

tcaacaatto ccttagcgc att 83

<210> 181

<211> 30

<212> ADN

<213> Sequência Artificial

<220>

<223> Iniciador

<400> 181

ctcgagtcaa ccaattgatg acactgaatc 30

<210> 182

<211> 41

<212> ADN

<213> Sequência Artificial

<220>

<223> Iniciador

<400> 182

aagcttcasa gttcatcctt gttgcaatag ttctccaatt g

41

<210> 183

<211> 21

<212> ADN

<213> Sequência Artificial

<220>

<223> Iniciador

<400> 183

aagcttcagt tgcattagtt c

21

<210> 184

<211> 26

<212> ADN

<213> Sequência Artificial

<220>

<223> Iniciador

<400> 184

gcattgcccc ccaattgatg acactg

26

<210> 185

<211> 34

<212> ADN

<213> Sequência Artificial

<220>

<223> Iniciador

<400> 185

gcacgcatgc ctttgtaat caacatctt gtgg

34

<210> 186

<211> 54

<212> ADN

<213> Sequência Artificial

<220>

<223> Iniciador

<400> 186

acattgtca acaattcctc tctttctct agtcttagga gtgtagaaa atcc

54

<210> 187

<211> 29

<212> ADN

<213> Sequência Artificial

<220>

<223> Iniciador

<400> 187

gcataagctt caaagctcat cctttgagc

29

<210> 188

<211> 387

<212> ADN

<213> Sequência Artificial

<220>

<223> Sequência de ácido nucleico da proteína de fusão insulina

<400> 188

atgaacttc ttaagtcttt ccctttctac gctttccttt gtttcggta atacttcgtt	50
gctgttacgc atgeccaacc aattgatgac actgaatccc agaccacgta agtgaacctc	120
atggcogatg atactgagag cgcgctttgct acacaaacaa attcgggagg tcttgacggt	180
gtcggattga tctccatggc tsagagagaa gaaggagagc ctaagtttgt taatcaacat	240
ctttgtggat ctcatcttgt tgaggctctc taccttgtgt gtggagaaag aggatttttc	300
tacactccta aggotgctaa gggcattggt gaacaatggt gcacttctat ttgctcaact	360
taccaattgg agaactattg caactga	387

<210> 189

<211> 128

<212> PRT

<213> Sequência Artificial

<220>

<223> Proteína factor insulina

<400> 189

Met Asn Phe Leu Lys Ser Phe Pro Phe Tyr Ala Phe Leu Cys Phe Gly
1 5 10 15

Gln Tyr Phe Val Ala Val Thr His Ala Gln Pro Ile Asp Asp Thr Glu
20 25 30

Ser Gln Thr Thr Ser Val Asn Leu Met Ala Asp Asp Thr Glu Ser Ala
35 40 45

Phe Ala Thr Gln Thr Asn Ser Gly Gly Leu Asp Val Val Gly Leu Ile
50 55 60

Ser Met Ala Lys Arg Glu Glu Gly Glu Pro Lys Phe Val Asn Gln His
65 70 75 80

Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu
85 90 95

Arg Gly Phe Phe Tyr Thr Pro Lys Ala Ala Lys Gly Ile Val Glu Gln
100 105 110

Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
115 120 125

<210> 190

<211> 18

<212> ADN

<213> Sequência Artificial

<220>

<223> Iniciador

<400> 190

tttgtgaacc aacacttg

18

<210> 191
<211> 20
<212> ADN
<213> Sequência Artificial

<220>
<223> Iniciador

<400> 191
aagcttttcag ttacagtagt 20

<210> 192
<211> 17
<212> ADN
<213> Sequência Artificial

<220>
<223> Iniciador

<400> 192
gcctgcctgt gttgagc 17

<210> 193
<211> 16
<212> ADN
<213> Sequência Artificial

<220>
<223> Iniciador

<400> 193

ggtagtgtgc tgcca

16

<210> 194

<211> 38

<212> ADN

<213> Sequência Artificial

<220>

<223> Iniciador

<400> 194

ggtgcccagc acactacett cgtgaaccaa cacitgtg

38

<210> 195

<211> 1020

<212> ADN

<213> Sequência Artificial

<220>

<223> Sequência de ácido nucleico da insulina proteína de fusão

<400> 195

```
atggcggata cagctagagg aaccatcac gatatcatcg gcagagacca gtaccgatg      60
atgggcccag accgagacca gtaccgatg tccggacgag gatctgacta ctccaagtct      120
aggcagattg ctaaagctgc aactgetgto acagctggcg gttccctctt tgtttctctc      180
agccttacco ttggttggac tgtcatagct ttgactggtg caacacctct gctogttatc      240
ttcagcccaa tctttgtccc ggetctcctc acagttgcac tctcctcac cggttttctt      300
tcctctggag ggtttggcat tgcctctata accgttttct cttggattta caagtaagca      360
cacctttatc atcttacttc ataattttgt gcaatctgtg catgcatgtg ttgagccagt      420
agctttggat caattttttt ggtagaataa caaatgtaac aataagaaat tgcataattct      480
agggaaacatt tggttaacta aatacgaat ttagcctagc tagcttgaat gtgtctgtgt      540
atatcatctc tataggtaaa atgcttggta tgatacctat tgattgtgaa taggtacgca      600
acggggagac acccacaggg atcagacaag ttggacagtg caaggatgaa gttgggaage      660
aaagctcagg atctgaaaga cagagctcag tactacggac agcaacatac tggtygggaa      720
catgaccgtg accgtactcg tggtyggcag cacactacct tctgtaacca acacttctgt      780
ggatctcctc tcttgaage totctacttg gtttytggtg agagaggatt cttctacact      840
cctaagacca gaaggggaagc tgaggacttg caggtgggac aagttgagtt ggttggaggt      900
cttggagcag gatctttgca acctctcctt ttggaaggtt ctttgcagaa gagaggaatc      960
gttgaacaat gttgcacttc aatctgttct ttgtatcagt tggagaacta ctgtaactga     1020
```

<210> 196

<211> 257

<212> PRT

<213> Sequência Artificial

<220>

<223> Insulina proteína de fusão

<400> 196

Met Ala Asp Thr Ala Arg Gly Thr His His Asp Ile Ile Gly Arg Asp
1 5 10 15

Gln Tyr Pro Met Met Gly Arg Asp Arg Asp Gln Tyr Gln Met Ser Gly
20 25 30

Arg Gly Ser Asp Tyr Ser Lys Ser Arg Gln Ile Ala Lys Ala Ala Thr
35 40 45

Ala Val Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Ser Leu Thr Leu
50 55 60

Val Gly Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile
65 70 75 80

Phe Ser Pro Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Leu Ile
85 90 95

Thr Gly Phe Leu Ser Ser Gly Gly Phe Gly Ile Ala Ala Ile Thr Val
100 105 110

Phe Ser Trp Ile Tyr Ala Thr Gly Glu His Pro Gln Gly Ser Asp Lys
115 120 125

Leu Asp Ser Ala Arg Met Lys Leu Gly Ser Lys Ala Gln Asp Leu Lys
130 135 140

Asp Arg Ala Gln Tyr Tyr Gly Gln Gln His Thr Gly Gly Glu His Asp
145 150 155 160

Arg Asp Arg Thr Arg Gly Gly Gln His Thr Thr Phe Val Asn Gln His
165 170 175

Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu
180 185 190

Arg Gly Phe Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu
195 200 205

Gln Val Gly Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu
210 215 220

REIVINDICAÇÕES

1. Método para a expressão de insulina em sementes de plantas compreendendo:

(a) proporcionar uma construção de ácido nucleico quimérica compreendendo na direcção 5' para 3' da transcrição como componentes ligados operacionalmente:

(i) uma sequência de ácido nucleico que compreende um promotor preferido das sementes;

(ii) uma sequência de ácido nucleico codificando um polipéptido da insulina; e

(iii) uma sequência de ácido nucleico codificando o polipéptido que é capaz de reter um polipéptido da insulina no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER;

(b) introduzir a construção de ácido nucleico quimérica numa célula vegetal; e

(c) cultivar a célula vegetal até ficar uma planta madura capaz de produzir sementes em que a semente expressa insulina, em que um polipéptido da insulina se acumula no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER e em que, pelo menos, 0,1% da proteína total de semente presente na semente é insulina; e

(d) obter insulina substancialmente pura das referidas sementes.

2. Método de acordo com a reivindicação 1, em que o referido polipéptido retendo um polipéptido da insulina no ER é seleccionado do grupo consistindo em KDEL, HDEL, DDEL, ADEL e SDEL.
3. Método de acordo com a reivindicação 2, em que o referido polipéptido retendo um polipéptido da insulina no ER é seleccionado do grupo consistindo em SEQ ID N°:150, SEQ ID N°:151, SEQ ID N°:152, SEQ ID N°:153 e SEQ ID N°:154.
4. Método de acordo com a reivindicação 1, em que a referida vesícula de armazenamento derivada de ER é um corpo oleoso.
5. Método de acordo com a reivindicação 1, em que o referido polipéptido retendo um polipéptido da insulina numa vesícula de armazenamento derivada de ER é uma proteína do corpo oleoso.
6. Método de acordo com a reivindicação 5, em que a referida proteína de corpo oleoso é seleccionada do grupo de proteínas de corpo oleoso consistindo em oleosina, caleosina e esteroleosina.
7. Método de acordo com a reivindicação 5, em que a referida proteína do corpo oleoso é seleccionada do grupo consistindo em SEQ ID N°: 156, SEQ ID N°:157 e SEQ ID N°:158, SEQ ID N°:159 e SEQ ID N°:160.

8. Método de acordo com a reivindicação 1, em que o referido ácido nucleico quimérico contém adicionalmente uma sequência de ácido nucleico codificando uma proteína de estabilização fundida em grelha de leitura com a sequência de ácido nucleico codificando insulina.
9. Método de acordo com a reivindicação 8, em que o referido ácido nucleico quimérico contém adicionalmente uma sequência de ácido nucleico codificando uma sequência de péptido sinal fundida em grelha de leitura com a sequência de ácido nucleico codificando insulina.
10. Método de acordo com a reivindicação 9, em que o referido péptido sinal é uma sequência sinal da proteína relacionada com a patogénese do tabaco (PR-S).
11. Método de acordo com a reivindicação 10, em que o referido péptido sinal é SEQ ID N°:161.
12. Método de acordo com a reivindicação 8, em que o referido ácido nucleico codificando a referida proteína estabilizante permite a associação de um polipéptido da insulina com o corpo oleoso após recolha e moagem da semente.
13. Método de acordo com a reivindicação 12, em que a referida proteína estabilizante codifica um anticorpo de cadeia simples com especificidade para um corpo oleoso.

14. Método de acordo com a reivindicação 8, em que a sequência de ácido nucleico codificando uma proteína estabilizante fundida em grelha de leitura com a sequência de ácido nucleico codificando insulina é seleccionada do grupo de polipéptidos consistindo num anticorpo de cadeia simples e na subunidade da toxina B da cólera.
15. Método de acordo com a reivindicação 1, em que a sequência de ácido nucleico quimérica é introduzida na célula vegetal em condições de integração genómica nuclear.
16. Método de acordo com a reivindicação 1, em que o promotor preferido das sementes é um promotor da faseolina.
17. Método de acordo com as reivindicações 1 a 16, em que a sequência de ácido nucleico codificando insulina é seleccionada do grupo de sequências de ácido nucleico consistindo em insulina humana, insulina porcina e insulina bovina.
18. Método de acordo com as reivindicações 1 a 17, em que o ácido nucleico codificando insulina é uma mini-insulina.
19. Método de qualquer uma das reivindicações 1 a 18, em que a sequência de ácido nucleico codificando insulina é optimizada para utilização de códons de planta.
20. Método para obter sementes de plantas compreendendo insulina, compreendendo:

(a) proporcionar uma construção de ácido nucleico quimérica compreendendo na direcção 5' para 3' da transcrição como componentes ligados operacionalmente:

(i) uma sequência de ácido nucleico que compreende um promotor preferido das sementes;

(ii) uma sequência de ácido nucleico codificando um polipéptido da insulina; e

(iii) uma sequência de ácido nucleico codificando um polipéptido que é capaz de reter um polipéptido da insulina no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER;

(b) introduzir a construção de ácido nucleico quimérica numa célula vegetal;

(c) cultivar a célula vegetal até ficar uma planta madura capaz de produzir sementes; e

(d) obter sementes da referida planta em que a semente compreende insulina, em que o referido polipéptido da insulina se acumula no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER e em que, pelo menos, 0,1% da proteína solúvel total presente na semente é insulina.

21. Planta capaz de produzir sementes compreendendo uma sequência de ácido nucleico quimérica compreendendo na direcção 5' para 3' da transcrição:

(a) uma primeira sequência de ácido nucleico que compreende um promotor preferido das sementes ligado operacionalmente a;

(b) uma segunda sequência de ácido nucleico codificando um polipéptido da insulina; e

(c) uma sequência de ácido nucleico codificando um polipéptido que é capaz de reter um polipéptido da insulina no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER, em que a semente contém insulina, em que o referido polipéptido da insulina se acumula no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER e em que, pelo menos, 0,1% da proteína solúvel total presente na semente é insulina.

22. Semente de planta compreendendo uma sequência de ácido nucleico quimérica compreendendo na direcção 5' para 3' da transcrição:

(a) uma primeira sequência de ácido nucleico que compreende um promotor preferido das sementes ligadas operacionalmente a;

(b) uma segunda sequência de ácido nucleico codificando um polipéptido da insulina; e

(c) a sequência de ácido nucleico codificando um polipéptido que é capaz de reter um polipéptido da insulina no retículo endoplásmico (ER) ou numa

vesícula de armazenamento derivada de ER.

23. Sequência de ácido nucleico codificando insulina ligada a uma sequência de ácido nucleico compreendendo um promotor preferido das sementes ligado a uma sequência de ácido nucleico codificando um polipéptido que é capaz de reter um polipéptido da insulina no retículo endoplásmico (ER) ou numa vesícula de armazenamento derivada de ER.

Lisboa, 22 de Outubro de 2008

FIGURA 1-1

1 atgaacttccttaagtctttccctttctacgctttcctttgtttcgggtcaatacttcggt 60
 1 M N F L K S F P F Y A F L C F G Q Y F V 20

 61 gctgttacgcatgctgacattgtgatgacacagtcctccatcctccctggctatgtcagtg 120
 21 A V T H A D I V M T Q S P S S L A M S V 40

 121 ggacagcgggtcactatgcgctgcaagtcagtcagagccttttaaaaagtaccaatcaa 180
 41 G Q R V T M R C K S S Q S L L K S T N Q 60

 181 aagaactatttggcctgggtaccagcagaaaccaggacagtcctcctaaacttctggatac 240
 61 K N Y L A W Y Q Q K P G Q S P K L L V Y 80

 241 tttgcatccactaggaatctggggtccctgatcgcttcataggcagtgatctgggaca 300
 81 F A S T R E S G V P D R F I G S G S G T 100

 301 gatttcactcttaccatcagcagtggtgaggctgaagacctggcagattacttctgtcag 360
 101 D F T L T I S S V Q A E D L A D Y F C Q 120

 361 caacattataaacactcctcccagttcgggtgctgggaccaagctggagcttaagcggctc 420
 121 Q H Y N T P P T F G A G T K L E L K R S 140

 421 ccgaacgggtgcttctcatagcggttctgcaccaggcactagctctgcatctggatctcag 480
 141 P N G A S H S G S A P G T S S A S G S Q 160

 481 gtgcacctgcagcagtcctggagctgagctgatgaagcctggggcctcaatgaagatccc 540
 161 V H L Q Q S G A E L M K P G A S M K I S 180

 541 tgcaaggctactggctacacattcagtagctactggatagagtgggtaagcagaggcct 600
 181 C K A T G Y T F S S Y W I E W V K Q R P 200

 601 ggacatggccttgagtgattggagagattttacctggcagtggtagtactacctacaat 660
 201 G H G L E W I G E I L P G S G S T T Y N 220

 661 gagaagttcaagggcaaggccacattcactgcagatacatcctccaacacagcctacatg 720
 221 E K F K G K A T F T A D T S S N T A Y M 240

 721 caactcagcagcctgacatctgaggactctgcccgtctattactgtgcaagattggatggt 780
 241 Q L S S L T S E D S A V Y Y C A R L D V 260

 781 gactcctggggccaaggcaccactctcacagtcctcgagtcacccaattgatgacactgaa 840
 261 D S W G Q G T T L T V S S Q P I D D T E 280

 841 tcccagaccagtcagtgaaacctcatggccgatgatactgagagcgcgtttgctacacaa 900
 281 S Q T T S V N L M A D D T E S A F A T Q 300

 901 acaaattcggggaggtcttgacgttgtcggattgatctccatggctaagagagaagaagga 960
 301 T N S G G L D V V G L I S M A K R E E G 320

 961 gagcctaagtttgtaatacaacatctttgtggatctcatcttgttgaggctctctacctt 1020
 321 E P K F V N Q H L C G S H L V E A L Y L 340

FIGURA 1-2

1021 gtgtgtggagaaagaggatTTTTctacactcctaaggctgctaaggaattggtgaacaa 1080
341 V C G E R G F F Y T P K A A K G I V E Q 360

1081 tgttgcaacttctatTTgctcactttaccaattggagaactattgcaacaaggatgaactt 1040
361 C C T S I C S L Y Q L E N Y C N K D E L 380

1041 tga
-

FIGURA 2

1 atggcgggatacagctagaggaaccatcacgatatcatcggcagagaccagtaccogtg 60
 1 M A D T A R G T H H D I I G R D Q Y P M 20

 61 atgggccgagaccgagaccagtaccagatgtccggacgaggatctgactactccaagtct 120
 21 M G R D R D Q Y Q M S G R G S D Y S K S 40

 121 aggcagattgctaaagctgcaactgctgtcacagctgggtggttccctccttgttctctcc 180
 41 R Q I A K A A T A V T A G G S L L V L S 60

 181 agccttacccttgttggaaactgtcatagctttgactggtgcaacacctctgctcgttatc 240
 61 S L T L V G T V I A L T V A T P L L V I 80

 241 ttcagcccaatccttgtccoggtctctcatcacagttgcactcctcatcaccggttttctt 300
 81 F S P I L V P A L I T V A L L I T G F L 100

 301 tcctctggagggttggcattgccgctataaccgttttctcttggatttacgcaacggga 360
 101 S S G G F G I A A I T V F S W I Y A T G 120

 361 gagcaccacagggatcagacaagttggacagtgcaaggatgaagttgggaagcaaagct 420
 121 E H P Q G S D K L D S A R M K L G S K A 140

 421 caggatctgaaagacagagctcagtactacggacagcaacatactgggtggggaacatgac 480
 141 Q D L K D R A Q Y Y G Q Q H T G G E H D 160

 481 cgtgaccgtactcgtgggtggccagcacactaccatggctgagatcaccgcattcctctc 540
 161 R D R T R G G Q H T T M A E I T R I P L 180

 541 taaaaaggttaagtctctccgtaaggcgctgaaggaacatggacttctagaagacttcttg 600
 181 Y K G K S L R K A L K E H G L L E D F L 200

 601 cagaaacaacagtatggcatctcgagcaagttccaaccaattgatgacactgaatcccag 660
 201 Q K Q Q Y G I S S K F Q P I D D T E S Q 220

 661 accacgtcagtgaacctcatggccgatgatactgagagcgcgtttgctacacaaacaat 720
 221 T T S V N L M A D D T E S A F A T Q T N 240

 721 tcgggaggtcttgacgttgtcggattgatctccatggctaagagagaagaaggagagcct 780
 241 S G G L D V V G L I S M A K R E E G E P 260

 781 aagtttgtaatacaacatctttgtggatctcatcttgttgaggctctctaccttgtgtgt 840
 261 K F V N Q H L C G S H L V E A L Y L V C 280

 841 ggagaaagaggatttttctacactcctaaggctgctaaggggaattgttgaacaatgttgc 900
 281 G E R G F F Y T P K A A K G I V E Q C C 300

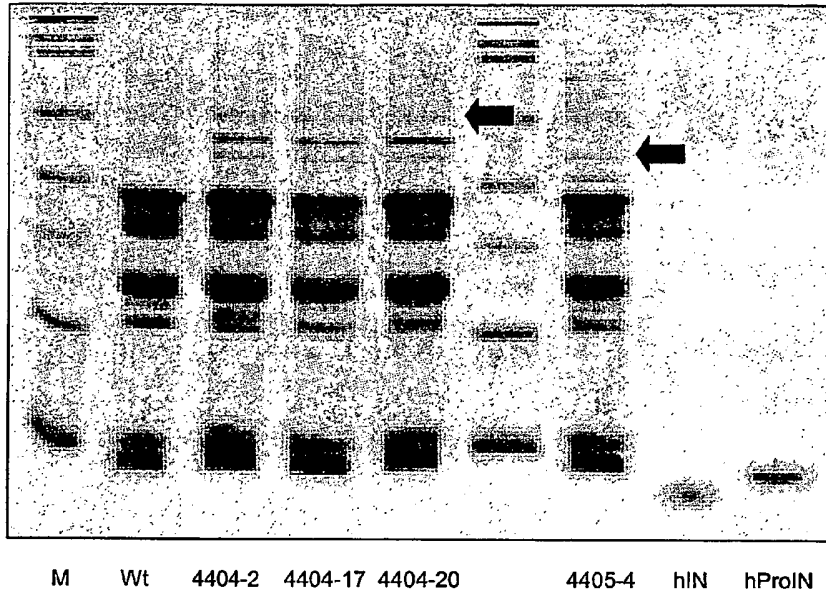
 901 acttctatgtgtcactttaccaattggagaactattgcaactga
 301 T S I C S L Y Q L E N Y C N -

FIGURA 3

1	atgaacttccttaagtctttccctttctacgctttcctttgtttcgggtcaatacttcggt	60
1	M N F L K S F P F Y A F L C F G Q Y F V	20
61	gctgttacgcatgcctttgtaatacaacatctttgtggatctcatctgttgaggctctc	120
21	A V T H A F V N Q H L C G S H L V E A L	40
121	taccttggtggtggagaaagaggatTTTTctacactcctaagactagaagaaagagagga	180
41	Y L V C G E R G F F Y T P K T R R K R G	60
181	attggtgaacaatggtgcacttctatttgctcactttaccaattggagaactattgcaac	240
61	I V E Q C C T S I C S L Y Q L E N Y C N	80
241	agaagaaagagagacattgtgatgacacagctccatcctccctggctatgtcagtgga	300
81	<u>R R K R</u> D I V M T Q S P S S L A M S V G	100
301	cagcgggtcactatgcgctgcaagtccagtcagagccttttaaaaagtaccaatcaaaag	360
101	Q R V T M R C K S S Q S L L K S T N Q K	120
361	aactatttggcctgggtaccagcagaaaccaggacagctctcctaaacttctgggtatacttt	420
121	N Y L A W Y Q Q K P G Q S P K L L V Y F	140
421	gcatccactaggaatctgggtccctgatcgcttcataggcagtggtatctgggacagat	480
141	A S T R E S G V P D R F I G S G S G T D	160
481	ttcactcttaccatcagcagtggtgcaggetgaagacctggcagattacttctgtcagcaa	540
161	F T L T I S S V Q A E D L A D Y F C Q Q	180
541	cattataacactcctcccacgttcgggtgctgggaccaagttggagcttaagcgggtctccg	600
181	H Y N T P P T F G A G T K L E L K R S P	200
601	aacgggtccttctcatagcgggtctgcaccaggcactagctctgcatctggatctcaggtg	660
201	N G A S H S G S A P G T S S A S G S Q V	220
661	cacctgcagcagctggagctgagctgatgaagcctggggcctcaatgaagatatcctgc	720
221	H L Q Q S G A E L M K P G A S M K I S C	240
721	aaggctactggctacacattcagtagctactggatagagtggttaaagcagaggcctgga	780
241	K A T G Y T F S S Y W I E W V K Q R P G	260
781	catggccttgagtggttggtgagagatTTTtacctggcagtggtagtactacctacaatgag	840
261	H G L E W I G E I L P G S G S T T Y N E	280
841	aagttcaagggcaaggccacattcactgcagatacatcctccaacacagcctacatgcaa	900
281	K F K G K A T F T A D T S S N T A Y M Q	300
901	ctcagcagcctgacatctgaggactctgccgtctattactgtgcaagattggatggtgac	960
301	L S S L T S E D S A V Y Y C A R L D V D	320
961	tcctggggccaaggcaccactctcacagtgagctcaaaggatgagccttga	
321	S W G Q G T T L T V S S <u>K D E L</u> -	

FIGURA 4

A)



B)

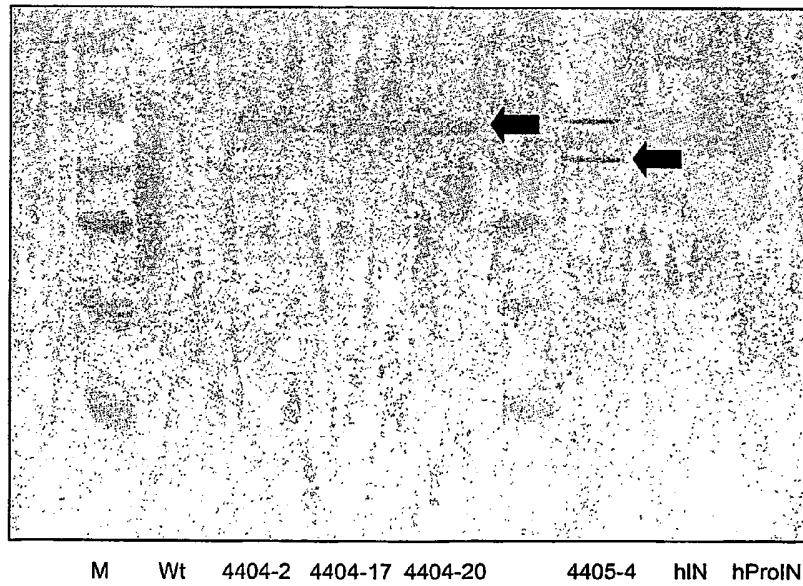
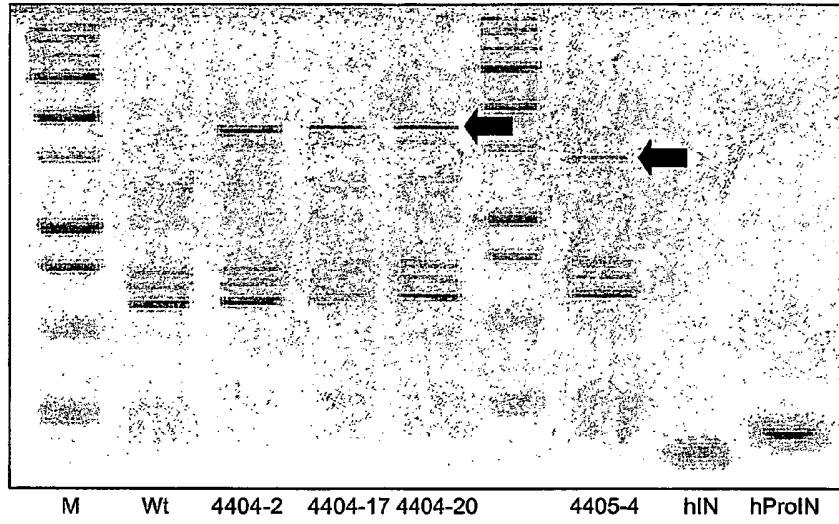


FIGURA 4-2

C)



D)

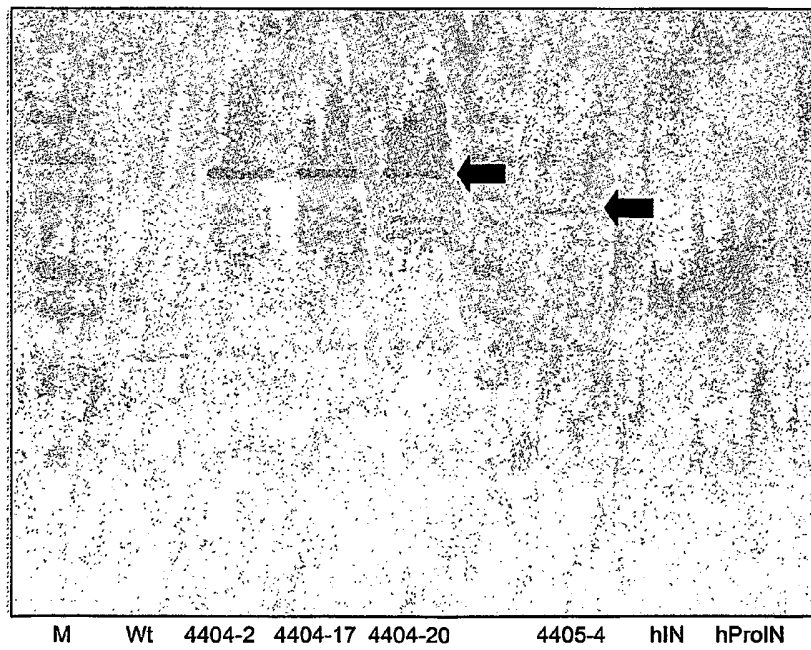
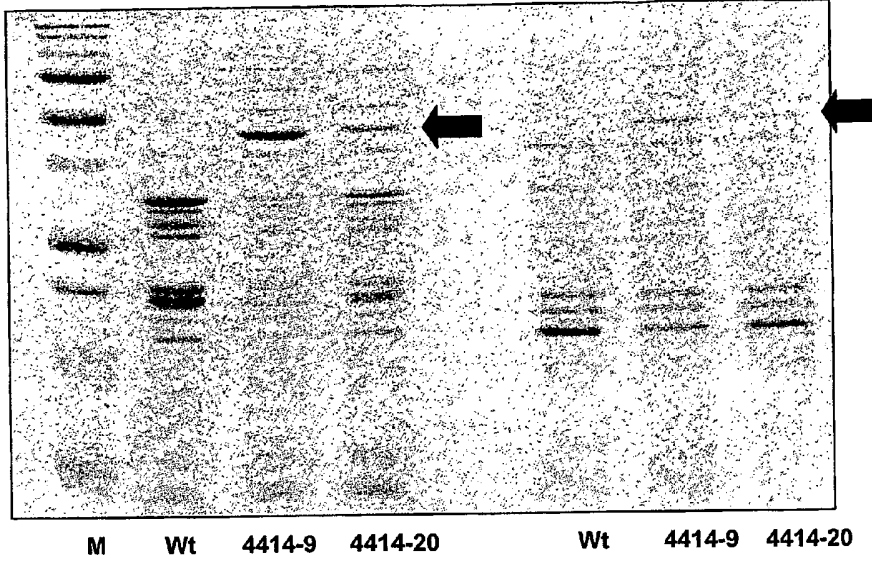


FIGURA 4-3

E)



F)

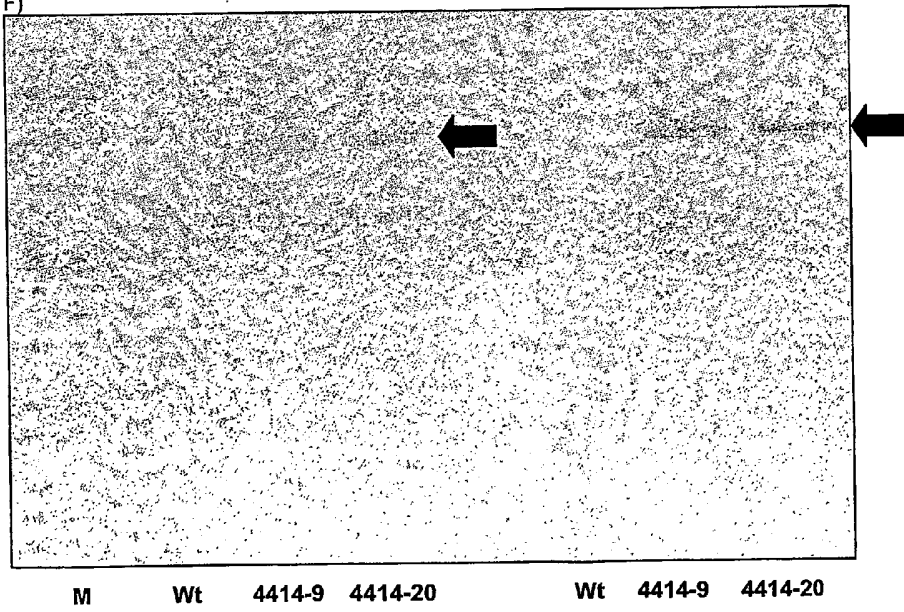
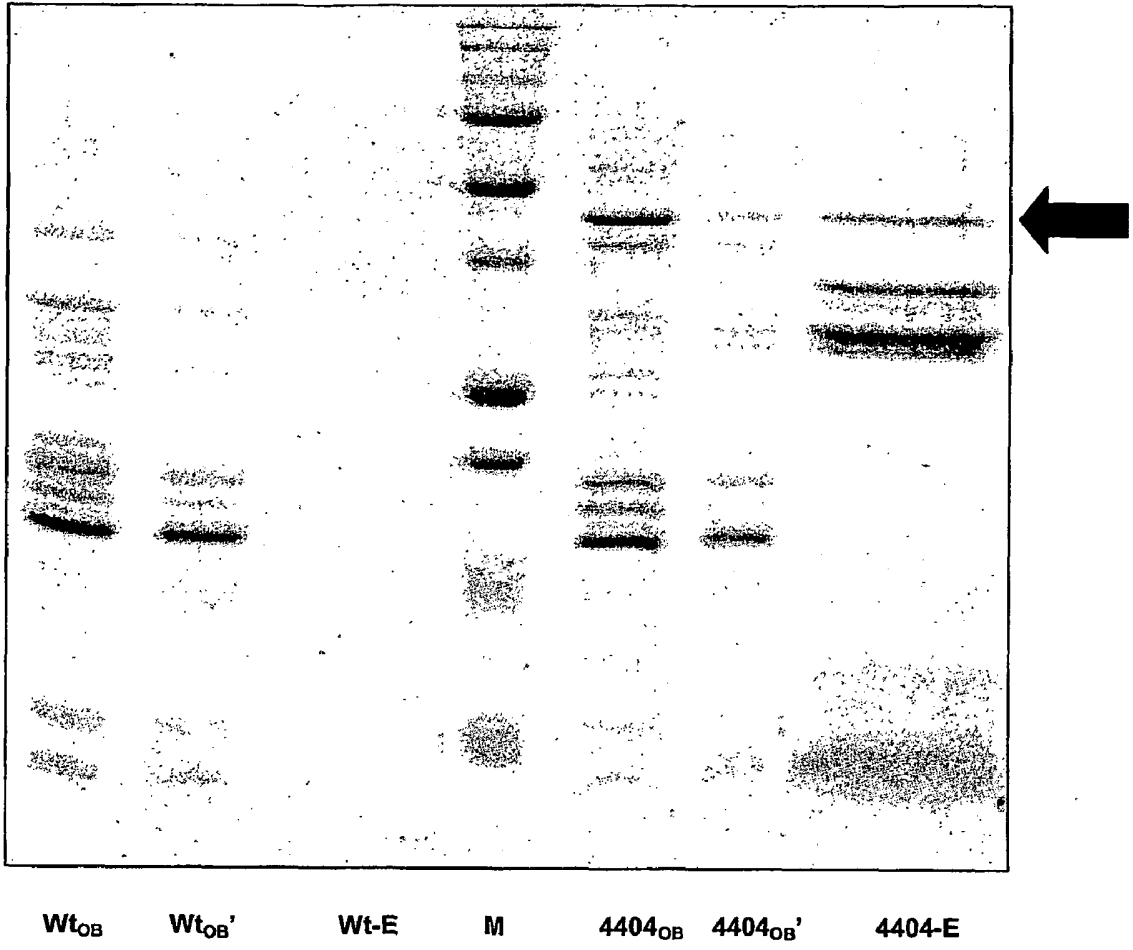


FIGURA 5

Construção	Linha	Transgene (% de proteína total de semente)	Mini-insulina (% de proteína total de semente)
4404	2	1.20	0.20
4404	17	1.09	0.19
4404	20	1.39	0.24
4405	4	0.63	0.11
4405	13	0.62	0.11
4405	19	0.75	0.13
4414	9	6.78	1.15
4414	20	2.50	0.42
Wt	C24	0	0

FIGURA 6



10/17

FIGURA 7

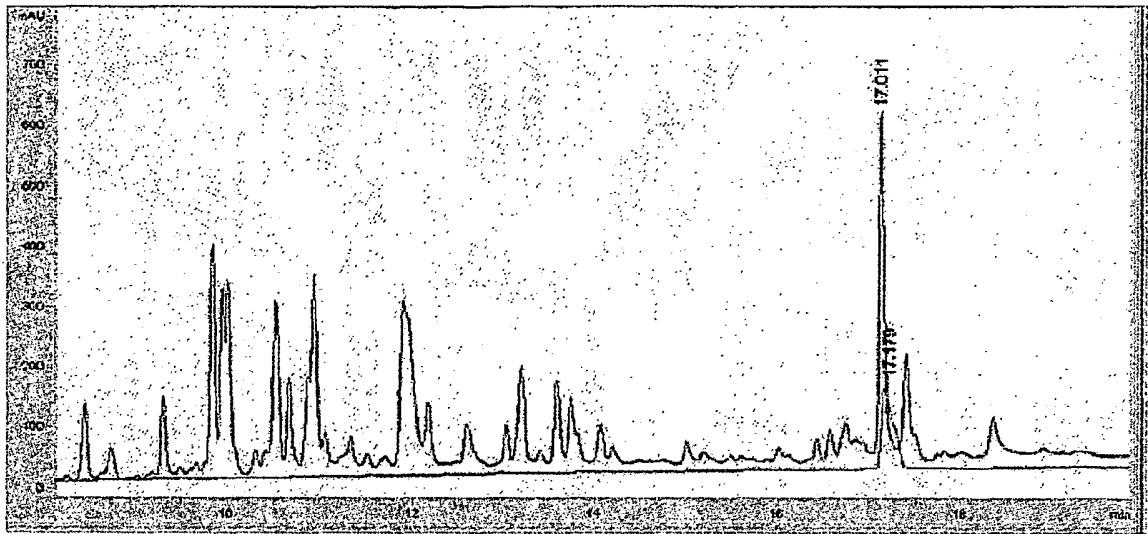


FIGURA 8

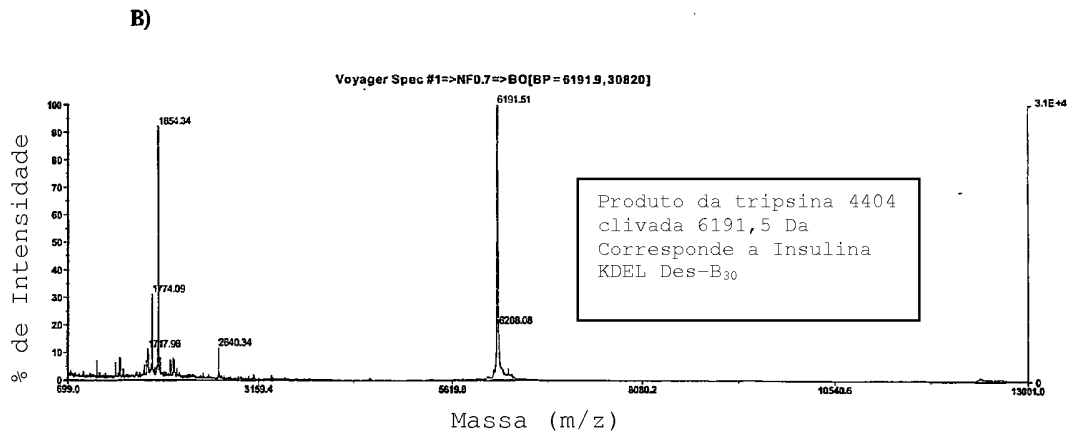
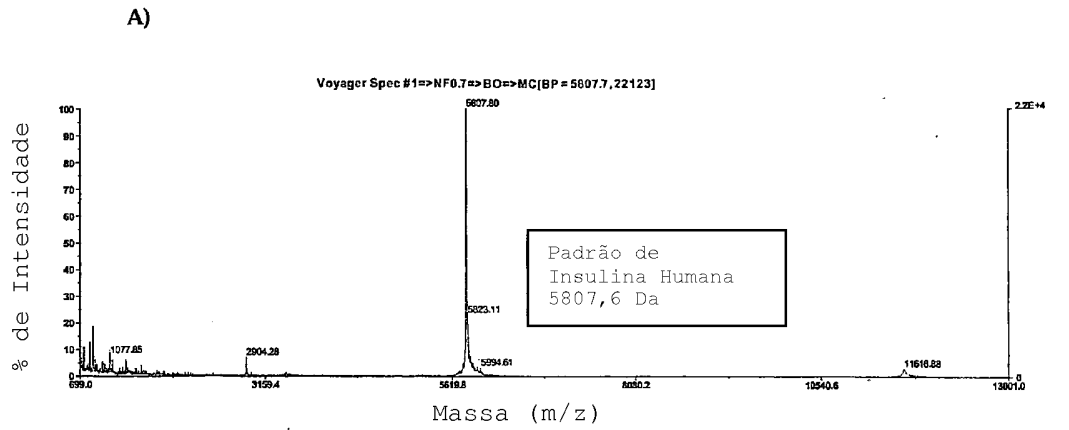


FIGURA 9

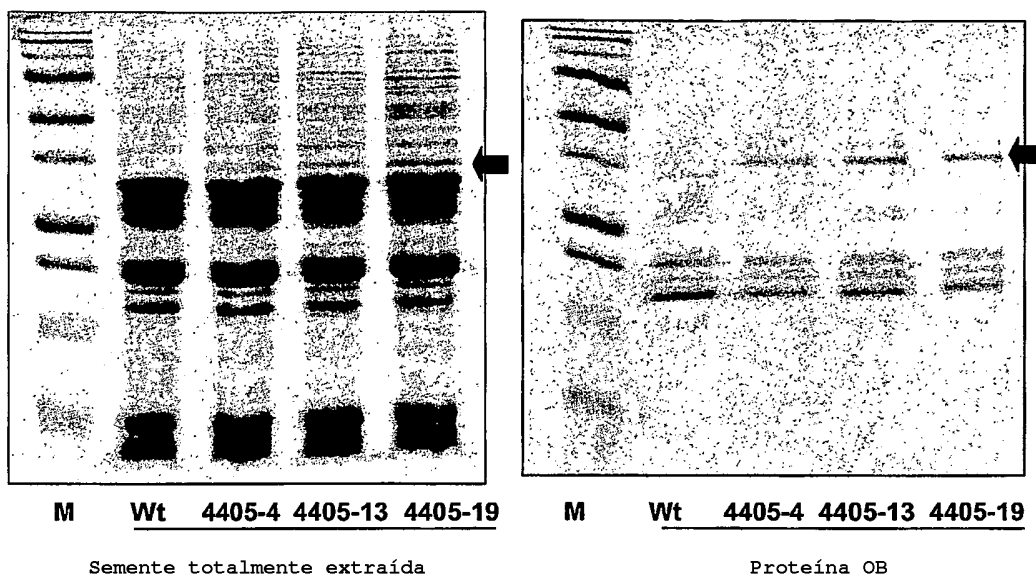


FIGURA 10

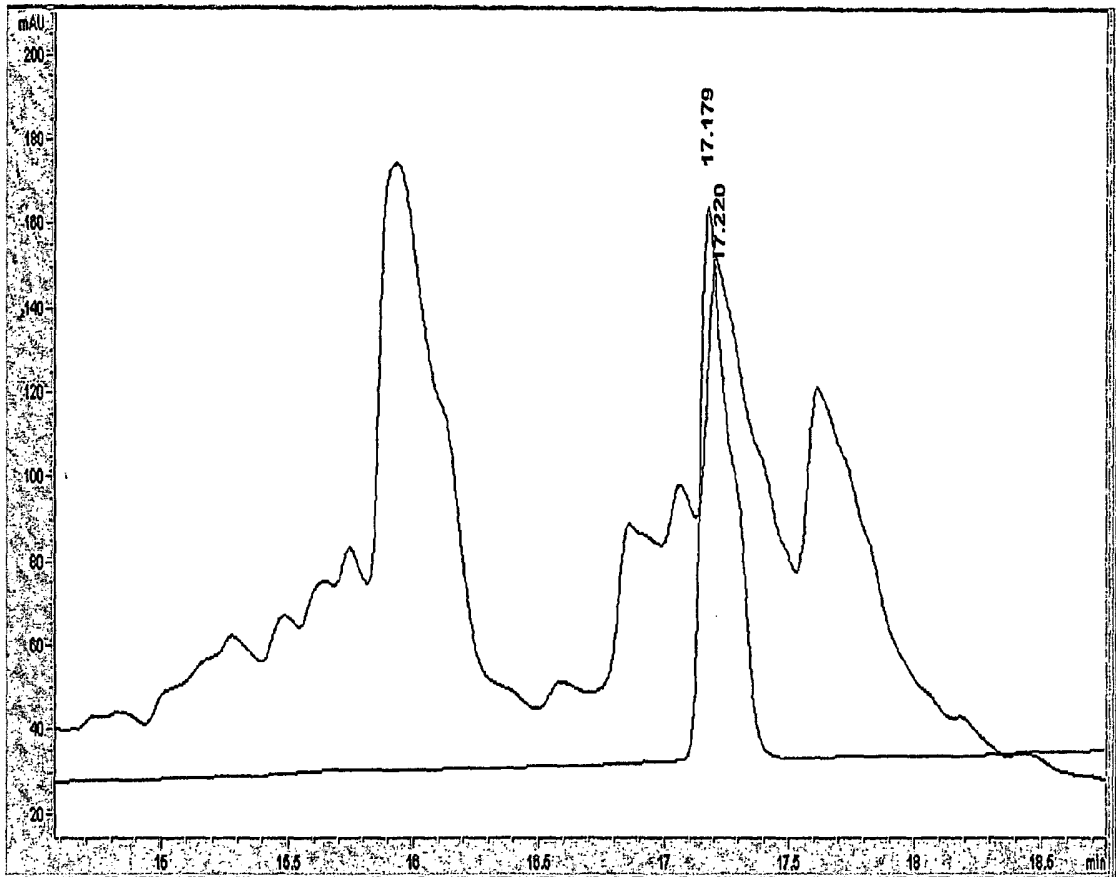


FIGURA 11

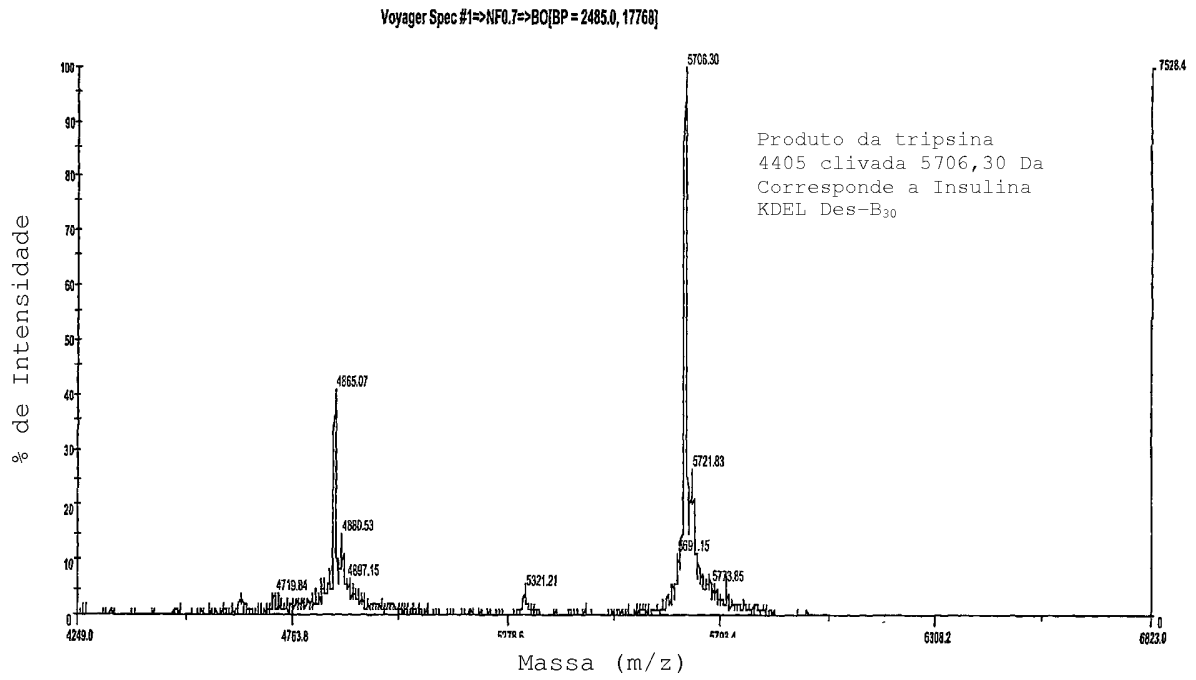


FIGURA 12

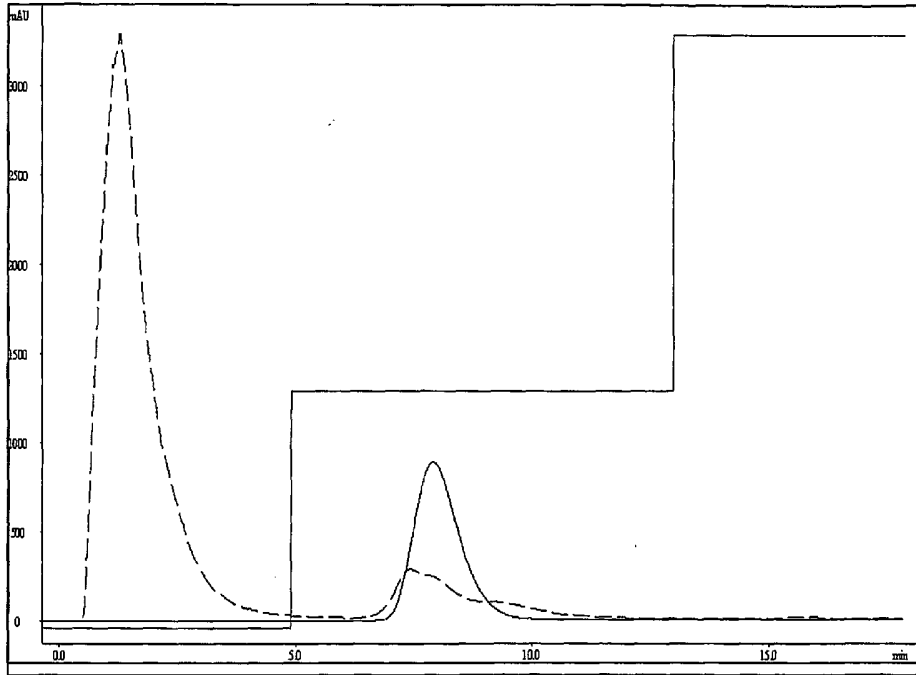
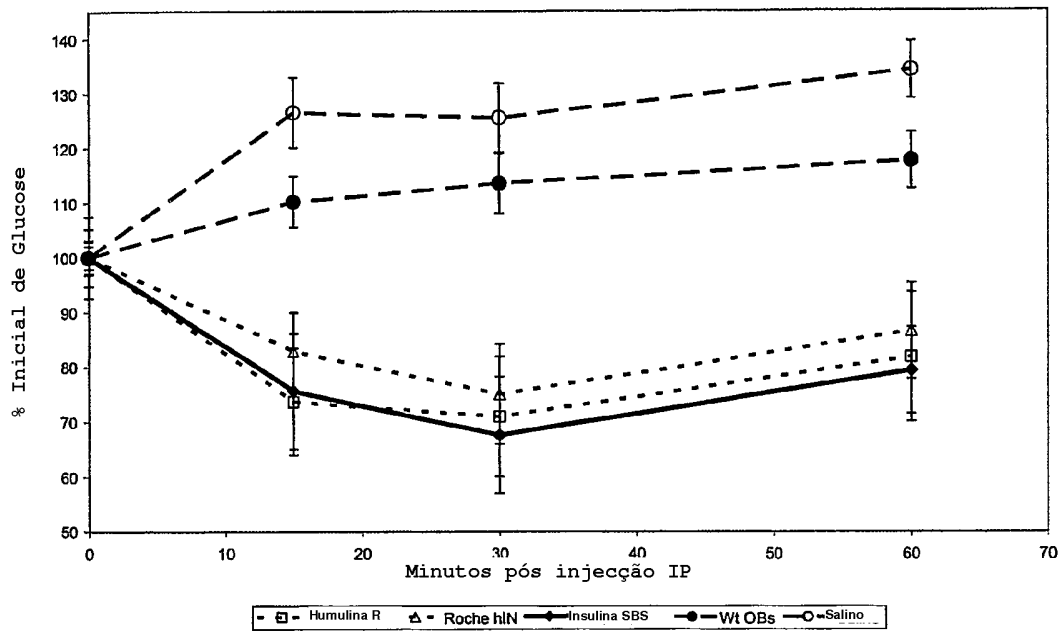


FIGURA 13

Testes de Tolerância a Insulina em
Murganhos C57BI/6 Machos



RESUMO

"MÉTODOS PARA A PRODUÇÃO DE INSULINA EM PLANTAS"

São aqui descritos métodos para a produção de insulina em plantas. Numa forma de realização, a presente invenção proporciona um método para a expressão de insulina em plantas compreendendo: (a) proporcionar uma construção de ácido nucleico quimérica compreendendo na direcção 5' para 3' da transcrição, como componentes ligados operacionalmente: (i) uma sequência de ácido nucleico capaz de controlar a expressão em células de sementes de plantas; e (ii) uma sequência de ácido nucleico codificando um polipéptido de insulina; (b) introduzir a construção de ácido nucleico quimérica numa célula vegetal; e (c) cultivar a célula vegetal até ficar uma planta madura capaz de produzir sementes em que a semente expressa insulina.