República Federativa do Brasil
Ministério da Economia
Instituto Nacional da Propriedade Industrial

(21) BR 112020015238-2 A2

(22) Data do Depósito: 01/02/2019

(43) Data da Publicação Nacional: 29/12/2020

(54) Título: GENE PRECURSOR OTIMIZADO POR CÓDON E GENE DE PEPTÍDEO SINAL DE ANÁLOGO DE INSULINA HUMANA

(51) Int. Cl.: C12N 15/62; C12N 15/17; C12N 15/81; C12P 21/00; C12N 1/16.

(30) Prioridade Unionista: 09/02/2018 CN 201810131794.3.

(71) Depositante(es): JIANGSU HENGRUI MEDICINE CO., LTD..

(72) Inventor(es): FEIFEI WANG; LEI CHEN; HONGWEI WANG.

(86) Pedido PCT: PCT CN2019074384 de 01/02/2019

(87) Publicação PCT: WO 2019/154311 de 15/08/2019

(85) Data da Fase Nacional: 27/07/2020

(57) Resumo: É fornecida uma molécula de ácido nucleico de um gene precursor otimizado por códon e gene do peptídeo sinal de um análogo da insulina humana. A molécula de ácido nucleico compreende uma molécula de ácido nucleico que codifica o precursor do análogo da insulina de fusão e uma molécula de ácido nucleico que codifica o peptídeo sinal de levedura secretor de fator a. A molécula de ácido nucleico melhora a expressão do precursor do análogo da insulina em Pichia pastoris e reduz o custo de produção do análogo da insulina humana.

Relatório Descritivo da Patente de Invenção para "**GENE PRECURSOR OTIMIZADO POR CÓDON E GENE DE PEPTÍDEO SINAL DE ANÁLOGO DE INSULINA HUMANA**"

CAMPO DE INVENÇÃO

[001] A presente invenção se refere a um gene precursor análogo de insulina humana otimizado por códon e um gene de peptídeo sinal de fator α otimizado por códon, e fornece um método para expressar o gene precursor de análogo de insulina humana.

FUNDAMENTOS DA INVENÇÃO

[002] A insulina humana é um polipeptídeo que consiste em 51 aminoácidos e compreende duas cadeias, Cadeia A e Cadeia B, respectivamente. A principal eficiência da insulina é regular o metabolismo da glicose. A intervenção com insulina é o método mais direto e eficaz como tratamento alternativo ou suplementar para o diabetes. A insulina também tem efeito na promoção da síntese de gordura, na inibição da decomposição da gordura e na redução da produção de corpos cetônicos, assim também é usada para corrigir vários sintomas de cetose e acidose relacionadas à insulina.

[003] A insulina foi previamente extraída do pâncreas de porcos, bovinos e outros animais, mas esses produtos são estruturalmente diferentes da insulina humana, por isso possuem imunogenicidade. Desde que a tecnologia geneticamente

recombinante para produção de insulina humana foi desenvolvida por Eli Lilly & Co. EUA e Novo Nordisk Dinamarca sucessivamente, no início e meados da década de 1980, a expressão genética de insulina humana e seus análogos se tornou o principal meio da indústria. No entanto, é extremamente inconveniente para pacientes que precisam ser injetados frequentemente com insulina humana devido ao efeito a curto prazo da insulina. Portanto, foram feitos esforços para obter análogos e derivados de insulina com efeito a longo prazo em humanos. Entre eles, a modificação da insulina humana ou análogo da mesma com o grupo acilado é um método eficaz para aumentar a meia-vida da mesma. Um análogo da insulina humana foi divulgado no documento WO2018024186, em que a posição B29 foi substituída por um ácido graxo de cadeia longa e o aminoácido na posição B30 foi deletado, e a estrutura e a atividade biológica do análogo da insulina humana foram divulgadas. Um análogo de insulina possuindo uma cadeia lateral de 14-acil ligada à posição B29 e uma deleção de aminoácidos na posição B30, e uma preparação dos mesmos, foram divulgados no documento WO9507931. Um análogo da insulina humana em que a posição B29 foi substituída por um ácido glutâmico e um ácido graxo de cadeia longa, e o aminoácido na posição B30 foi deletado, foi divulgado no documento WO2005012347. Atualmente, os sistemas de expressão mais comumente usados para expressar insulina humana e seu análogo são *Escherichia coli, Saccharomyces Cerevisiae* e *Pichia*

Pastoris, entre os quais a insulina humana e seu análogo são expressos na forma de corpo de inclusão em *E. coli* e clivagem e a renaturação do corpo de inclusão é necessária, o que torna o processo complicado e de baixo rendimento; *S. cerevisiae* e *P. pastoris* têm as vantagens devido à facilidade de operação, facilidade de cultivo, modificação de proteínas estranhas e capacidade de expressão de secreção, e assim por diante. No entanto, a eficiência de secreção para *Saccharomyces cerevisiae* é baixa e a cepa para expressão não é estável. Por outro lado, *Pichia Pastoris* é um sistema de expressão mais amplamente utilizado na produção industrial de proteínas recombinantes. Industrialmente, o rendimento da fermentação durante o processo de produção é um fator chave no controle do custo de produção. Devido à grande demanda de insulina comercial, na Novo Nordisk, grande produtora, a balança do tanque atinge dezenas de toneladas para produção no sistema de expressão de leveduras, o que envolve altos requisitos para instalações e equipamentos e, finalmente, resulta em alto custo. Assim, é de grande importância aumentar o rendimento de fermentação da insulina humana e o seu análogo na produção industrial.

[004] O códon genético é um código tripleto que consiste em três bases adjacentes no ácido ribonucleico mensageiro (mRNA). Existem 64 tipos de códons genéticos. No entanto, a frequência do uso de códons difere para diferentes organismos, mesmo para diferentes genes codificadores de proteínas do mesmo

organismo, ou seja, há preferência por códons. Códons de genes estranhos afetam principalmente a expressão gênica no nível traducional. Existem muitas literaturas mostrando que a otimização do códon tem efeito significativo no aumento da expressão de proteínas estranhas em *Pichia pastoris*. Os genes exógenos são expressos em *Pichia pastoris* na forma de expressão intracelular e expressão secretora. Para este último, são necessários peptídeos de sinal para direcionar a secreção de produtos expressos por genes exógenos. Atualmente, o peptídeo sinal mais comumente usado é derivado do peptídeo sinal de fator α de *Saccharomyces cerevisiae*, e a sequência nucleotídica do mesmo é também derivada de *Saccharomyces cerevisiae*. A sequência nucleotídica do peptídeo sinal do fator α otimizada para *Pichia pastoris* não foi relatada até o momento. Existem muitas literaturas relacionadas sobre otimização de códons de precursores de insulina, por exemplo, uma sequência de genes precursores de insulina humana otimizada e sua expressão em *Pichia pastoris* por Gurramkonda et al. (Gurramkonda et al. Application of simple fed-batch technique to high-level secretory production of insulin precursor using *Pichia pastoris* with subsequent purification and conversion to human insulin. Microbial Cell Factories, 2010, 9:31) e publicação de patente WO1998028429 divulga uma sequência de genes que expressa um precursor análogo de insulina humana, e a sequência de aminoácidos precursores de insulina codificada pelo gene é

EEGEPK-B (1-29)-AAK-A(1-21), em que EEGEPK é uma extensão do terminal N do precursor da insulina, referido como peptídeo espaçador ou peptídeo líder, capaz de proteger o N terminal do precursor da insulina da hidrólise via protease de levedura e é capaz de melhorar a eficiência da expressão do precursor da insulina; B (1-29) é a cadeia B da insulina humana com deleção de treonina B30; A (1-21) é a sequência de aminoácidos da cadeia A da insulina humana; AAK é um peptídeo ligante que liga a Cadeia B à Cadeia A, também conhecido como peptídeo C.

[005] A fim de aumentar ainda mais o rendimento de insulina humana e seus precursores análogos, os inventores otimizaram o gene precursor análogo de insulina e o gene do peptídeo sinal de fator α para expressão secretora em *Pichia pastoris*, de acordo com a preferência do códon em *Pichia pastoris*. Nossos resultados mostram que o rendimento do precursor análogo da insulina humana foi aumentado em quase duas vezes pela expressão do gene otimizado por códon, de acordo com a presente invenção, quando comparado com o gene precursor do análogo da insulina humana conhecido na técnica anterior (como controle). O custo da produção industrial de insulina humana e seus análogos será bastante reduzido no estágio final.

SUMÁRIO DA INVENÇÃO

[006] Em algumas modalidades da invenção, fornece uma molécula de ácido nucleico compreendendo a seguinte estrutura:

$5'- (PS)_a - (SP)_b - (LS)_c - GE - (P'S)_d - 3'$,

em que PS é uma molécula de ácido nucleico que codifica um sítio de processamento, a é 0 ou 1;

SP é uma molécula de ácido nucleico que codifica o peptídeo sinal, b é 0 ou 1;

LS é uma molécula de ácido nucleico que codifica o peptídeo espaçador, c é 0 ou 1;

GE é uma molécula de ácido nucleico que codifica polipeptídeo de interesse; e

P'S é uma molécula de ácido nucleico que codifica o sítio de processamento e d é 0 ou 1.

[007] Em algumas modalidades, fornece uma molécula de ácido nucleico compreendendo a seguinte estrutura:

$5'- (PS)_a - (SP)_b - (LS)_c - GE - (P'S)_d - 3'$,

em que PS é uma molécula de ácido nucleico que codifica o sítio de processamento, a é 0 ou 1;

SP é uma molécula de ácido nucleico que codifica o peptídeo sinal, b é 1;

LS é uma molécula de ácido nucleico que codifica o peptídeo espaçador, c é 1;

GE é uma molécula de ácido nucleico que codifica polipeptídeo de interesse; e

P'S é uma molécula de ácido nucleico que codifica o sítio de processamento e d é 0 ou 1.

[008] Em algumas modalidades, a molécula de ácido nucleico que codifica o peptídeo sinal compreende a sequência

mostrada como SEQ ID NO: 1.

[009] Em algumas modalidades, o polipeptídeo de interesse é um polipeptídeo precursor análogo da insulina humana; a molécula de ácido nucleico que codifica o polipeptídeo precursor do análogo da insulina humana compreende a sequência mostrada como SEQ ID NO:3.

[010] Em algumas modalidades, a sequência de ácido nucleico da molécula de ácido nucleico que codifica o peptídeo de sinal (SP) é mostrada como SEQ ID NO: 1, e sequência de aminoácidos do mesmo é mostrada como SEQ ID NO: 2:

ATGAGATTTCCTTCTATTTTCACTGCTGTTTTGTTTGCCGCTTCCTCTGCTT
TGGCAGCTCCAGTTAATACAACCACTGAAGATGAGACTGCTCAAATCCCAGCCGAAG
CAGTTATTGGTTACTCCGACTTGGAAGGAGATTTTGACGTCGCTGTTTTACCATTCT
CTAATTCCACTAATAACGGTCTGTTGTTTATTAATACTACCATTGCTTCTATCGCCG
CTAAGGAGGAAGGTGTGTCCCTCGAGAAAAGA

SEQ ID NO:1;

MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAV
LPFSNSTNNGLLFINTTIASIAAKEEGVSLEKR

SEQ ID NO:2;

[011] Em algumas modalidades, a molécula de ácido nucleico que codifica o polipeptídeo de interesse (GE) pode ser uma molécula de ácido nucleico que codifica o precursor análogo de insulina humana, em que o precursor análogo de insulina humana pode ser insulina humana com uma deleção de treonina na

posição B30. A sequência da molécula de ácido nucleico do precursor análogo da insulina humana é mostrada como SEQ ID NO:3 e a sequência de aminoácidos da mesma é mostrada como SEQ ID NO: 4:

TTCGTCAACCAGCACTTGTGTGGTTCCCATTTGGTTGAGGCTCTGTACTTGG TCTGTGGAGAAAGAGGTTTCTTTTACACCCCAAAGGCTGCTAAGGGTATCGTTGAGC AATGTTGCACCTCTATTTGTTCCCTGTATCAGTTGGAAAACTACTGCAACTAA

SEQ ID NO: 3;

FVNQHLCGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSICSLYQLENYC N

SEQ ID NO:4.

[012] Em algumas modalidades, as posições 88 a 96 da molécula de ácido nucleico que codifica o precursor análogo de insulina humana (ou seja, GCTGCTAAG) apresenta uma molécula de ácido nucleico que codifica o peptídeo ligante (também conhecido como peptídeo C), que pode ser substituído pela sequência a seguir, incluindo mas não limitado a: GCCGCTAAG, GCTGCCAAG, GCTGCTAAA, GCCGCCAAG.

[013] Em algumas modalidades, a sequência da molécula de ácido nucleico (LS) que codifica o peptídeo espaçador é mostrada como SEQ ID NO: 5.

gaagaaggtgaaccaaag

SEQ ID NO:5.

[014] Em algumas modalidades, o PS e/ou P'S são moléculas

de ácido nucleico que codificam o sítio de restrição.

[015] Preferivelmente, PS é uma molécula de ácido nucleico que codifica o sítio de restrição EcoRI e/ou P'S é uma molécula de ácido nucleico que codifica o sítio de restrição NotI.

[016] Em algumas modalidades, fornece uma molécula de ácido nucleico capaz de expressar o precursor análogo da insulina humana, em que a molécula de ácido nucleico compreende uma molécula de ácido nucleico que codifica o peptídeo espaçador e uma molécula de ácido nucleico que codifica o precursor análogo da insulina humana e é capaz de expressar o precursor análogo da insulina humana, após recombinação com um vetor compreendendo um peptídeo sinal.

[017] Em algumas modalidades, a sequência de aminoácidos do precursor análogo da insulina humana codificada pela molécula de ácido nucleico do precursor análogo da insulina humana é a seguinte:

EEGEPK-B(1-29)-AAK-A(1-21)

em que "EEGEPK (SEQ ID NO: 16)" pode ser uma extensão do N terminal do precursor da insulina, referido como peptídeo espaçador ou peptídeo líder; "B (1-29)" pode ser uma cadeia B de insulina humana com uma deleção de treonina na posição B30. "A (1-21)" pode ser uma sequência de aminoácidos da Cadeia A da insulina humana e "AAK" é um peptídeo ligante que liga a Cadeia B à Cadeia A, também conhecido como peptídeo C.

[018] Em algumas modalidades, a sequência da molécula de ácido nucleico precursor do análogo da insulina humana pode ser mostrada como SEQ ID NO: 6, e a sequência de aminoácidos da mesmas é mostrada como SEQ ID NO: 7:

GAAGAAGGTGAACCAAAGTTCGTCAACCAGCACTTGTGTGGTTCCCATTTGG TTGAGGCTCTGTACTTGGTCTGTGGAGAAAGAGGTTTCTTTTACACCCCAAAGGCTG CTAAGGGTATCGTTGAGCAATGTTGCACCTCTATTTGTTCCCTGTATCAGTTGGAAA ACTACTGCAACTAA

SEQ ID NO:6;

EEGEPKFVNQHLCGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSICSLY QLENYCN

SEQ ID NO:7.

[019] Em algumas modalidades, é fornecida outra molécula de ácido nucleico capaz de expressar o precursor análogo da insulina humana. A sequência da molécula de ácido nucleico compreende uma sequência de peptídeo sinal, uma sequência de peptídeo espaçador e uma sequência que codifica o precursor análogo de insulina humana, e a molécula de ácido nucleico é capaz de expressar o precursor análogo de insulina humana após a recombinação com um vetor que não compreende peptídeo de sinal.

[020] Em algumas modalidades, a sequência de ácido nucleico da molécula de ácido nucleico que expressa o precursor análogo da insulina humana é mostrada como SEQ ID NO: 8, e a

sequência de aminoácidos codificada é mostrada como SEQ ID NO: 9:

ATGAGATTTCCTTCTATTTTCACTGCTGTTTTGTTTGCCGCTTCCTCTGCTT
TGGCAGCTCCAGTTAATACAACCACTGAAGATGAGACTGCTCAAATCCCAGCCGAAG
CAGTTATTGGTTACTCCGACTTGGAAGGAGATTTTGACGTCGCTGTTTTACCATTCT
CTAATTCCACTAATAACGGTCTGTTGTTTATTAATACTACCATTGCTTCTATCGCCG
CTAAGGAGGAAGGTGTGTCCCTCGAGAAAAGAGAAGAAGGTGAACCAAAGTTCGTCA
ACCAGCACTTGTGTGGTTCCCATTTGGTTGAGGCTCTGTACTTGGTCTGTGGAGAAA
GAGGTTTCTTTTACACCCCAAAGGCTGCTAAGGGTATCGTTGAGCAATGTTGCACCT
CTATTTGTTCCCTGTATCAGTTGGAAAACTACTGCAACTAA

SEQ ID NO:8;

MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAV
LPFSNSTNNGLLFINTTIASIAAKEEGVSLEKREEGEPKFVNQHLCGSHLVEALYLV
CGERGFFYTPKAAKGIVEQCCTSICSLYQLENYCN

SEQ ID NO:9.

[021] Em algumas modalidades, a molécula de ácido nucleico que expressa o precursor análogo da insulina humana também pode compreender sítios de restrição, preferivelmente os sítios de restrição são sítios de restrição EcoRI e/ou sítios de restrição NotI.

[022] Em algumas modalidades, é fornecido um vetor capaz de ser expresso em uma célula eucariótica ou procariótica, capaz de expressar o precursor análogo da insulina humana em uma

célula procariótica ou eucariótica por meio de expressão secretora.

[023] Em algumas modalidades, uma célula hospedeira também é fornecida. Preferivelmente, a célula hospedeira é levedura, mais preferivelmente *Pichia*, que é capaz de expressar o precursor análogo da insulina humana via expressão secretora.

[024] Em algumas modalidades, também é fornecido um método para preparar um análogo de insulina humana, compreendendo a utilização da molécula de ácido nucleico, vetor e/ou célula hospedeira, como descrito acima.

[025] O método pode ainda compreender as seguintes etapas:

1) expressar um precursor análogo da insulina humana em uma célula eucariótica por uma molécula de ácido nucleico que codifica o precursor análogo da insulina humana;

2) obter um análogo de insulina humana digerindo enzimaticamente o precursor de análogo de insulina humana, a molécula de ácido nucleico que codifica o precursor de análogo de insulina humana pode ser mostrada como SEQ ID NO: 6, e o precursor análogo da insulina pode ser digerido enzimaticamente usando os métodos bem conhecidos dos especialistas na técnica.

[026] Em algumas modalidades, a etapa 1) compreende a expressão do precursor análogo da insulina humana por um vetor de expressão compreendendo uma sequência de peptídeo sinal e a referida sequência de peptídeo sinal é mostrada como SEQ ID

NO: 1.

[027] Em algumas modalidades, o análogo de insulina humana é uma insulina humana com deleção de B30, e o análogo de insulina humana é ainda substituído por um grupo acilado.

[028] Em algumas modalidades, dentro da referida insulina humana com deleção de B30, a lisina na posição B29 é substituída pelo grupo acilado.

[029] Preferivelmente, o produto obtido após a substituição acima é a insulina humana lisina B29 (N^{ε}-(N^{α}-hexadecano diácido graxo-L-lisina-N^{ε}-oxobutilil))des(B30).

[030] Abreviações e termos

[031] "Otimização por códons" se refere à síntese de genes com códons preferenciais em vez de códons com baixa frequência ou códons raros, de acordo com a regra preferivelmente de códons para células hospedeiras.

[032] "Controle 1" é mostrado como SEQ ID NO: 10 abaixo, em que uma molécula de ácido nucleico que codifica "EEGEPK" (GAAGAAGGTGAACCAAAG, duplo sublinhado) está ligada a uma molécula de ácido nucleico que codifica o gene precursor da insulina ensinado por WO1998028429:

<u>GAAGAAGGTGAACCAAAG</u>TTCGTTAACCAACACTTGTGCGGTTCCCACTTGG

TTGAAGCTTTGTACTTGGTTTGCGGTGAAAGAGGTTTCTTCTACACTCCTAAGGCTG

CTAAGGGTATTGTCGAACAATGCTGTACCTCCATCTGCTCCTTGTACCAATTGGAAA

ACTACTGCAACTAA

SEQ ID NO:10.

[033] "Controle 2" é mostrado como SEQ ID NO: 11 abaixo, em que uma molécula de ácido nucleico que codifica "EEGEPK" (sublinhado duas vezes) está ligada a uma molécula de ácido nucleico que codifica o gene precursor da insulina otimizado ensinado por Gurramkonda et al. (Microbial Cell Factories, 2010, 9:31):

GAAGAAGGTGAACCAAAGTTTGTTAACCAACATTTGTGTGGTTCTCATTTGG

TTGAAGCTTTGTACTTGGTTTGTGGTGAAAGAGGTTTCTTCTACACTCCAAAGGCTG

CTAAGGGTATTGTTGAACAATGTTGTACTTCTATTTGTTCTTTGTACCAATTGGAAA

ACTACTGTAACTAA

SEQ ID NO: 11.

[034] "IP-S" é uma molécula de ácido nucleico correspondente a um gene precursor da insulina após otimização por códon.

[035] O "fator α" é uma molécula de ácido nucleico correspondente ao gene do peptídeo sinal do fator α compreendido no vetor de expressão pPIC9K fornecido pela Invitrogen, que é derivado de *Saccharomyces cerevisiae*.

[036] "α-fator-S" é uma molécula de ácido nucleico correspondente a um gene de peptídeo sinal de fator a otimizado por códon.

[037] “Vetor" inclui molécula de ácido nucleico, que é capaz de transportar outro ácido nucleico ao qual está ligado, incluindo, mas não limitado a, vetor plasmídico e viral. Alguns vetores são capazes de se replicar autonomamente na célula hospedeira na qual são introduzidos, enquanto outros podem ser integrados no genoma da célula hospedeira após a introdução na célula hospedeira e, assim, serem replicados junto com o genoma hospedeiro. Além disso, alguns vetores são capazes de direcionar a expressão de genes operacionalmente ligados a eles, e tais vetores são aqui referidos como "vetores de expressão recombinantes" (ou simplesmente como "vetores de expressão"). Os vetores tradicionais são bem conhecidos na técnica.

[038] “Polipeptídeo de interesse" é um polipeptídeo que pode ser expresso em levedura, incluindo mas não limitado a enzima, anticorpo, interferon, insulina, interleucina e semelhantes, e variante, precursor, intermediário dos mesmos. Por exemplo, o polipeptídeo de interesse pode ser precursor da insulina.

[039] “Célula" e "célula hospedeira" podem ser usadas de forma intercambiável.

[040] “Molécula de polinucleotídeo", "molécula de ácido nucleico" podem ser utilizadas de forma intercambiável e sua sequência pode ser a sequência de DNA.

DESCRIÇÃO DETALHADA DA DIVULGAÇÃO

[041] Os exemplos a seguir são fornecidos para ilustrar

melhor a invenção, mas não pretendem limitar o escopo da invenção.

[042] Os vetores, bactérias hospedeiras e meios de cultura utilizados nos exemplos da presente invenção foram adquiridos à Invitrogen. O pPIC9K, um vetor de expressão de *Pichia pastoris*, contém um promotor de álcool oxidase AOX1, que pode ser induzido por metanol, e o vetor também contém a sequência peptídica do sinal de fator a e é capaz de expressar proteínas estranhas por expressão secretora. O pPIC3.5K, outro vetor de expressão de *Pichia pastoris*, contém promotor de álcool oxidase AOX1, que pode ser induzido por metanol, e o vetor não contém a sequência peptídica do sinal de fator α. A cepa de *Pichia pastoris* GS115 foi utilizada como bactéria hospedeira. A formulação do meio de cultura foi fornecida pela *Pichia pastoris* manual.

Exemplo 1. Construção do vetor de expressão recombinante para precursor de insulina

[043] Os sítios de restrição EcoR I e Not I foram adicionados às extremidades 5' e 3' do Controle 1 (SEQ ID NO: 10), Controle 2 (SEQ ID NO: 11) e IP-S (SEQ ID NO: 6), respectivamente e sintetizados por Nanjing Kingsray Biotech Co., Ltd. A sequência da molécula de ácido nucleico sintetizada foi ligada ao vetor T.

GAAGAAGGTGAACCAAAGTTCGTCAACCAGCACTTGTGTGGTTCCCATTTGG

TTGAGGCTCTGTACTTGGTCTGTGGAGAAAGAGGTTTCTTTTACACCCCAAAGGCTG
CTAAGGGTATCGTTGAGCAATGTTGCACCTCTATTTGTTCCCTGTATCAGTTGGAAA
ACTACTGCAACTAA

SEQ ID NO:6;

GAAGAAGGTGAACCAAAGTTCGTTAACCAACACTTGTGCGGTTCCCACTTGG
TTGAAGCTTTGTACTTGGTTTGCGGTGAAAGAGGTTTCTTCTACACTCCTAAGGCTG
CTAAGGGTATTGTCGAACAATGCTGTACCTCCATCTGCTCCTTGTACCAATTGGAAA
ACTACTGCAACTAA

SEQ ID NO:10;

GAAGAAGGTGAACCAAAGTTTGTTAACCAACATTTGTGTGGTTCTCATTTGG
TTGAAGCTTTGTACTTGGTTTGTGGTGAAAGAGGTTTCTTCTACACTCCAAAGGCTG
CTAAGGGTATTGTTGAACAATGTTGTACTTCTATTTGTTCTTTGTACCAATTGGAAA
ACTACTGTAACTAA

SEQ ID NO:11.

[044] O vetor T portador da molécula de ácido nucleico precursor da insulina e o vetor de expressão pPIC9K foram digeridos com EcoRI e NotI, e depois o fragmento alvo e o fragmento de vetor foram recuperados separadamente por um Kit de Recuperação em Gel. Após purificação dos fragmentos digeridos, e o fragmento alvo foi ligado ao vetor pPIC9K com T4 ligase.

[045] A solução de ligação acima mencionada foi transformada em células competentes E. coli TOP10 e plaqueadas

em uma placa com resistência à ampicilina. Após o cultivo, a colônia bacteriana foi coletada e o plasmídeo foi extraído e verificado por digestão com as duas enzimas de restrição. Finalmente, foram obtidos três vetores de expressão recombinantes compreendendo sequências de Controle 1, Controle 2 e IP-S, respectivamente.

Exemplo 2. Expressão do precursor da insulina pela cepa recombinante de Pichia Pastoris

[046] Os três vetores de expressão recombinante construídos no Exemplo 1 foram transformados em Pichia Pastoris GS115, respectivamente, as cepas recombinantes que expressam o Controle 1 e o Controle 2 foram usadas como cepas de controle, e a cepa recombinante que expressa IP-S foi usada como cepa de teste.

[047] As colônias das três bactérias recombinantes foram inoculadas em 5 mL de meio YPD e cultivadas a uma temperatura constante de 30 °C enquanto agitavam a 250 rpm, até o valor de OD_{600} atingir cerca de 10 (16-18 horas). As células foram coletadas e ressuspensas em 50 mL de meio BMGY e cultivadas durante a noite a uma temperatura constante de 30 °C enquanto agitavam a 250 rpm, até o valor de OD_{600} atingir cerca de 30. As células foram coletadas por centrifugação a 1500 rpm por 5 minutos e ressuspenso em 25 mL de meio BMMY. Foi adicionado 1/200 volume de metanol (concentração final de 0,5%) ao meio e cultivado a uma temperatura constante de 30°C enquanto agitava

a 250 rpm por 96 horas, enquanto 1/200 volume de metanol era suplementado a cada 24 horas. Terminada a expressão, o sobrenadante foi obtido após centrifugação a 10.000 rpm. O rendimento do precursor da insulina compreendido no sobrenadante foi medido por HPLC e convertido em um percentual relativo à quantidade de expressão do precursor da insulina expressa pela cepa de controle. O percentual do nível de expressão do precursor da insulina é mostrada na Tabela 1.

Tabela 1

Cepa bacteriana	vetor	Peptídeo sinal	Gene de expressão	Percentual de rendimento (%)
Controle bacteriano	pPIC9K	fator α	Controle 1	100
Controle bacteriano	pPIC9K	Fator α	Controle 2	125
Bactéria de teste	pPIC9K	Fator α	IP-S	225

[048] Os dados da Tabela 1 mostram que a quantidade de precursor da insulina expressa pelo gene precursor da insulina otimizada aumentou em 1,8 a 2,25 vezes em comparação com as dos dois grupos de controle. Pode-se observar que o gene precursor de insulina otimizado possui eficiência de expressão superior e pode melhorar significativamente o rendimento do precursor

de insulina expresso.

Exemplo 3. Construção de vetores de expressão recombinantes compreendendo gene precursor de insulina fundido com diferentes fatores αs

[049] Moléculas de ácido nucleico mostradas como SEQ ID NO: 13 e SEQ ID NO: 8, em que fator α (SEQ ID NO: 12) e fator α-S (SEQ ID NO: 1) foram fundidos separadamente ao sítio a montante do IP-S, foram sintetizadas. Os sítios de restrição EcoRI e NotI também foram incorporados nas extremidades 5' e 3' das moléculas de ácido nucleico sintetizadas. As moléculas de ácido nucleico sintetizadas foram ligadas ao vetor T.

ATGAGATTTCCTTCAATTTTTACTGCAGTTTTATTCGCAGCATCCTCCGCAT TAGCTGCTCCAGTCAACACTACAACAGAAGATGAAACGGCACAAATTCCGGCTGAAG CTGTCATCGGTTACTCAGATTTAGAAGGGGATTTCGATGTTGCTGTTTTGCCATTTT CCAACAGCACAAATAACGGGTTATTGTTTATAAATACTACTATTGCCAGCATTGCTG CTAAAGAAGAAGGGGTATCTCTCGAGAAAAGA

SEQ ID NO:12

ATGAGATTTCCTTCTATTTTCACTGCTGTTTTGTTTGCCGCTTCCTCTGCTT TGGCAGCTCCAGTTAATACAACCACTGAAGATGAGACTGCTCAAATCCCAGCCGAAG CAGTTATTGGTTACTCCGACTTGGAAGGAGATTTTGACGTCGCTGTTTTACCATTCT CTAATTCCACTAATAACGGTCTGTTGTTTATTAATACTACCATTGCTTCTATCGCCG CTAAGGAGGAAGGTGTGTCCCTCGAGAAAAGA

SEQ ID NO:1

ATGAGATTTCCTTCAATTTTTACTGCAGTTTTATTCGCAGCATCCTCCGCAT

TAGCTGCTCCAGTCAACACTACAACAGAAGATGAAACGGCACAAATTCCGGCTGAAG

CTGTCATCGGTTACTCAGATTTAGAAGGGGATTTCGATGTTGCTGTTTTGCCATTTT

CCAACAGCACAAATAACGGGTTATTGTTTATAAATACTACTATTGCCAGCATTGCTG

CTAAAGAAGAAGGGGTATCTCTCGAGAAAAGAGAAGAAGGTGAACCAAAGTTCGTCA

ACCAGCACTTGTGTGGTTCCCATTTGGTTGAGGCTCTGTACTTGGTCTGTGGAGAAA

GAGGTTTCTTTTACACCCCAAAGGCTGCTAAGGGTATCGTTGAGCAATGTTGCACCT

CTATTTGTTCCCTGTATCAGTTGGAAAACTACTGCAACTAA

SEQ ID NO:13

ATGAGATTTCCTTCTATTTTCACTGCTGTTTTGTTTGCCGCTTCCTCTGCTT

TGGCAGCTCCAGTTAATACAACCACTGAAGATGAGACTGCTCAAATCCCAGCCGAAG

CAGTTATTGGTTACTCCGACTTGGAAGGAGATTTTGACGTCGCTGTTTTACCATTCT

CTAATTCCACTAATAACGGTCTGTTGTTTATTAATACTACCATTGCTTCTATCGCCG

CTAAGGAGGAAGGTGTGTCCCTCGAGAAAAGAGAAGAAGGTGAACCAAAGTTCGTCA

ACCAGCACTTGTGTGGTTCCCATTTGGTTGAGGCTCTGTACTTGGTCTGTGGAGAAA

GAGGTTTCTTTTACACCCCAAAGGCTGCTAAGGGTATCGTTGAGCAATGTTGCACCT

CTATTTGTTCCCTGTATCAGTTGGAAAACTACTGCAACTAA

SEQ ID NO:8

[050] Moléculas de ácido nucleico mostradas como SEQ ID NO: 14 e SEQ ID NO: 15, em que fator α (SEQ ID NO: 12) e fator α-S (SEQ ID NO: 1) foram fundidos separadamente ao sítio a montante do controle 1, foram sintetizadas. Os sítios de

restrição EcoRI e NotI também foram incorporados nas extremidades 5' e 3' das moléculas de ácido nucleico sintetizadas. As moléculas de ácido nucleico sintetizadas foram ligadas ao vetor T.

ATGAGATTTCCTTCAATTTTTACTGCAGTTTTATTCGCAGCATCCTCCGCAT
TAGCTGCTCCAGTCAACACTACAACAGAAGATGAAACGGCACAAATTCCGGCTGAAG
CTGTCATCGGTTACTCAGATTTAGAAGGGGATTTCGATGTTGCTGTTTTGCCATTTT
CCAACAGCACAAATAACGGGTTATTGTTTATAAATACTACTATTGCCAGCATTGCTG
CTAAAGAAGAAGGGGTATCTCTCGAGAAAAGAGAAGAAGGTGAACCAAAGTTCGTTA
ACCAACACTTGTGCGGTTCCCACTTGGTTGAAGCTTTGTACTTGGTTTGCGGTGAAA
GAGGTTTCTTCTACACTCCTAAGGCTGCTAAGGGTATTGTCGAACAATGCTGTACCT
CCATCTGCTCCTTGTACCAATTGGAAAACTACTGCAACTAA

SEQ ID NO:14

ATGAGATTTCCTTCTATTTTCACTGCTGTTTTGTTTGCCGCTTCCTCTGCTT
TGGCAGCTCCAGTTAATACAACCACTGAAGATGAGACTGCTCAAATCCCAGCCGAAG
CAGTTATTGGTTACTCCGACTTGGAAGGAGATTTTGACGTCGCTGTTTTACCATTCT
CTAATTCCACTAATAACGGTCTGTTGTTTATTAATACTACCATTGCTTCTATCGCCG
CTAAGGAGGAAGGTGTGTCCCTCGAGAAAAGAGAAGAAGGTGAACCAAAGTTCGTTA
ACCAACACTTGTGCGGTTCCCACTTGGTTGAAGCTTTGTACTTGGTTTGCGGTGAAA
GAGGTTTCTTCTACACTCCTAAGGCTGCTAAGGGTATTGTCGAACAATGCTGTACCT
CCATCTGCTCCTTGTACCAATTGGAAAACTACTGCAACTAA

SEQ ID NO:15.

[051] O vetor T acima e o vetor de expressão pPIC3.5K foram digeridos com ambas as endonucleases EcoR I e Not I, e depois o fragmento alvo e o fragmento de vetor foram recuperados separadamente pelo kit de recuperação em Gel. Após purificação dos fragmentos digeridos, e o fragmento alvo foi ligado ao vetor pPIC3.5K com T4 ligase.

[052] A solução de ligação acima mencionada foi transformada em células competentes *E. coli* TOP10 e plaqueadas em uma placa com resistência à ampicilina. Após o cultivo, a colônia bacteriana foi coletada e o plasmídeo foi extraído e verificado por digestão com as duas enzimas de restrição. Finalmente, foram obtidos quatro vetores de expressão recombinante que incorporam separadamente o peptídeo sinal fator a ou fator α-S para expressar o gene precursor da insulina com diferentes sequências nucleotídicas.

Exemplo 4. Expressão do precursor da insulina pela cepa recombinante de *Pichia pastoris*, antes e depois da otimização

[053] Os vetores de expressão recombinantes construídos no Exemplo 3 foram transformados separadamente em *Pichia pastoris* GS115.

[054] As colônias de bactérias recombinantes foram inoculadas em 5 mL de meio YPD e cultivadas a uma temperatura constante de 30 °C enquanto agitavam a 250 rpm, até o valor de OD_{600} atingir cerca de 10 (16-18 horas). As células foram coletadas e ressuspensas em 50 mL de meio BMGY e cultivadas

durante a noite a uma temperatura constante de 30°C enquanto agitavam a 250 rpm, até o valor de OD_{600} atingir cerca de 30. As células foram coletadas por centrifugação a 1500 rpm por 5 minutos e ressuspensas em 25 mL de meio BMMY. Foi adicionado 1/200 volume de metanol (concentração final de 0,5%) ao meio e cultivado a uma temperatura constante de 30°C enquanto agitava a 250 rpm por 96 horas, enquanto 1/200 volume de metanol era suplementado a cada 24 horas. Terminada a expressão, o sobrenadante foi obtido após centrifugação a 10.000 rpm. O rendimento do precursor da insulina compreendido no sobrenadante foi medido por HPLC.

[055] A bactéria recombinante que expressa o gene de controle 1 fundido ao fator α foi utilizado como controle bacteriano. O rendimento do precursor da insulina por outras cepas foi convertido em um percentual relativo ao rendimento do precursor da insulina expresso pela cepa de controle, como mostra a Tabela 2.

Tabela 2

Cepa bacteriana	vetor	Peptídeo sinal	Gene de expressão	Percentual de rendimento (%)
Controle bacteriano	pPIC3.5K	fator α	Controle 1	100
Bactérias	pPIC3.5K	fator α -S	Controle	150

de teste			1	
Bactérias de teste	pPIC3.5K	fator α	IP-S	225
Bactérias de teste	pPIC3.5K	fator α -S	IP-S	275

[056] Os dados na Tabela 2 mostram que o rendimento do precursor da insulina foi aumentado em 1,5 vez após a otimização meramente do peptídeo sinal na molécula de ácido nucleico, e esse rendimento foi aumentado em 2,25 vezes após a otimização do gene precursor da insulina. Por outro lado, o rendimento do precursor da insulina foi aumentado em 2,75 vezes após a otimização do peptídeo sinal e do gene precursor da insulina. Em conjunto, a otimização do códon pode aumentar o nível de expressão do precursor da insulina para 1,5-2,75 vezes.

REIVINDICAÇÕES

1. Molécula de ácido nucleico **caracterizada** pelo fato de que compreende molécula ou estrutura com a seguinte fórmula geral:

$5'- (PS)_a - (SP)_b - (LS)_c - GE - (P'S)_d - 3'$,

em que PS é uma molécula de ácido nucleico que codifica o sítio de processamento, a é 0 ou 1;

SP é uma molécula de ácido nucleico que codifica o peptídeo sinal, b é 0 ou 1;

LS é uma molécula de ácido nucleico que codifica o peptídeo espaçador, c é 0 ou 1;

GE é uma molécula de ácido nucleico que codifica polipeptídeo de interesse; e

P'S é uma molécula de ácido nucleico que codifica o sítio de processamento, d é 0 ou 1; e

a molécula de ácido nucleico que codifica o peptídeo sinal compreende a sequência mostrada como SEQ ID NO: 1.

2. Molécula de ácido nucleico **caracterizada** pelo fato de que compreende molécula ou estrutura com a seguinte fórmula geral:

$5'- (PS)_a - (SP)_b - (LS)_c - GE - (P'S)_d - 3'$,

em que PS é uma molécula de ácido nucleico que codifica o sítio de processamento, a é 0 ou 1;

SP é uma molécula de ácido nucleico que codifica o peptídeo sinal, b é 1;

LS é uma molécula de ácido nucleico que codifica o peptídeo espaçador, c é 0 ou 1;

GE é uma molécula de ácido nucleico que codifica o polipeptídeo precursor análogo da insulina humana; e

P'S é uma molécula de ácido nucleico que codifica o sítio de processamento, d é 0 ou 1; e

a molécula de ácido nucleico que codifica o polipeptídeo precursor do análogo da insulina humana compreende a sequência mostrada como SEQ ID NO: 3.

3. Molécula de ácido nucleico, de acordo com a reivindicação 1, **caracterizada** pelo fato de que o polipeptídeo de interesse é precursor análogo de insulina humana e a molécula de ácido nucleico que codifica o precursor análogo de insulina humana compreende uma molécula de ácido nucleico que codifica a sequência de aminoácidos mostrada como SEQ ID NO: 4, preferivelmente, compreende a sequência de ácido nucleico mostrada como SEQ ID NO: 3.

4. Molécula de ácido nucleico, de acordo com a reivindicação 2, **caracterizada** pelo fato de que o peptídeo sinal de codificação da molécula de ácido nucleico compreende uma molécula de ácido nucleico que codifica a sequência de aminoácidos mostrada como SEQ ID NO: 2, preferivelmente compreende a sequência de ácido nucleico mostrada como SEQ ID NO: 1 ou SEQ ID NO: 12.

5. Molécula de ácido nucleico, de acordo com qualquer uma

das reivindicações anteriores 2 a 4, **caracterizada** pelo fato de que a molécula de ácido nucleico que codifica o precursor análogo da insulina humana compreende substituição nas posições 88-96 da SEQ ID NO: 3, preferivelmente, é substituído por GCCGCTAAG, GCTGCCAAG, GCTGCTAAA ou GCCGCCAAG.

6. Molécula de ácido nucleico, de acordo com qualquer uma das reivindicações anteriores, **caracterizada** pelo fato de que a sequência de aminoácidos do peptídeo espaçador compreende EEGEPK (Glu-Glu-Gly-Glu-Pro-Lys), preferivelmente, a molécula de ácido nucleico que codifica o peptídeo espaçador compreende a sequência mostrada como SEQ ID NO: 5.

7. Molécula de ácido nucleico, de acordo com qualquer uma das reivindicações anteriores, **caracterizada** pelo fato de que compreende qualquer uma selecionada do grupo que consiste em SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 13 e SEQ ID NO:15; ou consistindo em qualquer um dos selecionados do grupo que consiste em SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 13 e SEQ ID NO:15.

8. Molécula de ácido nucleico, de acordo com qualquer uma das reivindicações anteriores, **caracterizada** pelo fato de que o sítio de processamento é o sítio de restrição, preferivelmente, PS é uma molécula de ácido nucleico que codifica o sítio de restrição EcoRI e/ou P'S é uma molécula de ácido nucleico que codifica o sítio de restrição Not I.

9. Vetor **caracterizado** pelo fato de que compreende a

molécula de ácido nucleico, conforme definida em qualquer uma das reivindicações anteriores, preferivelmente o vetor é vetor de expressão eucariótico ou vetor de expressão procariótico.

10. Célula hospedeira **caracterizada** pelo fato de que compreende a molécula de ácido nucleico, conforme definida em qualquer uma das reivindicações 1 a 8 e/ou o vetor, conforme definido na reivindicação 9, preferivelmente a célula hospedeira é levedura; mais preferivelmente, *Pichia pastoris*.

11. Método para produzir um análogo de insulina humana **caracterizado** pelo fato de que compreende o uso da molécula de ácido nucleico, conforme definido em qualquer uma das reivindicações 1 a 8, o vetor, conforme definido na reivindicação 9 e/ou a célula hospedeira, conforme definida na reivindicação 10.

12. Método, de acordo com a reivindicação 11, **caracterizado** pelo fato de compreender ainda a etapa de:

1) expressar um precursor análogo da insulina humana; e

2) obter um análogo da insulina humana digerindo enzimaticamente o precursor do análogo da insulina humana obtido na etapa 1).

13. Método, de acordo com a reivindicação 11 ou 12, **caracterizado** pelo fato de que o análogo de insulina humana é insulina humana com deleção de B30, e/ou o análogo de insulina humana é ainda substituído por um grupo acilado; preferivelmente, a lisina na posição B29 é substituída por um

grupo acilado; mais preferivelmente, o produto após a substituição é a insulina lisina B29 (N^{ε}-(N^{α}-hexadecano diácido graxo-L-lisina-N^{ε}-oxobutilil))des (B30) humana.

Resumo da Patente de Invenção para "**GENE PRECURSOR OTIMIZADO POR CÓDON E GENE DE PEPTÍDEO SINAL DE ANÁLOGO DE INSULINA HUMANA**"

É fornecida uma molécula de ácido nucleico de um gene precursor otimizado por códon e gene do peptídeo sinal de um análogo da insulina humana. A molécula de ácido nucleico compreende uma molécula de ácido nucleico que codifica o precursor do análogo da insulina de fusão e uma molécula de ácido nucleico que codifica o peptídeo sinal de levedura secretor de fator α. A molécula de ácido nucleico melhora a expressão do precursor do análogo da insulina em *Pichia pastoris* e reduz o custo de produção do análogo da insulina humana.

Este anexo apresenta o código de controle da listagem de sequências biológicas.

Código de Controle

Campo 1



Campo 2



Outras Informações:

- Nome do Arquivo: Listagem de Sequencias G.15.019.PI-BR.txt
- Data de Geração do Código: 27/07/2020
- Hora de Geração do Código: 12:43:22
- Código de Controle:
  - Campo 1: 4AE3B0D68D19EE5B
  - Campo 2: 6439668CB1E2E4B0