



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

(11) PI 0811053-0 B1



(22) Data do Depósito: 28/04/2008

(45) Data de Concessão: 06/03/2019

(54) Título: VETOR DE EXPRESSÃO PARA EXPRESSÃO E SECREÇÃO DE PROTEÍNAS HETERÓLOGAS EM CÉLULAS DE INSETO CULTIVADAS

(51) Int.Cl.: C12Q 1/68.

(30) Prioridade Unionista: 26/04/2007 US 60/926,621.

(73) Titular(es): MERCK SHARP & DOHME CORP..

(72) Inventor(es): GORDON V. L. WANG; DAVID E. CLEMENTS; CAROLYN WEEKS-LEVY.

(86) Pedido PCT: PCT US2008005514 de 28/04/2008

(87) Publicação PCT: WO 2008/134068 de 06/11/2008

(85) Data do Início da Fase Nacional: 23/10/2009

(57) Resumo: VETOR DE EXPRESSÃO PARA EXPRESSÃO E SECREÇÃO DE PROTEÍNAS HETERÓLOGAS EM CÉLULAS DE INSETO CULTIVADAS. A presente invenção está direcionada a vetores 5 de expressão otimizados para a expressão de proteínas heterólogas semelhantes às nativas em células de inseto. As composições da invenção são sequências de nucleotídeo que representam elementos de um vetor de expressão que, quando combinado, resulta em expressão e secreção melhoradas de proteínas heterólogas. Os elementos 10 incluem sequências que definem ativadores transcricionais, promotores principais, sinais de secreção e regiões 3 não traduzidas que são funcionais em células de inseto. Os elementos contidos nos vetores otimizados são todos sinteticamente derivados ou são variantes modificadas de sequências de inseto de ocorrência natural. O vetores de expressão são úteis para a expressão de proteínas 15 semelhantes às nativas quando as sequências de nucleotídeo de codificação de proteína estão operativamente ligadas aos vetores. Estes vetores podem ser usados para transformar células de inseto, as quais podem então ser cultivadas para produzir o produto de proteína desejado. As proteínas semelhantes às nativas expressas podem ser usadas em diagnóstico, vacina ou outras 20 aplicações que requerem grandes quantidades de proteínas de alta qualidade.

Relatório Descritivo da Patente de Invenção para **"VETOR DE EXPRESSÃO PARA EXPRESSÃO E SECREÇÃO DE PROTEÍNAS HETERÓLOGAS EM CÉLULAS DE INSETO CULTIVADAS"**.

A presente invenção refere-se em geral ao desenho de vetores de expressão aperfeiçoados para a produção de proteínas heterólogas em célula de inseto. Os vetores de expressão aperfeiçoados são compostos de uma série de elementos reguladores que são de origem sintética ou otimizada e reunidos de uma maneira que proporciona altos níveis de expressão de proteína recombinante. Em adição, esta invenção se refere à produção de proteínas de subunidade recombinantes utilizando estes vetores de expressão aperfeiçoados para uso em formulações de vacinas.

Antecedentes da Invenção

Um número de sistemas de expressão de célula heteróloga foi desenvolvido para expressar e secretar proteínas recombinantes. Em geral, sistemas à base de células hospedeiras eucarióticas são empregadas para a expressão de proteínas eucarióticas que exigem dobramento adequado e modificações pós-tradução, desse modo permitindo a produção de proteínas "semelhantes às nativas". Existem quatro tipos de células hospedeiras eucarióticas primárias que são utilizados; fúngica, incluindo levedura; inseto, planta e mamífero.

A escolha de um sistema de expressão de proteína recombinante a ser empregado é dependente da aplicação desejada. O sistema de escolha deve atingir critérios chave, tal como dobramento adequado e processamento, consistência e produtividade (eficácia de custo) do produto de proteína desejado (Schmidt, *Appl. Microbiol. Biotechnol.* (2004) 65:363-372). Sistemas de expressão à base de células de inseto têm o potencial de alcançar requisitos com base na facilidade de cultura, tolerância superior à osmolalidade e concentrações de subproduto durante cultura em larga escala, e geralmente níveis de expressão superiores (Ikonomou *et al.*, *Appl. Microbial Biotechnol.* (2003) 62:1-20). Recentemente, o uso de sistemas de expressão à base de células de inseto tem se tornado mais comum. Estes sistemas proporcionam a maioria das características desejadas de sistemas eucarióticos, porém têm benefícios adicionados, tal como custo mais baixo de artigos.

Sistemas de célula de inseto são ou com base de infecção de células hospedeiras com insetos vetores de vírus (por exemplo, baculovírus) ou na geração de linhagens de células estáveis através da integração de plasmídeos de expressão ao genoma das células hospedeiras.

5 O sistema de expressão de baculovírus (BES) surgiu como o sistema de cultura de célula de inseto primário utilizado para expressão de proteína recombinante. Este sistema é com base no uso de vetores derivados dos vírus de inseto conhecidos como baculovírus. Estes vetores são usados para gerar vírus recombinantes que codificam o produto de proteína

10 desejado. Os vírus recombinantes são usados para infectar células de inseto hospedeiras que então expressam as proteínas recombinantes desejadas. Embora existam vantagens para este sistema em relação à facilidade de clonagem e "tempo para produto", também existem diversas desvantagens. O desafio primário no uso de BES é que ele é com base na infecção viral das

15 células hospedeiras. Isto resulta em lise celular e morte da célula 72 – 96 horas após a infecção (Farrell *et al.*, *Biotech. Biogen.* (1998) 60:656-663; Deo and Park, *Biotechnol. Appl. Biochem.* (2006) 43:129-135). Como resultado, durante os estágios finais da infecção, as máquinas de processamento das células de inseto estão comprometidas na medida em que o processamento

20 do produto desejado também está comprometido. Isto limita o tempo que a célula pode produzir o produto e, possivelmente, de maneira mais importante, leva a formas alteradas do produto a ser produzido. Além disso, a lise de células libera enzimas celulares que também podem afetar a qualidade do produto desejado.

25 O uso de células de inseto estavelmente transformadas para a expressão de proteínas recombinantes é uma alternativa para o uso de BES. Sistemas de expressão com base em linhagens de célula de inseto estavelmente transformadas são não líticos e proporcionam produção estável a longo prazo de produtos secretados que requerem dobramento adequado e

30 modificações pós-tradução. A secreção do produto no meio de cultura proporciona um material de partida mais limpo para o processo de purificação e permite que o produto de proteína final seja purificado com métodos básicos.

Isto leva a produtos que são de qualidade superior (Kirkpatrick and Shatzman em *Gene Expression Systems: Using Nature for the Art of Expression* (1999) pp 289-330).

O sistema de expressão de célula de *Drosophila melanogaster* ("sistema de expressão de *Drosophila*") é um sistema de expressão de proteína heteróloga estabelecido com base no uso de vetores de expressão contendo promotores de *Drosophila* e células S2 de *Drosophila* ("células S2") (Schneider, *Embryol. Exp. Morph.* (1972) 27:353-365). Células S2 são transformadas com estes vetores a fim de estabelecer linhagens de células estáveis que expressam proteínas correspondentes às sequências heterólogas introduzidas no vetor (Johansen, H. *et al*, *Genes Dev.* (1989) 3:882-889; Ivey-Hoyle, M., *Curr. Opin. Biotechnol.* (1991) 2:704-707; Culp, J.S., *et al.*, *Biotechnology* (NY) (1991) 9:173- 177; Patentes U.S. 5,550,043; 5,681,713; 5,705,359; 6,046,025). Este sistema de expressão de célula de inseto demonstrou produzir com êxito um número de proteínas a partir de diferentes fontes. Exemplos de proteínas que foram expressas com êxito no sistema de célula S2 de *Drosophila* incluem HIV gpl20 (Culp, J.S., *et al.*, *Biotechnology* (NY) (1991) 9:173-177; Ivey-Hoyle, M., *Curr. Opin. Biotechnol.* (1991) 2:704-707), dopamina β -hidrolase humana (Bin *et al.*, *Biochem. J.* (1996) 313:57-64), proteína de adesão de célula vascular humana (Bernard *et al.*, *Cytotechnol.* (1994) 15:139-144). Em cada um destes exemplos, os níveis de expressão foram maiores do que outros sistemas de expressão que tinham sido anteriormente utilizados.

Além de altos níveis de expressão, o sistema de expressão de *Drosophila* demonstrou ser capaz de expressar proteínas heterólogas que mantêm função biológica semelhante à nativa (Bin *et al.*, *Biochem. J.* (1996) 313:57-64), (Incardona and Rosenberry, *Mol. Biol. Cell.* (1996) 7:595-611). Exemplos mais recentes demonstraram por meio de estudos de cristalografia de raios-X que este sistema de expressão é capaz de produzir moléculas com estrutura semelhante à nativa (Modis *et al.*, *Proc. Natl. Acad. Sci. USA* (2003) 100:6986-6991), (Modis *et al.*, *Nature* (2004) 427:313-319), (Xu *et al.*, *Acta. Crystallogr. D Biol. Crystallogr* (2005) 61:942-950). Outras duas publi-

cações recentes também demonstraram a capacidade do sistema de expressão de *Drosophila* para produzir produtos de alta qualidade. No primeiro relatório, Schmetzer *et al.* (*J. Immun.* (2005) 174: 942-952) compara proteína EpCAM expressa por baculovírus à proteína EpCAM expressa por *Drosophila* quanto ao dobramento de proteína e conformação nativa. Especificamente, EpCAM expressa por BES e EpCAM expressa por *Drosophila* foram comparadas à EpCAM expressa por *Drosophila* desnaturada. Foi determinado que a EpCAM expressa por BES era um estado dobrado parcial em relação à proteína EpCAM expressa por *Drosophila* desnaturada e não desnaturada. Isto indica que a proteína expressa por BES é um estado dobrado de modo incompleto. A proteína EpCAM expressa por *Drosophila*, por outro lado, adotou um estado mais completamente dobrado. Os autores deste artigo consideraram a proteína expressa por *Drosophila* como estando no estado "natural", enquanto a proteína expressa por baculovírus não estava. No segundo relatório, Gardsvoll *et al.* (*Prot. Exp. Purif.* (2004) 34:284-295) demonstram que a expressão do receptor de ativador de plasminogênio do tipo uroquinase (uPAR) em células S2 resulta em um produto mais homogêneo em relação à glicosilação (5 sítios ligados a N) do que uPAR expresso em células CHO.

Com base no corpo de trabalho que utiliza células S2 de *Drosophila* como células hospedeiras para a expressão de proteínas heterólogas, é evidente que estas células têm muitas características que são desejáveis em um sistema de expressão. No levantamento dos relatórios publicados que utilizam essas células para produzir proteínas recombinantes, o nível de expressão dos produtos de proteína é tipicamente na faixa de 5 a 50 µg/ml. Para ter um sistema de produção de proteína que possa atender às exigências de produção de biotecnologia, níveis superiores de expressão são desejáveis. Para alcançar níveis de expressão consistentemente superiores, seria necessária a otimização de qualquer ou todos dos seguintes, linhagem de células hospedeiras, meio de crescimento ou vetores de expressão.

O desenvolvimento de qualquer sistema de expressão de proteína heteróloga requer a montagem de vários elementos de controle regulado-

res (daqui por diante, alternativamente chamados "elementos reguladores" ou "elementos de controle" ou simplesmente "elementos", incluindo a forma singular de cada um) em vetores de expressão que conduzem a expressão do produto de proteína recombinante desejado. Estes elementos de controle reguladores incluem ativadores e melhoradores transcrpcionais, elementos iniciadores e de término transcrpcionais, elementos de início e parada translacionais e sequências líderes sinal de secreção. Os cinco principais elementos de controle reguladores de um vetor de expressão para a secreção de produtos de proteína recombinante são 1) promotor proximal, 2) promotor principal 3) região não traduzida 5' 4) peptídeo de sinal de secreção e 5) região não traduzida 3'. Para cada um destes elementos, qualquer otimização deve primeiro ser feita independentemente. Uma vez que os elementos individuais são otimizados, eles devem ser montados e testados para assegurar que a montagem de dados elementos seja compatível e capaz de direcionar e expressão eficiente dos produtos de proteína recombinante desejados. No montagem das várias combinações de elementos, também é importante que os elementos sejam "operavelmente ligados". "Operavelmente ligados" se refere à ligação funcional entre os vários elementos de uma maneira que conserve a função de cada elemento individual, assim como a função dos elementos combinados, visto que muitas funções transcrpcionais e translacionais são o resultado do processamento a partir de um elemento para o próximo. Portanto, "operavelmente ligado" significa que as sequências de ácido nucléico dos vários elementos são ligadas e contíguas e, onde necessário, são contiguamente ligadas para manter uma estrutura de leitura de codificação de proteína apropriada.

Embora muitos vetores de expressão bem-sucedidos tenham sido desenvolvidos, sistemas totalmente otimizados são menos comuns. Além disso, sistemas mais desenvolvidos são com base no uso de sequências de ocorrência natural, isto é, os vários elementos reguladores são tomados a partir de sequências de gene existentes (de ocorrência natural). Ao longo dos últimos anos, o uso de sequências sintéticas tem sido empregado como um meio para desenvolver vetores de expressão otimizados.

Um "promotor" é composto de duas partes básicas, o promotor principal e o promotor proximal. O promotor está localizado a montante da sequência de codificação de um dado gene. O promotor principal é definido como a sequência de nucleotídeo mínima que é capaz de direcionar inicia-

5 ção de transcrição exata de um dado gene. O promotor principal em eucariotos é responsável por direcionar a iniciação através do complexo RNA polimerase II. O promotor principal é geralmente delineado como a sequência que abrange o sítio de iniciação de transcrição (INR), mais especificamente, os nucleotídeos 35 a 45 da sequência a montante e a jusante do INR (com-

10 primento total de 70 a 90 nucleotídeos). A região unida através do promotor principal pode conter um ou mais dos "motifs" reguladores conservados a seguir, elemento de reconhecimento TFIIB (BRE), TATA box, iniciador (INR), elemento "motif" ten" (MTE), elemento promotor a jusante (DPE) e elemento principal a jusante (DCE). Embora certas sequências de nucleotídeo dentro de

15 um elemento de controle regulador tenham nomes reconhecidos na técnica que incluem a palavra "elemento", tais sequências são aqui chamadas, em termos gerais, "motifs"; nomes específicos reconhecidos na técnica (por exemplo, BRE, MTE, DPE e DCE) têm o mesmo significado aqui como nas referências citadas, embora eles sejam chamados "motifs" aqui, por exemplo

20 "motif" MTE". O papel e a composição dos promotores centrais e seus "motifs" individuais constituintes foram revisados por Ohler *et al.* (*Genome Biol.*, (2002) 3:1-12), Smale and Kadonaga (*Ann. Rev. Biochem.* (2003) 72:449-479), FitzGerald *et al.* (*Genome Biol.* (2006) 7:R53), Gershenson *et al.* (*BMC Genomics* (2006) 7:161) e Juven-Gershon *et al.* (*Biochem. Soc.*

25 *Trans.* (2006) 34: 1047- 1050). Estudos definindo o "motif" DPE (Kutach and Kadonaga, *Mol. Cell. Biol.*(2000) 20:4754-4764) e o "motif" MTE (Lim *et al.*, *Genes Dev.* (2004) 18:1606-1617) também foram relatados. Embora os promotores centrais de muitos genes contenham um ou mais destes "motifs" em várias combinações, existe um pequeno percentual que não contém qual-

30 quer destes "motifs". No levantamento de promotores centrais de diversos organismos, é evidente que nenhum promotor principal, ou subconjunto universal de "motifs" compreendendo um promotor principal, existe; entretanto,

o "motif" INR é o "motif" mais comum (FitzGerald *et al.*, *Genome Biol.* (2006) 7:R53) encontrado em promotores centrais.

O promotor proximal em eucariotos geralmente é definido como a sequência que está imediatamente a montante do promotor principal. O comprimento do promotor proximal é altamente variável. O promotor proximal é composto de "motifs" ativadores transcrpcionais que recrutam fatores de transcrição que, por sua vez, ativam o complexo de iniciação de polimerase, o qual é unido à região do promotor principal. A natureza e o número dos "motifs" ativadores transcrpcionais são altamente variados para um dado promotor.

A otimização de promotores pode ser realizada por substituição sistemática e testagem de vários elementos ou "motifs", o uso de bibliotecas de promotor sintético, ou substituição randômica de elementos reguladores individuais ou "motifs". Exemplos destas diferentes abordagens para o desenvolvimento de promotor otimizado ou sintético foi relatado. No trabalho de Li *et al.* (*Nature Biotechnol.* (1999) 17:241-245), uma abordagem de avaliação de sítios de união de fator de transcrição aleatoriamente montados para conduzir a transcrição de promotores músculo-específicos foi relatada. Edelman *et al.* (*Proc. Natl. Acad. Sci.* (2000) 97:3038-3043) descreve um procedimento de seleção de alto rendimento para selecionar promotores proximais sintéticos que aumentem a atividade transcrricional de um promotor principal. Tornoe *et al.* (*Gene* (2002) 297:21-32) construiu um conjunto de promotores sintéticos para uso em células de mamíferos que combinou elementos promotores virais e humanos. Os promotores sintéticos de mamíferos foram adicionalmente otimizados através da substituição de sequências de consenso e randomização de outras sequências de não consenso para obter promotores com atividade variável. No desenvolvimento de promotores sintéticos para expressão de gene em *Lactobacillus*, Rud *et al.* (*Microb.* (2006) 152:1011-1019) utilizou sequências de consenso para elementos reguladores junto com randomização das sequências espaçadoras entre os elementos reguladores para melhorar o desempenho do promotor. Desta maneira, eles foram capazes de identificar promotores sintéticos que

tinham atividade aumentada. No trabalho de Juven-Gershon *et al.* (*Nature Methods* (2006) 3:917-922), o desenvolvimento de um promotor principal otimizado é alcançado através da combinação de "motifs" de promotor principal de diferentes genes virais e de *Drosophila*. Este trabalho focou no uso do "motif" MTE (Lim *et al.*, *Genes Dev.* (2004) 18:1606-1617) do promotor principal. Esta estratégia resultou em promotores centrais com atividade transcricional aumentada. Com base nestes relatórios, fica evidente que existe uma necessidade de determinar experimentalmente que elementos reguladores, e "motifs" constituintes, constituem um promotor funcional. Isto inclui quais combinações de "motifs" são usadas, em que ordem os "motifs" são montados e em que orientação os "motifs" e elementos são inseridos. Isso é necessário se os "motifs" ou elementos utilizados são com base nas sequências sintéticas ou com base nas sequências otimizadas. Embora a otimização de promotores tenha resultado em aperfeiçoamentos na expressão, o promotor apenas representa um aspecto dos elementos reguladores necessários em um vetor de expressão otimizado e totalmente funcional.

A sequência do região não traduzida 5' (5' UTR) de RNA mensageiro (mRNA) desempenha um papel importante na regulação pós-transcrição de expressão de gene a partir de mRNA eucariótico. A variabilidade de sequências e a importância de vários "motifs" e características da região 5'UTR de mRNA foram documentadas (Kozak, *J. Mol. Biol.* (1994) 235:95-110), e Kozak, *Gene* (2005) 361 : 13-37). A natureza da 5'UTR desempenha um papel em estabilidade de mensagem e eficiência de tradução. A estabilidade e a traduzibilidade de um dado mRNA irá impactar a capacidade de eficazmente expressar proteínas recombinantes. Portanto, o desenho de vetores de expressão para a produção ótima de proteínas recombinantes requer que a 5'UTR seja avaliada quanto a sua capacidade de funcionar de uma maneira eficiente em um dado sistema de célula hospedeira. A 5'UTR, como descrita, é uma parte importante do mRNA que é codificado pela sequência de DNA contida na extremidade 3' do promotores de gene. Por isso, no contexto da definição da sequência de DNA do promotor, é geralmente inclusive da sequência 5'UTR (a região do INR para o códon iniciador para

metionina).

As sequências da região não traduzida 3' (3'UTR) de mRNA, junto com a 5'UTR, desempenham importantes papéis na regulação pós-transcrição de expressão de gene. A natureza da 3'UTR desempenha um papel na estabilidade da mensagem, transporte do núcleo ao citoplasma e localização subcelular. Cada um destes fatores pode ter um impacto da eficiência da tradução de uma dada mensagem e, por fim, no nível de expressão de proteína. Portanto, o desenho de vetores de expressão para a produção ótima de proteínas recombinantes requer que a 3'UTR seja avaliada quanto à sua capacidade de funcionar de uma maneira eficiente em um dado sistema de célula hospedeira.

A maioria das proteínas que são secretadas a partir das células contém uma sequência sinal terminal em N que direciona a proteína para dentro do caminho de secreção da célula. Em células eucarióticas, o sinal de secreção ou peptídeo de sinal interage com a membrana endoplasmática para iniciar o processo de secreção. A sequência sinal eucariótica foi dividida em três regiões estruturais; básica, hidrofóbica e polar, iniciando a partir do término em N e prosseguindo para o término em C, respectivamente (von Heijne, *NMC Acids Res.* (1986) 14:4683-4690) e (Bendtsen *et al.*, *J. Mol. Biol.* (2004) 340:783-795). Ao longo dos anos, numerosos sinais de secreção foram identificados e usados para direcionar a secreção de proteínas recombinantes. Embora muitas sequências sinal diferentes tenham sido usadas e demonstrado serem funcionais, poucos estudos foram relatados definindo sequências ótimas para um dado tipo de célula. As características e regras gerais relacionadas às três regiões estruturais são bem estabelecidas, conforme detalhado por von Heijne (*Nuc. Acids Res.* (1986) 14:4683-4690) e por Bendtsen *et al.* (*J. Mol. Biol.* (2004) 340:783-795), entretanto, existem poucos dados experimentais comparativos quanto ao que constitui um sinal de secreção ótimo. A maioria dos relatórios publicados lida com a caracterização e otimização de sinais de secreção de levedura ou de bactéria gram positiva (Le Loir *et al.*, *Microb. Cell Fact.* (2005) 4:2 e Hofmann and Schultz, *Gene* (1991) 101:105-111). Um relatório que descreve a otimização

do sinal de secreção de IL-2 evidentemente demonstra os benefícios da otimização (Zhang *et al.*, *J. Gene Med.* (2005) 7:354-365).

O desenvolvimento de vetores de expressão otimizados para uso em células de inseto para gerar linhagens de células estáveis capazes de produzir grandes quantidades de proteínas recombinantes de alta qualidade requer a identificação de elementos reguladores apropriados, incluindo, porém não limitados ao elemento promotor principal, que pode ser usado para conduzir a transcrição e tradução das proteínas heterólogas a serem expressas. Além disso, elementos reguladores sintéticos podem ser designados e utilizados para adicionalmente otimizar a funcionalidade dos vetores de expressão. As divulgações de Lim *et al.*, *Genes Dev.* (2004) 18:1606-1617, Kutach and Kadonaga, *Mol. Cell. Biol.* (2000) 20:4754- 4764), e Juven-Gershon *et al.*, *Biochem. Soc. Trans.* (2006) 34:1047-1050 estão limitadas ao elemento promotor principal, e especificamente a sequências heterólogas ou novas e / ou de espaçamento do TATA box, "motifs" INR, MTE e DPE. Otimização total de controle regulador de expressão de proteínas recombinantes requer que múltiplos elementos reguladores, não apenas "motifs" no promotor principal, sejam otimizados.

Embora muitos elementos reguladores sejam conhecidos para *Drosophila*, assim como outros insetos, o que constitui elementos opcionais para a expressão de proteínas heterólogas em células de inseto não é conhecido. Tecnologia e métodos atuais proporcionam o potencial para montar elementos reguladores em um vetor de expressão com base no corpo de conhecimento atual. Embora o potencial exista, é de conhecimento comum que nem todas as tentativas de fazê-lo resultará em sucesso. Os elementos reguladores de expressão recombinantes que trabalham em um tipo de célula nem sempre trabalha em outro tipo de célula. Por exemplo, Olsen *et al.* (*Cytotechnol.* (1992) 10:157-167) avaliou a capacidade de uma série de promotores de conduzir a expressão de uma proteína heteróloga em células S2 e verificou que apenas o MtnA de *Drosophila* resultou em rendimentos de microorganismo de produto, apesar do fato de que todos os promotores testados tenham demonstrado trabalhar em outros tipos de célula. Em outro

exemplo, um vetor de expressão *Bombyx mori* com base nos promotores IE, que trabalha bem em células de Lepidópteros (Farrell *et al.*, Biotechnol. Bioeng. (1998) 60:656-663), não conduz adequadamente a expressão de proteínas heterólogas em células S2 (dados não publicados). Portanto, uma avaliação sistemática é necessária para determinar o potencial para consistentemente expressar altos níveis de proteínas heterólogas de alta qualidade usando células S2. No campo da biotecnologia, a capacidade de eficientemente produzir proteínas recombinantes a um custo favorável de artigos é a chave para o sucesso. A fim de alcançar este objetivo usando células S2 de *Drosophila*, é necessário o desenvolvimento adicional de vetores de expressão adequados.

A combinação de múltiplos elementos reguladores de uma maneira apropriada, de modo que um benefício aditivo seja alcançado pode também aumentar a utilidade do vetor de expressão. Portanto, os problemas técnicos a serem resolvidos são: (1) identificação de elementos reguladores para inclusão em vetores de expressão que são capazes de conduzir a expressão de grandes quantidades de proteínas recombinantes de alta qualidade em células de inseto, (2) o desenho de versões sintéticas de elementos reguladores funcionais que têm função aperfeiçoada, e (3) determinação da combinação ótima de múltiplos elementos reguladores, de modo que a combinação resulta em um aumento aditivo na produtividade da expressão da proteína. Aperfeiçoamentos adicionais em sistemas de expressão de inseto estáveis podem potencialmente proporcionar novas plataformas para a fabricação de proteínas, onde grandes quantidades de proteína de alta qualidade são necessárias, tal como em sistemas com base em célula para a produção de vacinas de subunidades contra doenças infecciosas, por exemplo, influenza ou organismos com potencial para bioterrorismo.

Sumário da Invenção

A presente invenção proporciona vetores de expressão compostos de elementos reguladores sintéticos e otimizados que são combinados para proporcionar altos níveis de expressão de proteínas recombinantes heterólogas em células de inseto estavelmente transformadas. De modo espe-

cífico, a invenção está direcionada à expressão de proteínas heterólogas quando células S2 de *Drosophila melanogaster* são usadas como a célula hospedeira.

Os vetores de expressão pHBI-10 (SEQ ID NO: 10) e pHBI-11 (SEQ ID NO: 11) descritos abaixo são compostos de cinco elementos reguladores: promotor proximal, promotor principal, 5'UTR, sequência sinal de secreção e 3'UTR. Os cinco elementos estão operavelmente ligados para criar um cassete de expressão. Os elementos de controle reguladores contidos nos cassetes de expressão são cada um otimizado para uso em células S2 de *Drosophila*; entretanto, as sequências de nucleotídeo que definem os elementos reguladores da presente invenção não são naturalmente encontrados na sequência genômica de *Drosophila* devido à natureza sintética ou otimizada destas sequências.

Os vetores de expressão da presente invenção são capazes de direcionar a expressão e secreção de alto nível de proteínas heterólogas para o meio de cultura de células transformadas. Especificamente, os vetores de expressão descritos contem 1) promotor proximal, otimizado induzível, sintético; 2) promotor principal de RNA polimerase II otimizado sintético; 3) sequência 5'UTR truncada e otimizada; 4) sequência sinal otimizada, sintética e 5) sequência 3'UTR otimizada.

A invenção também proporciona métodos para clonar sequências de gene heterólogas nos vetores de expressão otimizados e utilizando os vetores de expressão otimizados para transformar células de inseto, o que resulta na secreção de altos níveis de proteínas recombinantes que mantêm estrutura semelhante à nativa.

Breve Descrição dos Desenhos

Figura 1. Sequência do promotor sintético 3XR MRE-3XR MRE-SCP 1-TolloMTE-LII 11 (SEQ ID NO: 12). Cada um dos "motifs" reguladores nas regiões do promotor principal e central está indicado.

Figura 2. Sequência do promotor sintético 3XR MRE-3XR MRE-SCP7 (SEQ ID NO: 13). Cada um dos "motifs" reguladores nas regiões de promotor proximal e central está indicado.

Figura 3. Comparação dos vetores de expressão pMTtP A, pHBI-10 e pHBI-11. Os elementos de controle reguladores para cada vetor estão listados na Figura 3.

Figura 4. Mapa do plasmídeo pHBI-10. O mapa de plasmídeo do plasmídeo total é mostrado. O vetor pHBI-10 contém o promotor principal sintético TolloMTE- Δ 111.

Figura 5. Mapa do plasmídeo pHBI-11. O mapa de plasmídeo do plasmídeo total é mostrado. O vetor pHBI-11 contém o promotor principal sintético SCP7.

10 Descrição Detalhada da Invenção

A invenção proporciona uma série de elementos reguladores otimizados que foram combinados em vetores de expressão que são usados para conduzir a expressão de altos níveis de proteínas de alta qualidade de células de inseto que foram estavelmente transformadas com estes vetores de expressão que carregam a sequência de gene da proteína a ser expressa. O uso de elementos reguladores sintéticos e otimizados em um único vetor de expressão resulta na capacidade de segura e rapidamente produzir proteínas recombinantes em um sistema de expressão de inseto estável em níveis que proporcionam um custo favorável de artigos para o produto expresso.

O termo "elemento regulador" (ou "elemento de controle regulador" ou "elemento de controle" ou simplesmente "elemento") se refere a um segmento de um vetor de expressão, que tem uma função conhecida nos processos transcricional e translacional envolvidos na expressão do produto de proteína resultante de uma dada sequência de gene. Elemento regulador é distinto do termo "motif" regulador" (ou simplesmente "motif"), que se refere a uma sequência definida que atua ou como um sítio de união ou marca um sítio transitório. Em geral, elementos reguladores são compostos de múltiplos "motifs" reguladores.

30 O termo elemento regulador "sintético" se refere a sequências que não são encontradas como ocorrência natural. Mais especificamente, os elementos sintéticos descritos aqui não são encontrados na sequência

genômica de *Drosophila*. O termo elemento regulador "otimizado" se refere a sequências que foram derivadas de sequências de ocorrência natural e foram alteradas para melhorar suas funções. As sequências específicas dos elementos reguladores otimizados não representam variantes de ocorrência natural das sequências a partir das quais as sequências otimizadas são derivadas; portanto, os elementos reguladores otimizados (e "motifs" dos mesmos) podem ser referidos sintéticos.

"Cassete de expressão" significa a combinação de elementos promotores com outros elementos de controle reguladores transcricionais e translacionais que estão operavelmente ligados. Uma sequência de gene heteróloga pode ser inserida no cassete de expressão com a finalidade de expressão da dita sequência de gene. O cassete de expressão é capaz de direcionar a transcrição, o que resulta na produção de um mRNA para o produto de gene desejado. O cassete de expressão é inserido em um plasmídeo para produzir um vetor de expressão. Tal vetor de expressão direciona a expressão da proteína heteróloga em células hospedeiras.

O termo "transformado" se refere à transformação mediada por DNA de células. Isto se refere à introdução de DNA plasmídeo em células de inseto no processo de gerar linhagens de células estáveis após a integração do DNA introduzido ao genoma das células. Este termo é usado no lugar do termo "transfecção", o qual é frequentemente usado no mesmo contexto. Foi usado o termo transformação para a introdução de DNA plasmídeo em células cultivadas para distinguir da introdução de DNA viral em células cultivadas, que foi originalmente referido como transfecção. Visto que não existem sequências de DNA viral nos vetores de expressão da presente invenção e a introdução destes vetores de expressão nas células não resulta na produção de partículas semelhantes a vírus ou lise da célula, o termo transformado é preferido.

"Expressão" ou "expresso" significa a produção de proteínas usando vetores de expressão e células hospedeiras, por exemplo, células S2 de *Drosophila* para produzir um produto de proteína recombinante que é prontamente detectável como um produto associado à célula ou como um

produto secretado no meio de cultura.

"Secreção" significa secreção de uma proteína recombinante expressa a partir de células hospedeiras cultivadas em meio de cultura. A proteína expressa e secretada é o resultado de uma dada sequência de gene estando operavelmente ligada a um cassete de expressão, de modo
5 que a sequência codifica a dada proteína.

O termo "produto" se refere a qualquer proteína recombinante, comprimento total ou subunidade da mesma, que é expressa por uma célula hospedeira em que um vetor de expressão que carrega a sequência de gene
10 que codifica o produto foi introduzido.

Células de inseto são um sistema de expressão eucariótico alternativo que proporciona a capacidade de expressar proteínas devidamente dobradas e pós-translacionalmente modificadas enquanto proporcionando condições de crescimento simples e relativamente não dispendiosas. O uso
15 de sistemas de expressão de célula de inseto estavelmente transformadas proporciona benefícios sobre aqueles com base em infecção por baculovírus das células de inseto hospedeiras. Neste sentido, células S2 foram selecionadas como as células de inseto hospedeiras de escolha. Como resultado, os esforços para otimizar os vetores de expressão para células de inseto
20 estavelmente transformadas foram com base nos dados derivados da análise de genes específicos da *Drosophila*, assim como o genoma completo da *Drosophila*.

Em uma modalidade preferida da invenção, o promotor principal do vetor de expressão é definido como a sequência que abrange o sítio de
25 iniciação de transcrição (INR), mais especificamente os nucleotídeos 35 a 45 da sequência a montante e a jusante do INR (comprimento total de 70 a 90 nucleotídeos). O nucleotídeo "A" (adenina) do "motif" INR é designado como +1. Os promotores principais da invenção contêm uma combinação de "motifs" que não é encontrada na natureza. Estes promotores principais são
30 compostos de "motifs" reguladores conhecidos; entretanto, a combinação dos "motifs" é única e diversos dos "motifs" no promotor principal da invenção são com base na sequência de consenso e ainda ajustados para função

ótima, quando combinadas para formar o elemento regulador promotor principal. As regiões conhecidas de importância nos promotores principais dos vetores de expressão descritos aqui são o TATA box, sítio de iniciação transcricional (INR), elemento "motif" ten (MTE) e elemento promotor a jusante (DPE). Os promotores da presente invenção, com promotores principais sintéticos que contem pelo menos três destes quatro "motifs" reguladores principais, estão apresentados em SEQ ID NO: 1, SEQ ID NO: 2 e SEQ ID NO: 3. Na presente invenção, os promotores principais sintéticos (SEQ ID NO: 2, SEQ ID NO: 3) são capazes de conduzir altos níveis de expressão quando unidos com um elemento regulador a montante apropriado, tal como um promotor proximal.

Em ainda outra modalidade preferida da invenção, a sequência 5'UTR é definida como a sequência a partir do nucleotídeo "A" (adenina) do "motif" INR ao nucleotídeo na sequência que imediatamente precede o códon iniciador para metionina. As 5'UTRs da invenção têm uma 5'UTR comparativamente curta e contém uma sequência consenso de Kozak de *Drosophila* na extremidade 3'. As sequências que definem a 5'UTR do mRNA que resulta da transcrição direcionada pelos dois promotores principais sintéticos da presente invenção estão apresentadas em SEQ ID NO: 4 e SEQ ID NO: 5.

Em outra modalidade preferida da invenção, os elementos reguladores a montante contidos no promotor principal que conduzem a transcrição do promotor principal e resulta em altos níveis de expressão, e uma sequência sintética composta de múltiplos elementos responsivos a metal (MRE). Este promotor proximal sintético é composto de 2 conjuntos de 3XMREs separados por 11 nucleotídeos. Todos os MREs são orientados em uma direção reversa em relação ao início da transcrição, como ilustrado nas Figuras 1 e 2. A sequência deste promotor proximal sintético está apresentada em SEQ ID NO: 6.

Em ainda outra modalidade preferida da invenção, uma sequência de peptídeo de sinal de secreção otimizada e proporcionada. Variando-se o comprimento total, a região básica terminal em N e a compo-

sição da região hidrofóbica, uma sequência de peptídeo de sinal otimizada foi designada e sintetizada para uso no direcionamento da secreção de proteínas expressas no meio de cultura das células transformadas. Sítios de restrição que permitem a clonagem de sequências de codificação de proteína na mesma estrutura de leitura como sinal de secreção, e que não impactam negativamente o sítio de clivagem de sinal de secreção, também foram designados e incorporados à sequência sinal de secreção. A sequência de aminoácido do sinal de secreção sintético e a sequência de nucleotídeo que codifica o peptídeo estão apresentadas em SEQ ID NO: 7 e SEQ ID NO: 8.

Em ainda outra modalidade preferida da invenção, a sequência 3'UTR das transcrições do mRNA que são produzidas através dos vetores de expressão da invenção e uma versão otimizada de uma sequência 3'UTR nativa de *Drosophila* derivada do gene que codifica uma proteína altamente expressa e secretada chamada proteína semelhante à quitinase (Kirkpatrick *et al.*, *Gene* (1995) 153:147-154). A proteína semelhante à quitinase é uma das proteínas mais abundantemente expressas e secretadas em células S2. Por esta razão, o gene que codifica a proteína semelhante à quitinase (CLP) foi selecionada com uma base para a otimização da 3'UTR. A 3'UTR do gene da CLP não contém o "motif" sinal de poliadenilação típico, AATAAA. Modificações e truncações feitas para esta 3'UTR resultaram em níveis superiores de expressão. A sequência da 3'UTR da presente invenção esta apresentada em SEQ ID NO: 9.

Em uma modalidade preferida da invenção, a combinação dos cinco elementos reguladores descritos acima, que inclui ambos elementos transcrpcionais e translacionais, em um cassete de expressão funcional que direciona expressão de alto nível em células S2 e descrita. As sequências de dois cassetes de expressão montados que variam apenas nas sequências do promotor principal utilizado estão apresentadas em SEQ ID NO: 10 e SEQ ID NO: 11.

Desse modo, a presente invenção proporciona vetores de expressão compostos de elementos reguladores sintéticos e otimizados que são capazes de conduzir altos níveis de expressão de proteína heteróloga.

Os vetores de expressão sintéticos, quando usados para expressar proteínas heterólogas em células S2, possibilitam a produção econômica de grandes quantidades de proteínas de alta qualidade. Todos os elementos de controle reguladores sintéticos, ou uma combinação de sintéticos e tipo selvagem, ou de sintéticos e de consenso, ou de sintéticos, tipo selvagem e de consenso, podem ser usados em um cassete de expressão. Os exemplos abaixo mostram que o uso de todos os elementos sintéticos produz (ou "conduz") os níveis mais altos de expressão de proteína heteróloga. Onde menos do que todos os elementos de controle reguladores são sintéticos, os elementos restantes são assumidos como sendo elementos do tipo selvagem (não otimizados) conhecidos na técnica para trabalhar com um dado tipo de célula hospedeira.

Embora as descrições apresentadas acima e os exemplos que seguem sejam principalmente direcionados ao uso dos vetores de expressão otimizados com células S2 de *Drosophila*, os vetores e métodos podem ser aplicados a outras linhagens de células de inseto que resultam em linhagens de células estáveis após a transformação de células hospedeiras com DNA plasmídeo.

Os exemplos a seguir são oferecidos a título de ilustração e não por meio de limitação.

Exemplos

Os exemplos a seguir descrevem o desenvolvimento dos vetores de expressão sintéticos para uso em células de inseto. Os exemplos demonstram a capacidade de eficientemente expressar proteínas heterólogas em células S2 de *Drosophila* em níveis que são comercialmente adequados para desenvolvimento do produto.

Os exemplos demonstram a capacidade dos elementos reguladores individuais para aumentar a capacidade para expressar proteínas em células S2 e os esforços feitos para determinar que as mudanças contribuíram para a função aumentada destes elementos. Os resultados apresentados abaixo demonstram que elementos diferentes e a modificação destes elementos podem resultar em altos níveis de expressão ou em muito

pouca expressão ou nenhuma expressão detectável. Assim, a seleção de elementos reguladores funcionais e eficazes deve ser determinada através de experimentação. Por isso, a invenção descrita aqui é única pelo fato de que o cassete de expressão descrito é sobretudo sintético em composição e direciona altos níveis de expressão de proteína.

Exemplo 1

Uso de Promotores de não *Drosophila* para Conduzir a Expressão de Proteínas Heterólogas em Células S2 de *Drosophila*

Em um esforço para identificar vetores de expressão alternativos capazes de conduzir altos níveis de produtos de alta qualidade em células S2, vetores de expressão contendo promotores derivados de outros insetos foram avaliados. Para este padrão de trabalho, método de cultura e transformação de células S2 foram utilizados (Van der Straten, *Methods in Mol. and Cell Biol.* (1989) 1 :1-8; Culp *et al.*, *Biotechnology* (1991) 9:173-177; Kirkpatrick and Shatzman, *In Gene Expression Systems: Using Nature for the Art of Expression*, Eds. Fernandez and Hoeffler, Editora Academic, (1999) 289-330). Células S2 de *Drosophila* (Schneider, *J. Embryol. Exp. Morph.* (1972) 27:353-365) obtidas a partir de ATCC foram utilizadas. As células S2 foram adaptadas para crescer em meio Excell 420 (SAFC, St Louis, MO) e todos os procedimentos e cultura descritos aqui foram em meio Excell 420. Culturas são tipicamente semeadas a uma densidade de 1×10^6 células/ml e são passadas entre os dias 5 e 7. Todas as culturas foram incubadas a 26° a 27°C . Plasmídeos de expressão nos quais genes de interesse foram inseridos foram transformados em células S2 por meio do método de fosfato de cálcio. As células S2 foram cotransformadas com o plasmídeo pCoHygro para seleção com hidromicina B a uma proporção de 20 μg de plasmídeo de expressão para 1 μg de pCoHygro. Após transformação, células resistentes à higromicina, 0,3 mg/ml, foram selecionadas. Uma vez que linhagens de células estáveis foram selecionadas, elas foram avaliadas para expressão do produto apropriado. Para a avaliação de expressão, culturas de 5 ml de linhagens de célula selecionadas foram semeadas em 2×10^6 células/ml e cultivadas na

presença de sulfato de cobre a 0,2 mM a 26°C por 7 dias. As culturas foram avaliadas quanto à expressão de proteínas recombinantes tanto nas frações associadas da célula quanto no meio de cultura. Proteínas foram separadas por SDS-PAGE e ou tingidas com azul de Coomassie ou manchadas por nitrocelulose. Anticorpos específicos para a proteína a ser expressa foram usados para sondar as manchas de Western. Níveis de expressão de 1 µg/ml (1 mg/L) ou mais são prontamente detectadas em culturas de S2 através de tingimento com Coomassie de géis de SDS-PAGE.

O primeiro promotor de inseto testado foi o promotor A3 de actina citoplasmática *Bombyx mori* (mariposa da seda). Este promotor foi usado em um vetor de expressão designado para alto nível de expressão de células de *B. mori* estavelmente transformadas. O vetor de expressão, pIE1/153A, contém, além do promotor Ae da actina, o fator de transcrição imediatamente anterior (ie-1) do vírus da poliedrose nuclear *B. mori* (BmPNV) e a região de repetição 3 homóloga (HR3) de BmNPV (Farrell et al. (Biotech. Bioeng. (1998) 60:656-663). Com a transformação de células de lepidópteros, este vetor de expressão demonstrou conduzir altos níveis de expressão de proteínas recombinantes.

Vários genes que codificam proteínas de subunidade, tais como o ectodomínio da hemaglutinina (HA) de influenza (H3 HA-Ecto) e o fragmento terminal em C da proteína 1 p42 de superfície de merozoítos da malária (MSPI-p42) foram clonados no vetor pIE/153A, e então os plasmídeos recombinantes resultantes foram usados para transformar células S2. A sequência da H3 HA-Ecto era derivada da sequência de gene HA (HAO) de comprimento total A/Fujian/411/02 da cepa H3N2 influenza, que codifica uma proteína de 566 resíduos de aminoácidos. Especificamente, a sequência utilizada era derivada da sequência de nucleotídeo em número de acesso ISDN38157 (ISD, www.flu.lanl.gov). A sequência de proteína HAO contém uma sequência sinal de secreção de 16 aminoácidos no término em N e uma âncora de membrana terminal em C. Para a expressão da H3 HA-Ecto solúvel, uma molécula de N e C truncada foi expressa, que está contida na sequência de Gln₁₇ a Gly₅₂₆ (resíduo 175 de HA2, análogo ao término em C

da estrutura de cristal X31, Wilson *et al.*, *Nature* (1981) 289:366-373) da proteína de comprimento total. As sequências de aminoácidos e nucleotídeo H3 HA-Ecto e expressão são descritas em detalhes em WO 2007/022425. A sequência MSP1-p42 da malária era derivada da cepa FUP de *Plasmodium falciparum* (Número de acesso no Genbank M37213). A sequência de codificação MSP-I p42 obtida por amplificação de PCR codifica aminoácidos Ala₁₃₃₃ a Ser₁₇₀₅ da proteína MSP-I. As sequências de aminoácido e nucleotídeo MSP1-p42 e expressão são descritas em detalhes em WO 2006/026625.

10 O nível de expressão para ambas as proteínas de subunidade de influenza e malária foram menos do que 1 µg/ml em células S2, quando a expressão foi conduzida pela vetor pIE/153A. Com base nestes resultados, o vetor pIE/153A não direciona expressão de alto nível de proteínas de subunidade recombinantes, quando usado para transformar células S2.

15 O segundo promotor de inseto testado foi o promotor de metalotioneína 1 de *Anopheles gambiae* (mosquito). Para testar este promotor quanto à sua capacidade de conduzir expressão de alto nível em células S2, a região do promotor da metalotioneína de *Drosophila melanogaster* (MtnA) foi removida do vetor de expressão de *Drosophila* pMTtPA e substituída com
 20 o promotor de *Anopheles*. O vetor de expressão pMTtPA e métodos para seu uso para expressão de proteínas recombinantes em células de *Drosophila* cultivadas são descritos nas Patentes U.S: 5,550,043; 5,681,713; 5,705,359; 6,046,025. O plasmídeo pCoHygro descrito na Patente U.S. 5,681,713 foi usado para seleção de agramicina após cotransformação de
 25 células S2. O vetor de expressão pMTtPA contém os elementos a seguir: o promotor da metalotioneína de *Drosophila* (MtnA), a sequência sinal pre-pro do ativador de plasminogênio do tecido humano (tPA) e o sinal de poliadenilação prematuro SV40 (Culp *et al.*, *Biotechnol.* (1991) 9:173-177). O plasmídeo pCoHygro proporciona um marcador selecionável para
 30 higromicina (van der Straten *et al.*, *Methods in Mol. And Cell Biol.* (1989) 1 : 1-8). O gene da higromicina está sob o controle transcricional da repetição terminal longa do elemento transponível *Drosophila* COPIA. O vetor pMTtPA foi

modificado excluindo um fragmento *BamHI* de 15 pares de base que continha um sítio *XhoI* estranho. Este vetor modificado, referido como pMTtLIXho, permite a clonagem direcional de inserções utilizando locais *Bgl II* e *XhoI* únicos. O vetor foi adicionalmente

5 modificado através da adição de um sítio de restrição *HindIII* e adição de uma "sequência consenso de Kozak" de *Drosophila* (Cavener, *Nuc. Acids Res.* (1987) 15:1353-1361) imediatamente anterior ao códon iniciador para metionina da sequência sinal pre-pro tPA. Especificamente, a sequência de 12 nucleotídeos em pMTtPA, TGTGAAGCAATC (SEQ ID NO: 14), imediatamente anterior a ATG foi mudada para AAGCTTAACAAC

10 (SEQ ID NO: 15, os primeiros seis nucleotídeos representam o sítio de restrição *HindIII* e os segundos seis nucleotídeos representam a sequência consenso de Kozak de *Drosophila*). Este vetor adicionalmente modificado e referido como pMT- KtΔXho. O sítio de restrição *HindIII* permite a clonagem de sequências promotoras imediatamente anteriores ao códon de início de tradução da sequência sinal pre-pro tPA. O fragmento

15 promotor da metalotioneína de *Anopheles* era derivado da sequência na base de dados do genoma de *Anopheles* (www.ensembl.org/Anopheles_gambiae/). Especificamente, o fragmento promotor de 1833 nucleotídeos representa a sequência a montante do códon de início de tradução do gene da metalotioneína 1 de *Anopheles gambiae* (UniProtKB/TrEMBL nome de entrada Q52P92 ANOGA). Este gene pode ser encontrado

20 no Cromossomo 2R no local 11,911,992-11,912,384 e é o homólogo do gene da MtnA de *Drosophila*. Esta sequência de nucleotídeo foi quimicamente sintetizado (DNA 2.0, Menlo Park, CA) com a adição de um sítio de restrição *KpnI* na extremidade 5' e um sítio de restrição *HindIII* na extremidade 3'. O vetor pMT-KtΔXho foi digerido com *KpnI* e *HindIII* para remover o promotor de *Drosophila* e substituir com o promotor de

25 *Anopheles*. O plasmídeo resultante e referido como um pAgMT- KtΔXho. A sequência H3 HA-Ecto foi então clonada no vetor pAgMT- KtΔXho.

O plasmídeo de expressão pAgMT-KtΔXho que codifica a proteína de subunidade H3 HA-Ecto foi usado para transformar células S2. Com a seleção de linhagens

30 de células estáveis, as células foram triadas para

expressão da forma secretada da proteína H3 HA-Ecto, quando cultivada na presença de sulfato de cobre. A expressão da subunidade H3 HA-Ecto conduzida pelo promotor da metalotioneína de *Anopheles* resultou em um produto uniforme do peso molecular esperado. Embora a expressão da H3 HA-Ecto conduzida pelo promotor de *Anopheles* tenha sido maior do que aquela conduzida pelo promotor de *Bombyx*, o nível de expressão ainda foi considerado baixo, aproximadamente 1 µg/ml. Este nível é muito inferior ao nível de expressão de H3 HA-Ecto, quando o promotor da MtnA de *Drosophila* é utilizado, aproximadamente 30 µg/ml.

10 Exemplo 2

Desenho de Promotor Principal Sintético para Células S2 de *Drosophila*

O promotor principal é definido como a sequência de nucleotídeo mínima que é capaz de direcionar a iniciação transcricional de um dado gene. O promotor principal em eucariotos é responsável por direcionar a iniciação através do complexo RNA polimerase II. O promotor principal é em geral definido como a sequência que abrange o sítio de iniciação de transcrição (INR), mais especificamente, os nucleotídeos 35 a 45 da sequência a montante e a jusante do INR (comprimento total de 70 a 90 nucleotídeos). A sequência de codificação 5'UTR inicia no INR e continua até o códon ATG de iniciação de tradução. A sequência de codificação 5'UTR é variável em comprimento variando de relativamente curta, 50 nucleotídeos, até bastante longa, diversas centenas de nucleotídeos. Como anteriormente descrito, o promotor principal geralmente contém um ou mais dos "motifs" de sequência a seguir, elemento de reconhecimento TFIIB (BRE), TATA box, iniciador (INR), elemento "motif" TATA (MTE), elemento promotor a jusante (DPE) e elemento principal a jusante (DCE). Em levantamentos de promotores principais de diversos organismos incluindo *Drosophila*, fica evidente que nenhum elemento principal universal existe; entretanto, o "motif" INR é o mais comum (FitzGerald *et al.*, *Genome Biol.* (2006) 7:R:53). Em um esforço para fazer um promotor principal forte que seja capaz de direcionar a transcrição que leva a altos níveis de expressão de proteína recombinante, um promotor principal sintético para *Drosophila* foi designado, que incorporou todos esses

elementos. Sequências de consenso para os vários "motifs" reguladores de *Drosophila* que compõem o promotor principal foram utilizadas. Estes "motifs" reguladores de consenso são descritos por Ohler *et al.* (*Genome Biology* (2002) 3:1-12; Kutach and Kadonaga, *Mol. Cell. Biol.* (2000) 20:4754-4764; Smale and Kadonaga, *Ann. Rev. Biochem.* (2003) 72:449-479; Lim *et al.*, *Genes Dev.* (2004) 18:1606-1617; FitzGerald *et al.*, *Genome Biol.* (2006) 7:R53; Gershenzon *et al.*, *BMC Genomics* (2006) 7:161).

Para construir um promotor sintético de *Drosophila*, um promotor principal contendo os "motifs" reguladores de consenso de *Drosophila* a seguir, BRE compreendendo GGGCGCC, TATA compreendendo TATAAA, INR compreendendo TCAGTC, MTE compreendendo CGAACGGAAC e OPE compreendendo GGTTTCG, foi designado. O promotor principal sintético foi fundido a uma porção da 5'UTR da metalotioneína 1 de *Anopheles*. A fusão da 5'UTR da metalotioneína 1 de *Anopheles* ao promotor principal resulta na posição do INR em relação ao ATG de iniciação de tradução sendo análoga aquela do promotor da MtnA de *Drosophila*. Este promotor sintético foi quimicamente sintetizado e é referido como promotor principal sintético 1 (SCP1). A sequência do promotor SCP1 está listada como SEQ ID NO: 1. O SCP1 quimicamente sintetizado incluiu um sítio *Xba*I em sua extremidade 5' e um sítio *Hind*III em sua extremidade 3' para clonagem em vetores de expressão existentes.

O vetor de expressão pMT-KtΔXho descrito no Exemplo 1 foi adicionalmente modificado através da inserção de um sítio de restrição *Xba*I/sete nucleotídeos a montante da sequência do TATA box do promotor da MtnA. Este plasmídeo é referido como pMT-X-KtΔXho. Isto permitiu a remoção do promotor principal da MtnA (inclui a 5'UTR) através da digestão com enzimas de restrição *Xba*I and *Hind*III e inserção de promotor principal alternativo e sequências 5'UTR. Desta maneira, o SCP1 foi inserido no vetor de expressão pMT-X-KtΔXho.

Exemplo 3

Desenho de Promotor Proximal Contendo Elementos Ativadores Transcricionais Capazes de Conduzir a Expressão de Proteínas Heterólogas em Células S2 de *Drosophila*

5 Em um esforço para regular para cima o nível de transcrição em promotores principais, o que, por sua vez, resulta em altos níveis de expressão em células S2, promotores proximais sintéticos foram designados para conter vários elementos ativadores transcricionais. A escolha para desenvolver promotores proximais sintéticos em vez de utilizar promotores

10 proximais de ocorrência natural para uso em células S2 é baseada nos resultados no Exemplo 1 com os dois promotores de inseto e também no trabalho de Olsen *et al.* {*Cytotechnology* (1992) 10:157-167), em que foi demonstrado que promotores de origem de mamífero e vírus não funcionam bem em células S2. Portanto, os promotores proximais sintéticos foram

15 designados para conter "motifs" reguladores derivados de promotores de *Drosophila*. Os desenhos de promotor proximal envolveram a montagem de repetições de sequências de consenso de "motif" regulador individuais. Os promotores proximais sintéticos foram avaliados quanto à sua capacidade de direcionar altos níveis de transcrição, quando ligados a promotores centrais.

20 O primeiro tipo de elemento regulador transcricional avaliado foi o elemento relacionado à replicação de DNA de *Drosophila* (DRE). Os promotores de genes de *Drosophila* que codificam proteínas relacionadas à replicação de DNA contêm a sequência de DRE de consenso 5'-TATCGATA. Um fator de ligação a DRE específico, fator de elemento relacionado à replicação de

25 DNA ou DREF, se une ao DRE e regula positivamente genes sob o DRE de controle e resulta em altos níveis de expressão. Estudos computacionais de promotores de *Drosophila* (Ohler *et al.*, *Genome Biology* (2002) 3:1-12) demonstraram que o elemento DRE é encontrado em muitos promotores de *Drosophila*. Ele é, de fato, um dos elementos transcricionais mais comuns.

30 No trabalho de Hirose *et al.* (*Journal of Biological Chemistry* (1993) 268:2092-2099), múltiplas cópias de DRE foram montadas e ligadas a um promotor principal da MtnA de *Drosophila* que controla um gene repórter

luciferase. Os DREs regulam para cima a expressão de luciferase com o nível de expressão aumentando com o número de DREs adicionados. No desenho de um ORE sintético contendo promotor proximal, a sequência CTGCCTGC **TATCGATAGATT**CAGG (SEQ ID NO: 16, consenso de DRE está em negrito e itálico) foi usada. Os desenhos de promotor proximal de ORE sintético foram como a seguir: 1XDRE, 2XDRE e 4XDRE.

Os promotores proximais contendo DRE foram clonados no vetor pMT-X-KtΔXho descrito no Exemplo 2, através da remoção do promotor proximal da MtnA (KpnI -XbaI) e inserção de promotores proximais alternativos imediatamente a montante do promotor principal da MtnA. As repetições de DRE foram todas na mesma orientação para frente (5' para 3'). Estes vetores são referidos como pDRE-X-KtΔXho. Para avaliar a expressão de proteínas recombinantes com estes plasmídeos de expressão de DRE, a sequência do gene H2 HA-Ecto descrita no Exemplo 1 foi inserida nestes vetores e usada para estimar o nível de expressão.

Os plasmídeos de expressão pDRE-X-KtΔXho que codificam a proteína de subunidade H3 HA-Ecto foram usados para transformar células S2. Após seleção de linhagens de células estáveis, elas foram triadas para expressão da forma secretada da proteína H3 HA-Ecto. Todas as construções de promotor proximal de DRE não resultaram em qualquer expressão de H3 HA-Ecto mensurável.

O segundo tipo de elemento regulador transcricional avaliado foi o elemento GAGA de *Drosophila*. O fator regulador de transcrição de GAGA de *Drosophila* é denominado por sua capacidade de unir sequências alternadas (GA)ⁿ ou (CT)ⁿ (elementos GAGA). Imagina-se que este fator de transcrição aumenta a expressão de genes remodelando a estrutura da cromatina ao redor de genes afetados (Granok *et al.*, *Current Biology* (1995) 5:238-241). Elementos GAGA são encontrados nos promotores de uma variedade de genes de *Drosophila*. Ele é encontrado em promotores distais, proximais e até promotores principais destes genes (Soeller *et al.*, *Mol. Cell. Biology* (1993) 13:7961-7970). No trabalho de Soeller *et al.* (1993), foi demonstrado que múltiplos elementos GAGA podem conduzir a transcrição

de um promotor principal.

5 Dais promotores proximais foram designados, em que um (mica elemento GAGA foi combinado com elementos ORE. Estes dois desenhos foram como a seguir: GAGA-2XDRE e 2XDRE-GAGA-2XDRE. Como com os promotores proximais de DRE, estes promotores proximais foram inseridos a montante do promotor principal da MtnA no vetor de expressão pMT-X-KtΔXho. Para avaliar a expressão de proteínas recombinantes com
10 estes plasmídeos de expressão DRE-GAGA, a sequência de gene da He HA-Ecto descrita no Exemplo 1 foi usada como um repórter.

 Os plasmídeos de expressão com os promotores de combinação DRE-GAGA que codificam a proteína de subunidade H3 HA-Ecto foram usados para transformar células S2. Após a seleção das linhagens de células estáveis, as células foram triadas para
15 expressão da forma secretada da proteína H3 HA-Ecto. Essas construções também não resultam em qualquer expressão de proteína mensurável.

 O terceiro tipo de elemento regulador transcricional avaliado foi o elemento responsivo a metal de *Drosophila* (MRE). Um fator de união a MRE específico, fator de transcrição responsivo a metal ou MTF-1, se liga ao MRE e regula positivamente os genes
20 sob o MRE de controle e resulta em altos níveis de expressão. Os promotores de genes de *Drosophila* regulados por MTF-1 contém a sequência MRE de consenso 5'-TGCACAC. Foi demonstrado que quatro cópias de MREs podem conduzir a transcrição de um promotor mínimo (Zhang et. al., *Mol. Cell. Biol.* (2001) 21:4505-4514).

 Análise de exclusão do promotor proximal da MtnA no vetor de expressão
25 pMTtPA6Xho indicou que a exclusão de um MRE com a sequência 5' - TCTTT**TGCACAC**GCCGGC (SEQ ID NO: 17, sequência de consenso está em negrito e itálico) tinha o maior impacto sobre a força do promotor da MtnA. Portanto, promotores proximais sintéticos foram designados contendo uma ou mais cópias da sequência MRE TCTTT**TGCACAC**GCCGGC (SEQ ID NO: 17). Os desenhos do promotor principal
30 de MRE sintético foram como a seguir: 2X, 3X, 4X, 5X, 6X, 8X MREs. Os MREs nestes desenhos de promotor proximal foram todos arranjados de uma maneira cabeça à cauda 5' a 3' (para a

frente). Como com os promotores proximais de DRE, os promotores proximais de MRE foram inseridos a montante do promotor principal da MtnA no vetor de expressão pMT-X-KtΔXho. Para avaliar a expressão de proteínas recombinantes com estes vetores de expressão do promotor de MRE sintético, a sequência do gene da H3 HA-Ecto descrita no Exemplo 1 foi inserida nos vetores.

Os plasmídeos de expressão com os promotores proximais de MRE que codificam a proteína de subunidade H3 HA-Ecto foram usados para transformar células S2. Após a seleção de linhagens de células estáveis, as células form triadas para expressão da forma secretada da proteína H3 HA-Ecto. Visto que estas construções contêm MREs, a avaliação da expressão precisou que as culturas fossem induzidas com CuSO₄ a 0,2 mM e cultivadas a 26°C por 7 dias.

A análise destes transformantes de S2 revelou como o número de unidades de MRE aumentou, assim como o nível de expressão de gene até um máximo com 6XMRE. O nível de expressão diminuiu com 8X. Portanto, a análise adicional de expressão utilizando um promotor proximal de MRE sintético foi com base no desenho 6XMRE.

Usando o promotor proximal sintético 6XMRE, uma série de construções de expressão foram feitas, o que permitiu uma comparação do promotor proximal da MtnA nativo e o promotor proximal 6XMRE sintético. Nesta série de construções de expressão, os dois promotores proximais foram unidos a três promotores principais ou mínimos diferentes: 1) MtnA de *Drosophila*, 2) MtnA de *Anopheles* e 3) SCP1 descrito no Exemplo 2. As seis construções resultantes foram então unidas com o gene repórter da H3 HA-Ecto. Ambos promotores proximais 6XMRE e DmMtnA foram capazes de conduzir o nível de expressão do gene flu a níveis equivalentes, quando comparando cada um dos dados promotores principais individuais. Por exemplo, o proximal de 6XMRE-principal da MtnA e proximal de DmMtnA-principal da MtnA resultou em níveis equivalentes de expressão da proteína H3 HA-Ecto. Estes dados com o promotor proximal 6XMRE indicam que um promotor proximal sintético funcional pode ser designado, apesar das falhas

com os dois primeiros ativadores transcricionais testados neste exemplo.

A comparação do promotor proximal 6XMRE e DmMtnA também revelou que os níveis de expressão da H3 HA-Ecto em construções contendo o promotor principal da MtnA foram aproximadamente quatro vezes superiores às aquelas construções contendo o SCP1. Portanto, esforços foram feitos para avaliar promotores principais alternativos, como descrito no Exemplo 4, e também para otimizar adicionalmente o promotor principal sintético, como descrito no Exemplo 5.

Exemplo 4

10 Avaliação de um "Super Promotor Principal" para Conduzir a Expressão de Proteínas Heterólogas em Células S2 de *Drosophila*

Juven-Gershon *et al.* (*Nature Methods* (2006) 3:917-922) descrevem o desenho e análise de um promotor principal referido como um "super promotor principal", que direciona altos níveis de transcrição de RNA por polimerase de RNA II em células hospedeiras de metazoários. O uso dos mesmos "super promotores principais" também é descrito no pedido PCT N° PCT/US2006/020394 (Kadonaga and Gershon, data de prioridade de 25 de maio de 2005). Os "super promotores principais" descritos nestas publicações serão referidos como "promotores principais de Kadonaga" ou "KCP".

20 Os promotores principais de Kadonaga consistem em um promotor mínimo que contém elementos INR e MTE, e opcionalmente contendo elementos TATA e / ou DPE, que são derivados de diversos genes diferentes virais e de *Drosophila*. Existem similaridades no desenho para os promotores principais KCP e os promotores principais descritos desta Descrição, visto

25 que ambos dependem muito do artigo *Genes and Development*, primeiro documento que descreve o "motif" regulador MTE de *Drosophila* (Lim *et al.*, *Genes Dev.* (2004) 18:1606-1617). Os promotores principais de Kadonaga são descritos como sendo úteis para expressão de proteína heteróloga em células hospedeiras de metazoários, que incluem células S2 de

30 *Drosophila*. Portanto, Hawaii Biotech, Inc. criou três construções contendo o promotores principais de Kadonaga para testar a funcionalidade dos promotores principais de Kadonaga no contexto da expressão de proteínas

heterólogas em células S2 estavelmente transformadas.

O promotor principal de Kadonaga foi sintetizado com sítios de restrição de flaqueamento apropriados. A sequência sintética de KCP foi inserida no vetor de expressão HBI-10 descrito no Exemplo 9 e definido por SEQ ID NO: 10, do qual as sequências promotor-principais e proximais foram removidas. O vetor de expressão híbrido KCP/HBI-10 resultante contém KCP unido ao sinal de secreção 1R5L (SEQ ID NO: 7), o gene repórter da H3 HA-Ecto e a Op-CLP-3'UTR (SEQ ID NO: 9). O vetor KCP/HBI-10 não contém um promotor proximal. Portanto, duas construções adicionais foram criadas a partir deste "vetor híbrido KCP/HBI-10 de base" que, em adição ao KCP, continha, em uma primeira construção adicional, o promotor proximal 6xMRE ("6xMRE/KCP/HBI-10"), e em uma segunda construção, o melhorador do Citomegalovírus (CMV) ("CMV/KCP/HBI-10"); cada promotor proximal foi inserido imediatamente a montante da sequência de KCP. O melhorador do CMV é o mesmo usado por Juven-Gershon *et al.* (*Nature Methods* (2006) 3:917-922, para o qual Dr. Kadonaga é o autor correspondente). As três construções foram transformadas em células S2 de *Drosophila* em triplicata e o nível de expressão da H3 HA-Ecto foi determinado.

As duas construções CMV/KCP/HBI-10 e KCP/HBI-10 sozinhas (nenhum promotor proximal ou melhorador) mostraram um nível mínima de expressão da H3 HA-Ecto, isto é, <1µg/ml. Desse modo, estes dados indicam que 1) apenas KCP proporciona baixo nível de expressão de uma proteína secretada e 2) este nível de expressão não é melhorado ou aumentado pelo melhorador do CMV em células S2 de *Drosophila* transformadas. A construção 6xMRE/KCP/HBI-10 resultou na expressão e secreção da proteína H3 HA-Ecto em 2-3 µg/ml no sobrenadante bruto. Experimentos anteriores usando a construção análoga de 6xMRE-SCPI resultaram na expressão da mesma proteína H3-HA a 10 µg/ml ou aproximadamente 2-3 vezes superior ao KCP unido ao mesmo promotor proximal 6xMRE. Visto que a única diferença entre as duas construções e o promotor principal, pode ser concluído que SCP1 descrito nos Exemplos 2 e 3

proporciona níveis de expressão de proteína heteróloga secretada que são aproximadamente 2 – 3 vezes superiores àquela expressão que resulta do uso do promotor principal de Kadonaga e o promotor proximal 6xMRE em células S2 de *Drosophila* transformadas. O uso de um MRE ou promotor proximal com base no MRE não foi descrito nas referências de Lim *et al.* (2006) 3:917-922).

O promotor principal de Kadonaga foi otimizado para uso em células de mamífero. Os dados apresentados no uso de KCP em células S2 HeLa foram limitados a transfecções transitórias e ao uso de ensaios de fluorescência sensitiva; portanto, com base nos resultados obtidos com os promotores principais KCP após transformação estável de células de *Drosophila*, o uso de ensaios repórteres de transfecção e sensibilidade transiente não é indicativo da capacidade para expressar e secretar proteínas heterólogas em altos níveis em todos os tipos de célula. O desenho do KCP inclui elementos de promotores virais, tal como CMV e AdML, que têm sido classicamente usados para expressão de proteínas em células de mamíferos. Não obstante, um aumento de 2 – 3 vezes na expressão estável em células de *Drosophila* de proteína heteróloga secretada usando SCP1 versus KCP é notável e inesperado.

Exemplo 5

Otimização do Promotor Principal Sintético para Expressão de Proteínas Heterólogas em Células S2 de *Drosophila*

O desenho de SCP1 original descrito no Exemplo 2 contém um promotor principal com um BRE, TATA box, INR, MTE e DPE. Cada um destes "motifs" foi individualmente removido e o efeito de sua remoção sobre a expressão foi avaliado. Em cada caso, o nível de expressão pareceu ser um pouco reduzido. Tentativas foram feitas para otimizar o SCP trocando os "motifs" BRE, TATA, INR, MTE e DPE individualmente com as sequências alternativas ou contendo uma sequência alternativa que representa um dado "motif" regulador ou com a sequência que não tem um "motif" regulador na posição equivalente no promotor principal. Por exemplo, o promotor principal

da MtnA de *Drosophila* do tipo selvagem não tem os "motifs" BRE, MTE ou DPE, ainda, aom base nos dados apresentados no Exemplo 3, o promotor principal da MtnA de *Drosophila* é o mais forte promotor principal do tipo selvagem testado até agora. A remoção da sequência de BRE em SCP1 e substituição com sequências "neutras" não resultou em mudança no nível de expressão de um prodito de gene repórter indicando que a presença do BRE foi neutra, isto é, não teve efeito aparente na transcrição. No caso do TATA box, o espaçamento entre o SCP1 TATA e o INR foi reduzido em 3 nucleotídeos para igualar o espaçamento entre o TATA box da MtnA de *Drosophila* e seu INR correspondente. Os três nucleotídeos removidos (ACA) foram aqueles logo imediatamente 3' do TABA box. Esta mudança resultou em uma ligeira diminuição do nível de expressãodo gene repórter, indicando que o espaçamento no SCP1 foi preferido em relação àquele em MtnA. O INR de SCP1 (TCAGTC) foi trocado para GCATCA, uma mudança de três dos seis nucleotídeos. Esta mudança aumentou a expressão do produto do gene repórter em aproximadamente 50%.

O SCP1 contém uma sequência de codificação 5'UTR que é de 190 nucleotídeos de comprimento. Esta porção do SCP1 é derivada do gene Mtn de *Anopheles*. Não foi sabido se esta sequência foi benéfica ou prejudicial à transcrição / expressão. Para determinar a contribuição desta sequência, duas construções de exclusão foram geradas. A primeira removeu 60 nucleotídeos da sequência a partir da extremidade 3'. Esta mudança resultou em um leve aumento no nível de expressão. A segunda remover 127 nucleotídeos da sequência a partir da extremidade 3'. Esta mudança resultou em um aumento no nível de expressão de aproximadamente 2 vezes. Esta versão truncada de SCP1 é referida como SCP1 Δ 111 e foi usada para adicionalmetne avaliar as modificações ao promotor principal.

Os efeitos do Tollo MTE sobre os níveis de transcrição foram descritos por Lim *et al.* (*Genes Dev.* (2004) 18:1606-1617). Entretanto, o MTE de consenso usando em SCP1 Δ 111 foi trocado para aquele encontrado no promotor Tollo de *Drosophila*. A subatituição do MTE de consenso com a sequência de MTE Tollo resultou em um aumento no nível de

expressão de aproximadamente 2 a 3 vezes. A inserção do TolloMTE no principal SCP1
 5 evidentemente teve um impacto muito positivo nos níveis de expressão. O promotor principal com TolloMTE inserido é referido como "SCP1-TolloMTE Δ 111" e a sequência do promotor contendo este promotor principal é listada como SEQ ID NO: 2.

10 A otimização adicional do promotor principal foi testada. Um conjunto de três novos promotores principais foi designado, o qual incorporou os aspectos positivos descritos nos três parágrafos anteriores que resultaram em expressão aumentada. A combinação específica dos três "motifs" chave e o elemento 5'UTR são: 1) o "motif" TATA, compreendendo TATAAA, e o espaçamento definido pelo SCP1 original em relação
 15 ao INR, 2) o "motif" da sequência INR GCATCA, 3) o "motif" da sequência MTE derivada de *Drosophila* Tallo e 4) o elemento 5'UTR truncado (Δ 111). Os três novos promotores principais foram designados SCP5, SCP6 e SCP7.

Todos os três destes promotores principais do "motif" da sequência BRE
 20 (compreendendo GGGCGCC quando presente) do SCP1. Os três novos promotores principais diferiram quanto à sequência ocupada pelo "motif" DPE. SCP5 continha o DPE do gene Tallo de *Drosophila*, SCP6 continha o DPE de consenso em SCP1 (GGTTCG) e SCP7 não tinha um "motif" DPE e, ao contrário, continha uma sequência de "carga neutra" para proporcionar o mesmo espaçamento em relação às sequências
 25 SCP5 e SCP6. Quando SCP1, SCP5, SCP6 e SCP7 foram comparados em construções equivalentes de expressão de S2 de *Drosophila*, que diferiram apenas quanto às suas sequências SCP, o desenho de SCP7 resultou no nível mais alto de expressão. A sequência de SCP7 está listada como SEQ ID NO: 3.

30 Como anteriormente discutido, a sequência da região não traduzida 5' (5'UTR) do RNA mensageiro (mRNA) desempenha um papel importante na regulação pós-transcricional da expressão de gene do mRNA eucariótico. A 5'UTR é definida pela região do INR ao codon de iniciação para metionina; por isso, no contexto da definição da sequência de DNA do promotor, ele é geralmente inclusive da sequência 5'UTR. Este é o caso para as sequências listadas para os promotores SCP1-Tollo Δ 111 e SCP7. A

5'UTR que resulta do promotor SCP1-TolloMTE Δ 111 e referida como a Δ 111 5'UTR e listada como SEQ ID NO: 4; a 5'UTR que resulta do promotor SCP7 e referida como a Δ 111-7 5'UTR e listada como SEQ ID NO: 5.

Exemplo 6

Otimização do Ativador de Transcrição MRE Sintético para Expressão de Proteínas Heterólogas em Células S2 de *Drosophila*

Na natureza, diferentes combinações de ativadores transcripcionais são encontradas nos promotores proximais. Nessa base, combinações dos elementos GAGA e MRE descritas no Exemplo 3 foram criadas para determinar se esta combinação de "motifs" resulta em níveis superiores de transcrição e, finalmente, em níveis superiores de expressão. Dois desenhos de promotor proximal GAGA-MRE foram testados. Eles são como a seguir: GAGA-2XMRE e 2XMRE-GAGA-2XMRE.

Os dois promotores proximais GAGA-MRE foram clonados na área imediatamente anterior a vários promotores principais descritos no Exemplo 3, utilizando o sítio de restrição *Xba*I no vetor pMT-X-Kt Δ Xho descrito no Exemplo 3. Após a transformação e seleção da linhagem de célula estável com plasmídeos contendo os promotores proximais GAGA-DRE, a expressão do produto H3 HA-Ecto foi avaliada. A adição do elemento GAGA aos MREs utilizados não aumentou ou diminuiu os níveis de expressão em comparação às construções similares de MRE sem o elemento GAGA.

Em muitos promotores proximais, múltiplos sítios de união ao fator de transcrição do mesmo tipo podem frequentemente ser encontrados em ambas orientações para frente e reversa. Isto é verdade para os MREs endógenos encontrados nos promotores principais da família de gene Mtn de *Drosophila*. Portanto, em um esforço para determinar se a orientação dos MREs pode melhorar a função do promotor proximal, uma série de promotor proximal de MRE sintético foi criada usando várias combinações de sequência MRE para frente e reversas. A MRE reversa (R) é simplesmente o complemento reverso da sequência MRE para frente (F) descrita no Exemplo 3. A MRE reversa tem a sequência GCCGGC***GTGTGCAAA***AGA (SEQ ID NO: 18, a sequência MRE de consenso esta em negrito e italico). Esta série de novos

promotores proximais consiste em 3X repetições das MREs para frente e/ou reversas em
5 varias combinações. As combinações foram como a seguir: 3XF MRE, 3XR MRE, 3XF MRE-
3XR MRE, 3XR MRE-3XF MRE, 3XF MRE-3XF MRE e 3XR MRE-3XR MRE. O espaçamento
entre as unidades de repetição 3X MRE foi de 11 nucleotídeos. A sequência do
espaçador é ATCAAACCTAGA (SEQ ID NO: 19).

10 Como no Exemplo 3, os promotores proximais MRE contendo sequência para
frente e reversa foram inseridos a montante do promotor principal MtnA no vetor de
expressão pMT-X-KtΔXho. Plasmídeos de expressão similares contendo o principal SCP1
também foram construídos. Para avaliar a expressão de proteínas recombinantes com
estes vetores de expressão de promotor MRE sintético, a sequência do gene da H3 HA-
15 Ecto descrita no Exemplo 1 foi inserida nos vetores.

Os plasmídeos de expressão com os promotores proximais MRE¹⁵ que
codificam a proteína de subunidade H3 HA-Ecto foram usados para transformar células
S2. Após a transformação e seleção da linhagem de célula estável com plasmídeos
20 contendo as combinações de MRE para frente e reversa, a expressão do produto H3
HA-Ecto foi avaliada. Visto que estas construções contêm MREs, a avaliação da expressão
precisou que as culturas fossem induzidas com CuSO₄ a 0,2 mM, e cultivadas a 26°C
por 7 dias.

25 A análise destes transformantes de S2 revelou que quando as combinações
reversas e para frente diferentes foram usadas, elas resultaram em diversos níveis de
expressão, todos os quais foram maiores do que o 6XMRE do Exemplo 3. Mesmo a 3XF
MRE-3XF MRE resultou em expressão levemente maior do que o 6XMRE (cerca de 2
vezes maior). Isto pode ser um resultado do espaçador de 11 nucleotídeos que é a única
30 diferença entre as sequências do promotor proximal 6XMRE e 3XF MRE-3XF MRE. A
expressão direcionada pelo 3XF MRE-3XR MRE foi aproximadamente igual aquela do
3XF MRE-3XF MRE. A expressão direcionada pelo 3XR MRE-3XF MRE foi
proximadamente 3 vezes maior do que aquela do 6XMRE e a expressão direcionada pelo
3XR MRE-3XR MRE foi aproximadamente 4 a 5

vezes maior do que aquela do 6XMRE. Portanto, o 3XR MRE-3XR MRE foi selecionado como o promotor proximal para uso em construções adicionais, em que os elementos reguladores selecionados são combinados para criar um cassete de expressão completo composto de apenas elementos otimizados e sintéticos, como descrito no Exemplo 9. A sequência do promotor proximal sintético 3XR MRE-3XR MRE está listada em SEQ ID NO: 6.

Exemplo 7

Desenho e Otimização de Sinais de Secreção Sintéticos para Expressão de Proteínas Heterólogas em Células S2 de *Drosophila*

O peptídeo sinal de secreção desempenha um importante papel na expressão de proteínas que são marcadas para secreção da célula. Portanto, o uso de sequências ótimas em um vetor de expressão que direciona proteínas recombinantes para secreção no meio de cultura durante a produção é importante para a utilidade de tal vetor de expressão. O exemplo de otimização do sinal de secreção apresentado por Zhang et al. (*J. Gene Med.* (2005) 7:354-365) demonstra evidentemente os benefícios da otimização do sinal de secreção; entretanto, este exemplo específico se aplica às plantas e não está evidente que as mudanças para o sinal de secreção descritas se aplicam a outros tipos de células eucarióticas. Portanto, uma série de três sinais de secreção foi designada para estabelecer quais tipos de mudanças para a sequência resultariam em expressão aumentada de produtos de proteína no meio de cultura de células S2 estávelmente transformadas. Os desenhos dos sinais de secreção sintéticos seguiram a tabela matriz primeiramente descritos por von Heijne (*Nuc. Acids Res.* (1986) 14:4683-4690) e mais tarde aperfeiçoados por Bendtsen et al. (*J. Mol. Biol.* (2004) 340:783-795). O primeiro conjunto de três sinais de secreção testados, todos continham a mesma sequência na região de clivagem, comprimento de sequência variado na região hidrofóbica e um número variado de resíduos carregados na região básica. A sequência de aminoácido dos três peptídeos de sinal testados é conforme a seguir, com nomes entre parêntesis:

M R T I I A L L L L T V S G A Q G ("1R4L", SEQ ID NO: 20)

M R T I I A L L L L **L** T V S G A Q G ("1R5L", SEQ ID NO: 21)
 M R **R** T I I A L L L L L T V S G A Q G ("2R5L", SEQ ID NO: 22)

A sequência 1R5L é um aminoácido mais comprido (um "L" adicional mostrado em fonte em negrito e itálico) do que aquela da sequência 1R4L, visto que o comprimento do núcleo hidrofóbico aumentou, e a sequência 2R5L é um resíduo mais comprido (um "R" adicional mostrado em fonte em negrito e itálico) do que a sequência 1R5L, visto que um carregado adicional foi adicionado à região básica, embora mantendo o mesmo núcleo hidrofóbico. Os comprimentos das sequências 1R4L, 1R5L e 2R5L são 17, 18 e 19 aminoácidos, respectivamente. Os três sinais de secreção foram operativamente ligados ao gene do domínio da H3 HA- Ecto no vetor de expressão pMtaf. A expressão e secreção do produto H3 HA-Ecto utilizando estes sinais de secreção foram comparadas à uma construção do domínio da H3 HA-Ecto similar, onde o sinal de secreção tPA foi utilizado. Após transformação das células S2, seleção de linhagens de células estáveis e triagem para níveis de expressão do produto no meio de cultura como descrito no Exemplo 1, foi determinado que a 1R5L resultou no nível mais alto de expressão, aproximadamente 2 vezes mais do que a construção contendo tPA, 3 vezes mais do que a construção contendo 2R5L e 2 vezes mais do que a construção contendo 1R4L (equivalente ao tPA).

Além disso, uma mutação fortuita na sequência 1R5L, Q para H (posição -2), resultou em um aumento mais inesperado na expressão, aproximadamente 2 vezes, em relação à sequência 1R5L original. A sequência 1R5L contendo o resíduo H é referida como 1R5L(H) (SEQ ID NO: 8) e foi usada nos vetores de expressão finais descritos no Exemplo 9.

No desenho destas sequências sinal de secreção, a sequência de seis nucleotídeos AACAAAC foi colocada imediatamente anterior ao códon iniciador para metionina. Isto representa a sequência consenso de Kozak para *Drosophila* descrita por Cavener (*Nuc. Acids Res.* (1987) 15:1353-1361).

Os desenhos das sequências sinal de secreção também avaliaram o uso de um resíduo de aminoácido T ou S imediatamente após o G no sítio de clivagem. Quando o códon apropriado e selecionado para

estes resíduos, a sequência nos sítios de clivagem G/T e G/S codifica os sítios de clivagem de endonuclease de restrição *KpnI* e *BamHI*, respectivamente. Estas ambas sequências do sítio de clivagem funcionam bem. Ambas proporcionam a inserção conveniente da sequência que codifica a proteína desejada a ser fundida na mesma estrutura de leitura com o sinal de secreção.

A sequência do sinal de secreção otimizado 1R5L(H) e M R T I I A L L L L L T V S G A H G (SEQ ID NO: 8). O aminoácido preferido após o sítio de clivagem é S. A sequência de nucleotídeo que codifica o sinal de secreção otimizado é a sequência de aminoácido correspondente estão listadas como SEQ ID NO: 7 e SEQ ID NO: 8, respectivamente.

Exemplo 8

Avaliação de um Elemento 3'UTR Alternativo para Expressão de Proteínas Heterólogas em Células S2 de *Drosophila*

A 3'UTR do gene e o mRNA desempenha importantes papéis nos níveis transcricionais e translacionais de expressão de gene, respectivamente (revisado por Kloc et al., *Cell* (2002) 108:533-544). Portanto, o uso de sequências 3'UTR ótimas em um vetor de expressão que ajuda a melhorar a poliadenilação, melhora a translocação de mRNA do núcleo para o citoplasma, melhora a estabilidade do mRNA e não interfere com a localização do mRNA para o retículo endoplasmático rugoso e é importante para a utilidade de um vetor de expressão que direciona produtos para liberação no meio de cultura. É por estas razões que foi recomendado que, no desenho de plasmídeos de expressão, a 3'UTR seja considerada para otimização (van der Velden et al., *Biotechniques* (2001) 31 :572-582).

O plasmídeo de expressão pMTtPA de *Drosophila* utiliza a sequência 3'UTR anterior de SV40 viral para terminação e poliadenilação. Embora seja funcional, parece haver espaço para aperfeiçoar a utilização desta sequência SV40. Por exemplo, a otimização da 3'UTR para expressão heteróloga em sistemas de inseto foi relatada por Van Oers et al. (*J. Gen. Virol.* (1999) 2253-2262). Foi substituída a 3'UTR anterior de SV40 com uma 3'UTR de baculovírus e o nível de expressão foi aumentado. Embora eles

não tenham dissecado a nova 3'UTR para determinar quais partes contribuíram para o aumento na expressão, este trabalho demonstrou que mudanças na 3'UTR podem melhorar a expressão de proteínas heterólogas. Infelizmente, cada sistema de expressão é único e é necessário para determinar o que constitui uma 3'UTR ótima para expressão de produtos secretados em células S2.

A característica principal da maioria dos elementos 3'UTR eucarióticos é a sequência sinal de poliadenilação (sinal poliA), que é AATAAA. O sinal poliA está envolvido na definição do sítio de clivagem terminal 3' do mRNA recém transcrito no núcleo e no qual ocorre a poliadenilação. Em algumas 3'UTR's, sequências conservadas são observadas imediatamente a jusante do sítio de clivagem. Estas sequências são referidas como elementos de sequência a jusante (DSE) e acredita-se que auxiliam na definição do sítio de clivagem da 3'UTR.

A sequência 3'UTR anterior de SV40 que está presente no vetor pMTtPA, e derivados, possui comprimento de 138 bases e contém dois sinais poli-A putativos. O sítio de clivagem para a adição da cauda poli-A ocorre no nucleotídeo 96. Na sequência entre o nucleotídeo 96 e 13, existem sequências que são consideradas como representantes de "motifs" DSE que auxiliariam na clivagem.

Em um esforço para identificar uma sequência 3'UTR que resultaria em maiores níveis de expressão, as 3'UTRs em muitos genes de drosophila foram avaliadas. Os critérios primários foram que a 3'UTR seja de aproximadamente 100 nucleotídeos de comprimento e seja derivada de um gene que codifica um produto secretado. A proteína semelhante à Quitinase de Drosophila (CLP) descrita por Kirkpatrick *et al* (*Gene* (1995) 143:147-154) é uma das mais abundantes proteínas secretadas das células S2 e atende a estes critérios. Esta 3'UTR possui um sinal poli-A que se conforma à sequência AATAAA. Kirkpatrick *et al.* sugerem que o sinal poli-A é ou uma AATATA ou uma CATAAA.

Em Kirkpatrick *et al.* (1995), o sítio de clivagem da CLP 3'UTR é definido pela posição em que a sequência poli-A é adicionada ao mRNA,

núcleotídeo 106. Com base no clone CLP cDNA LD21619 da Coleção de Genes de *Drosophila* (Berkeley *Drosophila* Genome Project, www.fruitfly.org/DGC) o sítio de clivagem é de 7 a 11 nucleotídeos maior a jusante. O alinhamento sequencial da sequência de Kirkpatrick *et al.*, a sequência LD21619 e a sequência correspondente da sequência do genoma da *Drosophila* identificou um erro de nucleotídeo sinal na sequência de Kirkpatrick *et al.* no nucleotídeo 43, alteração de A para G. Esta alteração resulta em uma sequência de AAGAAAA ("sinal de poli-A variante") ao invés de AAAAAAA. Existe outra sequência AAGAAA na CLP 3'UTR. com base na posição destas sequências AAGAAA, os inventores postularam que uma destas poderia ser capaz de servir como o sinal poli-A, em que a segunda seria a mais provável. Para o teste inicial da CLP 3'UTR em um vetor de expressão, um fragmento de 117 nucleotídeos de comprimento foi selecionado. O fragmento contém ambos os sinais poli-A variantes de AAGAAA. Além disso, com base no alinhamento das diferentes fontes de sequências CLP, este fragmento 117 poderia permitir a clivagem em cada um dos sítios de clivagem sugeridos, isto é, próximo de cada um dos dois sinais de poli-A variantes.

A sequência CLP 3'UTR de 117 nucleotídeos foi adicionada ao vetor de expressão pMTtPAΔXho, do qual a 3'UTR anterior de SV40 foi removida. A inserção da CLP 3'UTR de 117 nucleotídeos resultou em um aumento de 1,5 vezes na expressão da proteína H3 HA-Ecto repórter. Em um esforço para estabelecer se as sequências AAGAAA servem como o sinal poli-A, os resíduos G nos dois sinais putativos foram mutagenizados em resíduos T para combinar com a sequência poli-A de consenso. Esta alteração resultou em um aumento de três vezes na expressão, quando comparada com o vetor de expressão equivalente com a 3'UTR anterior de SV40. Em outros experimentos, ambos os sítios de poliadenilação putativos na CLP 3'UTR sofreram mutação para romper as sequências (primeira AAGAAAA muda para ACGCAA e a segunda AAGAAA mudou para AATGAA). Estas alterações resultaram em uma redução dramática no nível de expressão. Isto ainda sustenta que pelo menos um dos sinais de polia-

denilização putativos serve como o sinal poli-A.

Para determinar se a sequência a jusante adicional era necessária para funcionar como um DSE, a 3'UTR foi estendida por outros 20 nucleotídeos com base na sequência genômica no local correspondente. A 3'UTR estendida diminuiu o nível de expressão de 2 a 3 vezes, o que era inesperado.

Com base nestes resultados, a CLP 3'UTR de 117 nucleotídeos com as duas mutações de G para T foi selecionada para uso na construção de um vetor de expressão que combina múltiplos elementos reguladores descritos no exemplo 9. A 3'UTR otimizada resulta em níveis mais altos de expressão e é referida como Op-CLP-3'UTR. A sequência da Op-CLP-3'UTR é listada como SEQ. ID. NO:9.

Exemplo 9

Avaliação de Vetores de Expressão Contendo Combinações de Elementos Reguladores para Expressão Melhorada de Proteínas Heterólogas em Células S2 de *Drosophila*

A fim de criar um vetor de expressão otimizado para expressão de produtos de proteína recombinante em células S2, cada um dos cinco elementos reguladores a seguir é necessário: promotor proximal, promotor principal, 5'UTR, sinal de secreção e 3'UTR. De modo ideal, cada um destes cinco elementos é otimizado e a combinação dos elementos irá resultar em um cassete de expressão que resulta em níveis mais altos de expressão de produto. Nos exemplos anteriores, os seguintes elementos otimizados foram identificados: o promotor proximal sintético 3XR MRE-3XRMRE, os promotores principais sintéticos SCP1-TolloMTE Δ 111 e SCP7, a 5'UTR truncada com a sequência consenso de Kozak, o sinal de secreção sintético 1R5L(H) e a 3'UTR Op-CLP-3'UTR otimizada.

Como uma primeira etapa para avaliar combinações destes elementos reguladores otimizados, o sinal de secreção 1R5L e a Op-CLP-3'UTR foram usados para substituir o sinal de secreção tPA e a 3'UTR anterior de SV40, respectivamente, no pMT-X-Kt Δ Xho descrito no Exemplo 3. O vetor de expressão contendo estes dois elementos otimizados é

referido como pHBI-5. Genes repórteres (por exemplo, H3 HA-Ecto) foram clonados em pHBI-5 e sua expressão foi avaliada. Geralmente, os níveis de expressão foram 1,5 vez maiores do que o nível de expressão vista para os mesmos genes repórteres clonados em pMT-X-Kt Δ Xho, indicando que as melhoras vistas individualmente pelo sinal de secreção 1R5L e Op-CLP-3'UTR não foram aditivas; entretanto, os níveis de expressão foram ainda mais altos do que aqueles conduzidos pelo vetor pMT-X-Kt Δ Xho.

Dois cassetes de expressão foram montados contendo cinco elementos. Os dois cassetes diferiram com relação a qual dos dois promotores principais foi usado. O primeiro cassette contém o promotor principal SCP1-TolloMTE Δ 111 e o segundo cassette contém o promotor principal SCP7. Estes cassetes de expressão são referidos como HBI-10 e HBI-11, respectivamente. As sequências dos cassetes inteiros são listadas em SEQ ID NO: 10 (HBI-10) e SEQ ID NO: 11 (HBI-11). As figuras 1 e 2 ilustram a sequência das regiões do promotor proximal e promotor principal de HBI-10 e HBI-11, respectivamente, e são totalmente anotadas. As diferenças do HBI-10 e HBI-11 em relação a pMTtPA são resumidas na tabela da figura 3.

Os cassetes de expressão HBI-10 e HBI-11 foram inseridos no vetor de clonagem p8R322 para criar os vetores de expressão pHBI-10 e pHBI-11 de células S2. Os mapas de plasmídeo para estes dois vetores de expressão de células S2 são mostrados nas figuras 4 e 5, respectivamente.

A avaliação da expressão do produto H3 HA-Ecto utilizando os vetores de expressão pHBI-10 e pHBI-11 revelou que o pHBI-11 resultou em um nível de expressão 3 vezes superior em relação ao vetor pMT-X-Kt Δ e o pHBI-10 resultou em um nível de expressão aproximadamente 2 vezes superior em relação à expressão do mesmo produto H3 HA-Ecto direcionado pelo vetor pMT-X-Kt Δ .

REFERÊNCIAS

- Apuya N, Kwok S, Alexandrov N, Tatrino T, Fang Y, Pennell R, Lu YP, Medrano L, Cook Z, Feldmann K. *Promoter, Promoter Control Elements, and Combinations, and uses Thereof*. U.S. Patent No.: US 2006/0090216 A1. Pub. Date; Apr. 27, 1006.
- Bendtsen JD, Nielsen H, von Heijne G, Brimal S. *Improved prediction of signal peptides: SignalP 3.0*. *J Mol Biol* 2004; 340:783-795.
- Bernard AR, Kost TA, Overton L, Cavegn C, Young J, Bertrand M, Yahia-cherif Z, Chabert C and Mills A. *Recombinant protein expression in Drosophila cell line: comparison with the baculovirus system*. *Cytotechnology*. 1994; 15:139-144.
- Bin L, Tsing S, Kosaka AH, Nguyen B, Osen EG, Bach C, Chan H and Barnett J. *Expression of human dopamine β -hydroxylase in Drosophila Schneider 2 cells*. *Biochem. J.* 1996; 313:57-64.
- Butler, JEF, Kadonaga JT. *The RNA polymerase II core promoter: a key component in the regulation of gene expression*. *Genes Dev.* 2002; 16:2583-2592.
- Cavener DR. *Comparison of the consensus sequence flanking translational start sites in Drosophila and vertebrates*. *Nucleic Acids Research*. 1987; 15(4):1353-1361.
- Culp JS, Johansen H, Hellmig B. *Regulated expression allows high level production and secretion of HIV gpl20 envelope glycoprotein in Drosophila Schneider cells*. *Biotechnology*. 1991; 9:173-177.
- Deo YK, Park EY. *Multiple co-transfection and co-expression of human β -1,3-N-acetylglucosaminyltransferase with human calreticulin chaperone cDNA in a single step in insect cells*. *Biotechnol. Appl. Biochem.* 2006; 43: 129-135.
- Edelman GM, Meech R, Owens GC, and Jones FS. *Synthetic promoter elements obtained by nucleotide sequence variation and selection for activity*. *Proc Natl Acad Sci.* 2000; 97: 3038-3043.
- Farrell PJ, Lu M, Prevost J, Brown C, Behie L, Iatrou K. *High-level expression of secreted glycoproteins in transformed lepidopteran insect cells using a novel expression vector*. *Biotechnology and Bioengineering*. 1998; 60(6):

- 656-663.
- FitzGerald PC, Sturgill D, Shyakhtenko A, Oliver B, Vinson C. *Comparative genomics of Drosophila and human core promoters*. *Genome Biology*. 2006; 7:R53.
- 5 Gardsvoll H, Werner F, Sondergaard L, Dano K, Ploug M. *Characterization of low-glycosylated forms of soluble human urokinase receptor expressin in Drosophila Schneider 2 cells after deletion of glycosylation-sites*. *Protein Expression and Purification*. 2004; 34:284- 295.
- Gershenzon NI, Trifonov ED, Ioshikhes IP. *The features of Drosophila core promoters revealed by statistical analysis*. *BMC Genomics*. 7:161.
- 10 Granok H, Leibovitch BA, Shaffer CD, Elgin SCR. *Ga-ga over GAGA factor*. *Current Biol*. 1995; 5: 238-241.
- Hirose F, Yamaguchi M, Handa H, Inomata Y, Matsukage A. *Novel 8-base pair sequence (Drosophila DNA replication-related element) and specific binding factor involved in the expression of Drosophila genes for DNA polymerase α and proliferating cell nuclear antigen*. *J Biol Chem*. 1993; 268: 2092-2099.
- 15 Hofmann KJ, Schultz LD. *Mutations of the α -galactosidase signal pepetide which greatly enhance secretion of heterologous proteins by yeast*. *Gene*. 1991; 101 :105-111.
- 20 Ikonomou L, Schneider Y-J, Agathos SN. *Insect cell culture for industrial production of recombinant proteins*. *Appl. Microbiol. Biotechnol*. 2003; 62:1-20.
- Incardona, J.P. and T.L. Rosenberry. *Construction and characterization of secreted and chimeric transmembrane forms of Drosophila acetylcholinesterase: a large truncation of the C- terminal signal peptide does not eliminate glycoinositol phospholipid anchoring*. *Mol. Biol, of the Cell* 1996; 7:595-611.
- 25 Ivey-Hoyle, M. *Recombinant gene expression in cultured Drosophila melanogaster cells*. *Curr. Opin. Biotechnol*. 1991; 2:704-707.
- Johansen HA, van der Straten R, Sweet R, Otto E, Maroni G, Rosenberg M.
- 30 *Regulated expression at high copy number allows production of a growth-inhibitory oncogene product in Drosophila Schneider cells*. *Genes and Development*. 1989; 3:882-889.

- Juven-Gershon T, Hsu JY, Kadonaga JT. *Perspectives on the RNA polymerase II core promoter*. Biochem Soc. Trans. 2006; 34:1047-1050.
- Juven-Gershon T, Cheng S, Kadonaga JT. *Rational design of a super core promoter that enhances gene expression*. Nature Methods. 2006; 3:917-922.
- 5 Kadonaga JT and T Gershon. *Optimized core promoters and uses therefor*. Application PCT/US2006/020394. App published 30 November 2006. Priority date 25 May 2005 (US provisional 60/684,482).
- Lim CY, Santoso B, Boulay T, Dong E, Ohler U, Kadonaga JT. *The MTE, a new core promoter element for transcription by RNA polymerase II*. Genes and Development. 2004; 18: 1606-1617.
- 10 Kirkpatrick RB, Matico RE, McNulty DE, Strickler JE, Rosenberg M. *An abundantly secreted glycoprotein from Drosophila melanogaster is related to mammalian secretory proteins produced in rheumatoid tissues and by activated macrophages*. Gene. 1995; 153: 147- 154.
- 15 Kirkpatrick RB and Shatzman A. *Drosophila S2 System for heterologous gene expression*. In *Gene Expression Systems: Using Nature for the Art of Expression*. Joseph M Fernandez and James Hoeffler editors. Academic Press. 1999; Pp. 289-330.
- Kloc M, Zearfoss NR, Etkin LD. *Mechanisms of subcellular mRNA localization*. Cell. 2002; 108:533-544.
- 20 Kozak M. *Features in the 5' non-coding sequences of rabbit α and β -globin mRNAs that affect translational efficiency*. J. Mol. Biol. 1994; 235:95-110.
- Kozak M. *Regulation of translation via mRNA structure in prokaryotes and eukaryotes*. Gene. 2005; 361:13-37.
- 25 Kutach AK, Kadonaga JT. *The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters*. Mol. Cell. Biol. 2000; 20:4754-4764.
- Le Loir Y, Azevedo V, Oliveira SC, Freitas DA, Miyoshi A, Bermudez-Humaran LG, Nouaille S, Ribeiro LA, Leclercq S, Gabriel JE, Guimaraes VD, Oliveira MN, Charlier C, Gautier M, Langella P. *Protein secretion in Lactococcus lactis: an efficient way to increase the overall heterologous protein production*. Microb. Cell Fact. 2005; 4:2.
- 30

- Li X, Eastman EM, Schwartz RJ, Draghia-Akli R. *Synthetic muscle promoters: activities exceeding naturally occurring regulatory sequences. Nature Biotechnology.* 1999; 17: 241-245.
- Lim CY, Santoso B, Boulay T, Dong E, Ohler U, Kadonaga JT. *The MTE, a new core promoter element for transcription by RNA polymerase II. Genes and Dev.* 2004; 18: 1606-1617.
- Lu M, Johnson RR, Iatrou K. *Trans-activation of a cell housekeeping gene promoter by the IE1 gene product of baculovirus. Virology.* 1996; 218:103-113.
- 10 Lu M, Farrell PJ, Johnson R, Iatrou K. *A baculovirus (BmNPV) repeat element functions as a powerful constitutive enhancer in transfected insect cells. J. Biol. Chem.* 1997;272:30724-30728.
- Modis Y, Ogata S, Clements D, Harrison SC. *A ligand-binding pocket in the dengue virus envelope glycoprotein. Proc. Natl. Acad. Sci. USA.* 2003; 15 100:6986-6991.
- Modis Y, Ogata S, Clements D, Harrison SC. *Structure of the dengue virus envelope protein after membrane fusion. Nature.* 2004; 427(6972): 313-319.
- Ohler U, Liao G, Niemann H, Rubin GM. *Computational analysis of core promoters in the Drosophila genome. Genome Biol.* 2002; 3(12): research0087.1-00087.12.
- 20 Olsen MK, Rockenbach SK, Fischer HD, Hoogerheide JG, Tomich CSC. *Stable production of an analog of human tissue plasminogen activator from cultured Drosophila cells. Cytotechnology.* 1992; 10: 157-167.
- Rud I, Jensen PR, Naterstad, Axelsson L. *A synthetic promoter library for constitutive gene expression in Lactobacillus plantarum. Microbiology.* 2006; 25 152: 1011-1019.
- Schmetzer O, Moldenhauer G, Riesenberger R, Pires JR, Schlag P, Pezzutto A. 2005. *Quality of recombinant protein determines the amount of autoreactivity against the tumor-associated epithelial cell adhesion molecule antigen: low frequency of antibodies against the natural protein. The Journal of Immunology.* 2005; 174: 942-952.
- 30 Schmidt FR. *Recombinant expression systems in the pharmaceutical indus-*

- try. *Appl Microbiol Biotechnol*. 2004; 65:363-372.
- Schneider IJ. *Cell lines derived from late embryonic stages of Drosophila melanogaster*. *J. Embryol. Exp. Morph.* 1972; 27:353-365.
- Smale ST, Kadonaga JT. *The RNA polymerase II core promoter*. *Annu. Rev. Biochem.* 2003; 72: 449-479.
- 5 Soeller WC, Oh CE, Romberg TB. *Isolation of cDNAs encoding the Drosophila GAGA transcription factor*. *Mol and Cell Biol.* 1993; 13: 7961-7970.
- Tornøe J, Kusk P, Johansen TE, Jensen PR. *Generation of a synthetic mammalian promoter library by modification of sequences spacing transcription factor binding sites*. *Gene*. 2002; 297: 21-32.
- 10 van der Straten AH, Johansen H, Rosenberg M, Sweet RW. *Introduction and constitutive expression of gene products in cultures of Drosophila cells using hygromycin B selection*. *Methods in Mol And Cell Biol.* 1989; 1:1-8.
- van der Velden AW, Voorma HO, Thomas AAM. *Vector design for optimal protein expression*. *BioTechniques*. 2001; 31: 572-582.
- 15 von Heijne G. *A new method for predicting signal sequence cleavage sites*. *Nuc Acids Res.* 1986; 14:4683-4690.
- van Oers MM, Vlak JM, Voorma HO, Thomas, AAM. *Role of the 3' untranslated region of baculovirus p10 mRNA in high-level expression of foreign genes*. *J Gen Virol.* 1999; 80:2253-2262.
- 20 Wilson IA, Skehel JJ, Wiley DC. *Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3Å resolution*. *Nature* 1981; 289:366-673.
- Xu T, Logsdon NJ, Water MR. *Structure of insect-cell-derived IL-22*. *Acta Crystallogr D Biol Crystallogr.* 2005; 61(pt 7): 942-50.
- Zhang L, Leng Q, Mixson AJ. *Alteration in the IL2 signal peptide affects secretion of proteins in vitro and in vivo*. *J. Gene Med.* 2005; 7:354-365.
- 25

REIVINDICAÇÕES

1. Vetor de expressão para expressão e secreção de proteínas heterólogas em células de inseto cultivadas, caracterizado pelo fato de compreender um cassete de expressão, em que o cassete de expressão contém um promotor sintético que é composto de elementos promotor-proximal e principal que estão operavelmente ligados dentro do cassete de expressão, de modo que a inserção de uma sequência de gene heteróloga irá resultar em expressão de proteína de células estavelmente transformadas com o vetor de expressão, em que o promotor-proximal compreende a sequência 3XR MRE-3XR MRE mostrada em SEQ ID NO: 6 e o promotor-central compreende a sequência SCP1-TolloMTE Δ 111 mostrada em SEQ ID NO: 2 ou a sequência SCP7 mostrada na SEQ ID NO: 3.

2. Vetor de expressão de acordo com a reivindicação 1, caracterizado pelo fato de compreender adicionalmente um elemento 5'UTR sintético operavelmente ligado ao promotor proximal e promotor central sintéticos, de modo que a combinação é capaz de conduzir à expressão de proteínas heterólogas.

3. Vetor de expressão de acordo com a reivindicação 2, caracterizado pelo fato de compreender adicionalmente uma sequência sinal de secreção sintética operavelmente ligada aos elementos sintéticos promotor-proximal, promotor-principal e 5'UTR, de modo que a combinação é capaz de conduzir a expressão e secreção de proteínas heterólogas.

4. Vetor de expressão de acordo com a reivindicação 3, caracterizado pelo fato de adicionalmente compreender uma sequência 3'UTR sintética operavelmente ligada aos elementos sintéticos promotor-proximal, promotor-central, 5'UTR e de sequência sinal de secreção, de modo que a combinação é capaz de conduzir à expressão e secreção de proteínas heterólogas.

5. Vetor de expressão de acordo com a reivindicação 2, caracterizado pelo fato de que o elemento 5'UTR compreende a sequência mostrada em SEQ ID NO: 4.

6. Vetor de expressão de acordo com a reivindicação 2, caracterizado pelo fato de que o elemento 5'UTR compreende a sequência

mostrada em SEQ ID NO: 5.

7. Vetor de expressão de acordo com a reivindicação 3, caracterizado pelo fato de que a sequência sinal de secreção sintética compreende a sequência que codifica 1R5L(H) mostrada em SEQ ID NO: 7.

5 8. Vetor de expressão de acordo com a reivindicação 4, caracterizado pelo fato de que a sequência 3'UTR sintética compreende a sequência Op CLP 3'UTR mostrada em SEQ ID NO: 9.

9. Vetor de expressão para expressão e secreção de proteínas heterólogas em células de inseto cultivadas, caracterizado pelo fato de
10 compreender o cassete de expressão mostrado em SEQ ID NO: 10.

10. Vetor de expressão para expressão e secreção de proteínas heterólogas em células de inseto cultivadas, caracterizado pelo fato de compreender o cassete de expressão mostrado em SEQ ID NO: 11.

FIG 1. 3XR MRE-3XR MRE-SCP1-TolloMTEΔ111 Promotor sintético

```

1                                     65
|                                     |
GCCGGCGTGTGCAAAAAGAGCCGGCGTGTGCAAAAAGAGCCGGCGTGTGCAAAAAGAATCAAACCTAGA
<----Rev MRE-----<----Rev MRE-----<----Rev MRE-----11nt spacer

66                                     130
|                                     |
GCCGGCGTGTGCAAAAAGAGCCGGCGTGTGCAAAAAGAGCCGGCGTGTGCAAAAAGAATCAAACCTAGA
<----Rev MRE-----<----Rev MRE-----<----Rev MRE-----11nt spacer

131                                     204
|                                     |
GGGCGCCTATAAAAACAGCGGCGCTCCAGGTGCAGGCATCAGTCGTTAGCCAGTCGCCGAGCCGAGCGGTTTCG
BRE      TATA                                     +1      Inr      MTE      DPE

204                                     236
|                                     |
AAGTGATTTTGCAAACCGAAGCTTAACAACATG
Kozak Met

```

FIG 2. 3XR MRE-3XR MRE -SCP7 Promotor sintético

```

1                                     65
|                                     |
GCCGGCGTGTGCAAAAAGAGCCGGCGTGTGCAAAAAGAGCCGGCGTGTGCAAAAAGAATCAAACCTAGA
<----Rev MRE-----<----Rev MRE-----<----Rev MRE-----11nt spacer

66                                     130
|                                     |
GCCGGCGTGTGCAAAAAGAGCCGGCGTGTGCAAAAAGAGCCGGCGTGTGCAAAAAGAATCAAACCTAGA
<----Rev MRE-----<----Rev MRE-----<----Rev MRE-----11nt spacer

131                                     204
|                                     |
CAAAATGTATAAAAACAGCGGCGCTCCAGGTGCAGGCAGCATCAGTTAGCCAGTCGCCGAGCCGAGCTGAATA
TATA                                     +1      Inr      MTE

204                                     236
|                                     |
AAGTGATTTTGCAAACCGAAGCTTAACAACATG
Kozak Met

```

FIG 3. Comparação de Vetores de Expressão pMTtPA, pHBI-10 e pHBI-11

ELEMENTO	pMTtPA	pHBI-10	pHBI-10
Promotor Proximal	MtnA de Drosophila	Repetições MRE sintéticas	Repetições MRE sintéticas
Promotor Principal	MtnA de Drosophila	SCP1-TolloMTE Δ 111 Sintética	SCP7 Sintética
5'UTR	MtnA de Drosophila	Δ 111	Δ 111-7
Sinal de Secreção	tPA humana	1R5L(H) Sintética	1R5L(H) Sintética
3'UTR	SV-40 Prematuro	CLP Otimizada	CLP Otimizada

FIG 4. Mapa do Plasmídeo pHBI-10

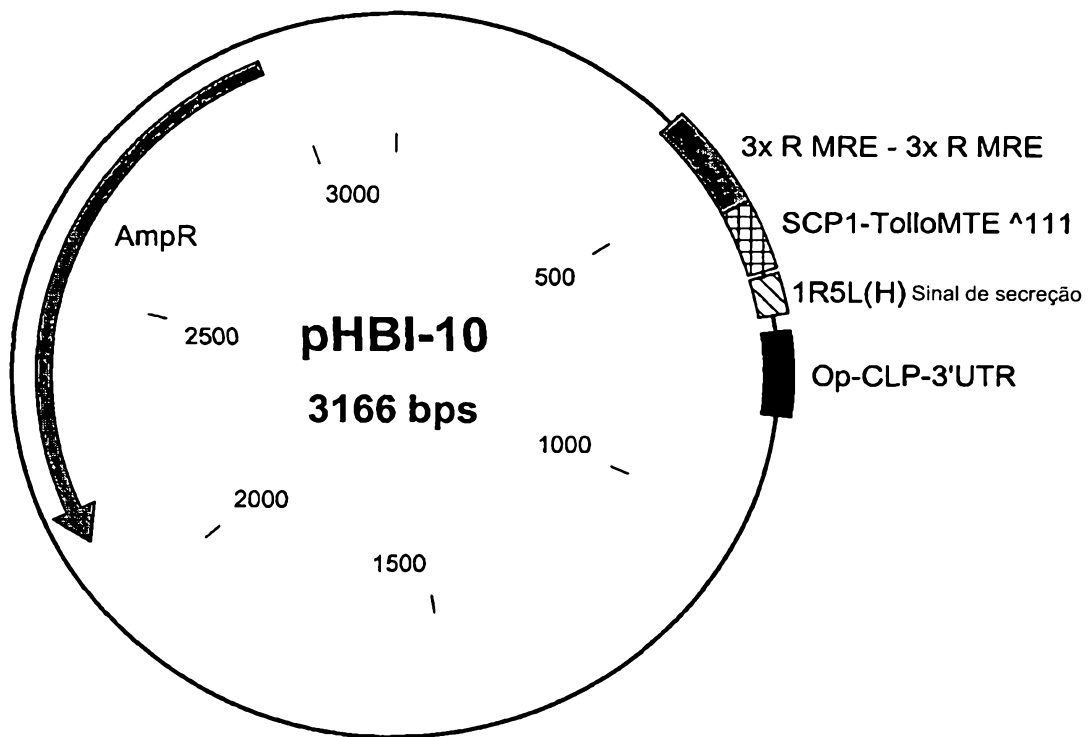


FIG 5. Mapa do Plasmídeo pHBI-11

