

República Federativa do Brasil
Ministério do Desenvolvimento, Indústria
e do Comércio Exterior
Instituto Nacional da Propriedade Industrial

(21) **PI0711207-6 A2**

(22) Data de Depósito: 22/05/2007
(43) Data da Publicação: 22/03/2011
(RPI 2098)



★ B R P I 0 7 1 1 2 0 7 A 2 ★

(51) Int.Cl.:

C12N 5/06
C12N 7/00
C12N 15/85
C12P 21/00

(54) Título: VETOR ALVO PARA INTEGRAÇÃO EM SÍTIOS-ESPECÍFICOS, MÉTODO DE SÍTIO-ESPECÍFICO QUE INTEGRA UM POLINUCLEOTÍDEO CODIFICANDO UMA PROTEÍNA DE INTERESSE EM UM GENOMA DE UMA CÉLULA EUCARIÓTICA, CÉLULA EUCARIÓTICA ISOLADA, KIT PARA USO NO SÍTIO-ESPECÍFICO QUE INTEGRA UM POLINUCLEOTÍDEO DENTRO DE UM GENOMA DE UMA CÉLULA IN VITRO E KIT PARA USO EM PRODUZIR UMA PROTEÍNA EM CÉLULA

(57) Resumo: VETOR ALVO PARA INTEGRAÇÃO EM SÍTIOS-ESPECÍFICOS, MÉTODO DE SÍTIO-ESPECÍFICO QUE INTEGRA UM POLINUCLEOTÍDEO CODIFICANDO UMA PROTEÍNA DE INTERESSE EM UM GENOMA DE UMA CÉLULA EUCARIÓTICA, CÉLULA EUCARIÓTICA ISOLADA, KIT PARA USO NO SÍTIO-ESPECÍFICO QUE INTEGRA UM POLINUCLEOTÍDEO DENTRO DE UM GENOMA DE UMA CÉLULA IN VITRO E KIT PARA USO EM PRODUZIR UMA PROTEÍNA EM UMA CÉLULA. Esta invenção fornece um sistema de integração local específico e métodos para a geração de linhagens de células eucarióticas para produção de proteína. O sistema fornecido inclui um primeiro vetor alvo integrando o local-específico e um segundo vetor doador integrando o local específico compreendendo o gene de interesse. Também são fornecidos linhagem de células mamíferas produzidas pelos métodos e sistemas revelados, bem como kits que incluem o sistema revelado.

(30) Prioridade Unionista: 22/05/2006 US 60/802,719

(73) Titular(es): HIPROCELL, LLC

(72) Inventor(es): JIMMY Z. ZHANG, MICHELLE P. CALOS,
WILLIAM J. RUTTER

(74) Procurador(es): David do Nascimento Advogados
Associados

(86) Pedido Internacional: PCT US2007069482 de 22/05/2007

(87) Publicação Internacional: WO WO2007/137267de
29/11/2007

VETOR ALVO PARA INTEGRAÇÃO EM SÍTIOS-ESPECÍFICOS, MÉTODO DE SÍTIO-ESPECÍFICO QUE INTEGRA UM POLINUCLEOTÍDEO CODIFICANDO UMA PROTEÍNA DE INTERESSE EM UM GENOMA DE UMA CÉLULA EUCARIÓTICA, CÉLULA EUCARIÓTICA ISOLADA, KIT PARA USO
5 NO SÍTIO-ESPECÍFICO QUE INTEGRA UM POLINUCLEOTÍDEO DENTRO DE UM GENOMA DE UMA CÉLULA IN VITRO E KIT PARA USO EM PRODUZIR UMA PROTEÍNA EM UMA CÉLULA

REFERÊNCIA CRUZADA

Este pedido reivindica o benefício do pedido
10 provisional US N° 60/802.719, depositado em 22 de maio de 2006, o qual é incorporado aqui por referência em sua totalidade.

ANTECEDENTES DA INVENÇÃO

Proteínas tais como, anticorpos, estão emergindo
15 como opções terapêuticas e/ou preventiva para uma grande variedade de doenças. Por exemplo, a administração de anticorpos terapêuticos fornece uma importante estratégia para o tratamento e/ou profilaxia dos indivíduos com câncer ou indivíduos que foram expostos, ou foram infectados por
20 agentes de doenças virais.

Entretanto, o processo atual de geração de linhagem celular que produz altos níveis de proteínas de recombinação, tais como anticorpos, requer trabalhos intensivos de etapas de seleção e clonagem. A identificação de uma linhagem
25 celular que seja capaz de produzir um rendimento elevado das proteínas é um processo fastidioso e demorado o qual exige a seleção de centenas de linhagens celulares. Esse processo de seleção atrapalha o potencial para selecionar numerosas proteínas terapêuticas ou candidatos profiláticos. Além
30 disso, o processo de seleção também retarda a manufatura das proteínas em uma maneira oportuna efetivamente econômica.

A maioria das linhagens celulares mamíferas atuais que expressam proteínas terapêuticas, tais como anticorpos,

pé desenvolvida pela integração genômica aleatória tem
significantes inconvenientes. Por exemplo, visto que a
expressão do transgene depende do contexto cromossômico no
local de integração, a integração do transgene em uma posição
5 não desejável resulta em uma expressão relativamente baixa do
transgene. Além disso, a integração é tendente para a excisão
durante a passagem de células transfectadas
"permanentemente". Além disso, a expressão do transgene
frequentemente torna-se "silenciosa" em consequência da
10 integração aleatória do transgene em uma posição indesejável
no cromossomo.

Portanto, um método para rapidamente gerar e
identificar as linhagens celulares estáveis que são capazes
de produzir altos níveis de proteínas de recombinação ára o
15 uso como terapêuticos e diagnósticos é necessário. A presente
invenção é endereçada a esta necessidade.

LITERATURA RELEVANTE

Thyagarajan et al, Mol Cell Bio 21, 3926-34 (2001);
Groth et al, Proc Natl Acad Sci USA 97,5995-6000 (2000); Groth
20 et al., J Mol Biol 335, 667-78 (2004); Olivares et al., Nat
Biotechnol 20, 1124-8 (2002); Ortiz-Urda et al., Nat Med 8,
1166-70 (2002); Ortiz-Urda et al., Hum Gene Ther 14,923-8
(2003); Ortiz-Urda et al. J Clin Invest 111, 251-5 (2003);
Thyagarajan et al., Methods Mol Biol 308, 99- 106 (2005);
25 Olivares et al., Gene 278, 167-76 (2001); Urlaub et al., Proc
Natl Acad Sci USA 11,4216-20 (1980); Traggiai et al., Nat Med
10, 871-5 (2004); Wurm et al., Nat Biotechnol 22, 1393-8
(2004); Andersen et al., CUIT Opin Biotechnol 13, 117-23
(2002); Wirth et al., Gene 73, 419-26 (1988); Kim et al.,
30 Biotechnol Bioeng 58, 73-84 (1998); Gandor et al., FEBS Lett
377, 290-4 (1995); Kito et al., Appl Microbiol Biotechnol 60,
442-8 (2002); Coquelle et al., Cell 89, 215-25 (1997); Stark
et al., Cell 57, 901-8 (1989); Wurm et al., Ann N Y Acad Sci

782, 70-8 (1996); Wurm et al., *Biologicals* 22,95-102 (1994); Kim et al., *Biotechnol Prog* 17,69-75 (2001); Chappell et al., *J Biol Chem* 278,33793-800 (2003); Owens et al., *Proc Natl Acad Sci USA* 98, \47\ -6 (2001); Chappell et al., *Proc. Natl. Acad. Sci. U.S.A.*, 97, 1536-1541 (2000); Weber et al., *Nat Biotechnol* 22, 1440-4 (2004); Weber et al., *Metab Eng* 1, 174-81 (2005); Chalberg et al., *J Mol Biol*, 357, 28-48 (2006); Jones et al., *Biotechnol Prog* 19, 163-8 (2003); Marks, et al., *J Mol Biol* 111, 581-97 (1991); Sblattero, et al., *Immunotech* 3, 271-8 (1998); and Yamanaka, et al., *J Biochem* 111, 1218-27 (1995).

SUMÁRIO DA INVENÇÃO

O objeto da invenção fornece um sistema de integração local-específico e métodos para gerar linhagens de células eucarióticas para a produção de proteína. O sistema fornecido inclui um primeiro vetor alvo de integração de local-específico e um segundo vetor doador de integração de local-específico compreendendo um gene de interesse também é fornecida linhagem celular eucariótica produzida pelos métodos e pelos sistemas do objeto do assunto, assim como kits que incluem os sistemas do objeto.

Uma característica da presente invenção fornece um vetor alvo de integração local-específico que inclui um primeiro local de recombinação do vetor que recombine com um local de recombinação genômica na presença de um primeiro local-específico unidirecional recombinante; um segundo local de recombinação do vetor que recombine com um local de recombinação do doador na presença de um segundo local-específico unidirecional recombinante que seja diferente do primeiro local-específico unidirecional recombinante; uma primeira porção de um primeiro marcador de seleção adjacente a extremidade 3' do segundo local da recombinação do vetor; e um segundo marcador de seleção

que seja diferente do primeiro marcador de seleção adjacente a extremidade 3' do segundo local de recombinação do vetor; e um segundo marcador de seleção que seja diferente do primeiro marcador de seleção.

5 Em algumas incorporações, o local de recombinação genômica é um local de recombinação genômica eucariótico. Em algumas incorporações, o primeiro local de recombinação do vetor é um local de recombinação genômica bacteriano (*attB*) ou um local de recombinação genômica do fago (*attP*). Em
10 outras incorporações, o primeiro local de recombinação do vetor é um local de recombinação genômica bacteriano (*attB*) e o local de recombinação genômica é um local de recombinação genômica de pseudo-fago (*pseudo-attP*). Em determinadas incorporações, o primeiro local de recombinação
15 do vetor é um local de recombinação genômica do fago (*attP*) e o local de recombinação genômico é um local de recombinação genômico pseudo-bacteriano (*pseudo-attP*). Em certas incorporações, o primeiro local de recombinação do vetor é um local de recombinação genômico pseudo-bacteriano
20 (*pseudo-attB*) ou um local de recombinação genômico pseudo-fago (*pseudo-attP*). Em algumas incorporações, o segundo local de recombinação do vetor é um local de recombinação genômico bacteriano (*attB*) ou um local de recombinação genômico fago (*attP*). Em algumas incorporações, o segundo
25 local de recombinação do vetor é um local de recombinação genômico pseudo-bacteriano (*pseudoattB*) ou um local de recombinação genômico do pseudo-fago (*pseudo-attl'*).

Em algumas incorporações, o primeiro local-específico unidirecional recombinante é um $\phi C31$ fago
30 recombinante, um TP901-1 fago recombinante, R4 fago recombinante, ϕFCI fago recombinante, $\phi Rv1$ fago recombinante ou ϕBTI fago recombinante. Em determinadas incorporações, o primeiro local-específico unidirecional recombinante é um

ϕ C31 fago recombinante. Em determinadas incorporações, o segundo local-específico unidirecional recombinante é um R4 fago recombinante. Em certas incorporações, um ϕ C31 fago recombinante inclui um ϕ C31 fago recombinante alterado, um
5 TP901-1 fago recombinante inclui um TP901-1 fago recombinante alterado e um R4 fago recombinante inclui um R4 fago recombinante alterado.

Outra característica da presente invenção fornece um método do local que integra especificamente um
10 polinucleotídeo que codifica uma proteína de interesse em um genoma de uma célula eucariótica pela introdução de um vetor alvo dentro de uma célula eucariótica que compreende um primeiro local-específico unidirecional recombinante e mantendo a célula sob condições suficientes para um evento
15 de recombinação mediado pelo primeiro local-específico unidirecional recombinante entre o primeiro local de recombinação do vetor e o local de recombinação genômica para integrar o local-específico ao vetor alvo dentro do genoma da célula; introduzindo um vetor doador dentro da
20 célula alvo que compreende um segundo local-específico unidirecional recombinante, em que o vetor doador compreende o polinucleotídeo que codifica uma proteína de interesse e um local de recombinação do doador e mantendo a célula alvo sob condições suficientes para um evento de recombinação
25 mediado pelo segundo local-específico unidirecional recombinante entre o local de recombinação do doador e o segundo local de recombinação do vetor do vetor alvo para integrar o local-específico ao polinucleotídeo que codifica uma proteína de interesse dentro do genoma da célula; em que
30 o primeiro local-específico unidirecional recombinante é diferente do segundo local-específico unidirecional recombinante. Em umas incorporações suplementares, o método inclui a seleção de uma célula que expressa a proteína de

interesse.

Em algumas incorporações, o primeiro local de recombinação do vetor é um local de recombinação genômico bacteriano (*attB*) ou um local de recombinação genômico fago (*attP*). Em outras incorporações, o primeiro local de recombinação do vetor é um local de recombinação genômico bacteriano (*attB*) e o local de recombinação genômico é um local de recombinação genômico pseudo-fago (*pseudo-attP*). Em determinadas incorporações, o primeiro local de recombinação do vetor é um local de recombinação genômico fago (*attP*) e o local de recombinação genômico é um local de recombinação genômico pseudo-bacteriano (*pseudo-attB*). Em outras incorporações, o primeiro local de recombinação do vetor é um local de recombinação genômico pseudo-bacteriano (*pseudo-attB*) ou um local *attP* de recombinação genômico pseudo-fago (*pseudo-attP*). Em algumas incorporações, o segundo local de recombinação do vetor é um local de recombinação genômico bacteriano (*attB*) ou um local de recombinação genômico fago (*attP*). Em outras incorporações, o segundo local de recombinação do vetor é um local de recombinação genômico pseudo-bacteriano (*pseudo-attB*) ou um local *attP* de recombinação genômico recombinação pseudo-fago (*pseudo-attP*). Em algumas incorporações, o local doador de recombinação é um local de recombinação genômico bacteriano (*attB*) ou um local de recombinação genômico fago (*attP*). Em algumas incorporações, o local doador de recombinação é um local de recombinação genômico pseudo-bacteriano (*pseudo-attB*) ou um local *attP* de recombinação genômico pseudo-fago (*pseudo-attP*).

Em algumas incorporações, o primeiro local-específico unidirecional recombinante é um ϕ C31 fago recombinante, um TP901-1 fago recombinante, R4 fago recombinante, ϕ FCI fago recombinante, ϕ Rv1 fago recombinante

ou um ϕ BTI fago recombinante. Em determinadas incorporações, o primeiro local-específico unidirecional recombinante é um ϕ C31 fago recombinante. Em determinadas incorporações, o segundo local-específico unidirecional é um R4 fago recombinante. Em algumas incorporações a proteína é uma enzima que possa ser usada para a produção de nutrientes ou realizar reações enzimáticas na química, ou um polinucleotídeo útil e disponível como um nutriente ou para o tratamento de uma doença humana ou animal ou para a prevenção deste, por exemplo, um hormônio, um polipeptídeo com atividade imunomoduladoras, atividade antiviral e/ou antitumoral (por exemplo, maspin), anticorpo, antígeno viral, vacina, fator de coagulação, inibidor de enzima, ingrediente dos gêneros alimentícios e similares. Em determinadas incorporações, a proteína é uma proteína segregada, tal como um anticorpo. Em algumas incorporações, a célula é uma célula mamífera. Em algumas incorporações, a célula mamífera é uma célula de roedor, tal como, uma célula CRO ou uma linhagem celular CRO-derivada deficiente de dihidrofolato redutase, tal como, DG44. Em outras incorporações, a célula mamífera é uma célula humana, tal como, uma célula PER.C6™.

Outra característica da presente invenção fornece uma célula isolada que inclui um cassete polinucleotídeo integrado genomicamente que compreende um primeiro local de recombinação híbrido e um segundo local de recombinação híbrido que flanqueando um local de recombinação do vetor que recombina com um local de recombinação do doador na presença de um local-específico unidirecional recombinante; uma primeira porção de um primeiro marcador de seleção adjacente a extremidade 3' dos locais de recombinação do vetor; e um segundo marcador de seleção que seja diferente do primeiro marcador de seleção.

Em algumas incorporações, o local de recombinação do vetor é um local de recombinação genômico bacteriano (*attB*) ou um local recombinação genômico fago (*attP*). Em algumas incorporações, o local de recombinação do doador é um local de recombinação genômico bacteriano (*attB*) ou um local de recombinação genômico fago (*attP*). Em algumas incorporações, o local-específico unidirecional recombinante é um $\phi C31$ fago recombinante, TP901-1 fago recombinante ou R4 fago recombinante. Em algumas incorporações, a célula é uma célula mamífera. Em algumas incorporações, a célula mamífera é uma célula de roedor, tal como, uma célula de CRO ou uma linhagem celular CRO-derivada deficiente de dihidrofolato redutase, tal como DG44. Em outras incorporações, a célula mamífera é uma célula humana, tal como, uma célula de PER.C6™.

Contudo, outra característica da presente invenção fornece um kit para uso no local especificamente integrando um polinucleotídeo dentro de um genoma de uma célula *in vitro*, incluindo: um vetor alvo; e um vetor doador que inclua dois promotores, duas seqüências de sinal se a proteína de interesse for segregada, 2 genes interruptores reguladores para controlar a expressão do gene, dois realçadores translacionais para aumentar a expressão, dois locais múltiplos de clonagem, um local doador de recombinação, e uma segunda porção de um primeiro marcador selecionável (por exemplo, promotor) adjacente a extremidade 5' do local de recombinação do doador. Em algumas incorporações, o kit também inclui um primeiro local-específico unidirecional recombinante ou um ácido nucléico que codificam os mesmos. Em umas incorporações suplementares, o kit também inclui um segundo local-específico uni direcional recombinante ou um ácido nucléico que codificam o mesmo que seja diferente do primeiro local-

específico unidirecional recombinante.

Em algumas incorporações o primeiro local-específico unidirecional recombinante é um ϕ C31 fago recombinante, TP90 1-I fago recombinante, R4 fago recombinante, ϕ FCI fago recombinante, ϕ Rv1 fago recombinante ou ϕ BTI fago recombinante. Em algumas incorporações, o segundo local-específico unidirecional recombinante é um ϕ C31 fago recombinante, TP901-1 fago recombinante, R4 fago recombinante, ϕ FCI fago recombinante, ϕ Rv1 fago recombinante ou ϕ BTI fago recombinante.

Contudo, outra característica da presente invenção fornece um kit para uso na produção de uma proteína dentro de uma célula eucariótica, incluindo: uma célula eucariótica isolada que inclui um cassete de polinucleotídeo genomicamente integrado que compreende um primeiro local de recombinação híbrido e um segundo local de recombinação híbrido flanqueando um local de recombinação do vetor que recombina com o local de recombinação do doador na presença de um local-específico unidirecional recombinante, uma primeira porção de um primeiro marcador selecionável adjacente a extremidade 3' do local de recombinação do vetor e um segundo marcador selecionável que seja diferente do primeiro marcador selecionável; e um vetor doador que inclua um múltiplos locais de clonagem, um local de recombinação do doador, e uma segunda porção de um primeiro marcador selecionável (por exemplo, promotor) adjacente a extremidade 5' do local de recombinação do doador.

Em algumas incorporações, o kit também inclui um local-específico unidirecional recombinante ou um ácido nucléico que codificam os mesmos. Em algumas incorporações o local-específico unidirecional recombinante é um ϕ C31 fago recombinante, TP90 1-I fago recombinante, R4 fago recombinante, ϕ FCI fago recombinante, ϕ Rv1 fago recombinante

ou um ϕ BTI fago recombinante.

Estes e outros objetos, vantagens, e características da invenção tornar-se-ão aparentes daquelas pessoas hábeis na técnica sob a leitura dos detalhes da invenção como descrito mais inteiramente abaixo.

BREVE DESCRIÇÃO DOS DESENHOS

A invenção é mais bem compreendida da seguinte descrição detalhada quando lida conjuntamente com os desenhos em acompanhamento. Enfatiza que, de acordo com a prática comum, as várias características dos desenhos não são para parâmetros. Ao contrário, as dimensões das várias características arbitrariamente são expandidas ou reduzidas para maior clareza. São incluídas nos desenhos as seguintes figuras:

FIG. 1 é uma representação esquemática de um vetor alvo exemplificativo. O vetor alvo exemplificativo inclui um primeiro local de recombinação do vetor (por exemplo, um local *attB* ϕ C31), um segundo local de recombinação do vetor (por exemplo, local *attP* R4), uma primeira porção de um primeiro marcador selecionável (por exemplo, primeiro marcador selecionável do promotor-menor (por exemplo, o gene de resistência à zeocina)) a jusante do local *attP* R4, e um segundo marcador selecionável (por exemplo, um gene de resistência à higromicina).

FIG.2 é uma representação esquemática de um vetor doador exemplificativo. O vetor doador exemplificativo inclui um local de recombinação do doador (por exemplo, local *attB* R4) um gene de interesse e um promotor (por exemplo, um promotor CMV) apenas a jusante ao local *attB* R4.

FIG.3 é uma representação esquemática de um evento de integração local específico inicial exemplificativo entre o local *attB* ϕ C31 presente no vetor alvo e local do *pseudo-attP* ϕ C31 presente no genoma da célula alvo. O evento de integração

é mediado pela $\phi C31$ integrase.

FIG. 4 é uma representação esquemática de um evento de integração local específico exemplificativo entre o local *attB* R4 presente no vetor doador e o *attP* R4 integrado dentro do genoma da célula como resultado da integração do vetor do alvo. O segundo evento de integração é mediado pela R4 integrase.

FIG. 5 é uma representação esquemática de um vetor DHFR-alvo exemplificativo. O vetor DHFR-alvo exemplificativo inclui um local *attP* R4, um local do *attB* $\phi C31$, um gene de resistência à higromicina, um gene de DHFR, e uma primeira porção (por exemplo, promotor-menor) de um gene de resistência à zeocina a jusante ao local *attP* R4.

FIG. 6 é uma representação esquemática de um vetor DHFR-doador exemplificativo. O vetor doador exemplificativo inclui um local *attB* R4, um gene de interesse, um gene de DHFR e um promotor CMV apenas a montante ao local *attB* R4.

FIG. 7 é uma representação esquemática de um vetor IRES-doador exemplificativo. O vetor doador exemplificativo inclui um local *attB* R4, um gene de interesse, um promotor CMV apenas a montante do local *attB* R4 e um IRES entre o local do início da transcrição e a região de codificação para o gene do interesse.

FIG. 8 é uma representação esquemática de um vetor alvo pR1. O vetor alvo pR1 inclui um primeiro local de recombinação do vetor (por exemplo, um local R4), um segundo local de recombinação do vetor *attB* 295 (por exemplo, um local *attP* 103 $\phi C31$), uma primeira porção de um primeiro marcador selecionável (por exemplo, marcador selecionável do promotor-menor (por exemplo, o gene de resistência à puromicina)) a montante ao local $\phi C31$ *attP* 103 e um segundo marcador selecionável completo (por exemplo, um cassete de resistência à higromicina). Também contém um origem ColEI de replicação

do DNA e um cassete do gene de resistência à ampicilina para a manutenção e a seleção em *E. Coli*, respectivamente. Os asteriscos designam únicos locais da enzima de restrição.

FIG.9 é uma representação esquemática de uma
 5 espinha dorsal do vetor de expressão doador exemplificativo (PHPC-4). A espinha dorsal do vetor de expressão doador exemplificativo inclui um local de recombinação do doador (por exemplo, um local AAA ϕ C31 attB 285), dois promotores CMV, duas seqüências de sinal para o secreção das proteínas,
 10 dois poliligantes para a inserção dos genes de interesse e dois sinais polis adenilação do hormônio do crescimento bovino. Também inclui um promotor mais fraco (por exemplo, um promotor SV40) apenas a montante do local AAA ϕ C31 attB 285 para selecionar a integração de um vetor de expressão doador
 15 dentro do vetor alvo. Além disso, o vetor também inclui uma origem ColEI de replicação de DNA e um cassete do gene de resistência à ampicilina para a manutenção e a seleção em *E. Co li*, respectivamente. Os asteriscos designam locais únicos da enzima de restrição.

FIG.10 é uma representação esquemática de um vetor
 20 de expressão doador exemplificativo (pD1- DTX-I). O vetor de expressão doador exemplificativo inclui um local de recombinação do doador (por exemplo, um local AAA ϕ C31 attB 285), dois promotores CMV, duas seqüências de sinal, cadeias
 25 pesadas e leves de um anticorpo da toxina anti-diftérica e dois sinais de poliadenilação do hormônio de crescimento bovino. O vetor também inclui um promotor mais fraco (por exemplo, um promotor SV40) apenas a montante do local AAA ϕ C31 attB 285 para selecionar a integração do vetor de
 30 expressão doador dentro do vetor alvo. Além disso, o vetor também inclui uma origem ColEI de replicação do DNA e de um cassete do gene de resistência à ampicilina para a manutenção e a seleção em *E. Coli*, respectivamente.

FIG. 11 é uma representação esquemática do procedimento de teste rápido usado para verificar a função de cada um dos quatro vetores usados para gerar linhagem celular para a produção de nível elevado de proteína. A primeira etapa usa R4 integrase codificada por um vetor da expressão R4 integrase e (por exemplo, pCMV sre para mediar a integração do vetor do alvo dentro de locais pseudo attP R4. Quarenta e oito horas são permitidas para a integração ocorrer sem seleção (por exemplo, seleção por higromicina).

A segunda etapa usa uma integrase mutante ϕ C31 codificada por um vetor de expressão da integrase mutante ϕ C31 (por exemplo, pCS-M3J) para mediar a integração do vetor doador dentro do vetor do alvo. Quarenta e oito horas são permitidas para a integração ocorrer e uma seleção por puromicina é então usada para isolar uma associação estável das células. Estas células são analisadas para a expressão da proteína. A expressão de nível elevado da proteína depende da função apropriada de cada um dos quatro plasmídeo usados. Mesmo se o vetor alvo integrado aleatoriamente ou local-específico em locais pseudo attP R4 na primeira etapa podem ser avaliados fazendo a experiência com ou sem o vetor de expressão R4 integrase. O nível de expressão da proteína será substancialmente mais baixo se o vetor de expressão da R4 integrase for omitido porque os vetores alvos não-integrados estarão diluídos para fora como células divisoras sobre o comprimento da experiência «17 dias).

FIG. 12 é uma representação esquemática de um primeiro evento de integração local-específico exemplificativo entre o local R4 attB 295 presente no vetor alvo e os locais R4 pseudo-attP presente no genoma da célula alvo. O evento de integração é mediado o pela R4 integrase e codificado pelo plasmídeo pCMV sre. A seleção pela Higromicina é usada para isolar clones estáveis (por exemplo,

linhagem celular *attP* PER.C6- ϕ C31 ou *attP* DG44- ϕ C31) com o vetor alvo integrado nos locais R4 *pseudoattP*.

FIG. 13 é uma representação esquemática de um segundo evento de integração local-específico exemplificativo que ocorre em linhagem celular *attP* ϕ C31 entre local AAA ϕ C31 *attB* 285 presente no vetor doador e local ϕ C31 *attP* 103 integrado dentro do genoma da célula como um resultado da integração do vetor alvo. O segundo evento de integração é mediado por um mutante ϕ C31 integras e (por exemplo, um mutante ϕ C31 integras e codificado pelo plasmídeo pCS-M3J). Um cassete de expressão reconstitutivo de resistência à droga é usado para selecionar os integrantes em que o vetor de expressão doador integrou no vetor alvo, e para selecionar contra aquelas linhagens celulares em que o vetor doador integrou nos locais ϕ C31 *pseudo-attP*.

FIG. 14 é um diagrama das seqüências dos locais *attB* ϕ C31, *attP* e *attL* 88. As seqüências do tipo selvagem de *attB* ϕ C31 e *attP* ϕ C31 são dadas na metade superior. A seqüência sublinhada na metade superior indica as seqüências do *attB* e do *attP* as quais poderiam formar um local *attL* após a recombinação. Pela convenção *attL* é nomeado de acordo com o lado da recombinação sobre o ponto que foi derivado de *attB*. Por exemplo, no *attL*, seqüências no lado esquerdo da recombinação sob o ponto são derivadas das seqüências do lado esquerdo (5') da recombinação sobre o ponto do *attB*. As seqüências no *attL* no lado direito de recombinação sob o ponto são derivadas das seqüências do lado direito (3') de recombinação sobre o ponto do *attP*.

A metade inferior da figura de diagramas como as seqüências *attB* e do *attP* foram modificadas para fazer os locais AAA ϕ C31 *attP* 103 e ϕ C31 *attB* 285 que foram usados nos vetores alvo e doador, respectivamente. Também indica a seqüência do local ϕ C31 *attL* 88 que resulta depois que o

local AAA $\phi C31$ attB 285 no vetor doador integra dentro do local $\phi C31$ attP 103 no vetor do alvo.

FIG. 15 é uma representação esquemática de um vetor alvo-DHFR exemplificativo (pR1 - DHFR). O vetor alvo-DHFR exemplificativo inclui um local $\phi C31$ attP 103, um local R4 attB 295, um gene de resistência à higromicina, um gene DHFR e uma primeira porção do gene de resistência à puromicina (por exemplo, promotor-menor) a montante do local de $\phi C31$ attPI03. O vetor também inclui uma origem de replicação de DNA ColEI e um cassete do gene de resistência à ampicilina para a manutenção e a seleção em *E. Coli*, respectivamente.

FIG. 16 é uma representação esquemática de um vetor de expressão do doador-DHFR exemplificativo (pD1-DHFR). O vetor de expressão do doador-DHFR exemplificativo inclui um local de recombinação do doador (por exemplo, um local AAA $\phi C31$ attB 285), dois promotores CMV, duas seqüências de sinal, cadeias pesadas e leves de um anticorpo da toxina da anti-difteria, dois sinais de poliadenilação do hormônio de crescimento bovino, um cassete de expressão DHFR e um promotor (por exemplo, um promotor SV40) apenas a montante do local AAA $\phi C31$ attB 285 para selecionar a integração do vetor doador dentro do vetor alvo. O vetor também inclui uma origem de replicação do DNA Co 1 EI e um cassete do gene de resistência à ampicilina para a manutenção e a seleção de *E. Coli*, respectivamente.

FIG. 17 é uma representação esquemática de um vetor de expressão IRES-doador exemplificativo (PD1-IRA). O vetor de expressão IRES-doador exemplificativo inclui um local de recombinação do doador (por exemplo, um local AAA $\phi C31$ attB 285), dois promotores CMV, dois locais internos da entrada da ribossomo (IRES) na região não-traduzida 5', duas seqüências de sinalização, cadeias pesadas e leves de um

anticorpo da toxina da anti-difteria, dois sinais de poliadenilação do hormônio de crescimento bovino e um promotor (por exemplo, um promotor SV40) apenas a montante do local AAA $\phi C31$ attB 285 para selecionar a integração do
 5 vetor doador dentro do vetor alvo. O vetor também inclui uma origem de replicação de DNA ColEI e um cassete do gene de resistência à ampicilina para a manutenção e a seleção em *E. Coli*, respectivamente.

FIG. 18 é uma representação esquemática de um
 10 vetor alvo de regulação exemplificativo (pR1reg). O vetor alvo de regulação exemplificativo inclui um primeiro local de recombinação do vetor (por exemplo, um local R4 attB 295), um segundo local de recombinação do vetor (por exemplo, um local $\phi C31$ attP 103), uma primeira porção de um primeiro
 15 marcador selecionável (por exemplo, marcador selecionável do promotormenor (por exemplo, o gene de resistência à puromicina)) ajustante ao local $\phi C31$ attP, um segundo marcador selecionável completo (por exemplo, um cassete do gene de resistência à higromicina) e um cassete que
 20 codifique as proteínas (por exemplo Rheoativador e RheoReceptor) capazes de conferir regulação do controlável um ou mais genes presente em um vetor de expressão de regulação doador (por exemplo pD1reg), que tenha os genes que são configurados em uma maneira, tal que são capazes de
 25 regulação. O vetor também inclui uma origem de replicação do DNA Col EI e um cassete do gene de resistência à ampicilina para a manutenção e seleção em *E. Coli*, respectivamente.

FIG.19 é uma representação esquemática de um vetor DHFR- alvo de regulação exemplificativo (pR1reg-DHFR). O
 30 vetor DHFR -alvo de regulação exemplificativo inclui um primeiro local de recombinação do vetor (por exemplo, um local R4 attB 295), um segundo local de recombinação do vetor (por exemplo, um local $\phi C31$ attB 103), uma primeira

porção de um primeiro marcador selecionável (por exemplo, marcador selecionável do promotor-menor (por exemplo, o gene da resistência à puromicina) a jusante ao local ϕ C31 attP 103, um segundo marcador selecionável completo (por exemplo, um cassete do gene de resistência à higromicina), um gene de DHFR e um cassete que codifiqueas proteínas (por exemplo RheoAtivador e Rheo receptor) capazes de conferir regulação ao gene controlável em um ou mais genes presentes em um vetor de expressão de regulação doador (por exemplo pDlreg, o qual tenha os genes que são configurados em uma maneira, tal que são capazes de regulação. O vetor também inclui uma origem de replicação de DNA Col E1 e um cassete do gene de resistência à ampicilina para a manutenção e a seleção em *E. Cole*, respectivamente.

FIG. 20 é uma representação esquemática de uma espinha dorsal do vetor de expressão de regulação doador (pDlreg). A espinha dorsal do vetor doador de expressão de regulação doador inclui um local de recombinação do vetor doador (por exemplo, um local AAA ϕ C31 att B 285) duas seqüências que mediem a regulação do gene (por exemplo, uma região de poliadenilação SV40), duas seqüências reguladoras do gene (por exemplo, 5x GAL4 UAS, TATA box, e um 5' UTR), duas seqüências de sinalização, um poliligante para introduzir genes de interesse, dois sinais de poliadenilação do hormônio de crescimento bovino, e um promotor (por exemplo, um promotor SV40) apenas a montante do local AAA ϕ C31 attB 285 para selecionar a integração do vetor doador dentro do vetor alvo. O vetor também inclui uma origem de replicação de DNA Col E1 e um cassete do gene de resistência à ampicilina para a manutenção e a seleção em *E. Coli*, respectivamente. Os asteriscos designam locais únicos e enzimas de restrições.

FIG. 21 é uma representação esquemática de um

vetor de expressão doador exemplificativo selecionável (pD1-DTXI-G418). a vetor de expressão doador exemplificativo selecionável inclui todos os elementos de um vetor de expressão doador (FIG. 10), mas também inclui um completo
5 gene marcador selecionável (por exemplo, G418).

FIG. 22 demonstra a recombinação local-específica de um vetor alvo com um vetor de expressão doador após o transfecção transiente.

FIG. 23 mostra a seqüência de um local R4 pseudo
10 att isolado das células em que um vetor alvo foi integrado especificamente ao local usando a R4 integrase. A seqüência do núcleo R4 em que a recombinação ocorre é mostrada em letras maiúsculas.

FIG. 24 mostra as seqüências dos locais ϕ C31 att
15 híbrido isolados das células DG44 em que um vetor de expressão doador foi integrado especificamente ao local em dentro de um vetor alvo. Pannel A mostra o local attL híbrido e o pannel B mostra o local attR híbrido. A seqüência superior do ácido nucléico mostra a seqüência prevista da região do vetor de expressão doador, seguida
20 pelo attL, e então a seqüência de resistência à puromicina, que originou do vetor alvo. A seqüência inferior é a seqüência real da linhagem celular. Como mostrado na figura a seqüência real do ácido nucléico corresponde exatamente
25 com a seqüência prevista.

FIG. 25 mostra seqüências dos locais ϕ C31 att
híbrido isolado das células PER.C6™ em que um vetor de expressão doador foi integrado especificamente ao local dentro de um vetor alvo. Pannel A mostra o local attL
30 híbrido e o pannel B mostra o local attR híbrido. A seqüência superior do ácido nucléico mostra a seqüência prevista da região do vetor de expressão doador, seguida pelo attL, e então a seqüência de resistência à puromicina,

a qual originou do vetor alvo. A seqüência inferior é a seqüência real da linhagem celular. Como mostrado na figura, a seqüência do ácido nucléico real corresponde exatamente com a seqüência prevista.

5 FIG. 26 mostra amplificação reação-mediada de cadeia de polimerase dos locais attB (Painel A) e attR (painel B) do DNA genômico das células com vetores de expressão doadores especificamente integrados ao local.

10 FIG. 27A mostra a expressão de um anticorpo da associação de CRa DRFR de clones após a integração do vetor da expressão doador especificamente ao local.

 FIG. 27B mostra a expressão de um anticorpo da associação PER.C6™ de clone após a integração do vetor da expressão doador especificamente ao local.

15 FIGURAS 28A e 28B mostram a expressão de um anticorpo de clone de célula única da associação CRa dhfr#2G7 que contém vetores de expressão doadores integrados especificamente ao local.

20 FIG. 29 mostra a expressão de um anticorpo (pg/célula/dia) de uma associação das células em que um vetor de expressão doador foi integrado especificamente ao local dentro de um vetor DHFR-alvo e populações de célula que foram expostas então às concentrações crescentes de metotrexato.

25 FIG. 30 é uma representação esquemática de um vetor de expressão doador repórter exemplificativo (pD3-DTXI). O vetor de expressão doador repórter exemplificativo inclui todos os elementos de um vetor de expressão doador (FIG. 10), mas também inclui um gene que codifica uma
30 molécula repórter, tal como, proteína fluorescente verde. A presença do gene repórter permite a fácil identificação das células individuais que expressam uma proteína de interesse.

 FIG. 31 mostra a comparação da atividade de

ligação específica do anticorpo da toxina anti-difteria expressado nas células DG44 e nas células de PERC6™.

FIG. 32 mostra biologicamente, a atividade de neutralização in vitro do anticorpo da toxina da anti-difteria expressado das células DG44 ou das células PERC6™ comparadas àquela da linhagem humana da célula-B (D2.2), de que os genes do anticorpo foram clonados.

FIGURAS 33A-33B mostram a sequência do ácido nucléico para o vetor pR1. [0058] FIGURAS 34A-34C mostram a sequência do ácido nucléico para o vetor pD1-DTX-1.

FIGURAS 35A-35C mostram a sequência do ácido nucléico para o vetor pR1-DHFR

FIGURAS 36A-36D mostram a sequência do ácido nucléico para o vetor pD 1-DTXI-G418.

FIGURAS 37A-37D mostram a sequência do ácido nucléico para o vetor pD3-DTX1.

DEFINIÇÕES

"Recombinases" é uma família das enzimas que media a recombinação local-específico entre seqüências específicas de DNA reconhecidas pela recombinase (Esposito, D., and Scocca, J. J., Nucleic Acids Research 25,3605-3614 (1997); Nunes-Duby, S. E., et al, Nucleic Acids Research 26,391-406 (1998); Stark, W. M., et al., Trends in Genetics 8, 432-439 (1992). Dentro deste grupo estão diversas subfamílias que incluem a "Integrase" ou tirosina recombinase (incluindo, por exemplo, integras e Cre e lambda) e "Resolvase/Invertase" ou serina recombinase (incluindo, por exemplo, ϕ C31 integras e, R4 integras, e TP-901 integrase). O termo também inclui as recombinases que são alteradas em comparação ao tipo-selvagem, por exemplo, como descritos na publicação de patente US 20020094516, a divulgação de que é incorporada aqui por este meio por referência em sua totalidade.

Uma "recombinase local-específica unidirecional" é uma recombinase natural, tal como, ϕ C31 integras e, uma recombinase transformada ou alterada, tal como, uma ϕ C31 integrase transformada ou alterada que retenha atividade de recombinação local-específica uni direcional ou uma recombinase bidirecional modificado para ser uni direcional, tal como, uma cre recombinase que foi modificada para se tomar unidirecional.

"Recombinases alteradas" e "recombinases mutantes"; são usadas permutavelmente aqui para referir-se às enzimas recombinas, e em que o gene recombinase nativo, tipo selvagem encontrado no organismo de origem tenha sido transformado em uma ou mais posições relativas a uma recombinase original (por exemplo, em um ou mais nucleotídeos que podem resultar nas alterações de um ou mais aminoácidos na recombinase alterada relativa a uma recombinase precursora). "Recombinase precursora" é usada para referir-se à seqüência do nucleotídeo e/ou de aminadoácido da recombinase de qual a recombinase alterada é gerada. A recombinase precursora pode ser uma enzima natural (isto é, uma enzima tipo selvagem ou nativa) ou uma enzima de ocorrência não natural (por exemplo, uma enzima geneticamente projetada). As recombinases alteradas de interesse na invenção exibem uma especificidade de DNA e/ou um nível de ligação de atividade que difere daquela da enzima do tipo-selvagem ou outra enzima precursora. Tal especificidade de ligação alterada permite a recombinase reagir com uma dada seqüência de DNA diferentemente do que poderia a enzima precursora, enquanto um nível alterado de atividade permite a recombinase realizar a reação em maior ou menor eficiência. Uma reação de recombinase tipicamente inclui ligação a uma seqüência de reconhecimento e a execução do corte e da ligação ajustados, tendo por

resultado nas trocas de fileiras entre dois locais de reconhecimento de recombinação.

"Integração local-específica" ou "integrando especificamente ao local", como usada aqui refere-se à recombinação e integração da sequência específica de um primeiro ácido nucléico com um segundo ácido nucléico, tipicamente mediado por uma recombinase. Geralmente, recombinação ou integração local-específica ocorre nas sequências definidas particulares reconhecidas pela recombinase. Em contraste com a integração aleatória, a integração local-específica ocorre em uma sequência particular (por exemplo, um local acessório da recombinase) em uma eficiência mais elevada.

Os locais de reconhecimento nativos *attB* e *attP* do fago ϕ C31 (isto é, bacteriófago ϕ C31) são geralmente aproximadamente 34 a 40 nucleotídeos do comprimento (Groth et al., Proc Natl de Acad Sci USA 97:5995-6000 (2000)). Estes locais são arranjados tipicamente como segue: *AttB* compreende uma primeira sequência de DNA *attB5'*, uma região do núcleo, e uma segunda sequência de DNA *attB3'*, na ordem relativa de 5' a 3' *attB5'* - região do núcleo-*attB3'*. *AttP* compreende uma primeira sequência de DNA *attP5'*, região do núcleo e uma segunda sequência de DNA *attP3'*, na ordem relativa de 5' a 3' *attP5'* - região do núcleo-*attP3'*. A região do núcleo de *attP* e *attB* de ϕ C31 têm a sequência 5'-TTG-3'. Outras integrases do fago (tais como, integrase e do fago R4) e suas sequências de reconhecimento podem ser adaptadas para o uso na invenção.

A ação da integrase em cima destes locais de reconhecimentos é unidirecional em que a reação enzimática produz os produtos de recombinação do ácido nucléico que não são substratos efetivos da integrase. Isto resulta na integração estável com pouca ou quase nenhuma excisão

mediada-recombinase detectável, isto é, recombinação que é "unidirecional". O produto da recombinação da ação do integrase em cima dos pares do local de reconhecimento compreende, por exemplo, na ordem de 5' a 3': attB5' -
 5 seqüência do local do produto de reconhecimento-attP3', e attP5' - seqüência do local do produto de reconhecimento-attB3'. Assim, onde o vetor alvo compreende um local attB e o genoma alvo compreende uma seqüência attP, um produto típico do recombinação compreende a seqüência (den5' a 3'):
 10 attP5'-TTG-attB3' {seqüência do vetor alvejado} attB5' - TTG-attP3'. Como os locais attB e do attP são seqüências diferentes, a recombinação resulta em um local de recombinação de local-específico híbrido (designado attL ou attR para esquerdo e direito) que é nem uma seqüência do
 15 attB ou nem uma seqüência do attP, e é funcionalmente não-reconhecida como um local de recombinação de local-específico (por exemplo, attB ou attP) para a recombinase local-específico uni direcional relevante, assim removendo a possibilidade que ao recombinase local-específico
 20 unidirecional catalisará uma segunda reação de recombinação entre o attL e o attR que inverteria a primeira reação de recombinação.

Um "local de reconhecimento nativo", como usado aqui, significa um local de reconhecimento que ocorra
 25 naturalmente no genoma de uma célula (isto é, os locais não são introduzidos no genoma, por exemplo, por meios de recombinação).

Um "local de recombinação tipo-selvagem", como usado aqui significa um local de recombinação usado
 30 normalmente por uma integrase ou por uma recombinase. Por exemplo, o lambda é um bacteriófago temperado que infecta E. Coli. O fago tem um local acessório para a recombinação (attP) e o genoma bacteriano de E. Coli tem um local

acessório para recombinação (*attB*). Ambos desses locais são locais de recombinação tipo-selvagem para a integrase lambda. No contexto da presente invenção, locais de recombinação tipo selvagem ocorre no sistema homólogo do fago/bactérias. Conformemente, locais de recombinação tipo-selvagem pode ser derivado do sistema homólogo e associado com as seqüências heteróloga, por exemplo, local *attB* pode ser colocado em outros sistemas para atuar como um substrato para a integrase.

Um "pseudo-local" ou um "local de pseudo-recombinação, como usado aqui, significa a seqüência do DNA que compreende um local de reconhecimento que seja limitado por uma enzima recombinase onde o local de reconhecimento difere-se em um ou mais nucleotídeos de uma seqüência de reconhecimento de recombinase tipo selvagem e/ou esteja presente como uma seqüência endógena em um genoma que difere-se da seqüência de um genoma onde a seqüência de reconhecimento tipo selvagem para resíduos da recombinase. Para uma dada recombinase, uma seqüência de pseudo-recombinação é funcionalmente equivalente a uma seqüência do reconhecimento tipo-selvagem ocorre em um organismo à exceção daquela em que a recombinase é encontrada na natureza, e pode ter a variação da seqüência relativa seqüência de recombinação do tipo selvagem. Em algumas incorporações um "pseudo-local *attP*" ou "pseudo local *attB*" refere-se aos pseudo locais que são similares ao local de reconhecimentos para o fago tipo selvagem (*attP*) ou seqüências de locais acessórios bacterianos (*attB*), respectivamente, para enzimas integrase do fago, tais como, o fago ϕ C3!. Em muitas incorporações da invenção, o pseudo local *attP* está presente no genoma de uma célula hospedeira, enquanto o local *ttB* tipo selvagem esta presente em um vetor alvo no sistema da invenção. "Pseudo local att" é um termo

mais geral que possa referir a um pseudo local *attP* ou um pseudo local *attB*. Compreende-se que os locais *att* ou os pseudo locais *att* podem estar presentes em moléculas lineares ou circulares do ácido nucléico. Em determinadas
5 incorporações, a presença de "pseudo-locais de recombinação" no genoma da célula alvo evita a necessidade para introdução de um local de recombinação dentro do genoma.

Um "local de recombinação-híbrido", como usado aqui, refere-se a um local de recombinação construído das
10 porções do tipo-selvagem e/ou locais de pseudo-recombinação. Como um exemplo, um local de recombinação do tipo-selvagem pode ter uma pequena, região de núcleo flanqueada por palindrômeros. Em uma incorporação de um "local de recombinação-híbrido" a seqüência 5' da seqüência de região
15 do núcleo do local de recombinação híbrida combina um local de pseudo-recombinação e a seqüência 3' do local de recombinação- híbrida do núcleo combina o local de recombinação tipo selvagem. Em uma incorporação alternativa, o local de recombinação híbrido pode ser compreendido da
20 região 5' do núcleo de um local *attB* tipo selvagem e da região 3' do núcleo de um local *attP* de recombinação tipo-selvagem ou vive-versa. Outras combinações de tais locais de recombinação híbrida serão evidentes àqueles que têm a habilidade ordinária na técnica, em virtude dos ensinamentos
25 do presente relatório.

Pelo "fragmento do ácido nucléico de interesse" significa qualquer fragmento do ácido nucléico adaptado para a inserção dentro de um genoma. Os exemplos apropriados de fragmentos do ácido nucléico de interesse incluem elementos
30 do promotor, genes terapêuticos, genes do marcador, regiões de controle, fragmentos de traços-produzidos, elementos do ácido nucléico para realizar o rompimento do gene, e similares.

Os métodos de transfecção de células são conhecidos na técnica. Por "transfectada" significa uma alteração em uma célula resultando da absorção do ácido nucléico estrangeiros, geralmente DNA. Uso do termo "transfecção" não é pretendido limitar a introdução do ácido nucléico estrangeiros a nenhum método particular. Os métodos apropriados incluem a infecção viral, conjugação, eletroporação, tecnologia de injeção de partícula, precipitação do fosfato de cálcio, microinjeção direta e similar. A escolha do método é geralmente dependente do tipo de célula transfectada e das condições sob a qual a transfecção está ocorrendo (isto é, in vitro, ex vivo ou in vivo). Um exame geral destes métodos pode ser encontrado Ausubel, et al, Short Protocols in Molecular Biology, 3rd ed., Wiley & Sons, 1995.

Os termos "molécula de ácido nucléico", "polinucleotídeo" são usados permutavelmente e referem-se a uma forma polimérica dos nucleotídeos de qualquer comprimento, tanto deoxiribonucleotídeos ou ribonucleotídeos, ou análogos destes. Os polinucleotídeos podem ter qualquer estrutura tridimensional e podem executar qualquer função, conhecida ou desconhecida. Os exemplos dos polinucleotídeos não limitantes incluem um gene, um fragmento do gene, éxons, íntrons, um RNA mensageiro (mRNA), um RNA transferente, RNA ribossomal, ribozimas, cDNA, polinucleotídeo de recombinação, polinucleotídeo ramificados, plasmídeo, vetores, DNA isolado de qualquer seqüência, regiões de controle, RNA isolado de qualquer seqüência, sondas de ácido nucléico e iniciadores. A molécula do ácido nucléico pode ser linear ou circular.

Um polinucleotídeo é tipicamente composto de uma seqüência específica de quatro bases dos nucleotídeos: adenina (a); cito sina (c); guanina (g); e timina (T)

(uracil (U) para a timina (T) quando o polinucleotídeo for RNA). Assim, o termo a seqüência do polinucleotídeo é a representação alfabética de uma molécula do polinucleotídeo. Esta representação alfabética pode ser entrada em bases de dados em um computador que tem uma unidade do processador central e ser usada para aplicações de bioinformática, tais como, a pesquisa funcional genômica e de homologia.

Uma "seqüência de codificação" ou uma seqüência que "codifica" um polinucleotídeo selecionado é uma molécula de ácido nucléico que é transcrita (no caso de DNA) e traduzida (no caso de mRNA) em um polinucleotídeo, por exemplo, in vivo quando colocado sob o controle de seqüências reguladoras apropriadas (ou de "elementos de controle"). Os limites da seqüência de codificação são determinados tipicamente por um códon inicial no 5' (amino) terminal e um códon de parada da tradução na 3' (carbóxi) terminal. Uma seqüência de codificação pode incluir, mas não sendo limitada a, cDNA de mRNA viral, procariótico ou eucariótico, seqüências de DNA genômico de DNA viral ou procariótico e mesmo seqüências sintéticas de DNA. Uma seqüência da terminação da transcrição pode ser localizada 3' para a seqüência de codificação. Outros "elementos de controle" podem também ser associados com uma seqüência de codificação. Uma seqüência do DNA que codifica um polipeptídeo pode ser aperfeiçoada para a expressão em uma célula selecionada usando os códons preferidos pela célula selecionada para representar a cópia do DNA da seqüência de codificação do polipeptídeo desejado.

"Codificado pelo" refere-se a uma seqüência do ácido nucléico a qual códons para uma seqüência do polipeptídeo, onde a seqüência do polipeptídeo ou uma porção do mesmo contém uma seqüência do aminoácido de pelo menos 3 a 5 aminoácidos, melhor pelo menos 8 a 10 aminoácidos, e

mais preferivelmente pelo menos 15 a 20 aminoácidos de um polipeptídio codificado pela seqüência do ácido nucléico. Também são abrangidas as seqüências do polipeptídio que são imunologicamente identificáveis com um polipeptídio
5 codificado pela seqüência.

"Operativamente ligados" refere-se a um arranjo dos elementos onde os componentes assim que descrito são configurados para executar sua função usual. Assim, um dado promotor que seja operativamente ligado a uma seqüência de
10 codificação (por exemplo, um cassete de expressão do repórter) é capaz de efetuar a expressão da seqüência de codificação quando as enzimas apropriadas estão presentes. O promotor ou outros elementos de controle não precisam ser contínuos com a seqüência de codificação, contanto que
15 funcione para dirigir a expressão do mesmo. Por exemplo, intervindo nas seqüências não-translatadas já transcritas podem estar presentes entre a seqüência do promotor e a seqüência de codificação e a seqüência do promotor pode ainda ser considerada "operativamente ligada" à seqüência de
20 codificação.

Pelo "domínio genômico" é significado uma região genômica que inclua um ou mais, tipicamente uma pluralidade de, éxons, onde os éxons são tipicamente agrupados junto durante a transcrição para produzir um mRNA, onde o mRNA
25 codifique freqüentemente um produto de proteína, por exemplo, uma proteína terapêutica, etc. Em muitas incorporações, o domínio genômico inclui os éxons de um dado gene, e pode também ser referido aqui como um "gene." A modulação da transcrição do domínio genômico conforme aos
30 resultados dos métodos sujeitados a pelo menos modulação de aproximadamente 2 dobras, às vezes pelo menos aproximadamente 5 dobras e às vezes pelo menos cerca de 10 dobras, por exemplo, aumento ou diminuição, da transcrição

do domínio genômico alvejado em comparação a um controle, para aqueles exemplos onde pelo menos alguma transcrição do domínio genômico alvejado ocorre no controle. Por exemplo, nas situações onde um dado domínio genômico é expresso a
5 somente baixos níveis em uma célula alvo não-modificada (usada como um controle), os métodos sujeitos podem ser empregados para obter pelo menos um aumento de 2 dobras na transcrição em comparação a um controle. Os níveis da transcrição podem ser determinados usando qualquer protocolo
10 conveniente, onde os protocolos representativos para determinar níveis da transcrição incluem, mas não são limitados: Hibridação de blot de RNA, PCR-RT, proteção do RNase e similares.

Pelo "constructo de ácido nucléico" significa uma
15 seqüência do ácido nucléico que seja construída para compreender uma ou mais unidades funcionais não encontradas junto na natureza. Os exemplos incluem filamentos circulares, lineares, duplos, moléculas de DNA extracromossômico (plasmídeo), cosmídios (plasmídeos que
20 contêm seqüências COS do fago lambda), genomas virais que compreendem seqüências inativas do ácido nucléico, e similares.

Um "vetor" é capaz de transferir seqüências do gene às células alvo. Tipicamente, "constructo de vetor",
25 "vetor de expressão" e "vetor de transferência do gene" significam qualquer constructo de ácido nucléico capaz de dirigir a expressão de um gene de interesse e que pode transferir seqüências do gene às células alvo. Assim sendo, o termo inclui veículos de clonagem e de expressão, assim
30 como vetores de integração.

Um "cassete de expressão" compreende qualquer constructo de ácido nucléico capaz de dirigir a expressão de uma seqüência do gene/codificação de interesse. Tais

cassetes podem ser construídos em um "constructo de vetor", "vetor de expressão" ou "vetor de transferência de gene" a fim de transferir a expressão do cassete dentro de células alvo. Assim, o termo inclui veículos de clonagem e de
5 expressão, assim como vetores virais.

Na presente invenção, quando uma recombinase for "derivada de um fago", a recombinase não precisa ser produzida explicitamente pelo próprio fago, o fago é considerado simplesmente sendo a fonte original das
10 seqüências da recombinase e da codificação do mesmo. Recombinases podem, por exemplo, ser produzidas recombinantemente ou sinteticamente, pelos métodos conhecidos na técnica, ou alternativamente, as recombinases podem ser purificadas do fago infectadas das culturas
15 bacterianas.

"Substancialmente purificada" refere-se geralmente ao isolamento de uma substância (composto, polinucleotídeo, proteína, polipeptídeo, composição de polipeptídeo) tal que a substância compreende a maioria do percentual da amostra
20 as quais resíduos.

Tipicamente em uma amostra um componente substancialmente purificado compreende 50%, preferivelmente 80%-85%, melhor 90-95% da amostra. As técnicas para purificação de polinucleotídeo e polipeptídios de interesse
25 são conhecidas na técnica e incluem, por exemplo, a cromatografia de troca iônica, cromatografia por afinidade e sedimentação de acordo com a densidade.

A termo "exógeno" é definido aqui como o DNA que é introduzido em uma célula pelo método da presente invenção, tal como com os constructos de DNA definidos aqui. a DNA
30 exógeno pode possuir seqüências idênticas ou diferentes do DNA endógeno presente na célula antes da transfecção.

Por "transgene" ou "elemento transgênico" é

significado uma seqüência artificial introduzida, cromossomicamente integrada do ácido nucléico presente no genoma de um organismo hospedeiro.

O termo "animal transgênico" significa um animal não-humano que tem um elemento transgênico integrado no genoma de uma ou mais células do animal. "Animais transgênicos" como usados aqui abrangem assim os animais que têm toda ou quase todas as células que contêm uma modificação genética (por exemplo, animais inteiramente transgênicos, animais particularmente transgênicos que têm um transgene hereditário) assim como os animais transgênicos, quiméricos, em que um subconjunto das células do animal é modificado para conter o transgene genomicamente integrado.

"Célula alvo" como usada aqui se refere a uma célula na qual uma modificação genética é desejada. As células alvo podem ser isoladas (por exemplo, em cultura) ou em um organismo multicelular (por exemplo, em um blastocisto, em um feto, em um animal pós-natal, e similar). As células alvo de interesse particular na presente aplicação incluem, mas não são limitadas a, células mamíferas cultivadas, incluindo células CRa, e células tronco (por exemplo, células tronco embrionárias (por exemplo, células que têm um fenótipo embrionário da célula tronco), células tronco adultas, células tronco pluripotente, células tronco hematopoiético, células tronco mesenquimal e similares).

DESCRIÇÃO DETALHADA DA INVENÇÃO

O objeto da invenção fornece um sistema de integração local-específico e métodos gerando linhagem eucariótica das células para a produção de proteína. a sistema fornecido inclui um primeiro vetor alvo que integra especificamente o local e um segundo vetor doador que

integra especificamente o local compreendendo um gene de interesse.

Também é fornecida linhagem celular eucariótica produzida pelos métodos e pelos sistemas sujeitos, assim
5 como kits que incluem os sistemas revelados.

Antes que a presente invenção seja descrita, deve ser compreendida que esta invenção não é limitada às incorporações particulares descritas, porque tais podem, naturalmente, variar. Deve também ser compreendido que a
10 terminologia usada aqui é com a finalidade de descrever as incorporações particulares somente, e não é pretendida para limitar, visto que o escopo da presente invenção será limitado somente pelas reivindicações adicionadas.

Onde uma escala dos valores é fornecida,
15 compreende-se que cada valor de intervenção, ao décimo da unidade do menor limite a menos que o contexto ditar claramente de outra maneira, entre os limites superiores e inferiores dessa escala é também especificamente divulgado. Cada escala menor entre qualquer valor indicado ou valor de
20 intervenção em uma escala indicada e em um valor indico ou de intervenção qualquer outro valor de intervenção ou indício na escala indicada é englobado dentro da invenção. Os limites superiores e inferiores destas escalas menores podem independente ser incluídos ou excluídos na escala e
25 cada escala onde, nem ou ambos os limites são incluídos nas escalas menores é também abrangida dentro da invenção, revelado alguns limite especificamente excluído na escala indicada. Onde a escala indicada inclui um ou ambos dos limites, as escalas com exclusão tanto uma ou ambas daqueles
30 limites incluídos são também incluídas na invenção.

A menos que definidos de outra maneira, todos os termos técnicos e científicos usados aqui têm o mesmo significado como geralmente compreendido por uma pessoa com

habilidade ordinária na técnica a qual esta invenção pertence. Embora todos os métodos e materiais similares ou equivalentes a aqueles descritos aqui possam ser usados na prática ou teste da presente invenção, alguns métodos e materiais potenciais e preferidos são agora descrito. Todas as publicações mencionadas aqui são incorporadas aqui pela referência para divulgar e descrever os métodos e/ou os materiais em relação da qual as publicações são mencionadas. Compreende-se que a presente descrição substitui qualquer divulgação de uma publicação incorporada à medida que existe uma contradição. [0093] Deve-se notar que como usado aqui e nas reivindicações adicionadas, a forma singular "um", "uma" e "o" inclui referentes plurais a menos que o contexto dita claramente de outra maneira. Assim, por exemplo, referência a "uma célula" inclui uma pluralidade de tais células e a referência para "o vetor" inclui a referência a um ou mais vetores e equivalentes conhecidos deste para aqueles hábeis na técnica, e assim por diante.

As publicações discutidas aqui são fornecidas unicamente para sua divulgação antes da data de depósito da presente aplicação. Nada aqui deve ser interpretado como uma admissão que a presente invenção não está intitulada para antedata da tal publicação em virtude da invenção prévia. Mais, as datas de publicação fornecidas podem ser diferentes das datas de publicação reais a qual podem precisar ser independentemente confirmada.

VISTA GERAL

Em geral, a presente invenção fornece um primeiro vetor alvo integrando especificamente ao local e um segundo vetor doador integrando especificamente ao local que compreendem um gene de interesse para o uso em gerar as linhagens de células mamíferas capazes da produção da proteína. Os elementos do vetor alvo são selecionados de

modo que uma primeira integrase local-específico uni-
direcional reconheça um primeiro local de recombinação
local-específico presente no vetor alvo e um local de
recombinação local-específico genômico no genoma da célula
5 alvo, resultando na integração do vetor alvo tendo um local-
específico de recombinação alvo para uma segunda integrase
local-específico unidirecional dentro do genoma da célula
alvo.

A linhagem celular resultante que tem um local de
10 recombinação local-específico alvo para a segunda integrase
local-específico unidirecional pode então ser usada
eficientemente para gerar uma linhagem celular capaz de
produzir uma proteína desejada. Um vetor doador que tem um
polinucleotídeo codificando uma proteína de interesse e um
15 local de recombinação local-específico do doador para a
segunda integrase local-específico unidirecional pode ser
introduzido na linhagem celular, resultando na integração do
vetor doador no genoma da célula alvo. Visto que a
integração do transgene pode ser dirigida em uma maneira
20 local-específica, a presente invenção é útil para fornecer a
integração de um transgene em uma posição desejável e evitar
a baixa expressão do transgene devido à integração em uma
posição indesejável.

A invenção será descrita agora em maior detalhe.

25

VETORES

Como observado acima, o sistema inclui um vetor
alvo para integrar um local de recombinação local-específico
dentro do genoma de uma célula alvo e um vetor doador para
integrar um polinucleotídeo que codifica uma proteína de
30 interesse dentro do local de recombinação local-específico
introduzido. Os vetores são tipicamente circulares e podem
também conter marcadores selecionáveis, uma origem de
replicação e outros elementos, tais como, um promotor,

seqüências do promotor-realçador, uma seqüência marcadora de seleção, uma origem de replicação, uma seqüência de elemento indutível, um epítipo de seqüência Tag e similares. Veja, por exemplo, a patente US N° 6.632.672, a divulgação de que
5 é incorporada pela referência aqui em sua totalidade.

A presente invenção fornece um vetor alvo que compreende (a) um primeiro local de recombinação local-específico do vetor capaz de recombinação com um local de recombinação genômico no genoma de uma célula eucariótica na
10 presença de uma primeira recombinase local-específico unidirecional; (b) um segundo local de recombinação local-específico do vetor capaz de recombinação com um local de recombinação local-específico em um vetor doador na presença de uma segunda recombinase local-específico unidirecional;
15 (c) uma primeira porção de um primeiro marcador selecionável (por exemplo, um primeiro marcador selecionável do promotor-menor) adjacente a um lado 3' do local de recombinação local-específico; e (d) um segundo marcador selecionável que seja diferente do primeiro marcador selecionável, e a primeira
20 recombinase local-específico unidirecional seja diferente da segunda recombinase local-específico unidirecional. Um vetor alvo exemplificativo é fornecido na FIG. 1.

A presente invenção também fornece um vetor doador que compreende (a) um local múltiplo de clonagem; (b) um
25 local de recombinação local-específico do doador que seja capaz de recombinação com o local de recombinação local-específico do segundo vetor do vetor alvo na presença de uma segunda recombinase local-específico unidirecional; e (c)
uma segunda porção de um primeiro marcador selecionável (por exemplo, promotor) adjacente ao lado 5' do local de
30 recombinação local-específico do doador. Em determinadas incorporações, o vetor doador também compreende um polinucleotídeo que codifica uma proteína de interesse

presente no local múltiplo de clonagem. Um vetor doador exemplificativo é fornecido na FIG. 2.

Duas famílias principais de recombinases local-específicos unidirecionais das bactérias e dos fermentos unicelular foram descritas: a família da recombinase de integrase ou de tirosine inclui aCre, Flp, R e a integrase lambda (Argos, et al, EMBO 1. 5:433-440, (1986)) e a família da recombinase resolvase/invertase ou serina incluindo algumas integrases do fago, tais como, aqueles dos fago ϕ C31, R4, e TP901-1 (Hallet e Sherratt, Microbiol de FEMS. Rev. 21: 157-178 (1997)). Para uma descrição suplementar de recombinases local-específicos apropriados, veja a patente U.S. Nº 6,632,672 e publicação de patente U.S. Nº 20030050258, as divulgações de que são incorporados aqui pela referência em suas totalidades.

Em determinadas incorporações, a recombinase local-específico unidirecional é uma serina integrase. As serina integrases que podem ser úteis para recombinação in vitro e in vivo inclui, mas não são limitados a, integrases dos fagos ϕ C31, R4, TP901-1, ϕ BTI, Bxb1, RV-I, AI 18, U153, e ϕ FCI, assim como outro na grande família de serina integras e (Gregory, Till and Smith, J. Bacteriol, 185:5320-5323 (2003); Groth and Calos, 1. Mol. Biol. 335:667-678 (2004); Groth et al. PNAS 97:5995-6000 (2000); Olivares, Hollis and Calos, Gene 278:167-176 (2001); Smith and Thorpe, Molec. Microbiol, 4:122-129 (2002); Stoll, Ginsberg and Calos, 1. Bacteriol, 184:3657-3663 (2002)). Além do que estas integrases do tipo selvagem, integrases alteradas que carregam mutações foi produzido (Scilimenti, Thyagarajan e Calos, NAR, 29:5044 - 5051 (2001)). Estas integrases podem ter alterado a atividade ou a especificidade comparada ao tipo-selvagem e são também úteis para reação de recombinação in vitro e a reação de integração no genoma eucariótico.

Em incorporações representativas, a primeira recombina-se local-específico unidirecional e a segunda recombina-se local-específico uni direcional são diferentes. Cada recombina-se local-específico unidirecional tem os
5 locais de recombinação local-específicos distintos (ATT ou locais acessórios) que não recombine com os locais do acessório de outras recombinações local-específicos unidirecionais. Usando duas recombinações local-específico unidirecionais diferentes em ordem, um para a integração do
10 vetor alvo e então a outra para a integração do vetor doador, não há nenhuma possibilidade para uma recombinação intramolecular não desejada dentro do vetor alvo inicial entre o local do acessório para a integração genômica do vetor alvo e o local do acessório para o uso na integração
15 do vetor doador. É desejável evitar tais eventos intramoleculares de recombinação porque não somente criariam os locais híbridos que não podem integrar-se no genoma da célula alvo, mas também podem conduzir a deleção de elementos importantes da sequência no vetor alvo.

20 Conformemente, as primeira e segunda recombinações local-específico unidirecionais devem ser derivados de fago diferentes, por exemplo, ϕ C31, R4, TP90 1-1, phiBTI, Bxb1, RV-I, A118, U153, e phiFCI, ou podem ser derivados do mesmo fago, mas pelo menos uma da primeira e segunda recombina-se
25 local-específico unidirecional é uma recombina-se local-específico uni direcional alterada como que reconhece um diferente local de recombinação local-específico do que o local de recombinação local-específico reconhecido pelo recombina-se local-específico uni direcional tipo-selvagem
30 correspondente.

Em geral, os locais de recombinação local-específicos reconhecidos por uma recombina-se local-específico em um genoma bacteriano são designados locais de

acessório bacterianos ("attB") e os locais de recombinação de locais- específicos correspondentes presentes no bacteriófago são designados locais acessório do fago ("attP"). Estes locais têm um comprimento mínimo de
 5 aproximadamente 34-40 pares de bases (bp) Groth, C.A., e outros, Proc. Natl. Acad. Sei, USA 97, 5995-6000 (2000)). Estes locais são arranjados tipicamente como segue: AttB compreende uma primeira sequência de DNA attB5', região do do núcleo, e uma segunda sequência de DNA attB3' de ordem
 10 relativa attB5' - região do núcleo-attB3'; o attP compreende uma primeira sequência de

DNA (attP5'), uma região de núcleo, e uma segunda sequência de DNA (attP3') na ordem relativa attP5' - região do núcleo-attP3'.

15 Por exemplo, para o attP do fago ϕ C31 (o local acessório do fago), a região de núcleo é 5'-TTG-3' seqüências flanqueada de cada lado são representadas aqui como attP5' e attP3' , a estrutura do local do recombinação do attP é, conformemente, attP5'-TTG-attP3'.
 20 Correspondentemente, para o local alvo (attB) genômico bacteriano nativo a região do núcleo é 5'-TTG-3' , e as seqüências flanqueadas de cada lado são representadas aqui como attB5' e attB3', a estrutura do local de recombinação do attB é, conformemente, attB5' - TTG-attB3'.

25 Como os locais attB e attP são seqüências diferentes, a recombinação conduz a um local de recombinação local-específico híbrido (designado attL ou attR para esquerdo e direito) que nem é uma seqüência attB ou nem uma seqüência attP, e é funcionalmente não-reconhecida como um
 30 local de recombinação local-específico (por exemplo, attB ou attP) para a recombinase local-específico unidirecional relevante, assim removendo a possibilidade que a recombinase local-específico unidirecional catalisará uma segunda reação

de recombinação entre o attL e o attR que inverteriam a primeira reação de recombinação. Por exemplo, após um único-local, mediado por ϕ C31 integrase, evento de recombinação ocorre no resultado de que é o seguinte produto de recombinação: attB5'-TTG-attP3' {seqüências do vetor ϕ C31} attP5'-TTG-attB3'. Tipicamente, depois da recombinação os locais de recombinação pós-recombinação já podem não atuar como um substrato para a ϕ C31 recombinase. Isto resulta na integração estável com pouca ou nenhuma excisão mediada por recombinase.

O local de recombinação nativo tem sido encontrado para existir nos genomas de uma variedade de organismos, onde o local de recombinação nativo não tem necessariamente uma seqüência de nucleotídeos idêntica a seqüências de recombinação do tipo-selvagem (para uma dada recombinase); mas tais locais de recombinação nativos são, todavia suficientes para promover a recombinação mediada pela recombinase. Tais seqüências do local de recombinação são referidas aqui como "seqüências de pseudo-recombinação" Para uma dada recombinase, uma seqüência de pseudo-recombinação é funcional equivalente a uma seqüência de recombinação do tipo selvagem, ocorre em um organismo à exceção daquele em que a recombinase é encontrada na natureza, e pode ter a variação da seqüência relativa a seqüência de recombinação do tipo selvagem.

A identificação de seqüências de pseudo-recombinação pode ser realizada, por exemplo, usando o alinhamento da seqüência e a análise, onde a seqüência em questão é o local de recombinação de interesse (por exemplo, attP e/ou attB).

O genoma de uma célula alvo pode ser procurado por seqüências tendo identidade de seqüência ao local de recombinação selecionado para uma dada recombinase, por

exemplo, *attP* e/ou *attB* de ϕ C31 ou R4. As bases de dados da sequência do ácido nucléico, por exemplo, podem ser procuradas pelo computador. O algoritmo dos moldes-fornecidos do Wisconsin software Package Version 9.0 desenvolvida pelo Genetics computer Group (GCG; Madison, Wis.), é um exemplo do programa usado para selecionar todas as seqüências na base de dados do GenBank (Benson et al, 1998, Nucleic Acids Res. 26, 1-7). Neste aspecto, ao selecionar locais de pseudo-recombinação em uma célula alvo, as seqüências genômicas da célula alvo podem ser procuradas por locais de pseudo-recombinação apropriados usando as seqüências *attP* ou *attB* associadas com uma recombinase particular ou recombinase alterada. Os tamanhos funcionais e a quantidade de heterogeneidade que pode ser tolerada nestas seqüências de recombinação podem empiricamente ser avaliados, por exemplo, próximo à eficiência de avaliação da integração de uma construção de escolha de objetivos usando uma recombinase alterada da presente invenção (para métodos exemplificativos de avaliar eventos de integração, veja, WO 00/11155, publicado em 02 de março de 2000).

Os pseudo-locais funcionais podem também ser encontrados empiricamente. Por exemplo, as experiências executadas na sustentação da presente invenção mostraram que após a co-transfecção em células humanas de um plasmídeo que carrega o *attB* ϕ C31 e o gene de resistência à neomicina, junto com um plasmídeo expressando a ϕ C31 integrase, um número elevado de colônias resistentes a neomicina é obtido, comparado aos co-transfecções em que o *attB* ou o gene de integrase foram omitidos. A maioria destas colônias refletiram a integração em locais *attP* pseudo nativos. Tais locais são recuperados, por exemplo, pelo resgate do plasmídeo e analisados a nível da seqüência de DNA, produzindo, por exemplo, a seqüência de DNA de um local

pseudo *attP* do genoma humano. Este método empírico para a identificação dos pseudos-locais pode ser usado, mesmo se um conhecimento detalhado dos locais de reconhecimento do recombina-se e a natureza do recombina-se que liga a eles são desconhecidos.

Em algumas incorporações, o primeiro local de recombinação do vetor do vetor alvo é um local de recombinação genômico bacteriano (*attB*) ou um local de recombinação genômico do fago (*attP*) reconhecido por uma primeira recombina-se local-específico. Em tais incorporações, o local de recombinação genômico presente no genoma da célula alvo é um local correspondente a pseudo-recombinação. Por exemplo, onde o primeiro local de recombinação do vetor do vetor alvo é um local de recombinação genômico bacteriano (*attB*), o local pseudo-recombinação genômico presente no genoma da célula alvo é um local de recombinação genômico do pseudo-fago (pseudo-*attP*). Do mesmo modo, onde o primeiro local de recombinação do vetor do vetor alvo é um local de recombinação genômica do fago (*attP*), o local de pseudo-recombinação genômica presente no genoma da célula alvo é um local de recombinação genômico pseudo-bacteriano (pseudo-*attB*).

Algumas recombina-ses local-específicos unidirecionais integram preferencial dentro de locais de recombinação pseudo-bacterianos (por exemplo, pseudo-*attB*), um pouco do que locais de recombinação do pseudo-fago (por exemplo, pseudo-*attP*). Nestes casos, o vetor alvo carrega um local de recombinação do fago (*attP*) e integrá-lo-á dentro do local pseudo-*attB*. Os exemplos das enzimas com esta preferência são *phiBTI* integras e *A118* integrase. Em tais incorporações, o primeiro local de recombinação do vetor do vetor alvo é um local *attP* e o local de recombinação genômico no genoma da célula alvo é um local pseudo-*attB*.

Outras recombinases unidirecionais, local-específicos, tais como ϕ C31 e R4, preferem integrar em locais acessório do pseudo-fago (locais pseudo-*attP*) preferencialmente do que locais de recombinação pseudo-bacterianos (locais pseudo-*attB*), então o vetor alvo carrega um local *attB* e integrá-lo-á em um local pseudo-*attP* (Groth et al., 2000; Olivares, Hollis e Calos 2001). Em tais incorporações, o primeiro local de recombinação do vetor do vetor alvo é um local *attB* e o local de recombinação genômico no genoma da célula alvo é um local pseudo-*attP*.

Além disso, em determinadas incorporações, o primeiro local de recombinação do vetor do vetor alvo é um local de pseudo-recombinação e o local de recombinação atual no genoma da célula de alvo é um local correspondente do pseudo-recombinação reconhecido por uma primeira recombinase local-específico. Por exemplo, onde o local de recombinação do vetor do vetor alvo é um local de recombinação genômico pseudo-bacteriano (pseudo-*attB*), o local de pseudo-recombinação presente no genoma da célula alvo é um local de recombinação genômico pseudo-fago (pseudo-*attP*). Do mesmo modo, onde o primeiro local de recombinação do vetor do vetor alvo é um local de recombinação genômico do pseudo-fago (pseudo-*attP*), o local de pseudo-recombinação presente no genoma da célula alvo é um local de recombinação genômico pseudo-bacteriano (pseudo-*attB*).

Em algumas incorporações, o segundo local de recombinação do vetor do vetor alvo é um local de recombinação genômico bacteriano (*attB*) ou um local de recombinação genômico do fago (*attP*) reconhecido por uma segunda recombinase local-específico. Em tais incorporações, o local de recombinação do doador no vetor doador é um local de recombinação correspondente. Por exemplo, nas incorporações onde o segundo local de recombinação do vetor

do vetor alvo é um local de recombinação genômico bacteriano (*attB*), o local de recombinação do doador presente no vetor doador é um local de recombinação genômico do fago (*attP*). Do mesmo modo, onde o segundo local de recombinação do vetor do vetor alvo é um local de recombinação genômico do fago (*attP*), o local de recombinação do doador presente no vetor doador é um local de recombinação genômico bacteriano (*attB*).

Como observado acima, o vetor alvo inclui uma primeira porção de um primeiro marcador selecionável adjacente a um lado 3' do segundo local de recombinação do vetor e o vetor doador inclui uma segunda porção do primeiro marcador selecionável adjacente a um lado 5' do local de recombinação do doador. Na presença de uma segunda recombinase local-específico unidirecional o segundo local de recombinação do vetor no vetor alvo recombina com o local de recombinação do doador presente no vetor doador para gerar um local de recombinação híbrido. Como um resultado da recombinação, a primeira porção do marcador selecionável no vetor alvo e a segunda porção do marcador selecionável no vetor doador são levadas dentro de grande proximidade para fornecer um primeiro marcador selecionável funcional reconstituído. Portanto, a seleção que usa o primeiro marcador de seleção pode ser usada para selecionar eventos bem sucedidos de recombinação entre um vetor alvo presente no genoma de uma célula alvo e o vetor doador que tem um polinucleotídeo codificando uma proteína de interesse.

Em uma incorporação do primeiro gene marcador selecionável reconstituído do promotor é fornecido pelo vetor doador e uma região de codificação para um sinal selecionável do gene e de poliadenilação do marcador é fornecida pelo vetor alvo. Em outra incorporação do gene marcador selecionável reconstituído o vetor doador pode

conter um promotor, uma parte N-terminal da região de codificação, e metade 5' de um íntron, enquanto o vetor alvo puder conter a metade 3' de um íntron, a parte C-terminal da região de codificação e um sinal de poliadenilação. Em uma
5 incorporação suplementar do gene marcador selecionável reconstituído o vetor doador pode conter um promotor e parte N-terminal da região de codificação enquanto o vetor alvo puder conter a parte C-terminal da região de codificação e um sinal de poliadenilação. Ainda em outra incorporação, o
10 vetor doador inclui um promotor e o vetor alvo inclui um marcador selecionável do promotor-menor. Em todas estas incorporações do gene marcador selecionável reconstituído, a característica chave é que os elementos genéticos presente no alvo separado e os vetores doadores são incapazes de
15 conferenciar resistência à droga independente de outro. Entretanto quando o vetor doador é integrado no vetor alvo um cassete de expressão funcional completo do gene é montada nas células as quais contêm tal configuração serão resistentes à droga que é usada para selecionar para a
20 presença do gene marcador selecionável reconstituído.

As seqüências do promotor e do promotor-realçador são as seqüências do DNA a qual a RNA polimerase liga-se e inicia a transcrição. O promotor determina a polaridade do transcrito pela especificidade a qual o filamento será
25 transcrito. Os promotores bacterianos consistem nas seqüências consenso, -35 e -10 nucleotídeos, relativo ao começo transcricional, que são limitados por um fator sigma e por uma RNA polimerase específica.

Promotores eucarióticos são mais complexos. A
30 maioria dos promotores eucarióticos utilizados em vetores de expressão é transcrita pela RNA polimerase II. Fatores de transcrição gerais (GTFS) primeiro ligam seqüências específicas perto do local do início da transcrição e

recrutam então o emperramento da RNA polimerase II. Além disso, estes mínimos elementos de promotor, pequena seqüência de elementos são reconhecidos especificamente pela ligação modular de DNA, proteínas de ativação-trans (por exemplo, AP-1, SP-1) que regulam a atividade de um dado promotor. Os promotores virais servem para a mesma função que os promotores bacterianos ou eucarióticos e requerem uma RNA polimerase promotor-específico no transporte (por exemplo, RNA polimerase do bacteriófago T7 nas bactérias) ou recrutam os fatores celulares e RNA polimerase II (em células eucarióticas). Os promotores virais (por exemplo, promotores SV40, RSV e CMV) podem ser preferidos porque são geralmente particularmente promotores fortes.

Os promotores podem ser, além disso, constitutivos ou reguláveis. Os promotores constitutivos expressam constantemente o gene de interesse. Ao contrário, genes expressos dos promotores reguláveis (isto é, de repressão ou indução) de interesse somente sob determinadas condições que podem ser controladas. Os elementos de repressão são os elementos da seqüência de DNA que atuam conjuntamente com promotores e ligam os repressores (por exemplo, sistema do repressor de lacO/lacIq em E. Coli). Os elementos induzíveis são os elementos da seqüência de DNA que atuam conjuntamente com promotores e ligam os indutores (por exemplo, sistema indutor gal/gal4 da levedura). Em um ou outro caso, a transcrição é virtualmente "desligada" até o promotor seja reprimido ou induzido pela alteração de uma condição no ambiente (por exemplo, adição de IPTG ao sistema lacO/lacIq ou adição de galactose ao sistema gal/gal4), em que o ponto de transcrição é "ligado".

Outro tipo de promotor regulado é uma "repressão" em que um gene é expresso inicialmente e pode então ser desligado alternando uma condição ambiental. Em sistemas de

repressão a transcrição é constitutivamente ligada até que o repressor ligue-se a uma pequena molécula reguladora em que o ponto de transcrição seja "desligado". Um exemplo deste tipo de promotor é o sistema de tetraciclina/repressor de tetraciclina. Neste sistema quando a tetraciclina liga-se ao repressor de tetraciclina, o repressor liga-se a um elemento do DNA no promotor e desliga a expressão de gene.

Os exemplos de promotores procarióticos constitutivos incluem o promotor λ do bacteriófago, promotor da sequência do gene β -lactamase de pBR322, promotor CAT da sequência do gene do cloranfenicol cetil transferase de pPR325, e similares.

Os exemplos de promotores procarióticos induzíveis incluem os principais promotores direito e esquerdo do bacteriófago (PL e PR), promotores trp, recA, lacZ, AraC e gal de E. Coli, a α -amilase (Ulmanen et al., J. Bacteriol. 162: 176-182, 1985) e promotores específicos sigma-28 de B. subtilis (Gilman et al., Gene sequence 32:11-20(1984)), os promotores dos bacteriófagos do Bacillus (Gryczan, In: The Molecular Biology of the Bacilli, Academic Press, Inc., NY (1982)), Promotores de Streptomyces (Ward et al., Mol. Gen. Genet. 203:468-478, 1986), e similares. Os promotores procarióticos exemplares são revistos por Glick (1. Ind. Microbiol. 1:277-282, 1987); Cienfuegos (Biochimie 68:505-516, 1986); e Gottesman (Ann. Rev. Genet. 18:415-442, 1984).

Os promotores eucarióticos constitutivos exemplificativos incluem, mas não são limitados aos seguintes: o promotor da sequência do gene de metalotioneína I de rato (Hamer et al., 1. Mol. Appl. Gen. 1:273-288, 1982); o promotor TK do vírus da herpes (McKnight, Cell 31:355-365, 1982); promotor SV40 adiantado (Benoit et al., Nature (London) 290:304-310, 1981); promotor da

seqüência do gene gall da levedura (Johnston e outros, Proc. Natl. Acad. Sei, (USA) 79:6971 - 6975, 1982); Silver *Et al.*, Proc, Natl. Acad. Sci. (USA) 81:5951 -5955, 1984), promotor CMV, promotor EF-1.

5 Os exemplos de promotores eucarióticos induzíveis incluem, mas não são limitados aos seguintes: promotores responsáveis-ecdisona, promotor tetraciclina-responsável, promotores regulados pelo "dimerizeres" que traz duas porções de um fator de transcrição junto, promotores
10 hormonas-responsáveis, promotores progesterona-responsáveis, promotores riboswitch-reguláveis, promotores antibiótico-regulados, promotores acetaldeído-regulados e similares.

Alguns promotores regulados podem mediar a repressão e a ativação. Por exemplo, no sistema RheoSwitch
15 uma proteína (o RheoReceptor) liga-se a um elemento do DNA (UAS, seqüência de ativação a montante) no promotor e media a repressão. Entretanto na presença de certo indutores como ecdisona uma outra proteína (RheoAtivador) ligará ao indutor. O indutor-liga o Rheoativador que é capaz de ligar-
20 se ao RheoReceptor ligado ao DNA. O RheoReceptor/indutor RheoActivator é então capazes de ativar a expressão de gene.

Os genes marcadores selecionáveis comuns incluem aqueles para a resistência aos antibióticos, tais como, ampicilina, tetraciclina, canamicina, bleomicina,
25 estreptomicina, higromicina, neomicina, puromicina, G418, bleomicina, blasticidina, Zeocina" e similares. Os genes auxotróficos selecionáveis incluem, por exemplo, hisD, que permite o crescimento em meios livres de histidine na presença do histidinol.

30 Um elemento suplementar útil em um vetor de expressão é uma origem de replicação. As origens de replicação são os segmentos originais do DNA que contêm múltiplas seqüências repetidas curtas que são reconhecidas

por proteínas de ligação de origem multiméricas e que possuem um papel chave na montagem nas enzimas de replicação do DNA no local de origem. As origens de replicação apropriadas para o uso nos vetores de expressão empregados aqui incluem a origem *oriC* de *E. Coli*, *CoIE1* de plasmídio, 2/-1 e o *ARS* (ambos útil em sistemas da levedura), *sfl*, *SV40*, *oriP* de *EBV* (útil em sistemas eucarióticos, tais como, um sistema mamífero), e similares.

Como observado acima, o vetor doador inclui múltiplos locais de clonagem ou um poliligante. Um múltiplo local de clonagem ou um poliligante é um DNA sintético que codifica uma série de locais de reconhecimento da endonuclease de restrição introduzidos em um vetor doador e permitem a clonagem conveniente dos polinucleotídeos que codificam a proteína de interesse no vetor doador em uma posição específica.

Proteínas úteis que podem ser produzidas pelas composições e os métodos da invenção são, por exemplo, as enzimas que podem ser usadas para a produção de nutrientes e executando reações enzimáticas na química, ou polipeptídios que são úteis e disponíveis como nutrientes ou para o tratamento de doenças humanas ou animais ou para a prevenção destes, por exemplo, hormônios, polipeptídios com atividade imunomodulatórias, propriedades antiviral e/ou antitumoral (por exemplo, maspin), anticorpos, antígenos virais, vacinas, fatores de coagulação, inibidores de enzima, gêneros alimentícios, e similares. Outros polipeptídios úteis que podem ser produzidos pelos métodos da invenção são, por exemplo, aqueles codificando hormônios, tais como, secretina, timosina, relaxina, hormônio luteinizante, hormônio da paratireóide, adrenocorticotropina, hormônio de estimulação, β -lipotropina, urogastrona ou insulina, fatores de crescimento, tais como, fator de crescimento epidérmico,

fator de crescimento semelhante-insulina (IGF), por exemplo, IGF- I e IGF-II, fator de crescimento da célula mastro, fator de crescimento do nervo, fator neurotrófico derivada da linhagem da célula glial (GDNF) ou fator de crescimento de transformação (TGF), como TGF- α ou TGF- β (por exemplo, TGF- β 1, β 2 ou β 3), hormônio de crescimento, tal como, hormônios de crescimento humanos ou bovinos, interleucinas, tais como, interleucina-1 ou -2, fator inibitório da migração humana do macrófago (FIM), interferons, tais como, u-interferon humano, por exemplo, interferon-uA, aB, aD ou nF, n-interferon, γ -interferon ou um interferon híbrido, um interferon, por exemplo, aA-aD- ou aB-aD-híbrido, interferon híbrido BDBB especial, inibidores de protease, tais como, o al-antitripsina, SLPI, al-antiquimotripsina, inibidor C1, antígenos do vírus de hepatite, tais como, o antígeno da superfície ou do núcleo do vírus da hepatite B ou o antígeno do vírus da hepatite A, ou da hepatite não-A e não-B (isto é, antígeno do vírus da hepatite C), ativadores plasminogênicos, tais como, o ativador plasminogênico de tecido ou uroquinase, fatores de necrose tumoral (por exemplo, TNF- α ou TNF- β), somatostatina, renina, β -endorfina, imunoglobulina, tais como as cadeias leves e/ou pesadas da imunoglobulina A, D, E, G ou M ou imunoglobulina híbridas de humano-rato, fatores de ligação da imunoglobulina, tais como, fator de ligação da imunoglobulina E, por exemplo sCD23 e similares, calcitonina, peptídeo relacionado a calcitonina humano, fatores de coagulação do sangue, tais como, o fator IX ou VIIIC, eritropoietina, eglina, tal como, eglin C, desulfatohirudina, tal como, variação HV1, HV2 ou PA do desulfatohirudina, dismutase humano do superoxide, quinase timidina viral, β -lactamase, glicose isomerase, proteínas de transporte, tais como, proteínas do plasma humano, por

exemplo, albumina de soro e transferrina. As proteínas de fusão acima podem também ser produzidas pelos métodos da invenção.

Além disso, os níveis de uma proteína de interesse expressada podem ser aumentados pela amplificação do vetor (veja Bebbington and Hentschel, "The use of vectors based on gene amplification for the expression of cloned genes in mammalian cells in "DNA cloning", Vol. 3, Academic Press, New York, 1987). Quando um marcador no sistema do vetor que expressa uma proteína é amplificável, um aumento no nível de um inibidor desse marcador, quando presente na cultura de célula hospedeiro, aumentará o número de cópias do gene do marcador. Desde que a região amplificada esteja associada com o gene da proteína-codificação, a produção da proteína de interesse aumentará concomitante (Crouse et al., 1983, MoI. Cell. Biol, 3:257). Um sistema exemplificativo da amplificação inclui, mas não é limitado a, dihidrofolato reductase (DHFR), que conferencia resistência a seu inibidor de metotrexato. Outros sistemas apropriados da amplificação incluem, mas não são limitados a, glutamina sintetase (e seu inibidor metionina sulfoximina), timidina sintase (e seu inibidor 5-fluoruridina), carbamil-P-sintetase/aspartato transcarbamilase/dihidro-orotase (e seu inibidor N(fosfonacetil) - L-aspartato), ribonucleosídeo reductase (e seu inibidor hidroxiiurea), ornitina decarboxilase (e seu inibidor ornitina diflurometil), adenosina deaminase (e seu inibidor deoxicoformicina), e similar.

Cada um destes sistemas exige o uso de uma linhagem celular que seja deficiente no gene do marcador que é amplificado. Por exemplo, o uso do gene DHFR como um gene amplificável usa uma linhagem celular DHFR-deficiente, tal como, uma célula DHFR-deficiente de CHO (por exemplo, DG44). Métodos são disponíveis para isolar tais linhagens celulares

gene-deficientes do marcador. Um sistema de amplificação de gene que não use linhagem celular gene-deficientes do marcador é um sistema que use a proteína rep (AAV-2) vírus tipo 2 adeno-associado e o local de ligação da proteína rep.

5 [00133] A maioria de genes amplificáveis do marcador pode também ser usada como genes marcadores selecionáveis. Por exemplo, a presença do gene DHFR pode ser selecionada em células deficientes de DHFR usando meios de crescimento da célula que falta a glicina, a timidina e a hipoxantina. A
10 presença do gene da glutamina sintetase pode ser selecionada em células glutamina sintetase-deficientes usando meios que falta a glutamina, e assim por diante. Desse modo um pode assegurar-se de que o gene amplificável do marcador esteja presente a fim de mediar a amplificação do gene,
15 especialmente antes de todos os procedimentos de amplificação do gene.

Conformemente, em determinadas incorporações, o vetor alvo também inclui um polinucleotídeo que codifica o gene marcador selecionável e amplificável DHFR. Um vetor
20 alvo exemplificativo que inclui DHFR é fornecido na FIG. 5. Em tais incorporações, o vetor alvo que é integrado no genoma da célula alvo é amplificado usando concentrações crescentes de metotrexato. Desde que o vetor alvo compreenda um segundo local de recombinase local-específico para a
25 integração do vetor doador, a amplificação da sequência do vetor alvo no genoma da célula alvo conduzirá à amplificação do número de segundos locais da recombinase local-específicos presentes no genoma da célula alvo. Isto fornece uma pluralidade de posições em que o vetor doador pode
30 integrar-se.

Em outras incorporações, o vetor de expressão doador é integrado opcionalmente no vetor alvo-DHFR antes da exposição às concentrações crescentes de metotrexato. Em

tais incorporações, o gene que codifica a proteína de interesse situada no vetor de expressão doador tornar-se-á ligada próxima (dentro de -4.000 pares de bases) ao gene de DHFR situado no vetor alvo-DHFR. Em consequência da
5 exposição do metotrexato, o número de cópia do gene que codifica a proteína de interesse será amplificado pela seleção das células em concentrações crescentes de metotrexato.

Em um método tradicional de amplificação do gene,
10 o gene de DHFR é co-transfectado com um vetor de expressão da proteína em tal excesso (geralmente 100-dobras) que toma-se geralmente ligado ao vetor da expressão da proteína, mas somente depois da fragmentação e da ligação de ambos os vetores por mecanismos celulares. Ao contrário de um método
15 tradicional de amplificação do gene, este método opcional fornece a vantagem de poder controlar o arranjo, a composição e a posição do gene DHFR relativo ao gene de expressão da proteína antes da exposição ao metotrexato. Em consequência, isto fornecerá uma frequência mais elevada da
20 amplificação e do resultado bem sucedidos do gene em poucas linhagens celulares instáveis que não expressam o gene de interesse ou perda da expressão do gene de interesse sobre o tempo.

Alternativamente, em outras incorporações, o vetor
25 doador que tem o polinucleotídeo codificando a proteína de interesse inclui mais um polinucleotídeo que codifica o gene marcador selecionável e amplificável DHFR. Um vetor doador exemplificativo que inclui DHFR é fornecido na FIG. 6. Em tais incorporações, a sequência inteira que é integrada no
30 genoma, incluindo o polinucleotídeo que codifica a proteína de interesse é amplificada usando concentrações crescentes de metotrexato.

Em determinadas incorporações, o vetor doador

também inclui um local interno da entrada de ribossomo (IRES) posicionado entre o local do início da transcrição e o códon da iniciação da tradução da proteína de interesse. Um vetor doador exemplificativo incluindo IRES é fornecido na FIG. 7. Tais vetores podem permitir a expressão de gene aumentada se eles são realçadores translacionais ou podem também permitir a produção de múltiplas proteínas de interesse em um único transcrito, contanto que os IRES estejam localizados em 5' a cada região de codificação de interesse.

Os vetores descritos aqui podem ser construídos utilizando as metodologias conhecidas na técnica da biologia molecular (veja, por exemplo, Ausubel ou Maniatis) em virtude dos ensinamentos do relatório. Um método exemplificativo de obter polinucleotídeo, incluindo seqüências reguladoras apropriadas (por exemplo, promotores) é PCR. Os procedimentos gerais para o PCR são ensinados em MacPherson e outros, PCR: A PRACTICAL APPROACH, (IRL Press at Oxford University Press, (1991)). As condições do PCR para cada reação da aplicação podem ser empiricamente determinadas. Um número de parâmetros influencia o sucesso de uma reação. Entre estes parâmetros estão a temperatura e o tempo de anelamento, tempo de extensão, concentração de Mg²⁺ e de ATP, pH, e a concentração relativa dos iniciadores, moldes e de deoxiribonucleotídeos. Após a amplificação, os fragmentos resultantes podem ser detectados pela eletroforese em gel de agarose seguida pela visualização com uma coloração de brometo de etídio e a iluminação ultravioleta.

30

MÉTODOS

A presente invenção também fornece métodos de gerar uma linhagem celular que produz uma proteína de interesse pelo local que integra especificamente um

polinucleotídeo que codifica a proteína de interesse no genoma de uma célula eucariótica, tal como, uma célula mamífera. Geralmente o método envolve primeiramente introduzir um vetor alvo como descrito aqui em uma célula eucariótica utilizando uma primeira recombinase local-específico uni direcional e mantendo a célula sob condições suficientes para um evento de recombinação mediado pela primeira recombinase local-específico unidirecional entre o primeiro local de recombinação do vetor e o local de recombinação genômico local-específico a fim de integrar o local-específico ao vetor alvo dentro do genoma da célula. Os eventos bem sucedidos de integração do vetor alvo mediado pela primeira recombinase local-específico uni direcional podem ser selecionados usando o gene marcador selecionável presente no vetor alvo.

Um vetor doador que compreende o polinucleotídeo que codifica uma proteína de interesse e de um local de recombinação do doador é introduzido então na célula alvo utilizando uma segunda recombinase local-específico uni direcional. A célula alvo é mantida então sob as condições suficientes para permitir um evento de recombinação mediado pela segunda recombinase local-específico unidirecional a ocorrer. Como resultado, um evento de recombinação entre o local de recombinação do doador e o segundo local de recombinação do vetor do vetor alvo permite a integração local-específica do polinucleotídeo que codifica uma proteína de interesse dentro do genoma da célula. Os eventos bem sucedidos da integração do vetor doador mediado pela segunda recombinase local-específico unidirecional podem ser selecionados usando um primeiro gene marcador selecionável reconstituído. Em uma incorporação do primeiro gene marcador selecionável reconstituído o promotor é fornecido pelo vetor doador e uma região de codificação para um gene marcador

selecionável e um sinal de poliadenilação é fornecida pelo vetor alvo. Em outra incorporação do gene marcador selecionável reconstituído o vetor doador pode conter um promotor, uma parte N-terminal da região de codificação, e a
5 metade 5' de um íntron, quando o vetor alvo puder conter a metade 3' de um íntron, a parte C-terminal da região de codificação e um sinal de poliadenilação. Em uma incorporação suplementar do gene marcador selecionável reconstituído o vetor doador pode conter um promotor e parte
10 N-terminal da região de codificação quando o vetor alvo puder conter a parte C-terminal da região da codificação e um sinal de poliadenilação. Em outra incorporação, o vetor doador inclui um promotor e o vetor alvo inclui um marcador selecionável do promotor-menor. Em todas estas incorporações
15 de gene marcador selecionável reconstituído, a característica chave é que os elementos genéticos presentes no alvo separado e os vetores doadores são incapazes de conferenciar resistência às drogas independentes um do outro. Entretanto quando o vetor doador é integrado no vetor
20 alvo um cassete de expressão funcional completo do gene é montada nas células que contêm tal configuração a qual será resistente à droga que é usada para selecionar para a presença do gene marcador selecionável reconstituído.

Em geral, a interação da integrase local-
25 específica unidirecional com os locais de recombinação local-específico local produz um produto de recombinação que não contém uma sequência que atue como um substrato eficaz para a integrase local-específico unidirecional. Assim, o evento de integração empregado nos métodos sujeitos é
30 unidirecional, com quase nenhuma excisão detectável do ácido nucléico introduzido mediado pela integrase e local-específico unidirecional. Esta característica assegura a maior estabilidade da expressão das proteínas de interesse

comparadas a outros sistemas da integração do que possa ser fornecido por uma recombinase local-específico bidirecional (por exemplo, sistema de integração lox/cre) ou que contém as seqüências diretamente repetidas (por exemplo, repetições 5 terminais longas) o qual pode resultar na deleção dos genes que codificam proteínas de interesse (por exemplo, sistemas de integração do retro vírus ou lentivírus).

Os vetores podem ser introduzidos na célula hospedeira por qualquer dos meios padrões praticados por uma 10 pessoa com habilidade na técnica para produzir uma linhagem celular da invenção. Os vetores do ácido nucléico podem ser obtidos, por exemplo, com lipídios catiônicos (Goddard, et al, Gene Therapy, 4:1231-1236,1997; Gorman, et al, Gene Therapy 4:983-992, 1997; Chadwick, et al, Gene Therapy 15 4:937-942, 1997; Gokhale, et al, Gene Therapy4:1289-1299, 1997; Gao, and Huang, Gene Therapy2:710-722, 1995, que são incorporados pela referência aqui), usando vetores virais (Monahan, et al, Gene Therapy 4:40-49, 1997; ano dera, et al, Blood 91 :30-36, 1998" que são incorporadas pela 20 referência aqui), pela tomada do "DNA despido", meios químico (por exemplo, fosfato de cálcio), meios eletroforéticos, e similares.

As primeira e segunda recombinases local-específico unidirecionais usadas na prática da presente 25 invenção podem ser introduzidas na célula alvo antes, simultaneamente com, ou após a introdução de um vetor alvo ou de um vetor doador. As primeira e segunda recombinases local-específico unidirecionais podem ser introduzidas sob a forma de DNA codificando a recombinase local-específico 30 unidirecional (Olivares, Hollis and Calos, Gene, 278:167-176 (2001); Thyagarajan et al. MCB 21 :3926- 3934 (2001)), ou mRNA que codifica a recombinase local-específico unidirecional (Groth et al. JMB 335:667-678 (2004); Hollis

et al. Repr. BioI. Endocrin. 1 :79 (2003)) ou como a proteína recombinase local-específica unidirecional.

A expressão das primeira e segunda recombinases local-específico unidirecionais é tipicamente desejada para ser transiente. Isto é porque a expressão a longo prazo das recombinases pode promover a recombinação entre os locais pseudo att presente em várias localizações no genoma. Isto poderia conduzir aos rearranjos cromossômicos e eventualmente à morte de célula. Conformem ente, os vetores e os métodos que fornecem a expressão transiente da recombinase são preferidos na prática da presente invenção. Entretanto, a expressão estável das primeira e segunda recombinases local-específico unidirecionais pode ser aceitável se for regulada, por exemplo, pela colocação da expressão da recombinase sob o controle de um promotor regulável (isto é, um promotor cuja expressão possa seletivamente ser induzida ou reprimida).

A introdução das primeira e segunda recombinases local-específico unidirecionais como proteínas tem diversas vantagens. A proteína tem uma meia-vida curta, assim que a exposição das células a recombinase local-específico unidirecional é limitada pelo tempo. Além disso, não há nenhuma possibilidade da integração do gene da recombinase local-específico unidirecional no genoma. As limitações com transcrição ou tradução de recombinase local-específico unidirecional são evitadas e a cinética de reação pode ser mais rápida. A introdução de proteína em células é geralmente menos tóxica do que a introdução do DNA. Conseqüentemente, a introdução de uma recombinase local-específico unidirecional do fago nas células eucariótica como uma proteína pode ser preferível.

As proteínas, tais como, a recombinase local-específico unidirecional do fago podem ser introduzidas em

células por muitos meios, incluindo a eletroporação, transportadores de peptídeo (Siprashvili, Reuter e Khavari, MoI. Ther., 9:721 - 728 (2004)), ou o acessório de domínios do transducção de proteína, tais como, aqueles derivados da
5 proteína VP22 do Vírus Simples da Herpes, pepetídeos derivados-antenapedia, vários peptídeos ricos em arginina, ou proteína tat do Vírus da Imunodeficiência Humana. O DNA ou o RNA que codificam um recombinase local-específico unidirecional podem também ser introduzidos em células por
10 muitos meios, incluindo eletroporação, complexação com agentes químicos, tais como, interação eletrostática com moléculas de transportador ou endocitoses.

As células apropriadas para o uso com os métodos sujeitos da presente invenção são geralmente qualquer célula
15 eucariótica superiora, tal como, células mamíferas e células de levedura. Em algumas incorporações, as células são uma linhagem celular mamífera facilmente manipulada, facilmente cultivada. Em outras incorporações, as células são uma linhagem celular facilmente manipulada, facilmente cultivada
20 de levedura. As células apropriadas que são capazes de expressar moléculas de DNA de recombinação, incluem, mas não são limitadas a, células mamíferas, tais como, uma célula de roedor, tal como, células (CHO) do ovário de hamster (CHO), células BHK, células de rato que incluem células das células
25 SP210 e do mieloma de NS-O, células de primata, tais como, COS e células Vera, células MDCK, células BRL 3A, hibridomas, células tumorais, células preliminares não imortalizadas, células humanas, tais como, W138, HepG2, HeLa, REK293, HT1080, ou PER.C6™, e similares.

30 Em algumas incorporações, a célula é uma célula PER.C6™. Em outras incorporações, a célula é uma célula CRO ou uma célula dihidrofolato deficiente de reductase, tal como, as células DG44. As células CRO transformaram-se numa

rotina e um sistema de produção conveniente para a geração de proteínas biofarmacêuticas e proteínas para finalidades diagnósticas. Um número de características faz dessas células CRO apropriada como uma célula hospedeira. Os níveis da produção que podem ser alcançados em células CRO são extremamente elevados. A linhagem celular fornece um sistema de produção seguro, que possa estar livre de agentes infecciosos e de partículas virais infecciosas. As células CRO extensivamente foram caracterizadas sendo capazes do crescimento na suspensão até altas densidades de alcance nos bioreatores, usando meios de cultura do soro livre e um mutante deficiente de DRFR de células de CRO (DG-44 clone. Urlaub et al., Cell. 33(2):405-12 (1983)) foi tomado para obter um sistema fácil de seleção e de amplificação introduzindo um gene exógeno de DRFR, selecionando para sua presença, e depois disso executando uma amplificação bem-controlada, por etapas do gene de DRFR e todos os genes de interesse ligados usando concentrações crescentes de metotrexato.

LINHAGEM CELULAR

A presente invenção também fornece linhagens celulares geradas integrando o vetor alvo descrito acima no local de recombinação genômico da célula alvo. Conformemente, as células sujeitas têm um cassete de polinucleotídeo integrado genomicamente compreendendo um primeiro local de recombinação híbrido e um segundo local de recombinação híbrido que flanqueiam um local de recombinação do vetor que recombina com um local de recombinação do doador na presença de uma recombinase local-específico unidirecional; um primeiro marcador selecionável do promotor-menor adjacente ao local de recombinação do vetor na extremidade 3' e um segundo marcador selecionável que seja diferente do primeiro marcador selecionável.

Em algumas incorporações, o local de recombinação do vetor é um local de recombinação genômico bacteriano (*attB*) ou um local de recombinação genômico do fago (*attP*). Em algumas incorporações, o local de recombinação do doador
5 é um local de recombinação genômico bacteriano (*attB*) ou um local de recombinação genômico do fago (*attP*). Em algumas incorporações, a recombinase local-específico unidirecional é uma recombinase do fago ϕ C31, um recombinase do fago TP901-1 ou um recombinase do fago R4. Em algumas
10 incorporações, a célula mamífera é uma célula de roedor. Em outras incorporações, a célula mamífera é uma célula CRO. Em outras incorporações, a célula mamífera é uma célula PER.C6™.

KITS

15 Também são fornecidos pelo objeto da invenção os kits para praticar os métodos sujeitos, como descritos acima. Em determinadas incorporações, os kits revelados incluem pelo menos um ou mais de, e geralmente todo o vetor alvo e vetor doador como descrito acima. Em algumas
20 incorporações, os kits também incluem uma primeira e segunda componente recombinase local-específico unidirecional, onde o componente recombinase possa ser fornecido em qualquer forma apropriada (por exemplo, como uma proteína formulada para a introdução em uma célula alvo ou em um vetor
25 recombinase que forneça para a expressão da recombinase desejado depois da introdução dentro da célula alvo).
[00153] Em outras incorporações, os kits revelados incluem pelo menos um ou mais de, e geralmente qualquer linhagem celular isolada que tem um vetor alvo integrado e um vetor
30 doador como descrito acima. Em algumas incorporações, os kits também incluem uma primeira e segunda componente recombinase local-específico unidirecional, onde o componente recombinase possa ser fornecido em qualquer forma

apropriada (por exemplo, como uma proteína formulada para a introdução em uma célula alvo ou em um vetor recombinase que forneça a expressão da recombinase desejado depois da introdução dentro da célula de alvo).

5 Outros componentes opcionais do kit incluem enzimas de restrição, plasmídeo de controle, tampões, materiais para a introdução de vetores em células, etc. Os vários componentes do kit podem estar presentes em recipientes separados ou determinados componentes
10 compatíveis podem ser pré-combinados em um único recipiente, como desejado.

 Além disso, para os componentes acima mencionados, os kits revelados tipicamente também incluem instruções para uso dos componentes do kit para praticar os métodos
15 revelados. As instruções para praticar os métodos revelados são geralmente gravadas em um meio apropriado de gravação. Por exemplo, as instruções podem ser impressas em uma estrutura, tal como, papel ou plástico, etc. como tal, as instruções podem estar presentes nos kits como uma inserção
20 no pacote, no rótulo do recipiente do kit ou dos componentes do mesmo (isto é, associado com o empacotamento ou sub-empacotamento) etc. Em outras incorporações, as instruções estão presentes como um arquivo de armazenamento de dados eletrônico atual em um suporte de memória legível por
25 computador apropriado, por exemplo, CD-ROM, disquete, etc. Em outras incorporações, as instruções reais não estão presentes no kit, mas em meios para obter as instruções de uma fonte remota, por exemplo, através da Internet, são fornecidas. Um exemplo desta incorporação é um kit que
30 inclua um endereço da web de onde as instruções possam ser vistas e/ou de quais as instruções possam ser transferidas. Como com as instruções, estes meios para obter as instruções são gravados em uma estrutura apropriada.

EXEMPLOS

Os seguintes exemplos são colocados assim por diante para fornecer aqueles com habilidade ordinária na técnica uma divulgação e uma descrição completas de como
5 fazer e usar a presente invenção, e não pretendidos para limitar o escopo do que os inventores consideram como sua invenção nem são pretendidos para representar que as experiências abaixo são tudo ou as únicas experiências executadas. Os esforços foram feitos para assegurar a
10 exatidão no que diz respeito aos números usados (por exemplo, quantidades, temperatura, etc.), mas alguns erros e desvios experimentais devem ser esclarecidos. Salvo indicação contrária, as partes são partes por peso, o peso molecular é peso molecular de média do peso, a temperatura
15 está em graus centígrado e a pressão está ou próxima da atmosférica.

EXEMPLO 1

CONSTRUÇÃO DO VETOR ALVO E DOADOR

A expressão de nível elevado dos transgenes foi
20 difícil de conseguir consistentemente em células CRO e em outras linhagens NS celulares de mamíferos por causa da natureza aleatória da integração e dos efeitos cromosomáticos associados do contexto em cima do transgene integrado. Usando integrases local-específicos dos fagos
25 ϕ C31 e R4, vetores de integração local-específico podem ser gerados a fim de prever a integração local-específica dos cassetes de expressão que codificam um gene de interesse no genoma de uma célula mamífera.

Os sistemas da integração ϕ C31 e R4 removem muitas
30 das limitações da integração aleatória fornecendo a integração em um número relativamente pequeno de localizações no genoma que são também caracterizadas pela expressão robusta do gene. A integração dos transgenes com a

integrases ϕ C31 ou R4 proporciona um método fácil para gerar as linhagens celulares mamíferas que indicam a expressão estável, de nível elevado do gene introduzido. O uso de integrases do fago para gerar produção de linhagem celular
5 reduz assim o tempo e o esforço exigido no isolamento de clones apropriados para a produção da proteína. Conseqüentemente, visto que a integração é pensada para mais favoravelmente ocorrer nos lugares nos cromossomas com cromatina aberta ou metilação reduzido, tais localizações
10 também serão mais favoráveis para um nível mais elevado da expressão de gene sustentada.

Vetor Alvo

Um mapa esquemático de um vetor alvo exemplificativo para o uso na introdução a um local
15 acessório de integras e local específico no genoma da linhagem celular é fornecido na FIG. 1 e na FIG. 8. Em geral o vetor alvo irá incluir um primeiro local acessório para uma primeira integrase local-específico e um segundo local acessório para uma segunda integras e local-específico (por
20 exemplo, integras e local-específica, alterada, com uma eficiência mais elevada de integração), onde a primeiro e a segunda integras e local-específica são diferentes. Os vetores alvo podem também incluir elementos suplementares, tais como um marcador selecionável bacteriano (por exemplo,
25 β -lactamase codificando resistência à ampicilina) que fornece a seleção das células procarióticas que contêm os vetores. Além disso, o vetor pode também incluir um marcador selecionável específico da célula mamífera (por exemplo, gene codificando higromicina B fosfotransferase codificando
30 resistência à droga higromicina) para selecionar as células mamíferas que têm o vetor alvo integrado com sucesso no genoma e uma origem para a replicação do vetor (por exemplo, origem ColE1 da replicação do DNA) em células bacterianas,

tais como *E. Coli*.

Como mostrado na FIG. 12 e na FIG. 13, o vetor alvo irá ser usado para introdução de uma sequência de ácido nucléico codificando o local 103 attP ϕ C31 dentro do genoma das células, tais como, células mamíferas. Uma vez que integrado, este local 103 attP ϕ C31 será usado para o local que integra especificamente um plasmídeo doador que inclui um cassete de expressão para um gene de interesse e de uma sequência de ácido nucléico que codifica o local AAA 285 attB ϕ C31. O vetor alvo inicial inclui as sequências do ácido nucléico para dois diferentes locais att para duas diferentes integrases locais-específicas. Em particular, o vetor alvo incluirá uma sequência de ácido nucléico que codifica o local R4 attB 295. O local R4 attB 295 media a integração do vetor alvo em locais R4 pseudo attP (R4 Ψ attP) no genoma da célula mamífera. Estima-se a cerca de 100 locais R4 Ψ attP em um genoma mamífero típico. O vetor alvo também incluirá uma sequência de ácido nucléico que codifica um local 103 attP ϕ C31. O local 103 attP ϕ C31 serve como um local alvo para a integração do vetor doador que inclui uma cassete de expressão projetado para dirigir a expressão dos genes de interesse.

A ordem de integração escolhida aqui, ou seja, integração R4 integrase-mediada seguida pela integração ϕ C31 integrase-mediada mutante, é escolhida para duas razões. A integração integrase-mediada R4 foi escolhida como a primeira etapa, em vez da integração ϕ C31 integrase-mediada, porque há poucos locais R4 Ψ attP comparados aos locais do ϕ C31 Ψ attP em genomas mamíferos. Conseqüentemente o número de locais em que a integração ocorrerá é menor e poucos clones irão ser selecionados para identificar aqueles com os níveis mais elevados de expressão da proteína. A integração ϕ C31 integrase-mediada mutante é escolhida como a segunda

etapa porque uma vez que os primeiros locais de integração são identificados uma vez que resulta em altos níveis de expressão de proteína após a integração do vetor doador, é desejável ter integração do vetor doador sendo tão eficiente
 5 como possível. Portanto, uma integrase ϕ C31 mutante será usada. As integrase ϕ C31 mutantes foram identificadas que resultam a até 75% dos eventos de integração ocorrendo no local *attP* tipo selvagem contido em um vetor integrado (tal como que contido no vetor alvo), enquanto os 25% restante
 10 ocorrerem em uma variedade de locais ϕ C31 Ψ *attP*. Estima-se a ser aproximadamente 370 (média=202-764 com um intervalo de confiança de 95%) locais ϕ C31 Ψ *attP* em células humanas, tais como, células 293, D407, e HepG2 (Chalberg, et al., 2006). O local em que a integração mais freqüentemente
 15 ocorre pode variar entre diferentes células, mas é tipicamente < 5-10% do número total de locais que pode servir como locais de integração. Se uma integrase menos eficiente é usada que conduz a um menor grau de seletividade para locais *attP* tipo selvagem sobre locais pseudo *attP*,
 20 então mais integração poderia ocorrer em locais ϕ C31 Ψ *attP* preferencialmente do que local *attP* tipo selvagem desejado em um vetor alvo integrado.

Além disso, o vetor alvo também inclui uma sequência de ácido nucléico que codifica marcador
 25 selecionável de o higromicina, que é usado para selecionar os clones resistentes à higromicina que têm um vetor alvo genomicamente integrado. O vetor alvo tem uma primeira porção de região de codificação do puromicina (por exemplo, promotor-menor) e um sinal poli A SV40 a jusante da
 30 sequência do ácido nucléico que codifica o local 103 *attP* ϕ C31. Sobre a integração do vetor doador, um promotor SV40 é introduzido a montante do gene de puromicina, desse modo reconstituindo um cassete de expressão completo do gene

capaz de fornecer a expressão do marcador selecionável. Conseqüentemente, o marcador selecionável reconstituído de puromicina pode ser usado para selecionar eficientemente eventos bem sucedidos de recombinação entre um local attB
5 ϕ C31 (por exemplo, um local AAA 285 attB ϕ C31) no vetor doador e um local attP ϕ C31 (por exemplo, local 103 attP ϕ C31) presente no vetor do alvo.

Um promotor mais fraco (por exemplo, SV40) e uma droga mais tóxica para a seleção (por exemplo, puromicina)
10 são escolhidos ao contrário de promotores mais fortes (por exemplo, CMV) e de drogas mais fracas para a seleção (por exemplo, G418) a fim de fornecer uma seleção mais forte para o evento de integração do vetor doador desejado.

Esta etapa, a integração do vetor doador no vetor
15 alvo integrado, é a etapa chave da invenção que permite uma integração local específica no vetor doador, que contém cassetes de expressão para genes de interesse. Entretanto, é possível que uma grande variedade de promotores (sem regiões de codificação) no vetor doador possa trabalhar como
20 eficiência. Além disso, uma grande variedade de regiões de codificação para genes de resistência à droga (sem promotores) presente no vetor alvo pode também trabalhar como eficiência. Os exemplos dados aqui, usando um promotor SV40 e uma região de codificação da puromicina, não são
25 significados a ser exclusivos.

Em uma maneira similar um promotor relativamente fraco (timidina quinase do vírus simples do herpes) é usado para conduzir à expressão do marcador a resistência à droga (higromicina) no vetor alvo. Foi reportado por alguns que
30 uma expressão mais fraca de um marcador co-selecionado pode conduzir a uma expressão mais elevada de genes de interesse ligados.

CONSTRUÇÃO DO VETOR DO ALVO

Para construir o vetor alvo (pR1; FIG. 8) as seguintes etapas foram executadas. A sequência do vetor pR1 é fornecida nas FIGS. 33A-33B. Um fragmento de 295 bp que contém o local *attB* R4 (*attB* 295 R4) foi amplificado pelo
 5 PCR de células rehidratadas de *Streptomyces parvulus* (ATCC 12434) usando iniciadores

5'-CGTGGGGACGCCGTACAG-3' (SEQ ID NO:01)

e

5'-CCCGGTCAACATCCAGTACACCT-3' (SEQ ID NO:02)

como descritos por Olivares e outros, 2001 e clonados em
 PCR2.1-TOPO (Invitrogen) para fazer PTA-R4*attB*. O *attB* 295
 10 R4 foi isolado de pTA-R4*attB* pela digestão com *EcoRI*. Este
 fragmento foi terminado pelo preenchimento das extremidades
 com DNA polimerase de Klenow e ligado então no pTK-Hyg
 (TaKaRa Clontech) no local *Hind* III, no qual tinham sido
 terminados também preenchendo nas extremidades com a DNA
 15 polimerase de Klenow para fazer o vetor pTK-R4B.
 Sequenciamento de DNA foi usado para confirmar se pTK-R4B
 teve a sequência correta e também se o local do *attB* 295 R4
 estava na orientação mostrada na FIG. 8, ou seja, que o lado
 direito do local de recombinação do núcleo *attB* R4 (indicado
 20 pelo ponto estreito do triângulo) era o mais próximo ao
 cassete de resistência à higromicina.

Duas reações em cadeia de polimerase foram feitas
 para amplificar o *attP* 103 ϕ C31 e a região de codificação de
 resistência à puromicina separadamente. Então foram fundidas
 25 junto precisamente usando um terceiro PCR. As condições do
 PCR foram de 95°C para 1 minuto de desnaturação, 60°C para
 15 segundos a anelamento e 72°C para 45 segundos para
 polimerização. As reações foram feitas com uma enzima de
 correção (Pfu Ultra) que gera produtos PCR terminados sem
 30 corte.

Uma região de 103 bp do local *attP* ϕ C31 (*attP* 103 ϕ C31) que contém as seqüências conhecidas para codificar um local *attP* funcional foi amplificado de Pta-*attP* (descrita por Olivares et al., 2001) usando iniciadores C31-*attP*-1

5'-AAAAAAGAATTCGTACTGACGGACACACCGAAGCCCC-3'

5 (SEQ ID No: 03) e C31-*attP*-2

5'-AAAAAACCTTTCGTCTTCAGACATGATAAGATACATTGATGAGTTTGG-3')

(SEQ ID No:04) tendo por resultado um produto de 186 bp. A extremidade 5' do primeiro iniciador C31-*attP*- 2 tinha 24 bases de extremidade 5' da ORF de resistência à puromicina.

[00168] A região de codificação de resistência à puromicina
10 junto com um sinal de poliadenilação SV40 foi amplificada pelo PCR de pPUR (TaKaRa Clontech) usando iniciadores Puro 1

5'- CAGTTGGGGGCGTAGGGTCGCCACCATGACCGAGTACAAGCCCCACGGTG -3'

(SEQ ID No: 05) e SV40poliA

5'-AAAAAACCTTTCGTCTTCAGACATGATAAGATACATTGATGAGTTTGG-3'

(SEQ ID No: 06) tendo por resultado um produto 1001 bp. A extremidade 5' do iniciador Puro1 teve 24 bases de
15 extremidade 3' do *attP* ϕ C31 e a extremidade 3' de SV40poliA tinha um local de reconhecimento da enzima de restrição BBS L As condições do PCR para os primeiros 10 ciclos foram 95°C por 1 minuto de desnaturação, 47°C por 30 segundos de
anelamento e 72°C por 75 segundos de polimerização. As
20 condições do PCR para os 15 ciclos seguintes foram 95°C por 1 minuto para desnaturação, 60°C por 30 segundos para anelamento e 72°C por 75 segundos para polimerização. As reações foram feitas com uma enzima de correção (Pfu ultra) que gerasse produtos sem corte-terminados do PCR.

25 Para fundir o DNA contendo *attP* 103 ϕ C31 ao DNA contendo a região de codificação resistente a puromicina e sinal de poliadenilação SV40 os produtos daqueles PCRs separado foram misturados em uma relação equimolar e

amplificados pelo PCR com iniciadores C31-attP-1 e poliA SV40 para produzir um produto de 1138 bp. As condições do PCR foram a 95°C por 30 segundos para desnaturação, 60°C por 20 segundos para anelamento e 72°C por 90 segundos para polimerização. As reações foram feitas com uma enzima de correção (Pfu Ultra) que gera produtos de PCR terminados-sem corte.

O produto de PCR de 1138 bp contendo attP 103 ϕ C31, estrutura de leitura aberta de resistência à puromicina e o sinal de poliadenilação SV40 foi digerido com BBS I e clonado em pTK-R4B que foi digerido com Sma I e BBS L Isto produziu o pR1 do vetor alvo. As seqüências e a orientação apropriada de attP 103 ϕ C31, a estrutura de leituraaberta da resistência à puromicina e o sinal de poliadenilação SV40 no pR1 foram confirmados por seqüenciamento do DNA.

Uma característica chave do projeto da fusão da região de codificação attP 103 ϕ C31-puromicina é diagramado na FIG. 14. As 221 pares de bases ao longo do local attP ϕ C31 221 que está presente em Pta-attP tem um ATG que poderia terminar acima a montante da região de codificação da puromicina uma vez o vetor doador é integrado no vetor alvo, para criar um local attL ϕ C31. Geralmente as seqüências A TG (locais potenciais de iniciação da tradução) que estão a montante de legítimas regiões de codificação são prejudiciais à expressão de gene. Conseqüentemente, no produto do PCR que funde attP 103 ϕ C31 à região de codificação da puromicina, esse A TG foi feito ao códon inicial da região de codificação da puromicina. Além disso, 2 bases antes daquela ATG foram trocadas para criar uma melhor seqüência (GCCACC) consenso de início da tradução (Kozak). Como mostrado na FIG. 14 estas mudanças são pelo menos dezoito bases na 3' ao mínimo, mas inteiramente

funcional, local attP ϕ C31 identificado por Groth et al., 2000. Conseqüentemente não devem afetar a habilidade o local AAA 285 attB ϕ C31 no vetor doador para integrar dentro do local attP 103 ϕ C31 no vetor alvo. Após a integração do
 5 vetor doador dentro do vetor alvo o 88 bases ao longo do local attL ϕ C31 (attL 88 ϕ C31) é localizado na região não-traduzida 5', imediatamente antes da região de codificação da puromicina. Precedendo o attL 88 ϕ C31 pode ser 57, 62 ou
 10 74 bases derivadas da região não traduzida 5' do promotor adiantado SV40 (transcrição dirigida pelo promotor adiantado SV40 sendo em 3 locais diferentes).

VETOR DOADOR

Um diagrama esquemático de um vetor de expressão doador exemplificativo é fornecido nas FIGURAS 2 e 10. O
 15 vetor de expressão doador exemplificativo contém uma seqüência de ácido nucléico codificando o local AAA attB 285 ϕ C31 e cassete de expressão de ácido nucléico codificando genes de interesse, tais como, um cassete que codifica as cadeias pesadas e leves de um anticorpo humano. O vetor
 20 doador também contém um promotor SV40 a montante da seqüência do ácido nucléico que codifica o local AAA attB 285 ϕ C31. Sob a integração do vetor doador dentro do vetor alvo previamente integrado, o qual é mediado pela recombinação local-específica entre local AAA attB 285 ϕ C31
 25 presente no vetor doador e attP 103 ϕ C31 presente no vetor alvo, o promotor SV40 conduzirá a expressão do gene de puromicina (FIG. 13). Conseqüentemente, o gene de resistência reconstituído da puromicina pode ser usado para
 30 vetor doador para expressar proteínas de interesse.

Esta etapa da seleção é crítica para conseguir um método de eficiência elevada porque o local AAA attB 285 ϕ C31 no vetor doador pode também integrar dentro de locais

attP Ψ ϕ C31 encontrados em 370 posições cromossômicas estimadas (Chalberg, et al., 2006). Entretanto todos os vetores de expressão doadores exemplificativos que integram em locais do attP Ψ ϕ C31 conterão somente o promotor SV40 e não irão reconstituir um gene funcional de resistência à puromicina. Algumas células resistentes a puromicina também resultam quando apenas a integrase é expressa em um clone do vetor alvo attP (isto é, na ausência de um vetor de expressão doador). Sem se prender à teoria, o mecanismo por qual este ocorre pode envolver a recombinação dos locais Ψ attB que estão perto de um promotor celular com o local attP 103 no vetor alvo. Transfecção da linhagem celular attP com um vetor de expressão doador selecionável e uma segunda integrase de vetor de expressão endereça este interesse porque as células sem nenhum vetor de expressão não serão resistentes a completa seleção do gene de resistência à droga no vetor de expressão doador selecionável. Em adição, caso necessário, a integração desejável dos vetores doadores dentro de vetores alvos cromossômicos pode facilmente ser distinta da integração aleatória indesejável ou da integração dos vetores doadores em locais attP Ψ ϕ C31 como descrita abaixo na seção "Métodos para a caracterização da linhagem celular".

CONSTRUÇÃO DO VETOR DE EXPRESSÃO DOADOR

O vetor de expressão doador (pD1-DTX-1) é baseado em pcDNA3002neo descrito por Jones et al., 2003. pcDNA3002neo é baseado no pcDNA3 (Invitrogen, Inc.). pcDNA3002neo contém dois promotores CMV seguidos por dois sinais de poliadenilação do hormônio de crescimento bovino para a expressão das proteínas em células mamíferas. pcDNA3002neo também inclui uma origem ColE1 e um gene de resistência à ampicilina para a manutenção e a seleção em E. Coli. Finalmente, o vetor pcDNA3002neo tem um gene de

resistência G418 expressado usando um promotor SV40 e um sinal de poliadenilação SV40. A sequência do vetor pD1-DTX-1 é fornecido nas FIGS. 34A-34C.

Para construir pD1-DTX-1, seis inserções foram
 5 clonadas no pcDNA3002neo que contém 1) um poliligante com os locais de reconhecimento para três enzimas de restrição que cortam dentro de oito pares ao longo da sequência de reconhecimento, 2) a região AAA attB 285 ϕ C31, 3) uma primeira sequência do sinal que media a secreção das
 10 proteínas, tais como, a cadeia pesada de um anticorpo humano e contém um único local de restrição, 4) uma segunda sequência do sinal que media o secreção das proteínas, tais como, a cadeia leve de um anticorpo humano e contenha um outro local único de restrição, 5) uma região de codificação
 15 para uma primeira proteína, tal como, a cadeia pesada de um anticorpo específico humano para a toxina da difteria, e 6) uma região de codificação para uma segunda proteína, tal como, a cadeia leve de um anticorpo específico humano para a toxina da difteria.

20 pcDNA3002neo perde o uso de poliligantes após um de seus promotores CMV. Conseqüentemente, em uma primeira etapa para criar o vetor doador pDI, um poliligante com os três locais de restrição de ocorrência rara foi inserido. Dois oligonucleotídeos sintéticos (BamBst-A e BamBst-B)
 25 foram anelados. A sequência de BamBst-A é:

5'- GATCCAAAAAATTAATTAATAAAAAAACACCGGCGAAAAAAGCGATCGCAAAAAACCAAGTGTG-3'
 (SEQ ID No: 07). A sequência BamBst-B é

5'-CTGGTTTTTTGCGATCGCTTTTTTCGCCGGTGTTTTTTTTAATTAATTTTTTG-3'
 (SEQ ID No: 08). Quando BamBst-A e BamBst-B são anelados

conterá seqüências complementares do Bam HI e do Bst XI em suas extremidades 5' e 3', respectivamente, para permitir a
 30 ligação a Bam HI/Bst XI-digerir pcDNA3002neo. As seqüências

também incluirão (na ordem 5' a 3') locais de reconhecimento da enzima de restrição para Pac I, SgrA I e AsiS I. As seqüências do espaçador de 6 adenosines separa cada local de restrição para permitir a digestão eficiente em dois locais adjacentes, se necessários. Os dois oligonucleotídeos sintéticos foram anelados (isto é, não-fosforilados). pcDNA3002neo foi digerido com Bam HI a 37°C e então com Bst XI a 55°C. O vetor digerido foi ligado ao poliligante anelado e a ligação foi transformada dentro de células de E. Coli XL-10 GOLD (Stratagene). O vetor resultante foi chamado pHPC-1.

Um elemento crítico da seqüência no vetor doador pD1 é o local AAA attB 285 ϕ C31. O local AAA attB 285 ϕ C31 foi amplificado pelo PCR do vetor Pta-attB descrita por Olivares, et al, 2001. O iniciador 5' foi chamado C31attB5' e tem uma seqüência de

5' - GTCGACGAAATAGGTCACGGTCTC-S'

(Seq ID No: 09). O iniciador 3' foi chamado C1attB-V e tem uma seqüência de

5' - TACGTCGACATGCCCGCCGTGACC -3'

(SEQ ID No: 10). As condições do PCR foram desnaturação a 95°C por 1 minuto, anelamento em 60°C por 15 segundos e extensão a 72°C por 30 segundos usando a polimerase Pfu Ultra (Stratagene). A concentração de outros componentes da reação foi a mesma que aquela de um PCR padrão (por exemplo, 200 μ M de dNTPs, 1 μ M cada iniciador, 1,5 mM de $MgCl_2$).

O iniciador 5' mudou uma seqüência ATG na extremidade 5' do local do attB ϕ C31 em Pta-attB a uma seqüência AAA. A razão para esta é similar àquela descrita acima para o local attP 103 ϕ C31 e diagramado na FIG. 14. A extremidade 5' do local attB 285 ϕ C31 que está presente em Pta-attP tem um ATG que poderia terminar sendo a montante da

região de codificação da puromicina uma vez que o vetor doador é integrada dentro do vetor alvo, para criar um local attL 88 ϕ C31. Geralmente as seqüências A TG (locais potenciais de iniciação da tradução) que está a montante de
 5 legítimas regiões de codificação são prejudiciais à expressão de gene. Conseqüentemente, o ATG na extremidade 5' do attB ϕ C31 foi mudada para AAA. Todas as variações de uma base de AUG foram encontradas para funcionar como códons alternativos de iniciação da tradução. Entretanto nenhuma
 10 variação de duas bases foi mostrada para função como códons alternativos de iniciação da tradução. Conseqüentemente a fim de impedir a 5' ATG no attB ϕ C31 sendo utilizado como um códon de iniciação da tradução, mas ao mesmo tempo introduz um número mínimo de mudanças à seqüência attB ϕ C31, o A TG
 15 foi mudado para AAA. Visto que este ATG está próximo da extremidade 5' da região attB ϕ C31 contida em PT A -attB foi mais conveniente incorporar a mudança A TG para AAA dentro do iniciador usado para PCR a seqüência de attB ϕ C31 de Pta-attB.

20 Amplificação de Pta-attB pelo PCR com iniciadores C3lattB-5' e C3 lattB-3' conduzido a um produto ao longo de 285 pares de bases chamado AAA attB 285 ϕ C31. pHPC-1 foi digerido com Sma I e Bst Z 17 I para produzir fragmentos de 1130 bp e 5718 bp. O produto do PCR AAA attB 286 ϕ C31 foi
 25 ligado ao fragmento de 5718 bp. Isto produziu um plasmídeo chamado pHPC-2. O plasmídeo com a seqüência AAA attB 286 ϕ C31 em uma orientação, tal que o lado esquerdo do attB estava ao lado do promotor SV40 foi chamado de pHPC-2(+) enquanto o plasmídeo com a seqüência AAA attB 286 ϕ C31 na
 30 orientação oposta foi chamado pHPC-2(-).

pHPC2 (+) e pHPC-2 (-) são úteis como vetores para integrar e expressar os genes que codificam as proteínas que não são segregadas. Entretanto, para segregar as proteínas

tais como, anticorpos, fatores hemofílicos, fatores de crescimento, fatores do soro ou receptores solúveis, um vetor doador que contivesse uma seqüência do sinal para a secreção seriam desejável. Conseqüentemente uma seqüência do

5 sinal (HAVT20; Boel et al, J Immunol Methods. 2000 May 26;239(1-2): 153-66) de uma cadeia do receptor alfa de célula T humano foi modificado para ter locais de restrição únicos. Uma versão com um único Pml I foi inserido em um dos dois poliligantes no pHPC2 (+) e uma outra versão com um

10 único PspX 1 foi inserido no outro poliligante no pHPC2 (+). Nenhuma versão mudou a seqüência de aminoácido da seqüência do sinal HA VT20 e as mudanças também utilizaram códons humanos freqüentemente usados. Tanto os locais Pml I e o PspX I ocorre imediatamente antes do local de clivagem da

15 seqüência do sinal. Conseqüentemente, uma fusão precisa entre o local de clivagem na seqüência do sinal HA VT20 e a região da codificação de uma proteína de interesse é facilmente alcançada projetando os iniciadores de PCR apropriadas para amplificar as regiões de codificação dos

20 genes de interesse. Alternativamente, é possível extirpar a seqüência do sinal HAVT20 (por exemplo, usando BamH I/Pac I em um local de clonagem e Asc I/Not I em outro local de clonagem) e introduzir outras seqüências do sinal. Aquelas seqüências poderiam ser heteróloga (por exemplo, a seqüência

25 do sinal IL-2) ou homóloga (por exemplo, uma seqüência humana do sinal de IgG1).

Para introduzir uma seqüência do sinal HA VT20 em pHPC-2 (+) um DNA duplex codificando um local Bam HI na extremidade 5', uma melhor seqüência consenso Kozak,

30 seqüência do sinal HAVT20 com um local Pml I e na extremidade 3' foi gerada por anelamento de 2 oligonucleotídeos: HAVT20-L-top

(5'-CGCGCCACCATGGCATGCCCTGGCTTCCTGTGGGGCACTTGTG
ATCTCCACCTGCCTCGAGTTTTCCATGGCTCG-3') (SEQ ID NO:11)

e HAVT20-L-bot

(3'-GGTGGTACCGTACGGGACCGAAGGACACCCGTGAACACTAG
AGGTGGACGGAGCTCAAAAGGTACCGAGC-5') (SEQ ID NO:12).

Este cassete anelado foi ligado a pHPC2 (+) que foi digerido com Bam HI e Pac I. O plasmídeo resultante foi chamado pHPC-3.

5 Para introduzir um segundo sequência de sinal HAVT20 dentro do pHPC-3 um DNA duplex que codifica um local Asc I na extremidade 5', uma melhor sequência consenso Kozak, sequência do sinal HA VT20 com um local PspX I e uma extremidade 3' sem corte foi gerada pelo anelamento de 2
10 oligonucleotídeos: HAVT20-H-top

5'GATCCGCCACCATGGCATGCCCTGGCTTCCTGTGGGGCACTTGTGATCTCC
ACGTGTCTTGAATTTTCCATGGCTTTAAT-3'

(SEQ ID NO: 13) e HA VT20-H-bot

(3'-GCGGTGGTACCGTACGGGACCGAAGGACACCCGTGAACACTAGAGGTG
CACAGAACTTAAAAGGTACCGAAAT-5')

(SEQ ID No: 14). Este cassete anelado foi ligado a pHPC3 que foi digerido com Asc I e Eco rv. O plasmídeo resultante é uma espinha dorsal do vetor de expressão doador que possa
15 ser usada, entre outras coisas, prontamente trocando vários elementos de expressão de gene, tais como, promotores. Esta espinha dorsal do vetor de expressão doador foi chamada pHPC-4

5'-AAAAAACACGTGTCTTGAATTTTCCATGGCTGAAGTGCAGCTGGTGGAGTCTGGG-3'

20 Para isolar genes IgG humanos, as linhagens humanas EBV-transformadas de célula-B que segregam os anticorpos que ligam a toxina da difteria foram derivadas como descritas por Traggiai, et al., 2004. Um anticorpo com afinidade elevada foi subtipado e conduzido para ter uma cadeia pesada IgG1 humano e uma cadeia leve kappa. O RNA foi

preparado das células produzindo este anticorpo e usado em reações de RT-PCR para gerar os cDNAs que codificam os genes do anticorpo de cadeia pesada e leve. Os iniciadores usados para a amplificação foram similares àqueles descritas por

5 Marca, et al (Transplantation, 1991 August; 52 (2): 340-5), Sblattero, et al (Immunotechnology, 1998 Jan; 3(4):271-8), e Yamanaka, et al (J Biochem (Tokyo), 1995 jun; 117(6): 1218-27) exceto que as extremidades tenham os locais apropriados de restrição para permitir a subclonagem. O cDNA da cadeia

10 leve foi clonado dentro do local Not I/Xba do pBK-CMV (Stratagene) para criar pBK-CMV-DTX-L. O cDNA de cadeia pesada foi clonado dentro do local Hind III/Sal I do pBK-CMV-DTX-L para criar pABMCI03. Os cDNAs foram seqüenciados e sua identidade como um IgG1k humano foi confirmado.

15 Para subclonar genes do anticorpo da toxina anti-difteria dentro de pHPC-4 o gene inteiro de cadeia pesada foi amplificado pelo PCR com iniciadores IMAGEM 6 (SEQ ID No: 15) e

5'-AAAAAATTAATTAATTATTTACCCGGAGACAGGGAGAG-S'

(SEQ ID No: 16) que usam pABMCI03 como um molde. O produto

20 resultante do PCR de cadeia pesada foi digerido com BbrP I (isoschizomer de PmI I) e de Pac I e clonado dentro de pHPC-4 que foi digerido com BbrP I e Pac I para criar pHPC4-DTX-H. O gene inteiro de cadeia leve foi amplificado com iniciadores

5'-AAAACCTCGAGTTTTCCATGGCTGAAACGACACTCACGCAGTCTCCAGS'

25 (SEQ ID No: 17) e

5'-AAAAAAGCGGCCGCTTAACACTCTCCCCTGTTGAAGCTCTTTG-3'

(SEQ ID No: 18) que usam o pABMC 103 como um molde. O produto resultante do PCR da cadeia leve foi digerido com PspX I e Not I e clonado dentro de pHPC4-DTX-H que foi

digerido com PspX I e Not I para criar pD1-DTX-1. As seqüências de ambos os genes da cadeia do anticorpo foram confirmadas para ambos os filemantos.

pHPC-2, pHPC-4, e pD1-DTX-I pode ser vetores de subclonagem e vetores de expressão. Embora as seqüências de cada um dos dois promotores as CMV, seqüência do sinal HAVT20, e sinais de poliadenilação do hormônio de crescimento bovino sejam quase idênticas, elas são separadas pelos poliligantes que são diferentes na seqüência. Conseqüentemente iniciadores de seqüenciamento específicos foram projetadas que são capazes de seqüenciamento de genes inseridos em cada cassete de expressão. Por exemplo, o iniciador

5'-GCTTGGTACCGAGCTCGGATCC-3'

(SEQ ID No: 19) pode ser usado para seqüenciar as regiões variáveis do anticorpo inseridas após o local Pml I de uma seqüência de sinal e o iniciador

5'-GAAGCTTGGT ACCGGTGAATTCCG-3'

(SEQ ID No: 20) pode ser usado para seqüenciamento das regiões variáveis do anticorpo inseridas após o local PspX I da outra seqüência do sinal. Conseqüentemente, não há nenhuma necessidade de clonar genes de interesse dentro de outros vetores para seqüenciamento antes da clonagem deles no pHPC-2, em pHPC-4 ou em pD1-DTX-I para a expressão.

Além disso, cada elemento no pHPC-4 ou no pD1-DTX-1 é flanqueado por locais únicos de restrição, tais que, todo o elemento (por exemplo, promotor, seqüência de sinal, cadeia variável do anticorpo, cadeia constante do anticorpo, região de codificação, local de poliadenilação, local attB ϕ C31) pode facilmente ser extirpado e substituído com outros elementos similares.

Por exemplo, a região variável da cadeia pesada

pode ser trocada digerindo pD1 -DTX-1 com Pml I/Xho I e substituindo a região variável da cadeia pesada do anticorpo da toxina da anti-difteria com outras regiões variável de cadeia pesada. A região variável de cadeia leve pode ser trocada digerindo pD1-DTX-1 com PspX I/BsiW I e substituindo a região variável de cadeia leve do anticorpo da toxina da anti-difteria com outras regiões variável de cadeia leve.

Similarmente a região constante da cadeia pesada de IgG1 pode ser trocada por aquelas de outros subtipos de anticorpo (por exemplo, IgG2, IgG3, IgG4) ou da outra classe de imunoglobulina (por exemplo, IgA1, IgA2, IgD, IgE ou IgM) trocando um fragmento de restrição Apa I/Pac I. A região constante de cadeia leve kappa no pD1-DTX1 pode ser trocada por uma região constante da cadeia leve kappa lambda trocando um fragmento de restrição BsiW I/Not I.

Um promotor CMV pode ser substituído com um outro promotor trocando um fragmento de restrição Mfe I/BamH I e o outro promotor CMV posso ser substituído trocando um fragmento de restrição BstZ 17 I/Asc I. Uma sequência do sinal HA VT20 pode ser substituída trocando um fragmento de restrição BamH I/Pml I e outro pode ser substituída trocando um fragmento de restrição Asc I/PspX I. Um sinal de poliadenilação do hormônio de crescimento bovino pode ser substituído trocando um fragmento de restrição Asis I/NgoM IV e outro pode ser substituído trocando um fragmento de restrição CIA IlPci L. O local attB ϕ C31 pode ser substituído com um local attB reconhecido por uma outra integrase e serina local-específico trocando um fragmento de restrição Stu I/BstZ17 I.

CONSTRUÇÃO DO VETOR ALVO-DHFR

O vetor alvo-DHFR (pR1 - DHFR) foi construído pela clonagem de um cassete de expressão DHFR de rato consistindo do promotor SV40, uma região de codificação DHFR de rato, o

3' UTR do DHFR cDNA de rato e o sinal de poliadenilatação do vírus da leucemia murina Moloney (MLV) dentro do vetor alvo pR1. A sequência do vetor do pR1-DHFR é fornecida nas FIGS. 35A-35C.

5 Um fragmento de DNA de 1.074 pares de base de pSV2dhfr (American Type Culture Collection) contendo o promotor SV40, uma região de codificação DHFR de rato e parte 3' UTR do cDNA do rato DHFR foi amplificado pelo PCR usando os iniciadores 5'

5' - CGAATCAGCACGGGGTGGCGCGCCCTGTGGAATGTGTGTCAGTTAGG -3'

10 (SEQ ID No: 21) e

5' - CGAATCAGCACGAAGTGCACCGGTGTTTAACTTAATTAAAGATCTAAA
GCCAGCAAAAGTCCCATGGT -3'

(SEQ ID No: 22). As circunstâncias usadas para o PCR foram a 95°C por 30 segundos, 60°C por 30 segundos, 72°C por 90 segundos para 10 ciclos, então 95°C por 30 segundos e 72°C por 90 segundos para 15 ciclos usando a polimerase Pfu. O
15 produto de PCR foi clonado então dentro PCR-BluntII-TOPO (Invitrogen), a seguir digerido com DraIII, e um fragmento de 1050 pares de bases foi isolado e o gel purificado. O pR1 foi digerido com Van91 I (isoschizomer de PflM I) e purificado usando um kit de purificação Qiagen PCR. O
20 fragmento de Dra III foi ligado a Van91 I cortado o pR1 para gerar o pR1-dHFR (noltr).

As 594 pb ao longo da repetição terminal ML V, as quais contêm um sinal de poliadenilação foram amplificadas pelo PCR do pLNXH (TaKaRa Clontech) usando os iniciadores

5' - AAAAAATTAATTAATAATGAAAGACCCACCTGTAGGTTTGG -S'

25 (SEQ ID No: 23) e

5' - AAAAAACACCGGTGAAAGTTTAAACAAACCTGCAGGAATGAAAGACCC
CCGCTGACGGGTAG -3'

(SEQ ID No: 24). As condições usadas incluíram 95°C por 30

segundos, 56°C por 30 segundos e 72°C por 45 segundos para 15 ciclos usando a polimerase de Pfu. O produto de PCR terminado-sem corte foi clonado então dentro de PCR-Blunt II-TOPO para criar o pCR-pLTR. O MLV LTR foi cortado do pCR-pLTR usando EcoRI, terminado-sem corte com Klenow e gel purificado. O pR1-dHFR (noltr) foi digerido com PmeI e tratado com CIP. O fragmento MLV LTR contendo o sinal poli A MLV foi ligado ao vetor Pme I-digerido para criar o pR1-DHFR. As orientações e as seqüências corretas da inserção wer confirmado pelas digestões de enzimas de restrições e seqüenciamento de DNA.

CONSTRUÇÃO DO VETOR DE EXPRESSÃO DOADOR-DHFR

O vetor de expressão doador-DHFR (pD1-DHFR) pode ser construído clonando um cassete de expressão DHFR de rato consistindo de um promotor SV40, uma região de codificação DHFR de rato, um 3' UTR do cDNA DHFR de rato e o sinal de poliadenilação do vírus da leucemia murina Moloney (MLV) no vetor de expressão doador pD1- DTX -1. Este cassete de expressão de 1626 pares de base é amplificada pelo PCR usando a polimerase Pfu do pR1-DHFR do vetor alvo-DHFR usando iniciadores DHFR-I

5'-TTTTTTGAAGACGAAAGGCTGTGGAATGTGTGTCAGTTAGGGTGTGGA-3'
(SEQ ID No: 25) e LTR-2

5'-AAAAAACCTGCAGGAATGAAAGACCCCCGCTGACGGGTAG-S'

(SEQ ID No: 26), e clonado como um fragmento terminado-sem corte no local BstZ17 I de pD1-DTX-I na orientação mostrada na FIG. 16.

CONSTRUÇÃO DO VETOR DOADOR-IRES

O vetor doador-IRES (pD1-IRES, FIG. 17) pode ser construído clonando duas cópias dos mesmos IRES (também conhecido como elemento realçador translacional (TEEs) dentro de tanto o local único BamHI ou tanto no local Asc I

de pD1-DTX-1. Diversos IRES podem ser escolhidos, tais como os Gtx IRES de ocorrência natural do gene do homeodomínio Gtx do rato (Chappell, et al., 2000), IRES de ocorrência natural no Rbm3 mRNA de rato (Chappell, et al., 2003), ou
 5 IRES sintéticos, tais como, ICS1-23b ou ICS2-17. 2 foram selecionados em um esquema enriquecido baseado em FACS (Owens, et al., 2001). Versões Multiméricas de alguns IRES frequentemente realçam várias traduções no melhor dobramento do que as versões monoméricas. As seqüências dos IRES, todas
 10 multiméricas, são pequenas e são facilmente introduzidas dentro de pD1-como vetores pela construção de oligonucleotídeos sintéticos que os codificam.

Um ICS1-23b IRES multimérico é montado pelo anelamento de 2 oligonucleotídeos sintéticos. Um par,
 15 consistindo das seqüências

5'-GATCCAGCGGAAACGAGCGAAAAAAAACAGCGGAAACGAGCGAAAAAAAACAGCG GAAACGAGCGAAAAAAAACAGCGGAAACGAGCGAAAAAAAACA
 .GCGGAAACGAGC GGACTCACAACCCCAGAAACAGACATG-3'

(SEQ IDNo: 27)

5'-GATCCATGTCTGTTTCTGGGGTTGTGAGTCCGCTCGTTTCCGCTGTTTTTTTTTC
 GCT CGTTTCCGCTGTTTTTTTTTCGCTCGTTTCCGCTGTTTTTTTTTCGCTCGTTCCG
 CT GTTTTTTTTTCGCTCGTTTCCGCTG-5'

(SEQ ID No: 28), a qual têm as extremidades complementares a um local de restrição BamH I e uns outros pares, consistindo nas seqüências

CGCGCCAGCGGAAACGAGCGAAAAAAAACAGCGGAAACGAGCGAAAAAAA
 AACAGCGG AAACGAGCGAAAAAAAACAGCGGAAACGAGCGAAAAAAAAC.
 AGCGGAAA CGAGCGGA CTCACAACCCCAGAAACAGACATGG-3'

20 (SEQ IDNo: 29) e

5'-CGCGCCATGTCTGTTTCTGGGGTTGTGAGTCCGCTCGTTTCCGCTGTTTTTTTTT
 CGCT CGTTTCCGCTGTTTTTTTTTCGCTCGTTTCCGCTGTTTTTTTTTCGCTCGTT
 'CCG CTG TTTTTTTTTTCGCTCGTTTCCGCTGG-5'

(SEQ ID No: 30), que têm as extremidades complementares a um

local de restrição Asc I. Estas seqüências contêm 5 cópias de 15 bases ao longo de ICS1-23b. Cada um é separado por quatro cópias de um espaçador de 9 bases ao longo do espaçador Poli A. Finalmente, a extremidade 3' contém uma
5 25 seqüência de base que precede imediatamente a região de codificação β -globina do rato (por exemplo, número de ascensão J00413 do GenBank). Estes oligonucleotídeos anelados são clonados nos locais BamH I e Asc I do pD1-DTX-1 para criar o pD1-IRES do vetor doador-IRES. Os clones são
10 seqüenciados para identificar aqueles com a orientação e a seqüência corretas.

CONSTRUÇÃO DO VETOR ALVO DE REGULAÇÃO

Quando algumas proteínas são expressas a níveis necessários para tomar-se comercialmente úteis, elas podem
15 ser tóxicas e conduzir para retardar o crescimento da célula ou mesmo a morte celular. Conseqüentemente, pode ser útil reprimir sua expressão até que seja necessário produzir grandes quantidades. Diversos métodos para regulação de genes estão disponíveis. Em algumas incorporações, é
20 desejável introduzir o sistema que regule os genes em células primeiramente antes que o cassete de expressão da proteína seja introduzido nas células. Desse modo, o sistema regulador do gene é estabelecido e reprimido a expressão de gene antes que um vetor de expressão seja introduzido.
25 Conseqüentemente, pode ser desejável ter um sistema regulador do gene pR1 do vetor alvo e não no vetor doador.

O sistema RheoSwitch (New England Biolabs) fornece a regulação do gene sobre uma ampla escala de expressão. A regulação do gene pelo sistema de RheoSwitch é mediado por
30 duas proteínas. O RheoReceptor consiste na proteína GAL4 de levedura fundida ao domínio de ligação do ligante de um receptor nuclear do hormônio estrogênico do inseto. O RheoReceptor liga-se às seqüências (DAS) de ativação a

montante (DAS) derivadas do gene GAL4 de levedura que é colocado a montante de um TATA-box. O RheoActivator consiste em um receptor de ligação ao ligante RXR mamífero/inseto híbrido fundidos ao domínio de ativação transcripcional VP
 5 16 do vírus simples da herpes. Os análogos de ecdisona podem dimerizar o RheoReceptor e o RheoActivator e quando isto ocorre os genes que estão corretamente ligados aos elementos de ligação GAL4 DAS DNA serão ativados. Além disso, na ausência do dimerizador o RheoReceptor liga-se às seqüências
 10 DAS e media a repressão da expressão de gene. A rede de resultado é que os níveis basais de expressão usando este sistema estão muito baixos e os níveis de indução que podem ser conseguidos são elevados.

Cassetes do gene codificando os dois componentes
 15 da proteína do sistema RheoSwitch (RheoReceptor e RheoAtivator) podem ser amplificados pelo PCR do pNEBR-R1 (New England Biolabs). Eles são clonados em uma orientação, como mostrado na FIG. 18, tais que, as regiões de codificação para o RheoReceptor e o RheoAtivator estão em
 20 uma orientação que seja a mesma que aquela da região de codificação da puromicina. Esta configuração é diferente da configuração no pNEBR-R1 (onde estão em orientações opostas) e esta é como os cassetes do gene RheoReceptor e RheoAtivator são clonados dentro do pR1 separadamente.

25 Mais especificamente, iniciadores do PCR consistindo nas seqüências

5' -AAAAAAACCCTGCAGGGGCCTCCGCGCCGGGTTTTGGCGCCT -3'
 (SEQ ID No: 31) e

5' -AAAAAAAACACCGGTGCTTATCGGATTTTACCACATTG-3'

(SEQ ID No: 32) são usados para amplificar o cassete de expressão do gene RheoActivator (que consiste um promotor de
 30 ubiquitina C (UbC), região de codificação de RheoAtivator, e

na sequência do sinal de poliadenilação da região atrasada SV40). As 2481 pares de base ao longo o produto é digerido com Sbf I e SgrA I e clonado dentro dos locais únicos SbfI/SgrA I do pR1-PU para criar o pR1-RA.

5 Iniciadores do PCR que consistem nas sequências

5'-AAAAAAAACACCGGTGCCGATATCGGGTGCCACGCCGTCCCG-3'

(SEQ ID No: 33) e

5'-AAAAAAAAGCCCGGGCGGCGGCCCGCCAGAAATCC-3'

(SEQ ID No: 34) são usados para amplificar o cassete de expressão do gene RheoReceptor (que consiste em um promotor de ubiquitina B (UbB), da região de codificação RheoReceptor e na sequência de sinal de poliadenilação TK).
10 As 3680 pares de base ao longo do produto é digerido com SgrA I e SrfI e clonado em locais únicos SgrA I/SrfI do pR1-RA para criar o pR1reg.

CONSTRUÇÃO DO VETOR ALVO-DHFR DE REGULAÇÃO

15 A fim de construir um vetor alvo que possa regular genes no vetor doador e ser sujeito à amplificação do gene, um vetor alvo-DHFR de regulação (FIG. 19) é construído. O cassete de regulação do gene pR1reg, consistindo nos genes RheoAtivator e de RheoReceptor, é
20 amplificado pelo PCR do pR1reg usando os iniciadores

5'-AAAAAAACCCTGCAGGGGCCTCCGCGCCGGGTTTTGGCGCCT-3'

(SEQ ID No: 35) e

5'-AAAAAAAAGCCCGGGCGGCGGCCCGCCAGAAATCC-3'

(SEQ ID No: 36), digerido com SbfI e Sfr I e clonado nos locais SbfI e Sfr I do pR1-DHFR para construir o pR1reg-DHFR de vetor do alvo-DHFR de regulação.

25 CONSTRUÇÃO DA ESPINHA DORSAL DO VETOR DE EXPRESSÃO DOADOR DE REGULAÇÃO

A espinha dorsal do vetor de expressão doador de regulação (FIG. 20) tem as sequências do DNA reconhecidas

pelo componente de proteína (por exemplo, RheoReceptor) do sistema regulador do gene codificado pelo pR1reg clonado a montante das regiões da codificação para proteínas de interesse. No caso do sistema RheoSwitch os elementos do DNAa em que o RheoReceptor liga-se são as seqüências (UAS) a de ativação a montante GAL4 (UAS). 722 pares de bases ao longo da seqüência de DNA, em ordem, locais de restrição (metade 3' de BstZ17 I, EcoR I), a região de sinal de poliadenilação SV40 (para impedir a transcrição oculta na região reguladora), cinco elementos de GAL4 DAS e uma TATA-box podem ser amplificados pelo PCR do pNEBR-XIHygro (New England Biolabs) que usa iniciadores

TACGAATTCATCAGCCATATCACATTTGTAGAG-3'

(SEQ ID No: 37) e

TTATATAACCCTCTAGAGTCTCCGCTCGGA -3'

(SEQ ID No: 38).

15 Duas seqüências de DNA longas de pares de base 173 ou 178 codificando duas versões da região não-traduzida (5'DTR) do promotor 5' adiantado 5 CMV com locais diferentes de enzima de restrição na extremidade 3' são geradas pelo anelamento de dois conjuntos de sobreposição

20 oligonucleotídeos e preenchendo em suas extremidades 3' usando a DNA polimerase de Klenow. A base 173 ao longo da versão é gerada pelo anelamento

CCGAGCGGAGACTCTAGAGGGTATATAAGCAGAGCTCGTTTAGTGAACCGTC
A GATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGAC -3'

(SEQ ID No: 39) e

AAAAAAGGATCCGAGCTCGGTACCAAGCTTCCAATGCACCGTTCCCGGCCGCG
GAGG CTGGATCGGTCCCGGTGTCTTCTATGGAGGTCAAAA-3'

(SEQ Id No: 40) e preenchendo com a polimerase de Klenow. A

25 base 178 ao longo da versão é gerada pelo anelamento

CCGAGCGGAGACTCTAGAGGGTATATAAGCAGAGCTCGTTTGTAGTGAACCGTC
AGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGAC-3'

(SEQ ID No: 41) e

AAAAAAGGCGCGCCGAATTCACCGGTACCAAGCTTCCAATGCAC CGTTC
CCGGCCGCGGAGGCTGGATCGGTCCCGGTGTCTTCTATGGAGGTCAAAA-3'

(SEQ ID No: 42) e preenchendo com a polimerase de Klenow.
Então elas são misturadas separadamente com o par de base
722 de produto de PCR (que contém o sinal poli SV40 A, cinco
5 GAL4 DAS, e uma TATA-box), e o PCR amplificou com dois
conjuntos de iniciadores do PCR: tanto

5' - TACGAATTCATCAGCCATATCACATTTGTAGAG-S

(SEQ ID No: 43) e

5' -AAAAAAGGATCCGAGCTCGGTACCAAGCTTCCAATGCACCGTTCCCGGC
CGCGGAGGCTGGATCGGTCCCGGTGTCTTCTATGGAGGTCAAAA-S'

(SEQ ID No: 44) ou

5' - TACGAATTCATCAGCCATATCACATTTGTAGAG-S'

(SEQ ID No: 45) e

AAAAAAGGCGCGCCGAATTCACCGGTACCAAGCTTCCAATGCACCGTTC
CCGGCCGCGGAGGCTGGATCGGTCCCGGTGTCTTCTATGGAGGTCAAAA -3'

10 (SEQ ID No: 46).

Desse modo, dois cassetes contendo uma região do
sinal de poliadenilação SV40 (para impedir a transcrição
oculta na região reguladora), cinco elementos GAL4 UAS, um
TATA-box, e um promotor CMV adiantado 5' são montados. Um é
15 digerido com EcoR I e BamH I e clonado dentro do local Mfe
I/BamH I do pHPC-4 para criar pHPC-4reg. O outro é digerido
com Asc I e clonado dentro do local BstZ 17 II Asc I de
pHPC-4reg para criar o pDlreg. Ambas as etapas de clonagem
removem os dois promotores CMV constitutivos em pHPC-4 que
20 poderia interferir com a expressão regulada. Como descrito
acima, os vários genes de interesse podem ser introduzidos

nas regiões de poliligante pDlreg, tais que podem ser integrados em um vetor alvo e sua expressão pode ser regulada.

Há duas características sobre a construção do pDlreg que pode ser importante para manter os altos níveis das versões de utilização possíveis da expressão de gene do vetor doador que não contêm componentes de um sistema regulador do gene (por exemplo, pD1, pD1-DHFR, pD1-IRA). Primeiro, o TATA-box do sistema regulador de gene foi precisamente fundido aos TATA-boxes dos promotores CMV do pD1. Segundo, 05' UTRs dos promotores CMV foram reconstituídos. A rede de resultado é que as seqüências entre o TATA-box e o códon do início da tradução (isto é, o local do início da transcrição e o 5' UTR) do pDlreg são os mesmos como são no pD1. Entretanto as seqüências antes do TATA-box no pDlreg consiste naquelas seqüências de DNA exigidas para obter a regulação do gene mediado pelos componentes de proteína do sistema regulador do gene que estão codificados pelo pDlreg.

20 CONSTRUÇÃO DE UM VETOR DE EXPRESSÃO DOADOR SELECIONÁVEL

O vetor de expressão selecionável (FIG. 21) é similar ao vetor de expressão doador exceto que também incluem um gene completo de resistência à droga, que seja diferente tanto do primeiro gene marcador selecionável do promotor menor e quanto do segundo gene marcador selecionável funcional no vetor do alvo. A construção de um vetor de expressão doador selecionável com um gene completo de resistência G418 (pD1-DTXI-G418, FIG. 21) é descrito como exemplo. A seqüência do vetor do pD1-DTXI-G418 é fornecida nas FIGURAS 36A-36D.

O vetor de expressão doador selecionável pDI-DTXI-G418 foi construído amplificando um completo, cassette funcional de resistência de droga G418 de pcDNA3002neo

(Crucell) que usa a reação em cadeia de polimerase e os iniciadores

GAGAGAGGATCCACGCGTCTGTGGAATGTGTGTCAGTTAGGG-3'

(SEQ ID No: 47) e

GAGAGAGAATTCTCTAGACAGACATGATAAGATACATTGATGAGTTTG-3'

(SEQ ID No: 48). O produto resultante do PCR contém um
 5 promotor SV40, o gene de resistência G418 e o sinal de
 poliadenilação SV40. O produto do PCR foi digerido com as
 enzimas de restrição BamH I e EcoR I e ligado dentro do
 vetor de expressão pD1-DTX-I, que tinha sido digerido com
 BgI II e Mfe I. A ligação foi digerida com BgI II e Mfe I
 10 (as quais são destruídas pela ligação da inserção) para
 reduzir a ligação da espinha dorsal do vetor sozinho e
 transformada dentro de células ultracompetentes de E. Coli
 Gold XL-10 (Stratagene). Os clones com as inserções na
 orientação desejado foram identificados pelo PCR e pela
 15 digestão da enzima de restrição. A sequência de DNA correta
 do gene de resistência G418 inteiro foi confirmada pelo
 sequenciamento.

CONSTRUÇÃO DE UM VETOR DE EXPRESSÃO DOADOR DO REPÓRTER

O vetor de expressão doador repórter (FIG. 30) é
 20 similar ao vetor de expressão doador exceto que também
 inclui um gene repórter, que possa ser detectado em células
 individuais tanto, por exemplo, microscopia de fluorescência
 ou um classificador celular ativado por fluorescência.
 Geralmente, o nível da expressão do gene repórter em um
 25 vetor de expressão doador repórter correlacionará ao nível
 da expressão de proteínas de interesse no mesmo vetor de
 expressão doador repórter. Conseqüentemente, depois que a
 transfecção do clone do vetor alvo com um vetor de expressão
 doador repórter, os clone do vetor alvo podem opcionalmente
 30 ser identificados que resultem em altos níveis da expressão
 de uma proteína de interesse identificando os clones que

expressam o gene repórter em altos níveis. Usando um instrumento elevado de produção, tal como, um classificador celular ativado por fluorescência um número muito maior de clone do vetor alvo (isto é, locais de integração) podem ser
5 selecionados para a expressão do que pode ser selecionado por métodos manuais de colheita de clone.

Em um esquema opcional um grande número de combinações de clone do vetor alvo será gerado. Por exemplo, as células transfectadas com um vetor alvo e um primeiro
10 vetor de expressão integram-se. As colônias estáveis serão selecionadas (por exemplo, pela resistência à higromicina). Por exemplo, tanto como 100 placas com 100 colônias por placa (isto é, 10.000 clones do vetor alvo) podem ser geradas. Cada combinação de clone do vetor alvo é então
15 transfectada separadamente com um vetor de expressão doador repórter e um segundo vetor de expressão integrase. A integração estável dos vetores de expressão doador repórter em vetores alvo é selecionada (por exemplo, pela resistência à puromicina). Cada associação individual dos clones do
20 vetor doador repórter é classificada usando um classificador celular ativado por fluorescência e as únicas células de cada combinação com a expressão de gene repórter mais elevada são coletadas. A expressão de nível elevado da proteína do interesse é confirmada então. O local da
25 integração do vetor alvo nas células com a expressão de gene repórter mais elevada é determinado então usando o resgate do plasmídeo ou as técnicas de PCR. Os iniciadores de PCR vetor alvo específico são projetados para serem específicos para os locais da integração do vetor alvo. Então, as
30 combinações dos clones do vetor alvo que fornecem níveis mais elevados de expressão são única célula clonada e o iniciador PCR vetor alvo-específico é usado para identificar que clone individual do vetor alvo causa os níveis mais

elevados de expressão após a transfecção com um vetor expressão doador repórter e um segundo vetor de expressão integrase. Isolando um pequeno número de clone do vetor alvo que conduz aos níveis mais elevados de expressão da
 5 proteína, outros vetores de expressão doadores podem ser transfectados nos clone identificados para expressar uma variedade de outras proteínas, ao invés de fazer a seleção da expressão da grande escala de cada vez.

Além disso, para um uso opcional descrito acima
 10 para a seleção elevada da produção de locais de integração, um simples, rápido método de monitoração do curso do tempo, frequência e uma estabilidade do vetor de expressão doador repórter em tempo real pela examinação de células transfectadas usando um microscópio de fluorescência. Como
 15 exemplo de construção de um vetor de expressão doador reportes com um gene protéico fluorescente verde (pD3-DTX1, FIG. 30) é descrito. O vetor de expressão doador repórter pD3-DTX1 foi construído primeiramente amplificando um promotor do vírus de Sarcoma de Rous (pRSV) do pLXRN do
 20 plasmídeo (Clontech) usando a reação em cadeia de polimerase e os iniciadores

TTTTCACCTGCATTGCGACAATTGTCATCCCCTCAGGATATAGTAGTTTC -3'

(SEQ ID No: 49) e

GACCAGCACGTTGCCCAGGAGTTGGAGGTGCACACCAATGTGGTG-3'

SEQ ID No: 50). Um DNA que contém a região de codificação (hrGFP) da proteína fluorescente verde Renilla reniformis
 25 humanizada e um sinal de poliadenilação do gene do hormônio do crescimento humano (hGH) foi amplificado pelo PCR do PAAV hrGFP (Stratagene) usando os iniciadores

CACCACATTGGTGTGCACCTCCAACCTCCTGGGCAACGTGCTGGTC -3'

(SEQ ID No: 51) e

5'-GAGAGAGCTAGCATTAAATAAGGACAGGGAAGGGAGCAGTGG-3'

(SEQ ID No: 52). Os 2 produtos do PCR foram misturados e amplificados com os iniciadores 5' -

TTTTCAGTGCATTCGACAATTGTCATCCCCTCAGGATATAGTAGTTTC-3'

(SEQ ID No: 53) e

GAGAGAGCTAGCATTAAATAAGGACAGGGAAGGGAGCAGTGG-S'

(SEQ ID No: 54) a fim de fundir o promotor do vírus de
5 Sarcoma de Rous à região da codificação do hrGFP e ao sinal de poliadenilação do gene do hGH. O produto terminado-sem corte resultante do PCR foi ligado a um local blunt Psi I do vetor de expressão doador pD1-DTX1. Os clones com inserções foram identificados pelo PCR usando os iniciadores

TTTTCAGTGCATTCGACAATTGTCATCCCCTCAGGATATAGTAGTTTC-3'

10 (SEQ ID No: 53) e

GAGAGAGCTAGCATTAAATAAGGACAGGGAAGGGAGCAGTGG-S'

(SEQ ID No: 54) e a orientação da inserção foi determinada pela digestão da enzima de restrição. A correta sequência de DNA da poli A pRSV-hrGFP-hGH inteiro inserida foi confirmada. A sequência do vetor pD3-DTXI é fornecida nas

15 FIGURAS 37A-37D.

TESTAGEM DOS VETORES

As funções do vetor alvo individual, do vetor de expressão doador e dos vetores de expressão integrase foram testadas. Por exemplo, a transfecção do transporte do vetor
20 alvo nas células DG44 ou nas células PER.C6™ pode conferenciar resistência à higromicina. Quando tanto o vetor expressando a integrase R4 ou o vetor expressando ϕ C31 integrase é transfectado com o vetor alvo aproximadamente 5 vezes tanto podem colônias resistentes a higromicina
25 resultaram comparado a transfecção do vetor alvo sozinho mostrando que a expressão de um ou outra integrase pode conduzir a um número aumentado de clones estáveis. A

transfecção transiente do vetor de expressão doador sozinho conduziu a uma produção de anticorpo de 300 ng/ml nas células DG44 e de 1 ug/ml em PER.C6™ (FIG. 31).

Outra função importante a demonstrar é a habilidade do local attP ϕ C31 em um vetor alvo para recombinar com o local attB ϕ C31 em um vetor de expressão doador. Isto é particularmente verdadeiro visto que os locais att tanto no vetor alvo e no vetor doador foram tanto transformados ou truncados para encontrar as demandas do sistema de expressão descrito aqui. As células DG44 (3e6) em placas de 10 em foram transfectadas com 500 ng de um vetor alvo (PRI) e 500 ng de um vetor de expressão doador (pD1-DTX-I) na presença ou na ausência de 4000 ng de uma integrase ϕ C31 que expressa o vetor (pCS-M3J) usando o Lipofectamine DC 2000. Quarenta e oito horas depois da transfecção, as células foram tripsinizadas e o DNA do plasmídeo foi isolado usando um kit Miniprep Spin QIAprep (QIAGEN). O DNA foi amplificado com iniciadores PCR

TGCCCCGGGGCTTCACGTTTTCC-S'

(SEQ ID No: 55) (de att P ϕ C31) e

GCCCGCCGTGACCGTCGAGAAC-3'

SEQ ID No: 56) (de att B ϕ C31), então com iniciadores

5'-CAGGTCAGAAGCGGTTTTCGGGAG-S'

(SEQ ID No: 57) (de att P ϕ C31) e

5'-CCGCTGACGCTGCCCCGCGTATC-3'

(SEQ ID No: 58) (de att B ϕ C31), todos dos quais foram projetados para especificamente amplificar o produto attR que poderia resultar somente da recombinação mediada por

ϕ C31 integrais e- de um local attP ϕ C31 em um vetor alvo com um local attB ϕ C31 em um vetor de expressão doador. Como um controle positivo 500 ng cada dos plasmídeo P_{Ta}-attB e P_{Ta}-attP o qual continha seqüências dos locais att ϕ C31 tipo

selvagem, mais longas foram transfectadas na presença ou na ausência de 4000 ng de um vetor ϕ C31 integras e (pCS-M3J). PTa-*attB* e PTa-*attP* têm 285 e 221 regiões longas dos pares de bases dos locais *attB* ϕ C31 e dos locais *attP* ϕ C31, respectivamente. Enquanto um controle negativo não-transfectado das células foi usadas. Como pode ser visto na FIG. 22 pR1 e pD1-DTX-I pode recombinar para gerar um local *attR* somente na presença da ϕ C31 integrase.

As funções do vetor alvo, do vetor doador e de ambos os vetores de expressão integras e foram testados de uma vez por transfecção e pela seleção das células PERC6™ ou DG44 como diagramada na FIG. 11, antes que um grande número de linhagem celular estáveis individuais fossem geradas. Esta experiência foi feita somente uma vez no curso do desenvolvimento da metodologia ou como necessário, por exemplo, se as variações do alvo, do doador ou dos plasmídeo integrase são construídas. Subseqüentemente somente os vetores de expressão doadores, os quais codificam outras proteínas de interesse transfectada transiente para testar a expressão da proteína de interesse e para confirmar se o vetor doador é capaz de expressão.

O pR1 do vetor alvo foi co-transfectado com um plasmídeo expressando R4 integras e (pCMV-sre) dentro de células PERC6™ ou DG44 por lipofecção usando Lipofectamine CD 2000 (Invitrogen) de acordo com as instruções do vendedor. As células foram incubadas então por quarenta e oito horas para permitir a expressão da proteína R4 integrase, a qual media a integração local-específica entre R4 o local *attB* 295 presente no vetor alvo e os locais pseudo *attP* R4 presentes no cromosoma (FIGS. 3 e 11). As colônias que contêm um vetor alvo integrado foram selecionadas então em higromicina contendo meios (por exemplo, DMEM, 10% de soro bovino fetal, 10 mM de MgCl₂ para

PERC6TM e o F-12, 5 % de soro bovino fetal, 30 /-1M de timidina para DG44). As colônias resistentes a higromicina, únicas, foram isoladas e selecionadas para a sensibilidade a puromicina.

5 A higromicina resistente, clones sensíveis do vetor alvo à puromicina foram co-transfectados outra vez com um vetor doador (por exemplo, pD1-DTX-1) contendo o local AAA attB 285 ϕ C31 e um cassete de expressão codificando genes de interesse, tais como, as cadeias pesadas e leves de
10 um anticorpo específico humano para a toxina da difteria e da expressão que codifica uma ϕ C31 integrase alterada (por exemplo, pCS-M3J). A proteína da integrase ϕ C31 alterada media a integração local-específica entre o local AAA attB 285 ϕ C31 presente no vetor doador e o local attP 103 ϕ C31
15 projetado no cromossoma da linhagem celular usando o vetor alvo (FIGOS. 4 e 11).

Uma combinação estável de células puromicina-resistentes é isolada como segue. Quarenta e oito horas após a segunda transfecção os meios regulares de crescimento da
20 célula foram substituídos com os meios de crescimento da célula que contêm a puromicina (1 ug/ml para PER.C6TM, 10/-1g/ml para DG44). Os meios de contenção foram trocado a cada 2-3 dias por 7 dias (células DG44) ou 14-21 dias (células de PER.C6TM), ou até o número de colônias crescentes tomar-se
25 estáveis.

Neste ponto, todas as colônias foram tripsinizadas e foram combinadas. As células foram recolocadas e permitidas a serem unidas por 24 horas. A seleção para a resistência à puromicina foi continuada por um total de pelo
30 menos 21 dias a permitir os vetores de expressão não-integrados serem diluídos. O nível de expressão da proteína de interesse (por exemplo, codificando um anticorpo) foi analisado então para confirmar a função dos vetores de

expressão integras e do vetor alvo e dos vetores doadores. Para a expressão de medição do anticorpo, um ensaio específico para IgG humano (por exemplo, o Easy Titer IgG Assay, Pierce, Inc.) foi usado.

5 O vetor alvo não pode integrar ou pode integrar aleatoriamente em diferentes posições dos locais pseudo *attP* R4. Mesmo nestes casos o vetor doador pode ainda integrar-se no vetor alvo para reconstituir um gene de resistência completo à puromicina. O número de colônias de puromicina
10 que seria esperado que resultasse destes eventos é muito mais baixo do que aqueles que ocorrem em consequência da integração de um vetor doador em um vetor alvo que seja, por sua vez, local-específico integrado usando a integrase R4. Isto é porque os vetores não-integrados seriam perdidos ao
15 longo do processo de seleção. A integração aleatória de um vetor alvo ocorrerá em uma frequência muito mais baixa do que a integração local-específica mediada pela integrase e R4. Para também documentar que os níveis de expressão da proteína medidos nesta experiência são primeiramente um
20 resultado da integração local-específica inicial do vetor alvo, uma experiência de controle é feita em que o vetor de expressão integras e R4 é omitido.

É desejável executar a etapa da seleção da resistência à puromicina para assegurar-se de que o trabalho
25 como essa etapa é o local-específico da chave que integra o vetor de expressão doador. A integração do local *attB* ϕ C31 dentro do vetor doador no local *attP* ϕ C31 no vetor do alvo conduz à criação de um local *attL* ϕ C31, em que este exemplo específico é 88 bases ao longo. Esta sequência adicional
30 estará presente na região 5' não-traduzida do mRNA codificando a resistência à puromicina. Visto que o efeito desta sequência adicional na transcrição, na estabilidade do mRNA, na tradução e daqui finalmente no nível de resistência

à puromicina que pode ser alcançado não podem ser previstos unicamente das seqüências do ácido nucléico, os vetores devem ser testados como descrito acima para assegurar as funções reconstitutivas do cassete de resistência à puromicina a um grau que permita a seleção eficiente das células em que o vetor doador tenha integrado dentro do vetor destinatário.

EXEMPLO 2

CONSTRUÇÃO DA LINHAGEM CELULAR EXPRESSANDO PROTEÍNA

10 A seguinte protocolo foi seguido para a construção de linhagem celular expressando proteína. As células de CHa/dhfr (por exemplo, células DG44 e células de PER.C6™) foram transfectadas usando Lipofectamine CD 2000 em placas de 10 em como segue:

15 1. A primeira transfecção foi feita com 500 ng do pR1-DHFR do vetor alvo e 5000 ng do pCMV-sre do plasmídeo integrais e R4 (FIG. 11) por placa de 10 cm.

20 2. As células foram crescidas por 48 horas no meio regular (Ham's F-12, 5% de soro bovino fetal, 30 μ M de timidina).

3. As células foram então tripsinizadas e plaqueadas em placas de 96 poços em meio seletivo, que foi um meio regular que contém 400 μ g/ml de higromicina B. Sob estas circunstâncias, aproximadamente 30 únicos clone da célula cresceram em cada um de cinco placas de poços.

4. Aproximadamente 7-8 dias após a transfecção, quando as colônias são primeiramente visíveis pelo olho, os clones individuais foram trisinizados e transferidos a um número mínimo para a placa de 96 poços. Um total de 165 clones foi selecionado e consolidado em duas placas de 96 poços.

5. As colônias selecionadas foram expandidas em um conjunto de triplicata da placa de 96 poços. Um conjunto

foi resfriado e estocado na fase de vapor de nitrogênio líquido. O terceiro conjunto foi para a segunda transfecção.

6. Um conjunto de colônias CHO foi expandido a 24 placas poços e co-transfectada com 15 ng de pD1-DTX1-G418, vetor de expressão doador selecionável, e 150 ng de pCS-M3J, o plasmídeo integrase ϕ C31 mutante (FIG. 11).

7. As células foram crescidas por 48 horas no meio regular que contém 400 ug/ml de higromicina B.

8. As células foram crescidas então em meio seletivo que contém 10 ug/ml de

puromicina. Após 7-21 dias de números variáveis da seleção de colônias crescidas, dependendo de que linhagem celular parental attP foi transfectada.

9. As colônias foram tripsinizadas e então combinadas. A metade foi plaqueada em meio contendo 10 ug/ml de puromicina e metade foi plaqueada em meio contendo 10ug/ml de puromicina e 400 ug/ml de G418.

10. Os meios seletivos foram trocados a cada 2-3 dias até que os poços estivessem confluentes. As combinações dos clones que cresceram em puromicina e em G418 foram expandidas a 6 placas de poços e testadas para a produtividade de IgG (pg IgG produzido/célula/dia).

11. Dos 165 clones parentais do vetor alvo-DHFR, 132 foram sensíveis à puromicina e foram usados para a segunda transfecção. Destes 96 clones restantes produzidos de puromicina e foram testados para a produção de IgG. Dos 96 clones, 14 produziram IgG a níveis detectáveis.

12. A combinação (2G7-G) com o nível mais elevado de expressão (β -8 pg/célula/dia) foi crescida em meios seletivos para tanto o gene DHFR e quanto para o vetor de expressão doador selecionável (MEMa-a, soro fetal bovino diluído a 7%, 400 ug/ml de G418) por 6 dias e então plaqueados em 1 célula por poço em duas placas de 96 a fim

de isolar o clone.

13. Um total de 56 clones foi obtido e a produtividade de IgG foi medida. Os resultados são mostrados nas FIGURAS 28A e 28B. Três clones foram identificados como tendo os níveis de produtividade médios que são considerados de extremidade alta (isto é, > 30 pg/célula/dia).

14. Outra combinação (2H9-G), em que o gene DHFR foi mostrado para ser ligado aos genes do anticorpo por métodos de resgate do plasmídeo, foi sujeita à amplificação do gene DHFR. As células crescidas em meios seletivos para o gene DHFR e o vetor de expressão doador selecionável (MEMa-, 7% dialisado do soro bovino fetal, 400 ug/ml de G418). O gene DHFR foi amplificado então adicionando quantidades crescentes de metotrexato ao meio. A concentração inicial foi de 2 uM e a concentração foi tipicamente aumentada 2 a 3 dobras sobre cada 10-14 dias.

15. As produtividades de IgG da combinação 2H9-G selecionada em várias concentrações de metotrexato foram medidas e os resultados são mostrados na FIG. 29. Em 200 nM de metotrexato um aumento acentuado na produtividade foi observado a um nível igual àquele dos clones 2G7-G expressando os mais elevados. Entretanto enquanto poderia tomar aproximadamente 1 mês para se isolar a maior expressão de clones 2G7-G usando a integração local- específica, ele tomaria aproximadamente 4 meses para isolar uma combinação 2H9-G de expressão elevada usando a amplificação do gene.

PRIMEIRA INTEGRAÇÃO

A fim de criar um local único específico para a integração de um vetor de expressão da proteína e identificar locais attP Ψ R4 nos genomas da linhagem celular que são apropriadas para a produção de nível elevado, reprodutível de proteínas o pR1 do vetor alvo ou o pR1-DHFR do vetor alvo- DHFR foram integrados em um grande número de

diferentes locais attP Ψ R4 nas células PERC6™ e DG44. O vetor alvo ou o vetor alvo-DHFR foram misturados com o pCMV-sre do vetor de expressão integrase R4 e transfectados nas células PERC6™ e DG44 por lipofecção de acordo com o as
5 instruções de vendedor. Os reagentes Lipossômicos apropriados para a lipofecção incluem Fugene 6 (Rouche Applied Science), Lipofectamine CD 2000 (Invitrogen) e similares. As células foram incubadas por quarenta e oito horas para permitir a expressão da integrase e da integração
10 do pR1 ou do pR1-DHFR nos locais attP Ψ R4 a ocorrer. O meio de crescimento regular da célula é substituído então com o meio de crescimento seletivo contendo 100 ug/ml (para células de PERC6™) de higromicina B de 400 ug/ml (para as células DG44) (Calbiochem). O meio do crescimento da célula
15 foi substituído a cada 2-3 dias por 7-14 dias ou até um número máximo as colônias são visíveis. Um total de 100 colônias, que fosse estimado para representar aproximadamente 50 locais diferentes do attP Ψ R4 foi escolhido e expandido para a segunda integração. Cada clone
20 da célula isolado nesta etapa é referido como uma linhagem celular attP de PERC6™ ou uma linhagem celular do attP de DG44.

As seqüências adjacentes aos vetores alvo integrados foram determinadas para mostrar que estiveram
25 integradas por um mecanismo R4 mediada-integrase. Para fazer isto um método de "plasmídeo resgatado" foi usado que envolve as seguintes etapas. O DNA genômico foi preparado dos clones do vetor alvo e digerido com Afl III ou Nsi I (New England Biolabs). Estas enzimas cortam o vetor alvo
30 perto da origem de replicação, mas não cortam em nenhum outro local entre a origem de replicação e um local attL Ψ R4 (veja FIG. 12). Mais importante ainda eles também não cortam dentro da origem de replicação e do gene de

resistência à ampicilina, que são exigidos para o resgate bem sucedido do plasmídeo em *E. Coli*. O DNA digerido foi ligado em baixa concentração (β -10 ng/ml) e eletroporado então em células TOP10 (Invitrogen). O DNA de Miniprep foi isolado das colônias resultantes e seqüenciado com um iniciador que corresponde ao filamento de anti-sentido da região de codificação da puromicina, tal que a seqüência obtida estenderia da região de codificação da puromicina através do local attP ϕ C31 e então no local attL Ψ R4. Como mostrado na FIG. 23, o plasmídeo resgatado de dois clones do vetor alvo conteve seqüências até a seqüência do núcleo do local att R4 e estendeu-as então no DNA cromossômico. A seqüência do núcleo do local att R4 foi suprimida em cada caso, como ocorre freqüentemente quando integrases de serine recombina um local att tipo selvagem com um local Ψ att.

Os métodos de PCR semi-aleatórios podem também ser usados para determinar seqüências nas junções entre vetores alvo e o DNA cromossômico. Por exemplo, o DNA Walking SpeedUp Kit (Seegene) pode ser usado com esta finalidade. O "iniciadores alvos-específicos" poderiam estar localizados no gene de resistência à puromicina para isolar uma seqüência que contém o local attL Ψ R4 ou na área poli A HSK TK para isolar uma seqüência que contém o local attR'P R4.

Alternativamente, métodos de "PCR inverso" podem ser usados. Nestes métodos o DNA genômico é digerido com uma enzima de restrição que não corta dentro da região de interesse. O DNA é ligado para formar um DNA circular. O DNA ligado é amplificado então pela reação em cadeia de polimerase usando iniciadores alojados em seqüências conhecidas. A orientação dos iniciadores é invertida relativa ao que poderia ser em um PCR normal, tais que, as seqüências através do ponto de ligação são amplificadas.

Antes da segunda integração, as linhagem celulares

do *attP* são selecionadas para a sensibilidade a puromicina. Uma seleção de resistência à puromicina é usada para selecionar a segunda etapa de integração e assim é útil para assegurar-se de que os clones do vetor alvo ou do vetor
 5 alvo-DHFR sejam obtidos na primeira integração sejam sensíveis à puromicina. Nós encontramos que cerca de 10% do vetor alvo ou do vetor alvo-DHFR os clone podem ser puromicina sensível, dependendo da linhagem celular. Visto que a eficiência da integração é aproximadamente 0, 1-1 % se
 10 um clone de resistência à puromicina foi tranfectada poderia ser prevista que somente 0, 1-1 % das células expressariam as proteínas de interesse e visto que as células já foram resistentes à puromicina poderia não ser possíveis a enriquecer para as células expressando as proteínas. Outra
 15 aproximação para contornar este problema, além da seleção dos clones do vetor alvo para a sensibilidade a puromicina após a primeira transfecção, seria usar um vetor de expressão doador selecionável na segunda transfecção.

SEGUNDA INTEGRAÇÃO

20 A fim de testar a habilidade de cada local *attP* Ψ R4 que o vetor alvo integrou dentro de uma primeira integração para permitir a expressão de nível elevado da proteína, uma segunda integração de um vetor de expressão doador é feita. Um vetor doador que codifica um anticorpo da
 25 toxina da anti-difteria (pD1- DTX-I) foi misturado com o vetor da expressão integrase ϕ C31 mutante (pCS-M3J) e transfectado em cada *attP* de PER.C6™ ou linhagem celular *attP* DG44 gerado na primeira transfecção por lipofecção de acordo com o instruções do vendedor. Os reagentes
 30 Lipossômicos apropriados para a lipofecção incluem Eugene 6 (Rouche Applied Science), Lipofectamine CD 2000 (Invitrogen), e similares. As células foram incubadas por quarenta e oito horas para permitir a expressão da integras

e ϕ C31 mutante e a integração de pD1-DTX-1 no vetor alvo a ocorrer. O meio de crescimento regular foi substituído então com o meio de crescimento seletivo contendo 1 μ g/ml de puromicina (para PER.C6™) ou 10 μ g/ml (para DG44) (Calbiochem). As células de meio de crescimento contendo puromicina foram substituídas a cada 2-3 dias por 7 -14 dias ou até um número máximo das colônias ficarem visíveis. As colônias levantadas de cada transfecção foi tripsinizada, expandida, congelada para o armazenamento da fase de vapor de nitrogênio líquido.

As seqüências ao redor da junção dos vetores de expressão alvo e doador foram determinadas para mostrar que eles foram recombinados por um mecanismo mediado por ϕ C31 integrase. Para fazer isto um método de "resgate de plasmídeo" foi usado envolvendo as seguintes etapas. O DNA genômico foi preparado das combinações transfectadas com os vetores de expressão integrase ϕ C31 mutante e doador. O DNA foi digerido com Tfi I (New England Biolabs). Esta enzima corta o vetor de expressão dentro do gene do anticorpo da cadeia pesada e o vetor alvo perto da origem de replicação, mas não corta em nenhum outro local entre estas áreas (veja FIG. 13). Mais importante ainda Tfi I não corta dentro da origem de replicação ou o gene de resistência à ampicilina, que são exigidos para o resgate bem sucedido do plasmídeo em E. Coli. O DNA digerido foi ligado em baixa concentração (β -1 0 ng/ml) e eletroporado então em células TOP10 (Invitrogen). O DNA minipreparado foi isolado das colônias resultantes e seqüenciado com um iniciador que corresponde ao filamento de anti-sentido da região de codificação de puromicina, tais que, a seqüência obtida estenderia da região da codificação de puromicina (do vetor alvo) através do local attL88 ϕ C31 (junção entre os vetores alvo e doador recombinado), e então no sinal de poliadenilação do hormônio

do crescimento bovino (vetor doador). Como mostrado na FIG. 24A e FIG. 25A a seqüência dos plasmídeos resgatados das células DG44 e PER.C6™ foram como previsto se a integrase ϕ C31 integrou corretamente o vetor de expressão doador dentro do vetor alvo. As seqüências que cercam os locais attR ϕ C31 foram determinadas em uma maneira similar e também encontradas para ser exatamente como previsto (FIG. 24B e FIG. 25B).

Os métodos PCR-baseados foram desenvolvidos também para permitir a determinação rápida dos tipos de integrações que possam estar presentes nos clones ou nas combinações dos clones. No que diz respeito a três tipos de integração vetor de expressão doador são possíveis: aleatório, vetor alvo ou local Ψ att. Para detectar a integração aleatória, os iniciadores do PCR específicos para o local attB ϕ C31 no vetor de expressão doador foram projetados. Na maioria dos casos de integração aleatória, o pequeno local attB (285 pares de bases) seria intacto, visto que se a integração do vetor doador dentro do vetor alvo ou um local Ψ att tenha ocorrido, o local attB seria interrompido. O DNA genômico de 6 combinações de clone em que o vetor doador tinha sido integrado foi preparado. Um micrograma de DNA foi sujeito à reação em cadeia de polimerase usando iniciadores

5'-CATCTCAATTAGTCAGCAACCATAGTC-3'

(SEQ ID No: 59) e

5'-AAGCTCTAGCTAGAGGTCGACGGTA-3'

(SEQ ID No: 60) para 30 ciclos e 1 % desse DNA da reação foram sujeitos então à reação em cadeia de polimerase usando iniciadores

5'-GTCGACGAAATAGGTCACGGTCTC-S'

(SEQ ID No: 61) e

TACGTCGACATGCCCGCCGTGACC-3'

(SEQ ID No: 62) para 30 mais ciclos. Os produtos do PCR foram separados em um gel do agarose a 4% e os resultados são mostrados na FIG. 26A. A evidência para a integração aleatória do vetor de expressão doador foi ausente de duas
 5 combinações (2G7, 2HI0), mas presente em quatro combinações (2B11, 2G11, 2H9G, 2H9P).

Para detectar a presença de integração em um vetor alvo, uma região que contém o local attR ϕ C31 híbrido foi amplificado por PCR diretamente nas células. Os vários
 10 números de células tripsinizadas da combinação 2H9G foram usados. A combinação 2H9G das células foi derivada da transfecção de um clone do vetor alvo DG44 (pR1 -DHFR) (2H9) com um vetor de expressão doador (pD1-DTXI-G418) e um vetor integrase ϕ C31 mutante (pCS-M3J). As células foram
 15 selecionadas em puromicina para um mês e então em G418 por um mês. As células tripsinizadas foram sujeitadas à amplificação do PCR usando iniciadores

5'-TGCCCCGGGGCTTCACGTTTTCC-S'

(SEQ Id No: 64) e

GCCCGCCGTGACCGTCGAGAAC-3'

(SEQ Id No: 65) para 30 ciclos e 1 % desse DNA da reação
 20 foram sujeitados então a um círculo subsequente de amplificação de PCR usando iniciadores

5'-CAGGTCAGAAGCGGTTTTTCGGGAG-S'

(SEQ Id No: 63) e

CCGCTGACGCTGCCCCGCGTATC-S'

(SEQ ID No: 66) para mais 30 ciclos. Os produtos do PCR foram separados em um gel do agarose a 4% e os resultados
 25 são mostrados na FIG. 26B. Um sinal específico do tamanho correto foi amplificado quando células 102, 103, ou 104 foram usadas.

Os métodos semi-aleatórios do PCR podem ser usados

para determinar se um vetor doador integrou em um local att
 Ψ φ C31. Por exemplo, o DNA Walking SppedUp Kit (Seegene)
 pode ser usado com esta finalidade. Alternativamente o
 método inverso do PCR pode ser usado.

5 Os níveis da produção do anticorpo foram testados
 como segue. Um número conhecido de células foi plaqueado em
 uma placa de 6 poços em meios de MEMu- (Invitrogen) com 7% de
 soro fetal bovino dializado (Invitrogen) para células de CHO
 DHFR- ou DMEM (Invitrogen), soro fetal bovino a 10% (JRH),
 10 10 mM de MgCl₂ para células de PER.C6™. As células foram
 colocadas para crescer por 1-4 dias. Os meios foram colhidos
 e ao mesmo tempo o número final de células foi determinado.

O número da célula foi determinado usando um
 hemocitômetro. Alternativamente, kit de ensaio MTT-baseado
 15 (Cell Titer 96 kit, Promega) ou kits similares podem ser
 usados para determinar o número de células na placa.
 Instrumentos, tais como, o ViaCount Assay (goiaba) que pode
 medir o número de células aderentes em uma placa também está
 disponível.

20 A concentração de IgG nos meios foi determinada
 usando o kit de ensaio (Pierce) Easy- Titer Human Ig (H+L)
 que mede especificamente todas as classes de IgG humano. A
 produtividade específica (anticorpo/célula/dia de picograms)
 foi calculada da seguinte equação:

$$25 \quad \frac{\text{Anticorpo pg/ml} \times \text{ml do meio colhidos (número de célula} \\ \text{final} + \text{número inicial de célula)} / 2}{\text{número de dias de anticorpo que foi produzido}}$$

Os resultados de linhagem celular do attP da
 seleção 100 PER.C6™ e de 100 linhagem celular do attP DG44
 30 são mostrados na FIG. 27 A e FIG. 27B, respectivamente.
 Dezesesseis linhagens celulares de attP DG44 causaram
 combinações de clone resistentes a puomicina com expressão
 detectável e da melhor combinação produzida aproximadamente

8 pg anticorpo/célula/dia (FIG. 27 A). Dezesete linhagens celulares attP de PER.C6™ causaram combinações de clone resistentes a puromicina com expressão detectável e da melhor associação produzida de aproximadamente 4 pg anticorpo/célula/dia (FIGO. 27B).

Freqüentemente, as combinações dos clones irão conter as células que variam extremamente nos termos de expressão da proteína. Conseqüentemente, nós subclonamos combinações de produção elevadas a fim de identificar linhagem celular específica dentro das combinações que fornecem um nível elevado da expressão da proteína. A combinação derivada da transfecção de linhagem celular do attP DG44 com o vetor de expressão doador exibiu o nível mais elevado da expressão (2G7) subseqüentemente foi clonada limitando a diluição em 96 placas de poços e analisada para a produtividade do anticorpo como descrita acima. Os resultados são mostrados na FIG. 28. Dentro da combinação, que produziu 7.6 pg/célula/dia, são os clones que variam na produtividade de 0,2 a 38 pg/célula/dia. Três clones produziram mais de 30 pg/células/dias.

As células que expressam altos níveis das proteínas são freqüentemente uma desvantagem de crescimento e podem conseqüentemente ser perdidos ou sub-representados quando expandidos como descritas acima como parte de uma combinação. Um método para contornar este problema é como segue. Após a transfecção com o vetor de expressão doador e o vetor integrase ϕ C31, as células são incubadas por 48 horas para permitir que a integração ocorra. As células transfectadas são tripsinizadas e plaqueadas então em placas de 96 poços, tais que, as únicas colônias crescerão em aproximadamente 30% dos poços. O número de células transfectadas que são plaqueadas por poço depende da eficiência de plaqueamento e da eficiência da integração do

vetor doador. Geralmente para obter o número máximo de clones únicos da célula em uma placa de 96 poços aproximadamente 0,3 células com viabilidade 100% são plaqueadas por poço. Assim, por exemplo, se a eficiência de plaqueamento de uma célula é 50% e 0,1 % das células submetem-se a um evento de integração em que os resultados em uma célula resistente a puromicina plaqueariam as células $0,3/0,5/0,001=6000$ por poço depois da transfecção a fim de obter clone. Se a eficiência de integração é muito alta pode ser preciso transfectar poucas células.

As linhagens celulares attP PER.C6™ parental ou attP DG44 que resulta ao número mais elevado de clone com os níveis mais elevados da expressão da proteína são escolhidos para ser usados como as linhagens celulares attP integrando outros vetores de expressão doadores e produzindo outras proteínas em altos níveis. Aquelas linhagens celulares são usadas repetidamente e somente um pequeno número «50) dos clones são gerados e selecionados para identificar aqueles com os níveis mais elevados de expressão. Este esquema trabalhará para a expressão de uma variedade de proteínas, mostrando que a habilidade de conseguir níveis elevados de expressão pela integração em um local não é específica à expressão do anticorpo. Este método conserva uma quantidade de tempo substancial comparada aos métodos que são usados atualmente que pode exigir centenas ou milhares de clone de todo tempo na proteína diferente são produzidos. Além disso, integrando cassetes de expressão nos mesmos locus cada vez que a estabilidade dos genes e da expressão das proteínas codificados por aqueles genes é mais previsível comparada aos métodos que são usados atualmente em que a estabilidade da expressão do gene e da proteína é freqüentemente altamente variável, e em consequência pode exigir a seleção de clones adicionais e de ensaios demorados para identificar

aquelas linhagens celulares que são bastante estáveis para ser úteis. Este método também elimina os métodos de amplificação do gene que são freqüentemente usados para impulsionar a expressão se uma linhagem celular que tem um
5 nível elevado da expressão da proteína não é obtida. Tais métodos de amplificação do gene, tais como, aqueles que utilizam o gene dihidrofolato reductase ou o gene da glutamina sintetase, tomam freqüentemente 3-6 meses para conseguir níveis elevados de expressão e em muitos casos a
10 expressão não pode ser estável.

Diversas características da configuração cromossômica que resulta quando o vetor doador é integrado no vetor alvo valem a pena observar (FIGS. 11-13). Primeiramente, todos os promotores estão no mesmo ou nas
15 orientações de oposição para evitar gerar os transcritos e o siRNA antisentido que possa reduzir a expressão de gene. Em segundo, uma configuração dupla do promotor CMV iguala a expressão das cadeias pesadas e leves de um anticorpo. Isto é importante porque freqüentemente quando há um
20 desequilíbrio na expressão da cadeia pesado ou o conjunto apropriado da cadeia leve não ocorre ou são degradadas. Em terceiro lugar, os locais AAA attB 285 ϕ C31 e attP 103 ϕ C31 foram projetados de modo que quando recombina-se uma pequena base 88 ao do local attL ϕ C31, não contendo nenhum
25 códons de início a montante da tradução, resultasse. O comprimento curto attL 88 ϕ C31, que está presente no 5' UTR da resistência à puromicina da codificação do mRNA, minimiza a interferência com expressão da resistência à puromicina.

Outra configuração exemplificativa inclui um em
30 que o local attL ϕ C31 termina acima de um íntron. Para gerar esta configuração o vetor doador é construído para conter (em ordem) um promotor, metade do N-terminal da região da codificação de um gene de resistência da droga e 5' metade

de um íntron que precede um local *attB* ϕ C31. O vetor alvo é construído então para conter (em ordem) o 3' metade de um íntron, a metade C-terminal da região de codificação de um gene de resistência à droga, e um sinal poli A seguido de um
5 local *attP* ϕ C31. Após a integração do tal vetor doador dentro do tal vetor alvo inteiramente - um cassete funcional de expressão de resistência à droga é reconstituído em que consiste em um promotor, na região completa de codificação de um gene de resistência à droga e em um sinal poli A. O
10 local *attL* ϕ C31 estará presente no íntron.

A informação extensiva está disponível sobre as seqüências de nucleotídeo em um íntron são exigidas para que a emenda apropriada ocorra. Por exemplo, as seqüências aproximam as junções 5' e 3' do éxonlíntron e um intervalo
15 de polipirimidina que seja localizado tipicamente a cerca de 30 bases da extremidade 5' a 3' do íntron é exigida para que a emenda eficiente ocorra. Conseqüentemente, nas configurações descritas acima o *attB* no vetor doador e *attP* no vetor alvo são colocados no meio de um íntron pelo menos
20 100 bases de uma ou outra extremidade do íntron de modo que o local resultante *attL* seja no meio do íntron longe de todas as seqüências e nucleotídeos que forem críticas para que a emenda apropriada ocorra. Isto assegurar-se-á de que o local resultante *attL* seja muito pouco susceptível de
25 interferir com a emenda. Além disso, o íntron pode ser longo (> 1KBP) para minimizar mais o potencial que o local *attL* interferirá com a emenda.

Métodos para a caracterização da linhagem celular.

Diversos procedimentos podem ser executados para
30 caracterizar o cassete do gene que está presente dentro e as proteínas que são produzidas pelas linhagens celulares derivadas usando os métodos descritos acima. O cassete do gene é caracterizado para determinar onde o cassete

integrado e para assegurar a estrutura prevista é atual e estável sobre o tempo. A proteína que estão sendo produzidas pela linhagem celular é caracterizada também para assegurar-se de que seja atual, ativa, e que a produção de nível
5 elevado seja estável sobre o tempo.

Para caracterizar o número de locais de integração e de sua posição um número de métodos está disponível. Em algumas incorporações, a hibridação de fluorescência in situ (FISH) é usada para determinar o número de locais da
10 integração no genoma inteiro. A posição de locais de integração é determinada isolando e seqüenciando o DNA cromossômico que flanqueia o cassete integrado e comparado à seqüência do genoma humano inteiro (veja, por exemplo, Chalberg, et al, 2006).

15 O cassete integrado inteira é isolado em dois fragmentos por um método "resgate de plasmídeo" cada mês de modo que o cassete seja arquivado caso seja desejável fazer uma análise retrospectiva. Em breve, o resgate do plasmídeo envolve preparar o DNA genômico das linhagens celulares, digerindo-o com as enzimas de restrição que cortam uma vez
20 no cassete integrado e uma vez no DNA genômico, tais que, o fragmento de DNA terá uma origem de replicação e um marcador selecionável apropriado para a manutenção e a seleção em E. Coli. O DNA digerido é ligado e usado para transformar E.
25 Coli. Todo o DNA que contiver uma origem de replicação em E. Coli (por exemplo, CoIE1) e de réplicas selecionáveis de um marcador (por exemplo, resistência à ampicilina) é então "resgatado". O cassete de DNA que resulta da integração do vetor alvo em um local attP Ψ R4 e então subsequente em
30 uma integração do pDI do vetor doador dentro do vetor alvo integrado terá duas origens de replicação de E. Coli e dois marcadores selecionáveis. Diversas enzimas de restrição cortam entre estas seqüências uma vez e permitem assim o

resgate de DNAs que contêm os vetores alvo e doador separadamente. Usando este método a integridade e a estabilidade do cassete de expressão sobre o tempo pode ser determinada. Por exemplo, o cassete inteiro (kbp β -14) pode
5 ser seqüenciado para confirmá-lo se tem a seqüência e o arranjo pretendidos de elementos de DNA.

Se o local de restrição no DNA cromossômico é demasiadamente longe do cassete integrada para gerar um DNA
pequeno o bastante para ser replicado em E. Coli, o resgate
10 do plasmídeo pode ser mal sucedido. Em tais incorporações, a reação em cadeia de polimerase é usada para analisar o cassete integrado. Diversas enzimas e circunstâncias estão disponíveis, tais que, o cassete inteiro integrado β - 14 kbp pode ser amplificado e armazenado sem uma clonagem
15 suplementar. Se for desejável obter as seqüências de flaqueamento, o DNA cromossômico tem um número de métodos disponíveis, como o PCR ou as aproximações inversas que usam iniciadores aleatórios para amplificar as seqüências cromossômicas flaqueadas.

20 Além disso, para a determinação de quais genes estão presente é também desejável assegurar-se de que os vetores integrase não integrem-se no genoma. Isto é porque a expressão persistente da integrase poderia conduzir à instabilidade dos cassetes integrados do vetor alvo e doador
25 ou à instabilidade do DNA cromossômico pela recombinação da mediação entre os locais att Ψ presentes no genoma. Os vetores integrase estáveis foram observados após uma transfecção transiente, mas são raros. Entretanto, em algumas incorporações pode ser desejável comandar a presença
30 de vetores integrase nas linhagens celulares. Todos os métodos apropriados para detectar a presença ou a ausência de ácidos nucleicos específicos, tais como, Southern blotting ou a reação em cadeia de polimerase, podem ser

usados para determinar se os vetores integrase estão presentes. Alternativamente os métodos, tais como, Western blotting ou ELISA, os quais detectam a presença de uma proteína integrase, podem ser usados.

5 CARACTERIZAÇÃO DE PRODUÇÃO DA PROTEÍNA

Além disso, para a caracterização dos cassetes do gene integrado, a qualidade, a estabilidade e o nível da produção da proteína (por exemplo, produção do anticorpo) são também caracterizados. Inicialmente, um grande número de
 10 combinação de linhagem celular (> 100) da segunda integração foi selecionado para a produção da proteína em uma placa de 96 poços. Uma variedade de métodos apropriados para a seleção do anticorpo pode ser usada. Por exemplo, ELISA é usado para medir a quantidade total presente do anticorpo.
 15 Se o nível de anticorpo que está sendo feito é produzido a nível apropriado, o gel de SDS-poliacrilamida pode também ser usado para selecionar níveis de produção. Se as células são crescidas em meios de soro livre, é possível carregar os sobrenadantes da cultura de célula diretamente em um gel de
 20 SDS-PAGE. Se as células são crescidas em meios de contenção o anticorpo pode ser detectado especificamente e dosado perto, por exemplo, Western blotting ou ELISA.

ATIVIDADE DE LIGAÇÃO ESPECÍFICA DO ANTICORPO

PRODUZIDO POR CÉLULAS

25 DG44 ou PER.C6™ foi transfectado com pD1 - DTXI (usando Lipofectamine CD 2000 como descritos em outra parte). Vinte quatro horas após a transfecção os meios foram colhidos. IgG total foi determinado usando um kit de ensaio Easy- Titer (H+L) IgG (como descrito em outros lugares na
 30 patente). A toxina IgG da Anti-difteria foi determinada usando um kit ELISA IgG Diphtheria (IBL Hamburgo) exatamente de acordo com o as instruções do vendedor.

A FIG. 31 mostra a atividade de ligação específica

do anticorpo da toxina anti-difteria expressada nas células DG44 ou nas células PER.C6™. O anticorpo produzido de cada célula tem a mesma atividade de ligação específica. Além disso, os resultados mostram que o anticorpo de ambas as
5 linhagens celulares tem a especificidade correta do antígeno e que β -250mg deste anticorpo seria necessário para uma dose típica de 10.000 IU.

ATIVIDADE BIOLÓGICA DO ANTICORPO PRODUZIDO POR CÉLULAS

Um ensaio de neutralização pode também ser usado
10 para medir a atividade funcional de um anticorpo. Por exemplo, a toxina do antraz e outras toxinas, tais como, a células de culturas mortas da toxina da difteria. Conseqüentemente, a atividade de um anticorpo da toxina anti-difteria pode ser determinada medindo sua habilidade de
15 neutralizar as propriedades de extermínio celular da toxina da difteria purificada. A relação da atividade funcional para a proteína total (atividade específica) é uma medida útil ao nível de anticorpo ativo ou de outra proteína que segrega uma linhagem celular particular produzida.

20 A atividade de neutralização do anticorpo da toxina anti-difteria produzido de DG44 ou de PERC6™ foi determinada e comparada ao anticorpo da linhagem celular D2.2, de qual os genes do anticorpo da toxina anti-difteria foram clonados. O anticorpo de DG44 ou de PERC6™ foi gerado
25 pelo transfecção transiente de células usando o Lipofectamine CD 2000 como descrito em outra parte. A quantidade de anticorpo presente nos sobrenadantes das células D2.2 ou de DG44 e as PERC6™ transfectados foram determinadas por ELISA usando a toxina da difteria pura como
30 o antígeno. As várias quantidades de anticorpos foram adicionadas então à 10ng/ml de toxina da difteria. Após uma incubação de 15 minutos em 37°C. As misturas do anticorpo/toxina foram adicionadas às células de Jurkat, que

são sensíveis a morte pela toxina da difteria. A divisão de célula foi medida pela incorporação 3H-timidina. Os resultados são mostrados na FIG. 32. Controle das células que foi tratado com somente toxina e nenhum anticorpo morto
5 como indicado pela falta da incorporação significativa de 3H-timidina. As células tratadas com as quantidades crescentes de anticorpo da toxina anti-difteria produzidos por células D2.2, DG44 ou de PERC6™ sobreviveram. O EC50 para proteger células de Jurkat da matança pela toxina da difteria foi
10 5,8, e 11 ng/ml para os anticorpos da toxina anti-difteria produzidos por células de D2.2, DG44 ou de PERC6™, respectivamente.

Aproximadamente dez linhagens celulares que produzem níveis mais elevados de anticorpo em uma pequena
15 escala são adaptadas à cultura em suspensão de soro-livre em uma escala maior (por exemplo, 100 ml - 1 litro). Diversos clones são adaptados, visto que alguns não podem se adaptar, crescer rapidamente, ou reter níveis da expressão do anticorpo em níveis mais elevados. Após a adaptação das
20 linhagens celular à produção do anticorpo da cultura em suspensão os níveis são testados outra vez. A produção do anticorpo exemplificativa em uma escala de laboratório é aproximadamente 10-100 mg/L dos meios por dia ou aproximadamente 10-100 pg/célula/dia que suporta uma
25 densidade de célula máxima de 1×10^9 células por litro.

Uma variedade de métodos foi descrita para a purificação do anticorpo de IgG humano em grande escala. Tipicamente, pelo menos três resinas de cromatografia são usadas. Uma coluna de proteína A é usada como uma primeira
30 etapa da afinidade para capturar o IgG pela ligação da sua região de Fc. A segunda coluna é projetada para remover a endotoxina, as proteínas celulares permanecendo, e toda proteína A que lixivie da primeira coluna. As resinas

exemplificativas incluem hidroxapatita, interação hidrofóbica ou resinas de troca catiônica o qual o Ca é usado para a segunda etapa da cromatografia. Uma coluna da troca aniônica é usada como a terceira etapa para remover o DNA.

Aproximadamente 100 mg do anticorpo são purificados e testados em um ensaio de atividade apropriado. Para anticorpos da toxina anti-difteria in vivo um ensaio apropriado é um teste de pele feito nas cobaias. O anticorpo é misturado com a toxina da difteria purificada e injetado na pele. Toxina que não é neutralizada resulta em uma resposta inflamatória. Para anticorpos da toxina anti-difteria in vitro um ensaio apropriado é utilização de células Vera. Tão pouco quanto uma molécula da toxina difteria (Sigma) é provavelmente capaz de matar células através de um ADP-ribosilação covalente da proteína acessória ribossomal do fator de alongamento (EF -2). Em consequência toda a síntese da proteína na célula é inibida e as células morrem. Assim, todo o ensaio que mede a viabilidade da célula ou o metabolismo da célula, tal como um ensaio MTT-baseado é usado para determinar o título do anticorpo de encontro a uma dada quantidade de toxina da difteria purificada. Tais ensaios são feitos a cada mês por 12 meses para estabelecer uma vida útil e para estudar a estabilidade do anticorpo purificado.

Um gel de SDS-poliacrilamida é usado para avaliar algumas características básicas do anticorpo. Por exemplo, a eletroforese do gel SDS de uma amostra reduzida do anticorpo pode ser usada para confirmar a quantidade, pureza e para corrigir o peso molecular das cadeias pesadas (β -50 kDal) e leves (β -25 kDal), mas o mais importante confirmar que a relação da cadeia pesada a leve é aproximadamente 1:1. A eletroforese do gel de SDS de uma amostra desnaturada, mas

não-reduzida é usada para determinar se o anticorpo é primeiramente monomérico ou multimérico. Isto é importante porque a presença de anticorpo agregado pode indicar problemas de produção ou de purificação. Os anticorpos
5 agregados podem ter efeitos indesejáveis, tais como, a toxicidade do rim, quando usados como terapêutico humano. Finalmente, os anticorpos agregados são também freqüentemente inativos no que diz respeito a sua atividade biológica desejada. Outros métodos bioanalíticos podem
10 também ser usados para avaliar o estado da agregação de um anticorpo que inclui a dispersão de luz ou a filtragem de gel.

EXEMPLO 3

LINHAGEM CELULAR CHO PARA A PRODUÇÃO DA PROTEÍNA USANDO

UM VETOR DE EXPRESSÃO DOADOR SELECIONÁVEL

Nós encontramos que transfecção de clones celulares DG44 pR1-DHFR com o vetor de expressão pCS-M3 J integras e ϕ C31 mutante sozinho poderia conduzir às células resistentes a puromicina sem transfecção do vetor de
20 expressão doador. Este parece ser o resultado de rearranjos integrase-mediado ϕ C31 do DNA cromos somático no plasmídeo integrado pR1-DHFR nas áreas 5' ao gene de resistência de puromicina. Tais DNAs cromos somáticos translocados podem conter os promotores que conduzem a expressão da resistência
25 à puromicina. Em algumas experiências o número destes eventos foi até 30% do número de eventos desejados da integração em que o vetor de expressa doador integrou dentro do vetor alvo.

Um método para contornar este problema era ter um
30 gene de resistência funcional completo da droga, tal como, uma codificação de resistência G418, no vetor de expressão doador. Depois que a transfecção do clone do vetor alvo com um gene G418 que contém o vetor de expressão doador e o

vetor integrase ϕ C31, seguido pela seleção para a puromicina haverá duas classes de integrantes. Em uma classe de recombinação do vetor de expressão doador dentro de locais att P tipo selvagem no vetor alvo terá ocorrido e em uma
5 outra classe os rearranjos do DNA cromossômico no vetor alvo terão ocorrido. Entretanto se uma seleção G418 é aplicada depois que a seleção a puromicina somente os recombinantes com um vetor de expressão doador completo permanecerá. Células em que rearranjos cromossômico de DNA
10 dentro do vetor alvo tenham ocorrido não conterão o vetor expressão doador-G418 e será eliminado.

Observe que a ordem das seleções da resistência à droga é importante. Se a seleção G418 foi feita primeiramente, a seguir as células com o vetor da expressão
15 G418-doador integrado aleatoriamente, no vetor alvo, e em locais att Ψ podem ser obtidas. Então, se uma seleção a puromicina foi feita subsequente, as células com integrações do local att Ψ ou aleatório seriam eliminadas, mas rearranjos cromossômicos no vetor alvo podem ainda
20 ocorrer como nas células em que a integração do vetor de expressão doador dentro do vetor alvo não tenha ocorrido. Por razões similares é indesejável fazer simultaneamente seleções de puromicina e G418.

Para determinar se faz uma seleção G418 depois que
25 a seleção de puromicina foi benéfica, pD1-DTXI-G418 foi transfectado dentro de clones 1AI de DG44 RI-DHFR, 2B1 1, 2E8, 2G7, 2HI, 2H9 como descrito no exemplo 2. Dois dias após a transfecção as células foram selecionadas em 10 ug/ml de puromicina por 7 dias. Então as colônias foram separadas
30 em meios de crescimento contendo 10 ug/ml de puromicina somente ou 10 ug/ml de puromicina e 400 ug/ml de G418. A seleção sob estas circunstâncias continuou por 21 dias. Os meios foram analisados então para a produção do anticorpo.

Os resultados destes ensaios são mostrados na tabela 1. A seleção G418 aumentou a produtividade específica pela dobra 30 a 73 em 4 casos e não teve nenhum efeito em dois casos. Mesmo se a seleção G418 tivesse um efeito pode depender da eficiência da integração do vetor de expressão doador em cada clone do vetor alvo, e também da frequência dos eventos vetor de expressão-independentes que conduzem à resistência à puromicina.

Tabela 1: Efeito do uso de um vetor de expressão doador selecionável na produção da proteína

Clone de vetor alvo transfectado	Produção de IgG (depois da seleção de puromicina e G418)	Produção de IgG (depois da seleção somente de puromicina)	Razão de produção (com seleção G418/sem seleção de G418)
1A1	15 ng/ml	19 ng/ml	0,8
2B11	1795 ng/ml	56 ng/ml	32
2E8	585 ng/ml	10 ng/ml	59
2G7	1017 ng/ml	34 ng/ml	30
2H1	815 ng/ml	658 ng/ml	1,2
2H9	1688 ng/ml	26 ng/ml	73

Genes de resistência à droga completos, à exceção de um codificação de resistência G418, podem opcionalmente ser incorporados em um vetor de expressão doador selecionável. A única limitação é que deve ser diferente dessa usada para selecionar a integração do vetor alvo (por exemplo, resistência à higromicina), para selecionar a integração do vetor doador (por exemplo, resistência à puromicina) ou para amplificar o número de cópia do vetor alvo (por exemplo, dihidrofolato reductase). Assim sendo, por exemplo, os genes que codificam a resistência à zeocina ou a blasticidina podem ser utilizados.

Um outro benefício de usar um vetor de expressão doador selecionável é aquele após a integração ϕ C31-mediada

de um vetor de expressão doador selecionável dentro de um vetor alvo, tal como, o pR1-DHFR, o gene selecionável estará localizado entre as regiões de codificação das cadeias pesadas e leves do anticorpo. Daqui a seleção contínua
5 impedirá a recombinação homóloga entre elementos repetidos do vetor de expressão (por exemplo, promotor, sequência do sinal, sinal poliadenilação) que poderia conduzir a deleção das regiões de codificação da cadeia pesada ou leve.

EXEMPLO 4

10 ENGENHARIA DA LINHAGEM CELULAR CHO PARA A PRODUÇÃO DE PROTEÍNA DE ELEVADO RENDIMENTO

O método de cultivar e de transfecção das células CHa seguirá o procedimento como descrito em Thyagarajan et al, Methods Mol. Bio., 308:99 - 106 (2005). Resumidamente,
15 células "CHa/dhfr-" (por exemplo, células DG44) irão ser transfectadas usando Eugene 6 em placas de 24 poços. a seguinte protocolo é seguido:

1. A primeira transfecção é feita com o vetor alvo e o plasmídeo integras e ϕ C31 (FIG. 3).

20 2. 24 horas após a transfecção, as células são transferidas a placa de 100 mm.

3. 48 horas após a transfecção, as células são selecionadas para clone resistentes a higromicina.

4. Aproximadamente 12-14 dias após a transfecção
25 quando as colônias formadas aparecem nos poços, os clones individuais são escolhidos e transferidos a uma placa de 24 poços. Da experiência precedente com utilização da integrase ϕ C31, somente 30-50 clone precisam ser selecionados para obter clone da elevado-expressão.

30 5. As colônias selecionadas serão mantidas em dois conjuntos de 24 placas poços. Um conjunto é para a manutenção. a outro conjunto é para a seleção.

6. a conjunto da seleção de colônias CHa nas

placas de 24poços é co-transfectado com o vetor doador expressando um gene repórter (por exemplo, CIP, GFP ou luciferase), e o plasmídeo integrase R4 (FIG. 4).

7.48 horas após a segunda transfecção, o meio não-seletivo é removido das placas e meio contendo zeocina é aplicado diversas vezes por aproximadamente 2 semanas.

8. As células são colhidas então para ensaios apropriados do gene repórter.

9.3-5 clones são selecionados que expressam os níveis mais elevados de gene repórter, e os clones correspondentes são expandidos do conjunto da manutenção.

10. As linhagens celulares resultantes, contendo um local acessório do fago integras e R4 (*attP*), são referidas como células CHO-R4*attP*.

15 TESTANDO A LINHAGEM CELULAR CHO-R4ATTP

Um anticorpo SARS ou antraz é usado para testar a linhagem celular CHO-R4*attP*. A maioria dos anticorpos SARS e antraz são IgG1. As regiões variáveis de VH e de VL dos anticorpos são clonadas e montadas então em um vetor que contenha regiões constantes IgG1 para produzir anticorpos completos. Os cDNAs para a cadeia pesada e a cadeia leve podem qualquer um ser clonados em dois plasmídeo doadores separados ou em um único plasmídeo doador no que tandem ser conduzido por dois promotores idênticos ou dois diferentes.

25 Uma vantagem de usar uma integras e de fago é que não há nenhuma limitação do tamanho no gene de interesse. Um sistema de dois plasmídeo e um sistema de um plasmídeo serão usados para expressar os anticorpos de comprimento completo.

A expressão de anticorpos monoclonais na escala da pesquisa foi descrita extensivamente (Wurm et al., Nat Biotechnol 22, 1393-8 (2004); Andersen et al, CUIT Opin Biotechnol 13, 117-23 (2002); Wirth et al., Gene 73, 419-26 (1988); Kim et al., Biotechnol Bioeng 58, 73-84 (1998);

Gandor et al., FEBS Lett 377,290-4 (1995); and Kito et al., Appl Microbiol Biotechnol 60, 442-8 (2002)). Estes procedimentos comuns são seguidos no que diz respeito à linhagem celular CHO-K4attP. O meio de soro-livre e o processo de cultura celular são desenvolvidos para aperfeiçoar a produção do anticorpo para a fermentação em grande escala.

A linhagem celular parental, um subclone de "CHO/dhfr-", é selecionado para produzir proteína com um rendimento elevado de 30-50 pg/célula/dia no meio de soro-livre. A taxa de produção prevista usando a linhagem celular projetada attP CHO-R4 será aproximadamente pelo menos 30 pg/célula/dia no meio do soro-livre. Uma vez que a linhagem celular e o vetor doador são desenvolvidos, todo o gene do anticorpo de interesse pode convenientemente ser clonado no cassete de expressão do vetor doador (FIG. 2). Visto que a seleção para clone de expressão de alto rendimento requer somente a seleção de 30-50 colônias, uma linhagem celular estável que expresse altos níveis de um anticorpo possa rapidamente ser gerada em uma maneira de baixo-custo.

CARACTERIZAÇÃO DA LINHAGEM CELULAR CHO-R4ATTP

O memorando "Points to Consider in the Characterization of Cell Lines Used to Produce Biologicals (1993)" publicado pelo Center for Biologies Evaluation and Research (CBER) do FDA é seguido para caracterizar a linhagem celular CHO- RAattP.

Além disso, o local de integração do attP R4 é caracterizado inteiramente, por exemplo, no que diz respeito ao número de cópias e ao locus da integração, por métodos convencionais, por exemplo, FISH, Southern blots, seqüenciamento de DNA e PCR. Visto que a integração futura de um gene de interesse será alvejada especificamente ao local que tenha sido projetado previamente no cromossoma, a

caracterização do local de integração de cada gene individual de interesse é trivial. Conseqüentemente, a caracterização futura das linhagens celulares estáveis que expressam o gene de interesse é significativamente
5 simplificada, tempo de regate e custo.

EXEMPLO 5

ENGENHARIA DA LINHAGEM CELULAR CHO DHFR -AMPLIFICÁVEL PARA A PRODUÇÃO DE PROTEÍNA DE ELEVADO RENDIMENTO

O sistema da DHFR-amplificação é amplamente
10 utilizado em sistemas de expressão CHa a fim de aumentar o número de cópia de um DHFR associado a um cassete de expressão. a sistema de expressão utiliza células hospedeiras CHa deficiente de dihidrofolato reductase (DHFR) conjuntamente com um gene transfectado DHFR como um marcador
15 selecionável. a sistema amplifica os genes e as seqüências ligadas ao DHFR, o qual conduz ao melhorado nível de expressão da proteína (Wurm e outros, Nat. Biotechnol 12, 1393-8 (2004)). As células transfectadas desenvolvem a resistência ao metotrexato (MTX), a um inibidor DHFR com a
20 amplificação do gene DHFR de até 100-10.000 kilobases da região circunvizinha (Coquelle et al., Cell 89, 215-25 (1997); e Stark et al., Cell 57, 901-8 (1989)). Após 2-3 semanas da exposição à MTX, a maioria das células morre. Entretanto, as células de sobrevivência contêm
25 freqüentemente várias centenas a algumas mil cópias do plasmídeo integrado (Wurm et al., Ann N Y Acad Sci 782, 70-8 (1996); e Wurm et al., Biologicals 22, 95-102 (1994)). A maioria das células "amplificadas" produzem até 10 - 20
30 dobras a mais de proteínas recombinantes (Wirth et al., Gene 73, 419 - 26 (1988)). Diversos ciclos de amplificação do gene são executados freqüentemente e tipicamente a concentração de metotrexato é aumentada 3-5 dobras após cada ciclo da amplificação do gene. Três opções alternativas são testadas

para a melhor amplificação DHFR.

Para testar se a amplificação DHFR do gene de interesse permitiria a expressão aumentada da proteína, o gene DHFR foi colocado no vetor alvo. Um diagrama esquemático de um vetor alvo que inclui um gene DHFR é fornecido na FIG. 15. A sequência do vetor resultante é fornecida nas FIGURAS 35A-35C. A FIG. 29 mostra que a expressão de um anticorpo (pg/célula/dia) de uma associação das células em que um vetor de expressão doador foi local-
10 específico integrado dentro de um vetor alvo-DHFR e populações da célula foram então expostas às concentrações crescentes de metotrexato.

Há pelo menos três vantagens da ligação do gene DHFR com o local *attP* R4 no vetor alvo. Primeiramente, após
15 a amplificação do DHFR, o cromossoma também terá múltiplas cópias do local *attP* R4. Depois que o vetor doador é transfectado dentro da linhagem celular CHO-R4*attP* (DHFR), o gene de interesse pode ser integrado dentro de múltiplo locais recebimento *attP* R4, mediado pela integrase R4. Em
20 segundo, se a linhagem celular previamente amplificada CHO-R4*attP* (DHFR) já tem a capacidade de expressar um nível suficientemente elevado do gene de interesse, uma segunda amplificação DHFR não pode ser exigida depois que o gene de interesse é transfectado, assim tendo tempo e esforço
25 significativos de economia. Em terceiro lugar, visto que a linhagem celular CHO-R4*attP* (DHFR) terá sido bem caracterizada, após a integração do gene de interesse do vetor doador, a linhagem celular de expressão que produz o gene de interesse não pode precisar de outra amplificação
30 longa DHFR e caracterização adicional, ganhando uma quantidade significativa de tempo e de custo.

Em um segundo exemplo, o gene DHFR está presente no vetor doador. Um diagrama esquemático do vetor doador

incluindo um gene DHFR é fornecido na FIG. 6. Em um terceiro exemplo, o gene DHFR está presente no vetor alvo (FIG. 5) e no vetor doador (FIG. 6). Após a amplificação de DHFR, a linhagem celular projetada CHO-R4attP (DHFR) é esperada para
5 produzir um rendimento bem acima de 30 pg/célula/dia no meio livre de soro.

EXEMPLO 6

LINHAGEM CELULAR PROJETADA DE CHO PARA A PRODUÇÃO ELEVADA DA PROTEÍNA DO RENDIMENTO COM A UTILIZAÇÃO

10

REALCADA DA TRADUÇÃO IRES

A possibilidade e a necessidade de usar um elemento otimizado IRES junto com a integras e ϕ C31 para também aumentar o nível da expressão são também testadas. O IRES-elemento aperfeiçoado é clonado no vetor doador, a
15 montante da região de codificação para a proteína de interesse e a jusante do local do começo da transcrição (FIG. 7). Este IRES-elemento aumentará significativamente a produção da proteína realçando a eficiência da tradução do mRNA alvo (Chappell et al., J Biol Chem 278,33793-800
20 (2003); Owens et al., Proc Natl Acad Sci USA 98, 1471-6 (2001); and Chappell et al., (2000) Proc. Natl. Acad. Sci. U.S.A., 97, 1536-1541).

Para obter grandes quantidades de proteínas e de anticorpos terapêuticos, as linhagens celulares
25 superexpressantes são desenvolvidas que usam novas tecnologias baseadas em translação que são capazes de níveis muito mais elevados de produção da proteína do que é possível usando transcrição tradicional dos métodos baseados que aumentam a quantidade do gene mRNA do alvo, por exemplo,
30 com o uso de promotores fortes, duplicação cromossômica e da seleção da elevação expressando linhagem celular.

Os realçadores translacionais foram desenvolvidos recentemente usando as seqüências curtas do RNA que

funcionam enquanto locais internos da entrada do ribossomo (IRESes) esses recrutam a maquinaria da tradução e facilitam a iniciação da tradução. Embora a atividade do IRES-elemento individual fosse relativamente fraca, mostrou-se que a
5 atividade dos IRES poderia ser aumentada sinergisticamente quando os elementos particulares IRES fossem ligados junto (Owens et al., Proc Natl Acad Sci USA 98, 1471-6 (2001); e Chappell et al., (2000) Proc. Natl. Acad. Sci. U.S.A., 97, 1536-1541). Nestes estudos, IRESes sintético foram testados
10 na região intercistrônica de mRNAs dicistrônicos para a sua habilidade de melhorar a tradução do segundo cístron. Entretanto, mostrou-se recentemente que um destes IRESes poderia também funcionar como um realçador translational poderoso quando colocado no 5' líder de um mRNA
15 monocistrônico. Este IRES sintéticos contiveram múltiplas cópias ligadas de um 9-nt IRES-módulo de líder 5' de um homeodomínio Gtx mRNA.

O objetivo é para identificar os elementos IRES que funcionam eficientemente em células CRO e usam estes
20 elementos individuais para gerar os realçadores translacionais sintéticos que funcionam eficientemente em células CRO. Os realçadores translacionais são também desenvolvidos funcionando eficientemente no humano-híbrido e nas linhagens celulares humanas que são usadas para a
25 produção em grande escala.

Os elementos individuais IRES que funcionam eficientemente nestas linhagens celulares são obtidos usando uma metodologia da seleção em que um cassete que contém 18 nucleotídeos aleatórios é clonado em um vetor de seleção e
30 transfectado na linhagem celular de interesse (Owens e outros, Proc Acad Nat Sei USA 98, 1471-6 (2001)). As experiências de seleção são executadas usando um vetor retroviral dicistrônico GFP/CFP. As células que contêm

elementos ativos IRES são selecionadas por F ACS. As seqüências selecionadas são recuperadas e reexaminadas em um vetor duplo luciferase RenillaPhotinus (RPh) para mostrar as funções IRES em um outro contexto e não é dependente ou influenciado pelas seqüências presentes nos vetores GFP/CFP 5 usados para selecioná-los. Os vários elementos IRES são testados para que sua atividade sinérgica pela ligação conjunta de múltiplas cópias do mesmo ou de elementos-IRES diferentes. As combinações de elementos que mostram IRES 10 realçados a atividade são testadas para que sua habilidade funcione como realçadores translacionais na 5' líder de um RNA monocistrônico repórter.

Os realçadores translacionais sintéticos que são gerados são testados então no 5' líderes dos mRNAs que 15 codificam proteínas ou anticorpos terapêuticos para determinar que combinações do realçador/gene funcionam mais eficientemente. Uma vez que combinações particulares eficientes são identificadas, os constructos são testados em condições de cultura escalonadas e também otimizados caso 20 necessário para maximizar a produção do anticorpo.

EXEMPLO 7

ENGENHARIA DA LINHAGEM CELULAR CHO PARA A PRODUÇÃO DA PROTEÍNA INDUTÍVEL DE RENDIMENTO ELEVADO

As linhagens celulares apropriadas para a escala 25 acima e a fabricação deve ter a capacidade combinada para o crescimento rápido e a produtividade-específica elevada. Devido ao nível elevado da expressão do vetor de expressão, as células da produção podem ter dificuldades para crescer ao expressar altos níveis de proteínas estrangeiras ou as 30 proteínas estrangeiras podem agregar durante uma fase de crescimento prolongada. Se esta dificuldade é encontrada, um interruptor ligado-desligado está adicionado ao vetor doador para prever a expressão induzível do gene de interesse. Como

tal, o elemento funcionaria para desligar a expressão do transgene durante o crescimento da célula e giraria somente sobre a expressão quando as células vissem uma quantidade crítica e estão prontas para a produção da proteína. Estes

5 interruptores são atuados pelos ligantes que interage com um sistema receptor apropriado que condicionalmente interfere com ou ativa a transcrição. Diversos interruptores propriamente foram desenvolvidos para estudos da terapia de gene e podem ser usados no sistema de produção

10 previsto, incluindo, mas não limitado a, o sistema ARGENT, o sistema GENE SWITCH, riboswitches, proteínas do dedo de zinco, sistemas baseados receptor de ecdisone, e similares. Além disso, os sistemas tetraciclina-indutíveis e gás-indutíveis podem também ser utilizados (Weber et al., Nat

15 Biotechnol22, 1440-4 (2004); e Weber et al., Metab Eng 7, 174-81 de Metab (2005)).

EXEMPLO 8

ENGENHARIA DA LINHAGEM CELULAR PER.C6™ PARA A PRODUÇÃO DE PROTEÍNA DE RENDIMENTO ELEVADO

20 O método de cultura e transfecção de células PER.C6™ seguirá o procedimento como descrito em Thyagarajan et al, Methods MoI. Bio., 308:99 - 106 (2005). Resumidamente, as células PER.C6™ serão transfectadas usando Fugene 6 em placas de 24 poços. O seguinte protocolo é

25 seguido:

1. A primeira transfecção é feita com o vetor alvo e o plasmídeo integrase ϕ C31 (FIG.3).

2. 24 horas após a transfecção, as células são transferidas a placas de 100 mm.

30 3. 48 horas após a transfecção, as células são selecionadas para clone resistentes a higromicina.

4. Aproximadamente 21 dias após a transfecção quando as colônias aparecem bem-formadas, os clones

individuais são escolhidos e transferidos a uma placa de 24 poços.

Da experiência precedente usando a integrase ϕ C31, somente 30-50 clones precisam ser selecionados para obter
5 clone de elevada-expressão.

5. As colônias selecionadas são mantidas então em dois conjuntos de placas de 24 poços. Um conjunto é para a manutenção. O outro conjunto é para a seleção.

6. O conjunto da seleção de colônias PERC6™ na
10 placa de 24 poços é o plasmídeo co- transfectado com o vetor doador expressando um gene repórter (por exemplo, SEAP, CIP, GFP ou luciferase), e integrase R4 (FIG. 4)

7.48 horas após a segunda transfecção, o meio não-seletivo é removido das placas e o meio contendo zeocina é
15 aplicado diversas vezes por aproximadamente 3 semanas.

8. As células são colhidas então para ensaios apropriados do gene repórter.

9.3-5 clones que expressam os níveis mais elevados do gene repórter é selecionado e os clones correspondentes
20 do conjunto de manutenção são expandidos.

10. As linhagens celulares resultantes, contendo um local acessório do fago da integrase R4 (*attP*), são referidas como células PERC6™ - R4*attP*.

TESTANDO A LINHAGEM CELULAR PER.C6™-R4ATTP

25 Um anticorpo SARS ou antraz é usado para testar e caracterizar a linhagem celular PERC6™- R4*attP*. A maioria dos anticorpos SARS e antraz são IgGI. As regiões variáveis de VH e de VL dos anticorpos são clonadas e montadas então em um vetor que contenha regiões constantes IgGI para
30 produzir anticorpos completos. Os cDNAs para a cadeia pesada e a cadeia clara podem qualquer um ser clonado em dois plasmídeos doadores separados ou em um único plasmídeo doador no que tende ser conduzido por dois promotores

idênticos ou dois diferentes. Uma vantagem de usar uma integras e do fago é que não há nenhuma limitação do tamanho no gene de interesse. Um sistema do dois-plasmídeo e um sistema de um plasmídeo será usado para expressar os anticorpos de comprimento total.

A expressão de anticorpos monoclonais na escala da pesquisa foi descrita extensivamente (Wurm et al., Nat Biotechnol 22, 1393-8 (2004); Andersen et al, Curr Opin Biotechnol 13, 117-23 (2002); Wirth et al., Gene 73, 419-26 (1988); Kim et al., Biotechnol Bioeng 58, 73-84 (1998); Gandor et al., FEBS Lett 311, 290-4 (1995); and Kito et al., Appl Microbiol Biotechnol 60, 442-8 (2002)), e também em células de PERC6™ (Urlaub et al, Proc Natl Acad Sci USA 77, 4216-20 (1980)). Estes procedimentos comuns são seguidos no que diz respeito à linhagem celular CHO-R4attP. O meio livre de soro e o processo celular é desenvolvido para aperfeiçoar a produção do anticorpo para a fermentação em grande escala.

A taxa de produção prevista usando a linhagem celular projetada de PERC6™-R4u#P será aproximadamente de pelo menos 30 pg/célula/dia no meio livre de soro. Uma vez que a linhagem celular e o vetor doador são desenvolvidos, todo o gene do anticorpo de interesse pode convenientemente ser clonado no cassete de expressão do vetor doador (FIG. 2). Visto que a seleção para clone de nível elevado de expressão requer somente a seleção de 30-50 colônias, uma linhagem celular estável que expresse altos níveis de um anticorpo possa rapidamente ser gerada em uma maneira de custos-efetivos.

CARACTERIZAÇÃO DA LINHAGEM CELULAR PER.C6™-R4ATTP

O memorando "Points to Consider in the Characterization of Cell Lines Used to Produce Biologicals (1993)" publicado pelo Center for Biologics Evaluation and Research (CBER) do FDA é seguido para caracterizar a

linhagem celular PERC6™-R4uffP. [00284] Além disso, o local da integração attP R4 é caracterizado inteiramente, por exemplo, no que diz respeito ao número de cópias e ao locus da integração, por métodos convencionais, por exemplo, FISH, Southem blots, seqüenciamento de DNA e PCR. Visto que a integração futura de um gene de interesse será alvejada especificamente ao local que tem sido projetado previamente no cromossoma, a caracterização attP R4 do local de integração de cada gene individual de interesse é trivial. Conseqüentemente, a caracterização futura das linhagens celulares estáveis que expressam o gene de interesse é significativamente simplificada na economia de tempo e custo.

EXEMPLO 9

15 ENGENHARIA DA LINHAGEM CELULAR PER.C6™ PARA A PRODUÇÃO DE PROTEÍNA DE RENDIMENTO ELEVADO COM A UTILIZAÇÃO A TRADUÇÃO MELHORADA USANDO UM IRES

A possibilidade e a necessidade de usar um elemento-IRES junto com uma integrase ϕ C31 para também aumentar o nível da expressão são também testados. O elemento-IRES aperfeiçoado é clonado no vetor doador, a jusante do promotor e a montante da região de codificação para o gene de interesse (FIG. 7). Este elemento-IRES aumentará significativamente a produção da proteína realçando a eficiência da tradução do mRNA alvo (Chappell et al., J Biol Chem 278,33793- 800 (2003); Owens et al., Proc Natl Acad Sci USA 98, 1471-6 (2001); and Chappell et al., (2000) Proc. Natl. Acad. Sci. U.S.A., 97, 1536-1541).

Para obter grandes quantidades de proteínas e de anticorpos terapêuticos, as linhagens celulares superexpressantes são desenvolvidas usando novas tecnologias baseadas na tradução em que são capazes de níveis muito mais elevados de produção da proteína do que é possível usando

transcrição tradicional dos métodos baseados que aumentam a quantidade do gene mRNA alvo, por exemplo, com o uso de promotores fortes, duplicação cromossômica e da seleção da linhagem celular de alta expressão. [00287] Os realçadores translacionais foram desenvolvidos recentemente usando as seqüências curtas do RNA que funcionam enquanto locais internos da entrada do ribossomo (IRESes) que recrutam a maquinaria da tradução e facilitam a iniciação da tradução. Embora a atividade de elemento-IRES individual fosse relativamente fraca, mostrou-se que a atividade dos IRES poderia ser aumentada sinergisticamente quando os elementos IRES particulares fossem ligados juntos (Owens et al., Proc Natl Acad Sci USA 98, 1471-6 (2001); and Chappell et al., (2000) Proc. Natl. Acad. Sci. U.S.A., 97, 1536-1541). Nestes estudos, IRESes sintéticos foram testados na região intercistônica de mRNAs dicistônico para que sua habilidade para realçar a tradução do segundo cístron. Entretanto, mostrou-se recentemente que um dos IRESes poderia também funcionar como um realçador translacional poderoso quando colocado no 5' líder de um mRNA monocistônico. Este IRES sintéticos contiveram múltiplas cópias ligadas de um NT-9 IRES-módulo do 5' líder do homeodomínio Gtx do mRNA.

Um objetivo é para identificar os elementos IRES que funcionam eficientemente em células PER.C6™ e usar estes elementos individuais para gerar os realçadores translacionais sintéticos que funcionam eficientemente em células PER.C6™. Os realçadores translacionais são desenvolvidos também funcionando eficientemente no humano-híbrido e nas linhagens celulares humanas que são usadas para a produção em grande escala.

Os elementos individuais IRES que funcionam eficientemente nestas linhagens celulares são obtidos usando

uma metodologia da seleção em que um cassete contendo 18 nucleotídeos aleatórios é clonado em um vetor de seleção e transfectado na linhagem celular de interesse (Owens et al., Proc Acad Natl Sei USA 98, 1471-6 (2001)). Seleções das
5 experiências são executadas usando um vetor retroviral dicistrônico de GFP/CFP. As células que contêm elementos ativos IRES são selecionadas por FACS. As seqüências selecionadas são recuperadas e reexaminadas em um vetor duplo do luciferase RenillaPhotinus (RPh) para mostrar as
10 funções IRES em um outro contexto e não são dependentes ou influenciadas pelas seqüências presentes nos vetores GFP/CFP usados para selecioná-los. Os vários elementos IRES são testados para que sua habilidade a atividade sinérgica ligada junto a múltiplas cópias do mesmo ou de elementos-
15 IRES diferentes. As combinações de elementos que mostram a atividade IRES realçada são testadas para que sua habilidade funcione como realçadores translacionais no 5' líder de um RNA monocistrônico repórter.

Os realçadores translacionais sintéticos que são
20 gerados são testados então no 5' líderes dos mRNAs que codificam proteínas ou anticorpos terapêuticos para determinar que combinações do realçador/gene funcionam mais eficientemente. Uma vez que as combinações eficientes particulares são identificadas, os constructos são testados
25 na escala acima das condições de cultura e também otimizados caso necessário para maximizar a produção do anticorpo.

EXEMPLO 10

ENGENHARIA DA LINHAGEM CELULAR PER.C6™ PARA A PRODUÇÃO DE PROTEÍNA INDUTÍVEL DE RENDIMENTO ELEVADO

30 As linhagens celulares apropriadas para a escala acima e a fabricação deve ter a capacidade combinada para o crescimento rápido e a produtividade-específica elevada. Devido ao nível elevado da expressão do vetor de expressão,

as células da produção podem ter dificuldades de crescimento ao expressar altos níveis de proteínas estrangeiras ou as proteínas estrangeiras podem agregar durante uma fase de crescimento prolongado. Se esta dificuldade é encontrada, um interruptor ligado-desligado é adicionado ao vetor doador para prever a expressão induzível do gene de interesse na linhagem celular PER.C6™. Como tal, o elemento funcionaria para desligar a expressão do transgene durante o crescimento da célula e giraria somente sobre a expressão quando as células vissem uma quantidade crítica e estão prontas para a produção da proteína. Estes interruptores são atuados pelos ligantes que interagem com um sistema receptor apropriado que condicionalmente interfere com ou ativa a transcrição. Diversos interruptores proprietariamente foram desenvolvidos para estudos da terapia de gene e podem ser usados no sistema de produção previsto, incluindo, mas não limitado a, o sistema ARGENT, o sistema GENE SWITCH, riboswitches, proteínas do dedo de zinco, sistemas receptor-baseados em ecdisone, e similares. Além disso, sistema tetracycline-induzível e gás-induzível podem também ser utilizados (Weber et al., Nat Biotechnol 22, 1440-4 (2004); and Weber et al, Metab Eng 7, 174-81 (2005)).

A precedência ilustra meramente os princípios da invenção. Apreciar-se-á que aqueles hábeis na técnica poderão planejar os vários arranjos que, embora descritos não explicitamente ou mostrados aqui, personificam os princípios da invenção e são incluídos dentro de seus espírito e escopo. Além disso, todos os exemplos e linguagem condicional relatados aqui são pretendidos principalmente para ajudar ao leitor na compreensão dos princípios da invenção e dos conceitos contribuídos pelos inventores a promover a arte, e devem ser interpretados como sendo sem limitação a tais exemplos e circunstâncias especificamente

relatados. Além disso, todas as indicações relatadas aqui dos princípios, aspectos e incorporações da invenção assim como exemplos específicos dos mesmos, são pretendidas para abranger equivalentes estruturais e funcionais dos mesmos.

- 5 Adicionalmente, pretende-se que tais equivalentes incluem ambos equivalentes atualmente conhecidos e equivalentes desenvolvidos no futuro, isto é, todos os elementos desenvolvidos que executarem a mesma função, não obstante a estrutura. O escopo da presente invenção, conseqüentemente,
- 10 não é pretendido ser limitado às incorporações exemplificativas mostradas e descritas aqui. Particularmente, o escopo e o espírito da presente invenção são personificados pelas reivindicações adicionadas.

REIVINDICAÇÕES

1. VETOR ALVO PARA INTEGRAÇÃO EM SÍTIOS-ESPECÍFICOS, caracterizado por compreender:

5 (a) um primeiro sítio de recombinação do vetor que recombina com um sítio de recombinação genômica na presença de uma primeira recombinase de sítio-específico unidirecional;

10 (b) um segundo sítio de recombinação do vetor que recombina com um sítio de recombinação doador na presença de uma segunda recombinase de sítio-específico unidirecional que é diferente da primeira recombinase de sítio-específico unidirecional;

15 (c) uma primeira porção de um primeiro marcador selecionável adjacente ao segundo sítio de recombinação do vetor na extremidade 3'; e

(d) um segundo marcador selecionável que é diferente do primeiro marcador selecionável.

20 2. VETOR ALVO, de acordo com a reivindicação 1, caracterizado pelo fato de que o sítio de recombinação genômica é um sítio de recombinação genômica de mamífero.

25 3. VETOR ALVO, de acordo com a reivindicação 1, caracterizado pelo fato de que o primeiro sítio de recombinação do vetor é um sítio de recombinação genômico bacteriano (*attB*) ou um sítio recombinação genômico de fago (*attP*).

30 4. VETOR ALVO, de acordo com a reivindicação 1, caracterizado pelo fato de que o primeiro sítio de recombinação do vetor é um sítio de recombinação genômico bacteriano (*attB*) e o sítio de recombinação genômico é um sítio de recombinação genômico de pseudo-fago (*pseudo-attP*).

5. VETOR ALVO, de acordo com a reivindicação 1, caracterizado pelo fato de que o primeiro sítio de recombinação do vetor é um sítio de recombinação genômico de

fago (*attP*) e o sítio de recombinação genômico é um sítio de recombinação pseudo-bacteriano (*pseudo-attB*).

6. VETOR ALVO, de acordo com a reivindicação 1, caracterizado pelo fato de que o primeiro sítio de recombinação do vetor é um sítio de recombinação genômico pseudo-bacteriano (*pseudo-attB*) ou um sítio de recombinação genômico de pseudo-fago (*pseudo-attP*).

7. VETOR ALVO, de acordo com a reivindicação 1, caracterizado pelo fato de que o segundo sítio de recombinação do vetor é um sítio de recombinação genômico bacteriano (*attB*) ou um sítio de recombinação genômico de fago (*attP*).

8. VETOR ALVO, de acordo com a reivindicação 1, caracterizado pelo fato de que o segundo sítio de recombinação do vetor é um sítio de recombinação genômico pseudo-bacteriano (*pseudo-attB*) ou um sítio *attP* de recombinação genômica pseudo-fago (*pseudo-attP*).

9. VETOR ALVO, de acordo com a reivindicação 1, caracterizado pelo fato de que a primeira recombinase de sítio-específico unidirecional é uma fago recombinase ϕ C31, uma fago recombinase TP901-1, uma fago recombinase R4, uma fago recombinase ϕ FC1, uma fago recombinase ϕ Rv1 ou uma fago recombinase ϕ BT1.

10. VETOR ALVO, de acordo com a reivindicação 1, caracterizado pelo fato de que a primeira recombinase de sítio-específico unidirecional é uma fago recombinase ϕ C31.

11. VETOR ALVO, de acordo com a reivindicação 1, caracterizado pelo fato de que a segunda recombinase de sítio-específico unidirecional é uma fago recombinase R4.

12. MÉTODO DE SÍTIO-ESPECÍFICO QUE INTEGRA UM POLINUCLEOTÍDEO CODIFICANDO UMA PROTEÍNA DE INTERESSE EM UM GENOMA DE UMA CÉLULA EUCARIÓTICA, caracterizado por:

(a) introduzir o vetor alvo de acordo com a

reivindicação 1 em uma célula de mamífero compreendendo uma primeira recombinase de sítio-específico unidirecional e mantendo a célula de mamífero sob condições suficientes para um evento de recombinação mediado pela primeira recombinase
5 de sítio-específico unidirecional entre o primeiro sítio de recombinação do vetor e o sítio de recombinação genômico para o sítio-específico integrar o vetor alvo dentro do genoma da célula de mamífero;

(b) introduzir um vetor doador na célula alvo que
10 compreende uma segunda recombinase de sítio-específico unidirecional, em que o vetor doador compreende o polinucleotídeo que codifica uma proteína de interesse e um sítio de recombinação doador e mantendo a célula alvo sob condições suficientes para um evento de recombinação mediado
15 pela segunda recombinase de sítio-específico unidirecional entre o sítio de recombinação doador e o segundo sítio de recombinação do vetor do vetor alvo ao sítio-específico para integrar o polinucleotídeo que codifica uma proteína de interesse no genoma da célula de mamífero;

20 onde a primeira recombinase de sítio-específico unidirecional é diferente da segunda recombinase de sítio-específico unidirecional.

13. MÉTODO, de acordo com a reivindicação 12, caracterizado por também compreender a seleção de uma célula
25 que expresse a proteína de interesse.

14. MÉTODO, de acordo com a reivindicação 12, caracterizado pelo fato de que o primeiro sítio de recombinação do vetor é um sítio de recombinação genômico bacteriano (*attB*) ou um sítio de recombinação genômico de
30 fago (*attP*).

15. MÉTODO, de acordo com a reivindicação 12, caracterizado pelo fato de que o primeiro sítio de recombinação do vetor é um sítio de recombinação genômico

bacteriano (*attB*) e o sítio de recombinação genômico é um sítio de recombinação genômico de pseudo-fago (*pseudo-attP*).

16. MÉTODO, de acordo com a reivindicação 12, caracterizado pelo fato de que o primeiro sítio de recombinação do vetor é um sítio de recombinação genômico fago (*attP*) e o sítio de recombinação genômico é um sítio de recombinação genômico pseudo-bacteriano (*pseudo-attB*).

17. MÉTODO, de acordo com a reivindicação 12, caracterizado pelo fato de que o primeiro sítio de recombinação do vetor é um sítio de recombinação genômico pseudo-bacteriano (*pseudo-attB*) ou um sítio *attP* de recombinação genômico de pseudo-fago (*pseudo-attP*).

18. MÉTODO, de acordo com a reivindicação 12, caracterizado pelo fato de que o segundo sítio de recombinação do vetor é um sítio de recombinação genômico bacteriano (*attB*) ou um sítio de recombinação genômico de fago (*attP*).

19. MÉTODO, de acordo com a reivindicação 12, caracterizado pelo fato de que o segundo sítio de recombinação do vetor é um sítio de recombinação genômico pseudo-bacteriano (*pseudo-attB*) ou um sítio *attP* de recombinação genômico de pseudo-fago (*pseudo-attP*).

20. MÉTODO, de acordo com a reivindicação 12, caracterizado pelo fato de que o sítio de recombinação doador é um sítio de recombinação genômico bacteriano (*attB*) ou um sítio de recombinação genômico de fago (*attP*).

21. MÉTODO, de acordo com a reivindicação 12, caracterizado pelo fato de que o sítio de recombinação doador é um sítio de recombinação genômico pseudo-bacteriano (*pseudo-attB*) ou sítio *attP* de recombinação genômico de pseudo-fago (*pseudo-attP*).

22. MÉTODO, de acordo com a reivindicação 12, caracterizado pelo fato de que a primeira recombinase de

sítio-específico unidirecional é uma fago recombinase ϕ C31, uma fago recombinase TP901-1, uma fago recombinase R4, uma fago recombinase ϕ FC1, uma fago recombinase ϕ Rv1 ou uma fago recombinase ϕ BT1.

5 23. MÉTODO, de acordo com a reivindicação 12, caracterizado pelo fato de que a segunda recombinase de sítio-específico unidirecional é uma fago recombinase ϕ C31, uma fago recombinase TP901-1, uma fago recombinase R4, uma fago recombinase ϕ FC1, uma fago recombinase ϕ Rv1 ou uma fago
10 recombinase ϕ BT1.

24. MÉTODO, de acordo com a reivindicação 12, caracterizado pelo fato de que a primeira recombinase de sítio-específico unidirecional é uma fago recombinase ϕ C31.

25. MÉTODO, de acordo com a reivindicação 12,
15 caracterizado pelo fato de que a segunda recombinase de sítio-específico unidirecional é uma fago recombinase R4.

26. MÉTODO, de acordo com a reivindicação 12, caracterizado pelo fato de que a proteína é uma proteína segregada.

20 27. MÉTODO, de acordo com a reivindicação 12, caracterizado pelo fato de que a proteína segregada é um anticorpo.

28. MÉTODO, de acordo com a reivindicação 12, caracterizado pelo fato de que a célula é uma célula de
25 mamífero.

29. MÉTODO, de acordo com a reivindicação 28, caracterizado pelo fato de que a célula de mamífero é uma célula de roedor.

30. MÉTODO, de acordo com a reivindicação 29,
30 caracterizado pelo fato de que a célula do roedor é uma célula CHO.

31. MÉTODO, de acordo com a reivindicação 28, caracterizado pelo fato de que a célula de mamífero é uma

célula humana.

32. MÉTODO, de acordo com a reivindicação 31, caracterizado pelo fato de que a célula humana é uma célula PER.C6™.

5 33. CÉLULA EUCARIÓTICA ISOLADA, caracterizada por compreender um polinucleotídeo cassete integrado geneticamente compreendendo,

um primeiro sítio híbrido de recombinação e um segundo sítio híbrido de recombinação flanqueado:

10 (a) um sítio de recombinação do vetor que recombina com um sítio de recombinação doador na presença de uma recombinase de sítio-específico unidirecional;

(b) uma primeira parcela de um primeiro marcador selecionável adjacente aos locais de recombinação do
15 vetor na extremidade 3'; e

(c) um segundo marcador selecionável que seja diferente do primeiro marcador selecionável.

34. CÉLULA EUCARIÓTICA ISOLADA, de acordo com a reivindicação 33, caracterizado pelo fato de que o sítio de
20 recombinação do vetor é um sítio de recombinação genômico bacteriano (*attB*) ou um sítio de recombinação genômico de fago (*attP*).

35. CÉLULA EUCARIÓTICA ISOLADA, de acordo com a reivindicação 33, caracterizado pelo fato de que o sítio de
25 recombinação do doador é um sítio de recombinação genômico bacteriano (*attB*) ou um sítio de recombinação genômico de fago (*attP*).

36. CÉLULA EUCARIÓTICA ISOLADA, de acordo com a reivindicação 33, caracterizado pelo fato de que a
30 recombinase de sítio-específico unidirecional é uma fago recombinase ϕ C31, uma fago recombinase TP901-1, uma fago recombinase R4, uma fago recombinase ϕ FC1, uma fago recombinase ϕ Rv1 ou uma fago recombinase ϕ BT1.

37. CÉLULA EUCARIÓTICA ISOLADA, de acordo com a reivindicação 33, caracterizado pelo fato de que a célula é uma célula de mamífero.

5 38. CÉLULA EUCARIÓTICA ISOLADA, de acordo com a reivindicação 37, caracterizado pelo fato de que a célula de mamífero é uma célula de roedor.

39. CÉLULA EUCARIÓTICA ISOLADA, de acordo com a reivindicação 38, caracterizado pelo fato de que a célula de roedor é uma célula CHO.

10 40. CÉLULA EUCARIÓTICA ISOLADA, de acordo com a reivindicação 37, caracterizado pelo fato de que a célula de mamífero é uma célula humana.

41. CÉLULA EUCARIÓTICA ISOLADA, de acordo com a reivindicação 40, caracterizado pelo fato de que a célula
15 humana é uma célula PER.C6™.

42. KIT PARA USO NO SÍTIO-ESPECÍFICO QUE INTEGRA UM POLINUCLEOTÍDEO DENTRO DE UM GENOMA DE UMA CÉLULA IN VITRO, caracterizado por compreender:

- 20 (a) um vetor de acordo com a reivindicação 1; e
(b) um vetor doador compreendendo:
(i) um múltiplo sítio de clonagem;
(ii) um sítio de recombinação doador; e
(iii) uma segunda parcela de um primeiro
marcador selecionável adjacente aos locais de recombinação
25 doador de extremidade 5'.

43. KIT, de acordo com a reivindicação 42, caracterizado por também compreender uma primeira recombinase de sítio-específico unidirecional ou um ácido nucléico codificando a mesma.

30 44. KIT, de acordo com a reivindicação 43, caracterizado por também compreender uma segunda recombinase de sítio-específico unidirecional ou um ácido nucléico codificando a mesma que seja diferente da primeira

recombinase de sítio-específico unidirecional.

45. KIT, de acordo com a reivindicação 43, caracterizado pelo fato de que a primeira recombinase de sítio-específico unidirecional é uma fago recombinase ϕ C31, uma fago recombinase TP901-1, uma fago recombinase R4, uma fago recombinase ϕ FC1, uma fago recombinase ϕ Rv1 ou uma fago recombinase ϕ BT1.

46. KIT, de acordo com a reivindicação 44, caracterizado pelo fato de que a segunda recombinase de sítio-específico unidirecional é uma fago recombinase ϕ C31, uma fago recombinase TP901-1, uma fago recombinase R4, uma fago recombinase ϕ FC1, uma fago recombinase ϕ Rv1 ou uma fago recombinase ϕ BT1.

47. KIT PARA USO EM PRODUZIR UMA PROTEÍNA EM UMA CÉLULA, caracterizado por compreender:

(a) uma célula eucariótica isolada de acordo com a reivindicação 43; e

(b) um vetor doador compreendendo:

(i) um múltiplo sítio de clonagem;

(ii) um sítio de recombinação do doador; e

(iii) uma segunda parcela de um primeiro marcador selecionável adjacente aos locais de recombinação do doador de extremidade 5' .

48. KIT, de acordo com a reivindicação 47, caracterizado por também compreender de um sítio-específico de recombinação unidirecional ou de um ácido nucléico que codifica o mesmo.

49. KIT, de acordo com a reivindicação 48, caracterizado pelo fato de que a recombinase de sítio-específico unidirecional é uma fago recombinase ϕ C31, uma fago recombinase TP901-1, uma fago recombinase R4, uma fago recombinase ϕ FC1, uma fago recombinase ϕ Rv1 ou uma fago recombinase ϕ BT1.

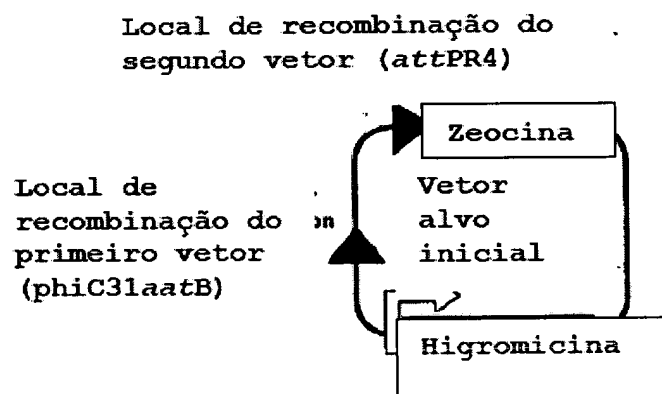


FIG. 1

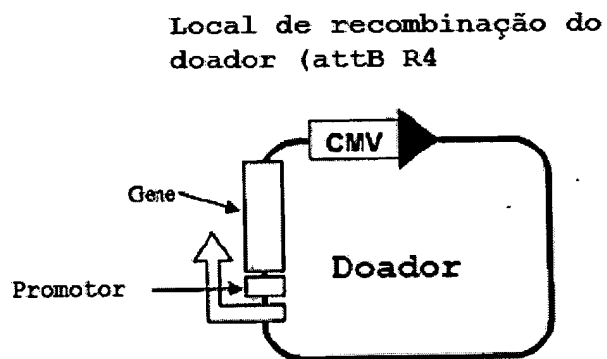


FIG. 2

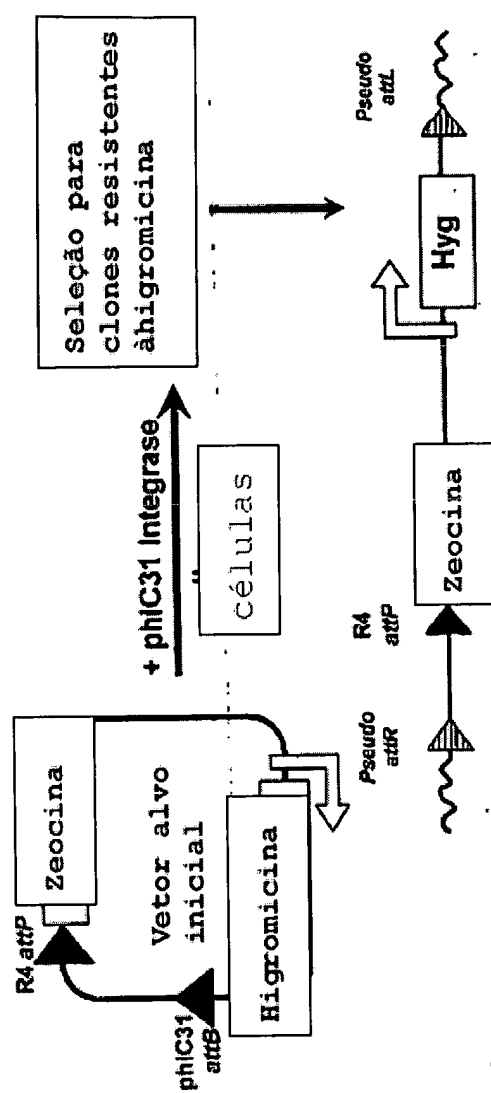


FIG. 3

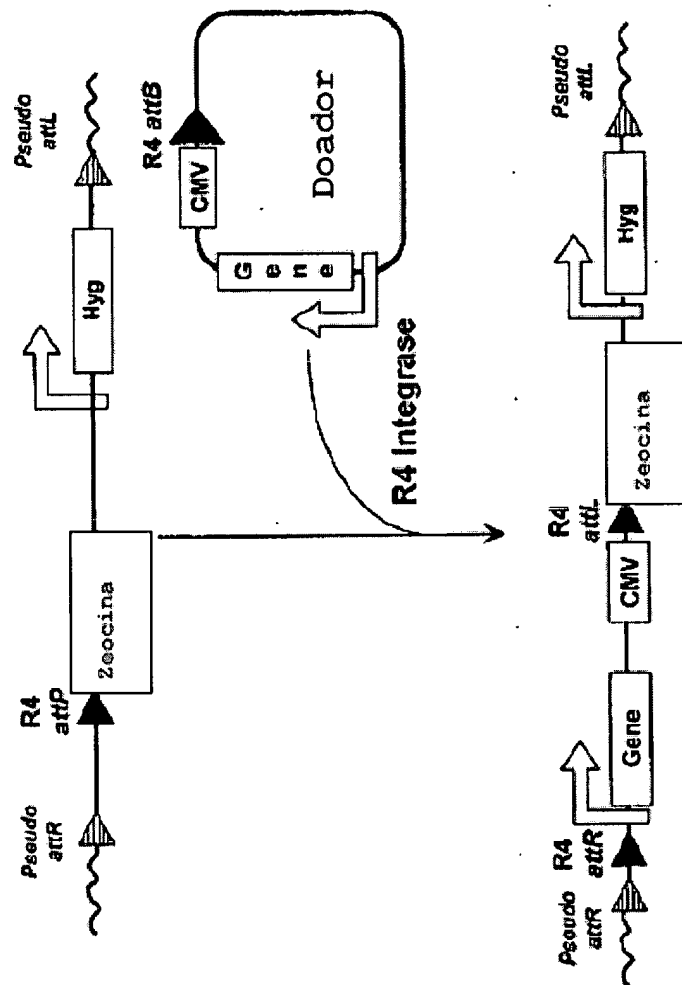


FIG. 4

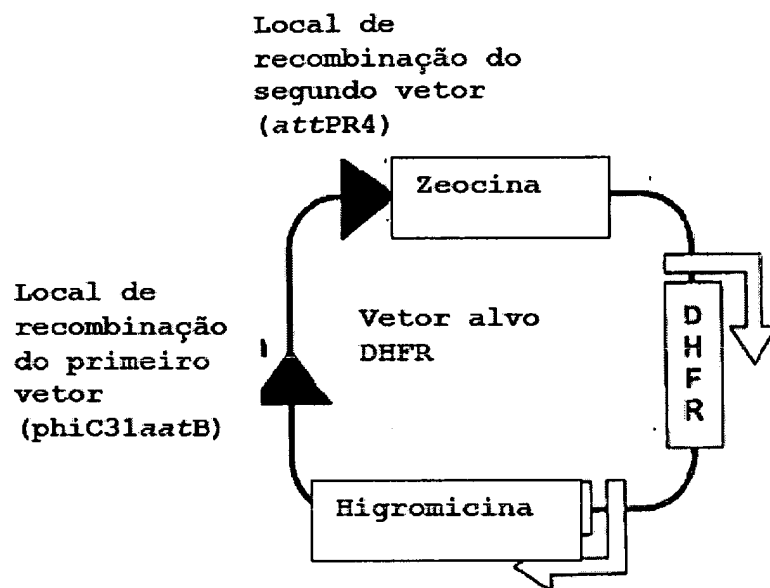


FIG. 5

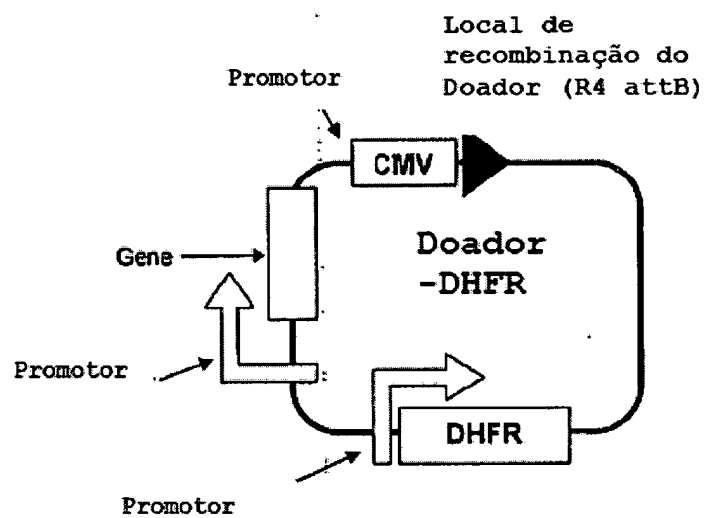


FIG. 6

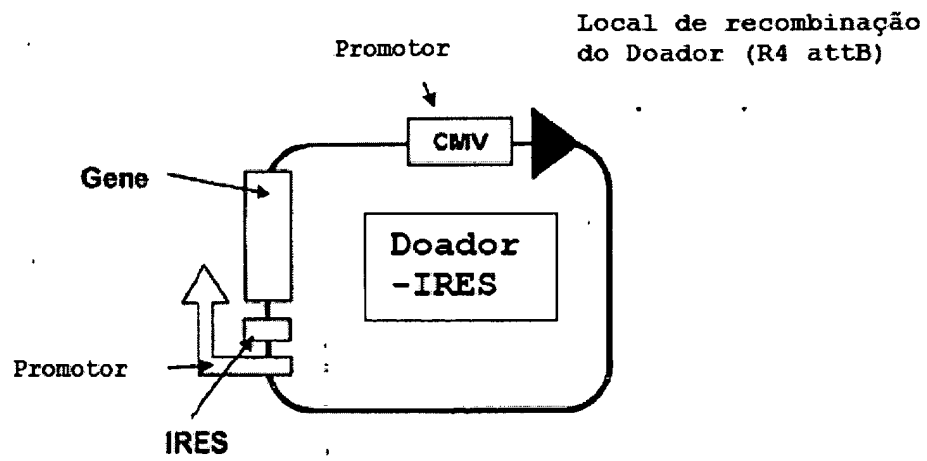


FIG. 7

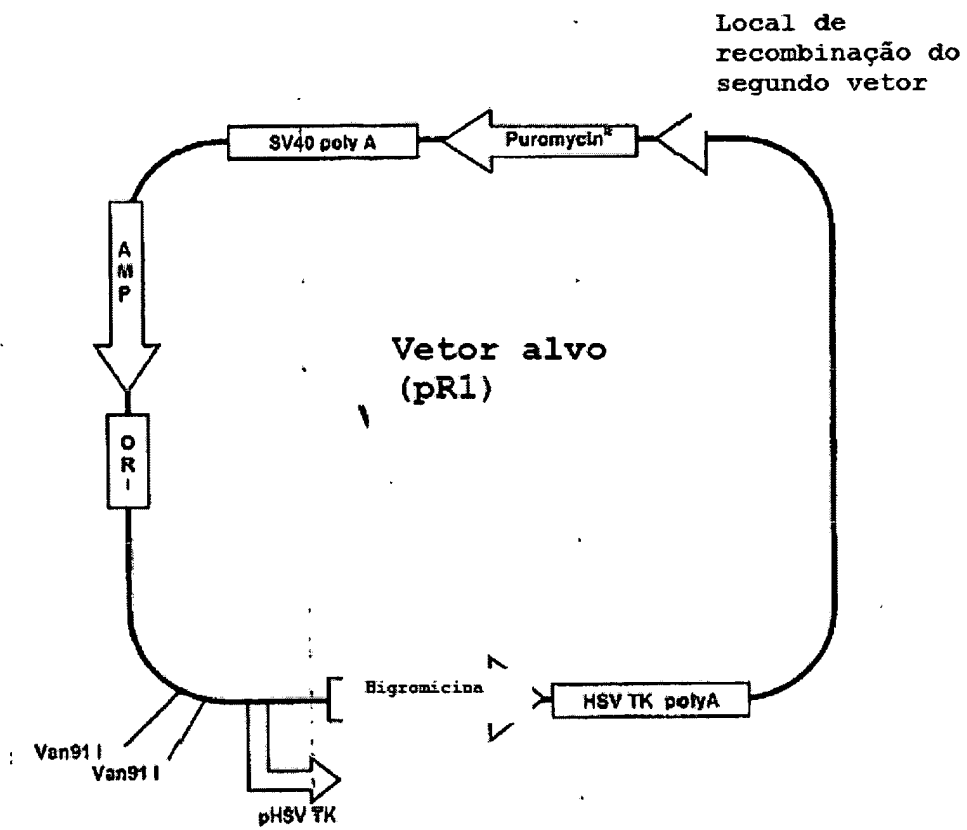


FIG. 8

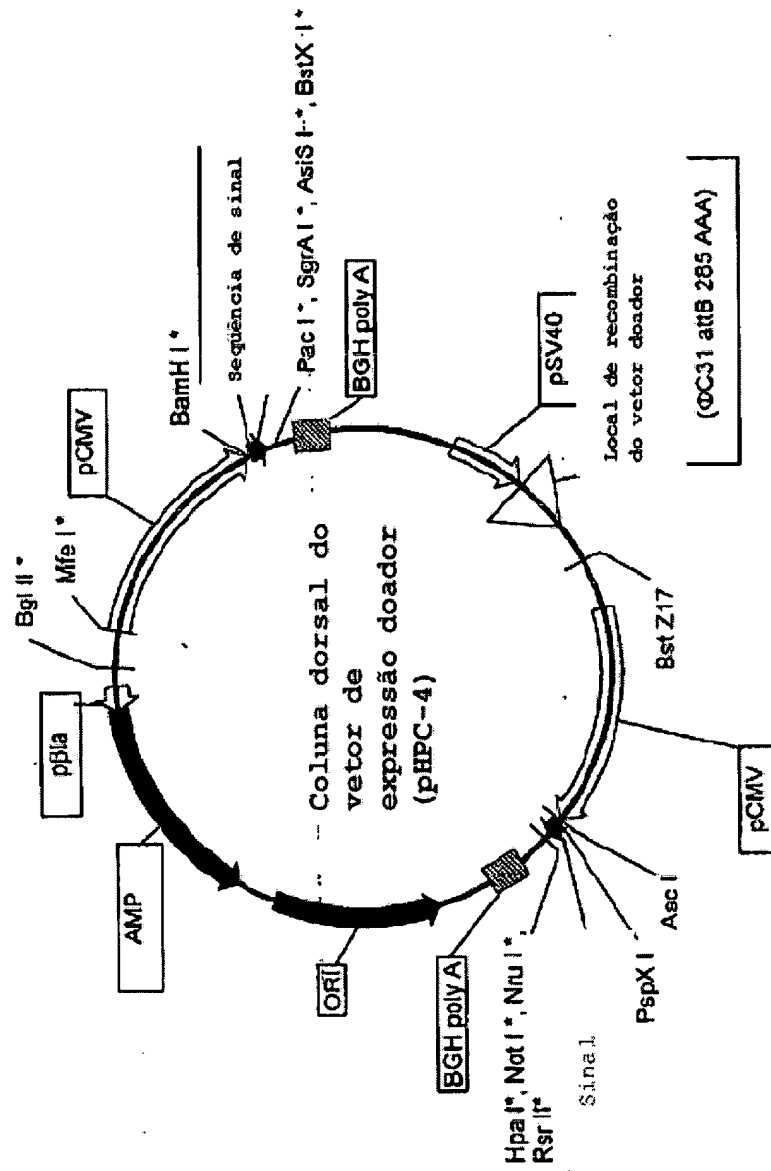


FIG. 9

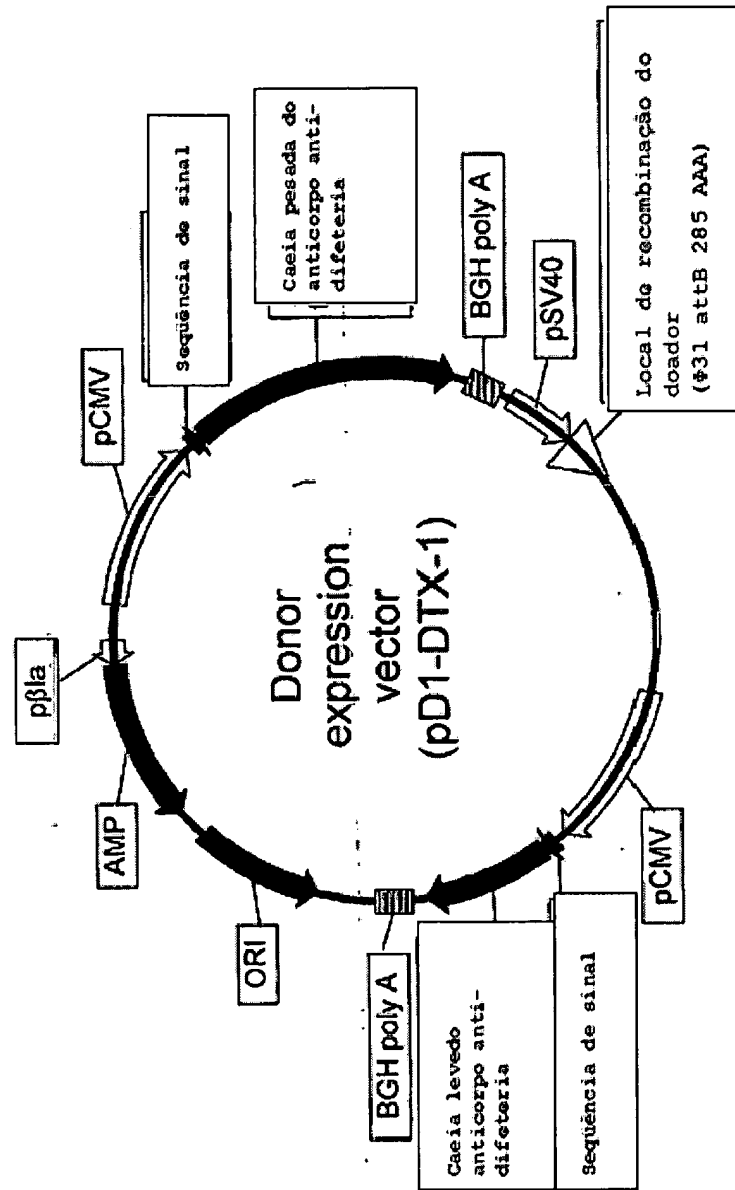
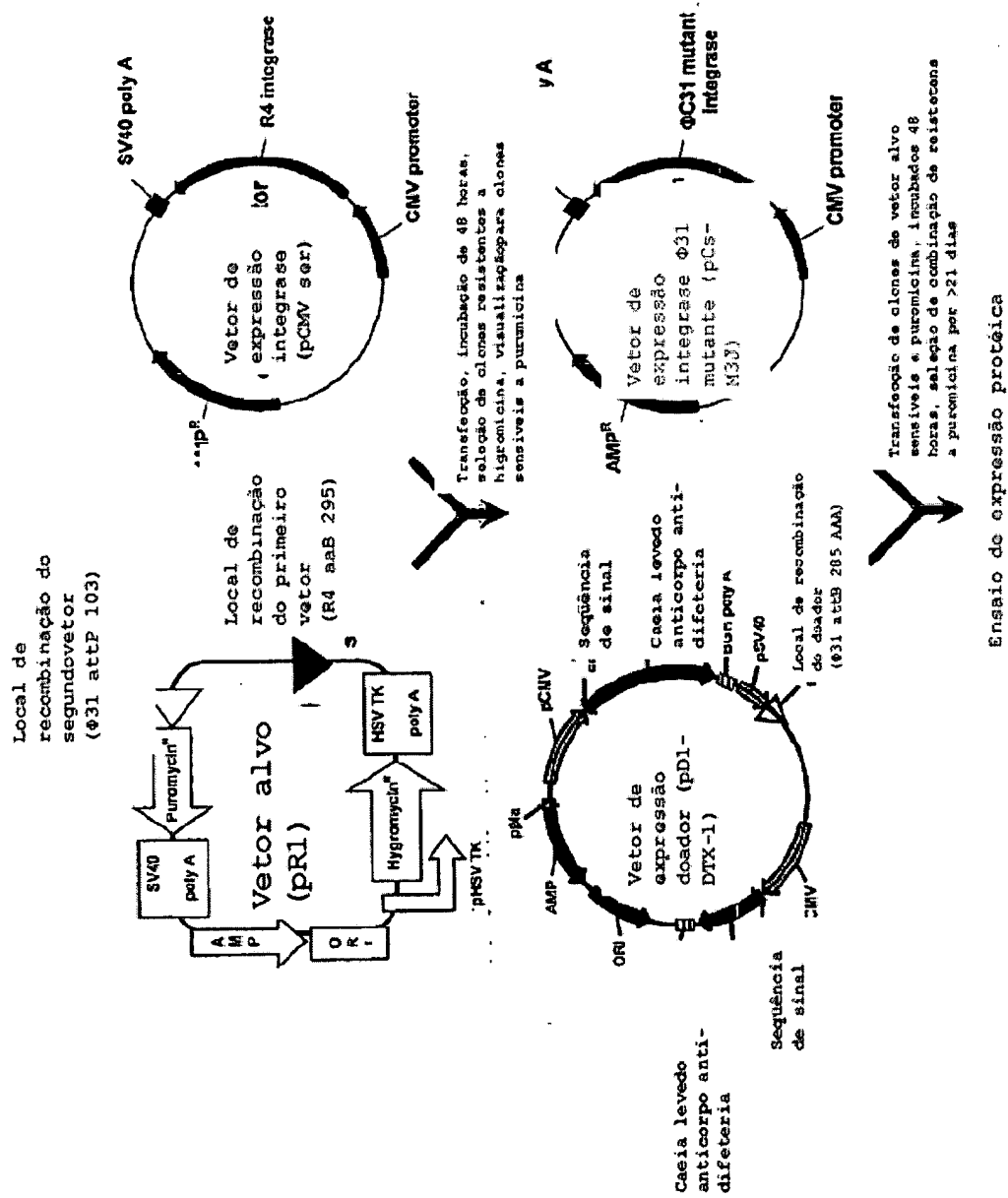
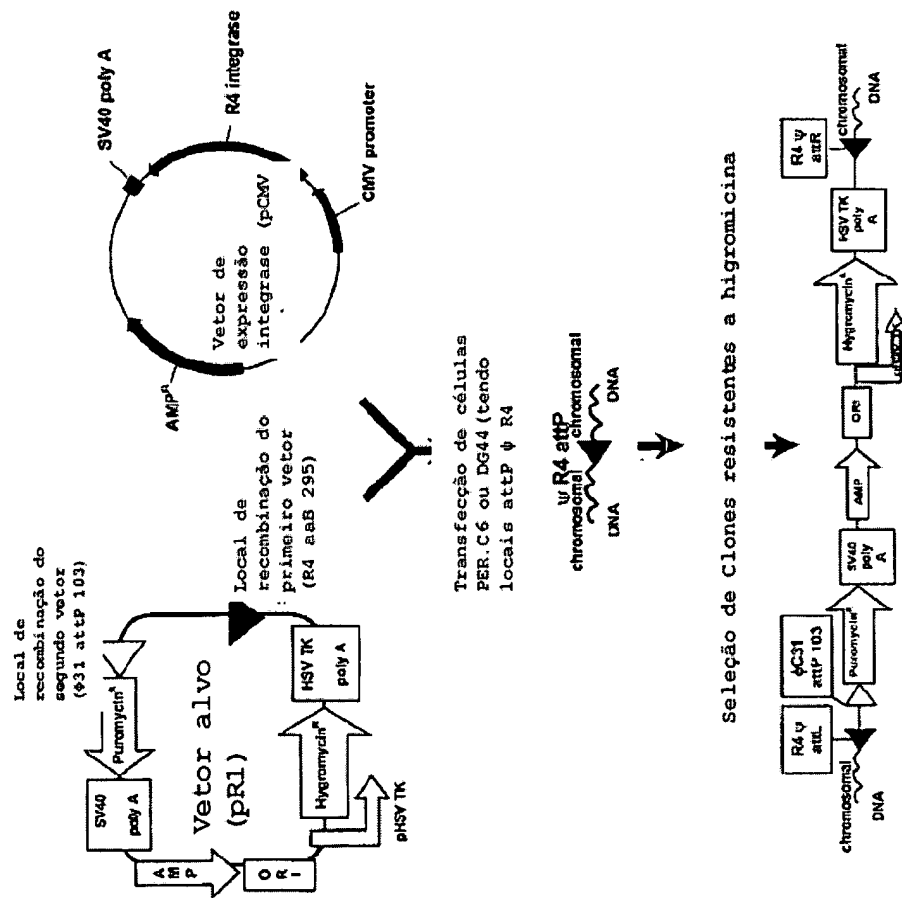


FIG. 10





Linagem celular attP DG44-φ31 ou attP PER.C6 φ31

FIG. 12

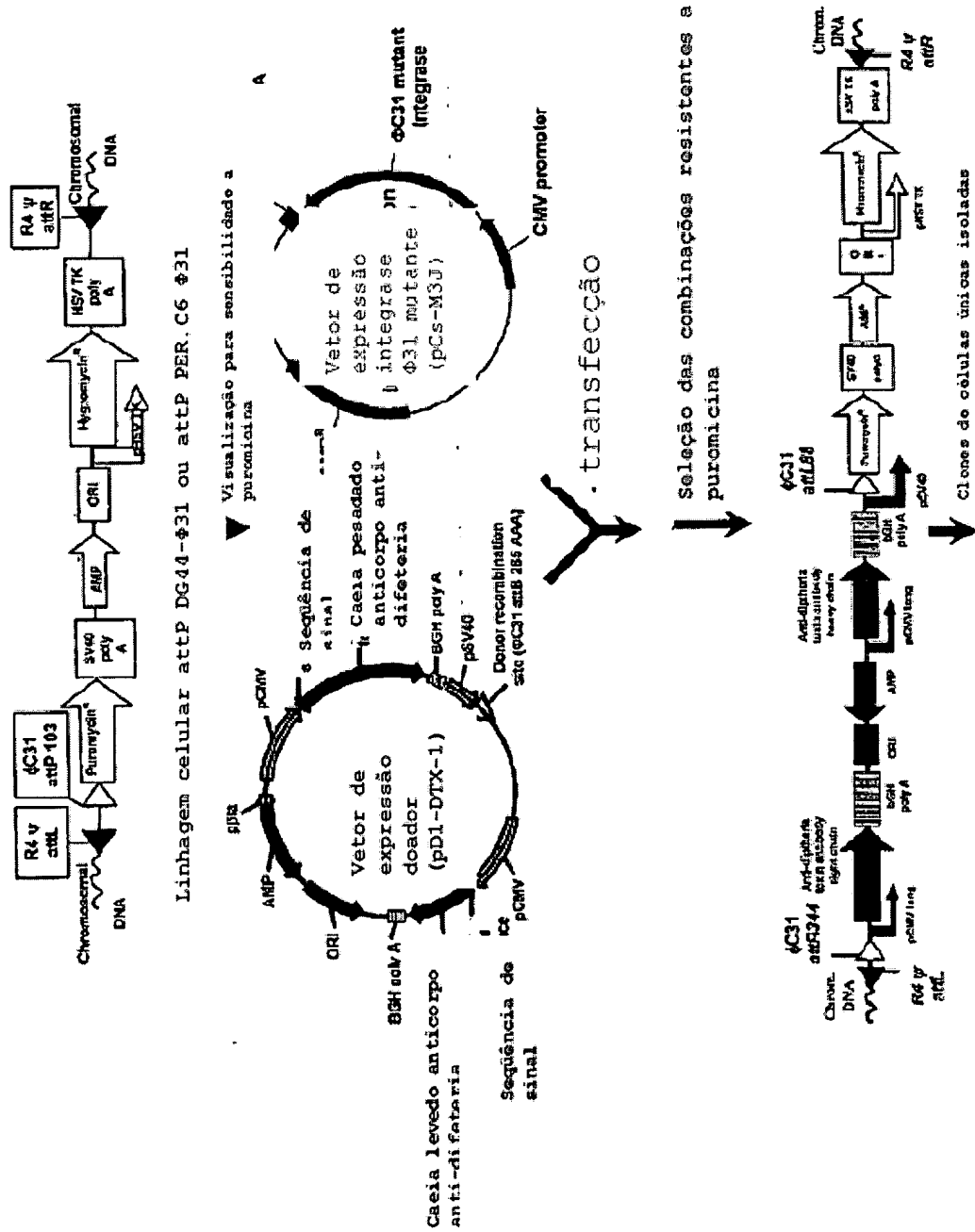


FIG. 13

Linagem celular DG44-ou PER.C6

Sequência de phiC31 attB and attP:

phiC31 attB
gtcgacgATGt

phiC31 attP 221:
 gtactgacggacacacccgaagcccgcggaaccctcagcggATGccccggggcttcacgttttccaggtc
 (SEQ ID NO:68)

phiC31 attB 285:
tcacgggtctcgaagccggtcgccgtccagggcgtgccTTGggctccccggggcgctactccacctcaccca
tATGATCaacgggtc (SEQ ID NO:69)

phiC31 attP 221:
 agaagcggttttcgggagtgtgccccaaactggggtaacctTTGagttctctcagttggggggcgtagggtcggcG
AcATGacacaaagggg (SEQ ID NO:70)

Recombination cross-over point: TTG

Minimal attB: gtgcccgggcgtgccTTGggctccccggggcg (SEQ ID NO:71)

Minimal attP: ccctaaactggggtaacctTTGagttctctcagttggggg (SEQ ID NO:72)

phiC31 attB 285:
 gaggtggcggttagttgatcccgggcgaacgtcgcgcgccacgggaagccctcgccctcgaaacgggtggggcggt
 gctgggtccatcgtca (SEQ ID NO:73)

phiC31 attB 285:
 cgggtgagcacgggacgtgcgacgggctcggggggtcgggatacgggggchagcgtcagcgsggtctcgacgggtca
 cggcgggcatgtcgac (SEQ ID NO:74)

phiC31 attP 221:
ttgtgaccgggtggacacgtacgggggtgcttaacyaccgtcagtcgcgcgagcgca (SEQ ID NO:75)

Sequência of phiC31 attL 88 in cell lines:

SV40 promoter.....57, 62, or 74 bases of SV40 5' untranslated region.....

gtcgacgAAAtaggtcacggtctcgagcccggtgagggtgcccaggggcgtgccctTGG (SEQ ID NO:76)

<-----attL derived from attB----->

ATG in attB changed to AAA TGG: recombination cross over point

agttctctcagttggggcgtagggtcgccACcATG (SEQ ID NO:77)

<-----derived from attP----->

GA changed to AC to make
 consensus Kozak sequence
ATG in attP made start codon for
 puromycin resistance coding region

accgagtagaagccacgggtgcgcctcgccaccggcgacgagtcggggggcgt (SEQ ID NO:78)

<-----puromycin resistance coding region----->

FIG. 14

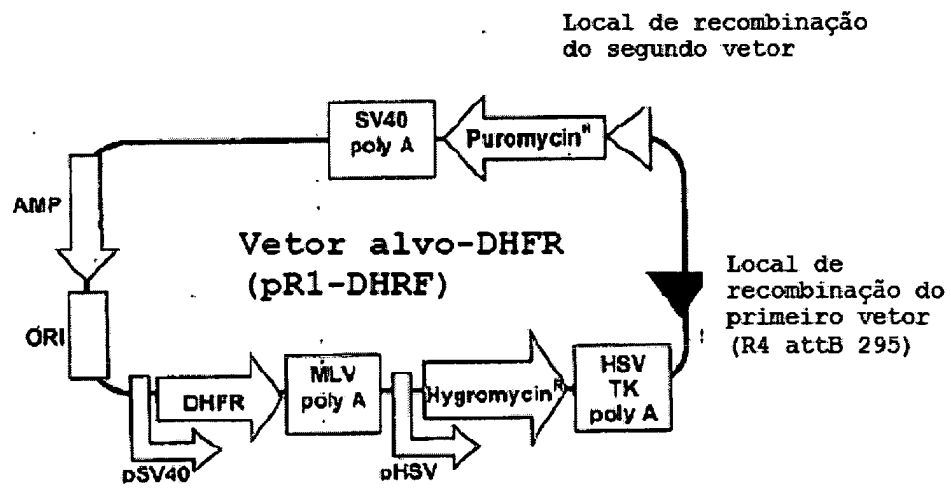


FIG 15.

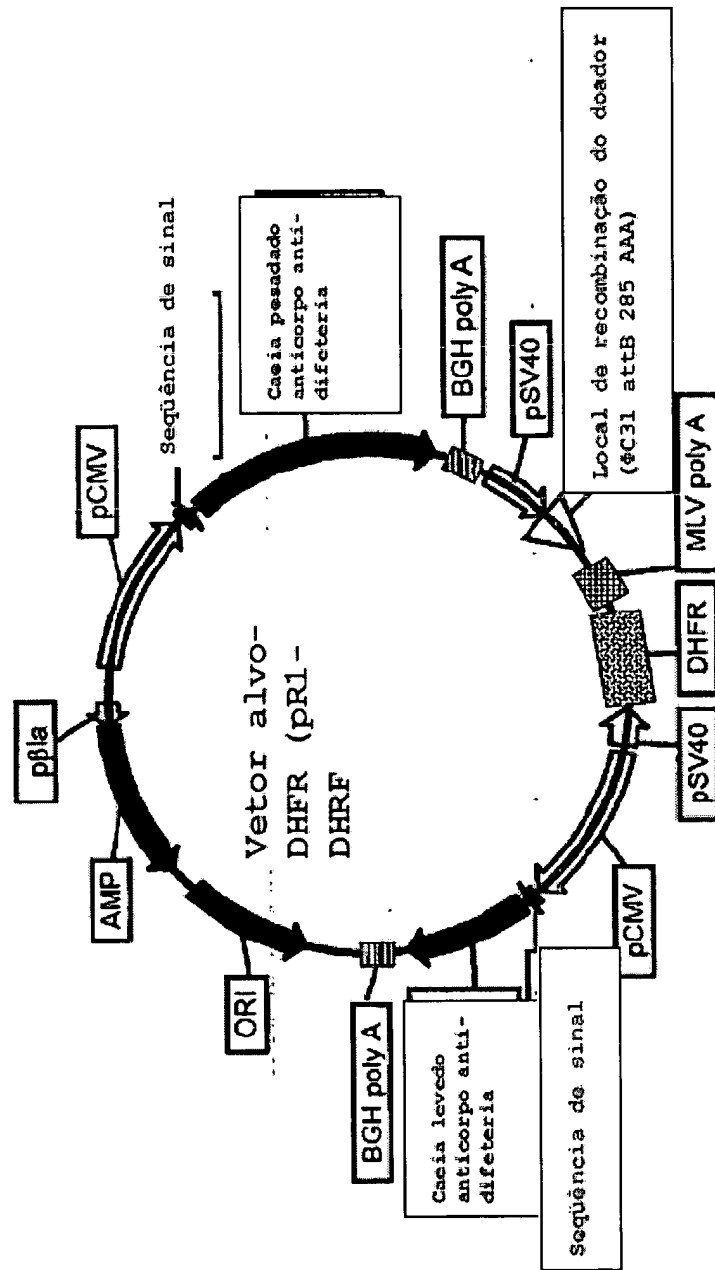


FIG. 16

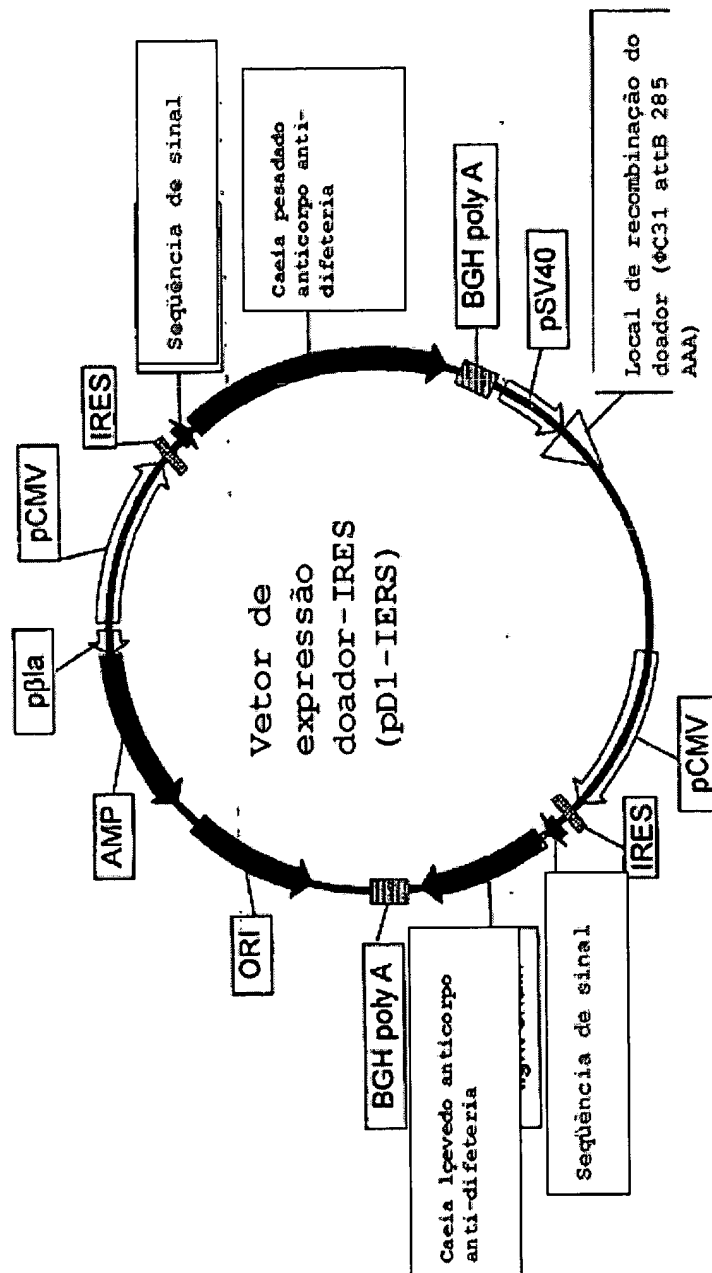


FIG. 17

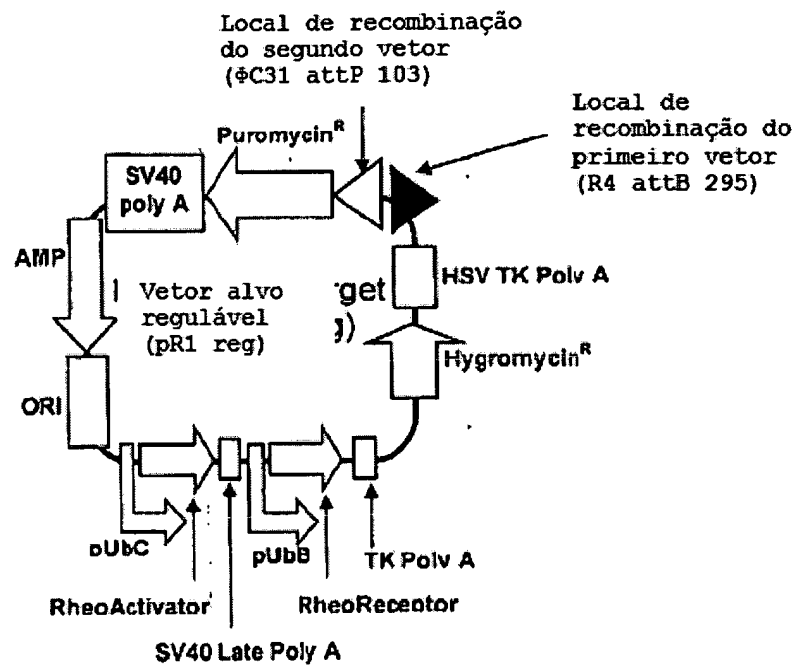


FIG. 18

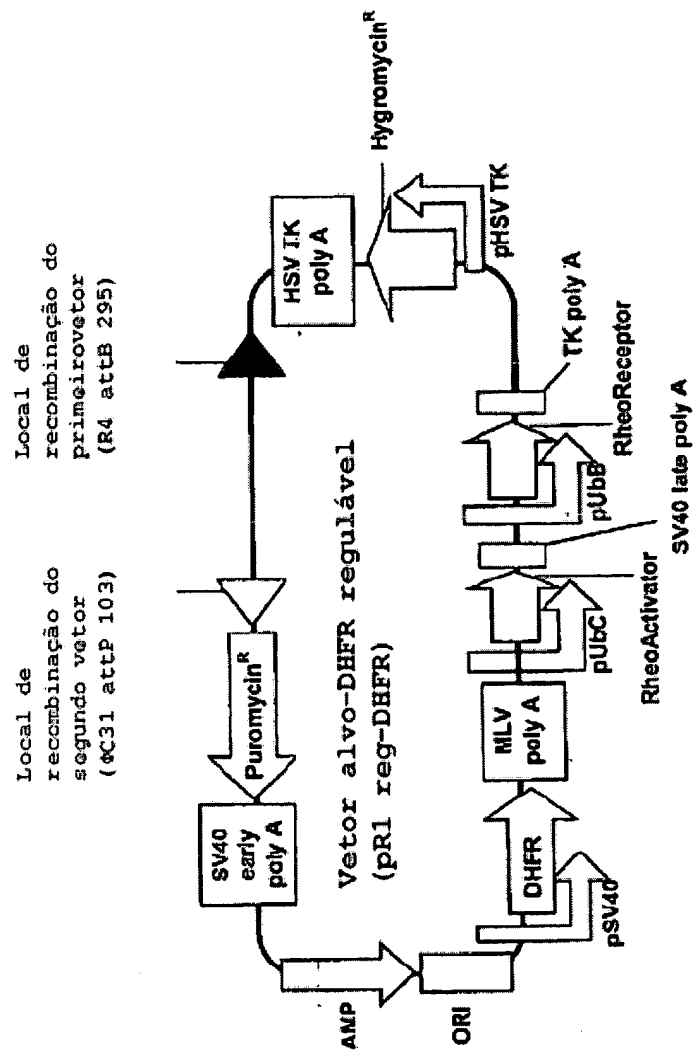


FIG. 19

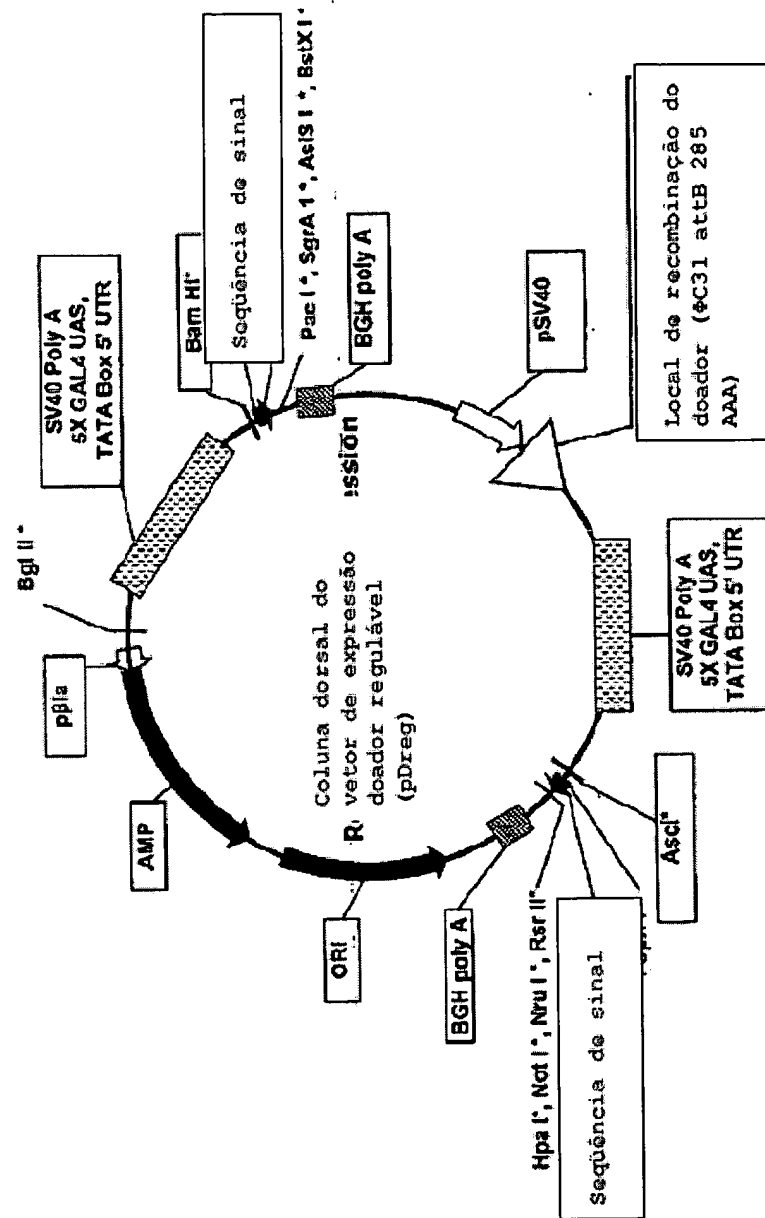


FIG. 20

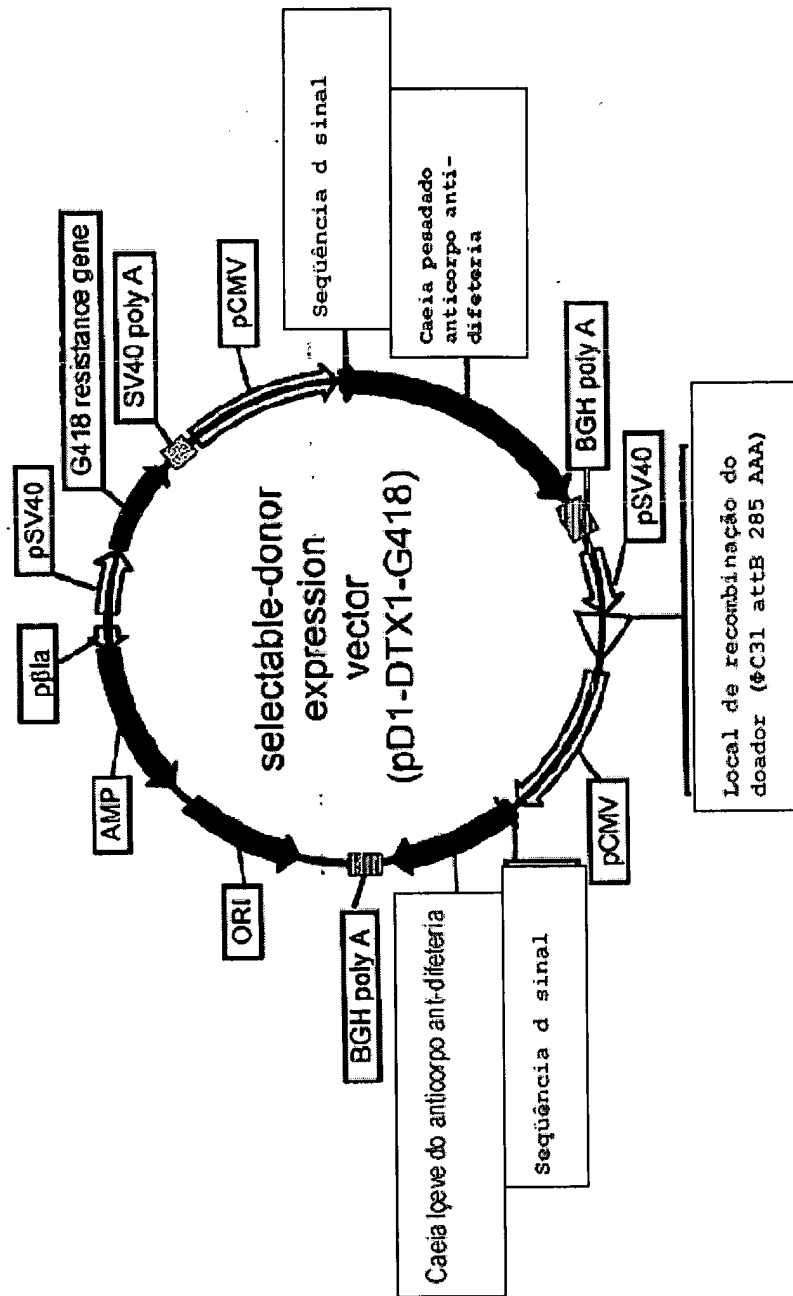


FIG. 21

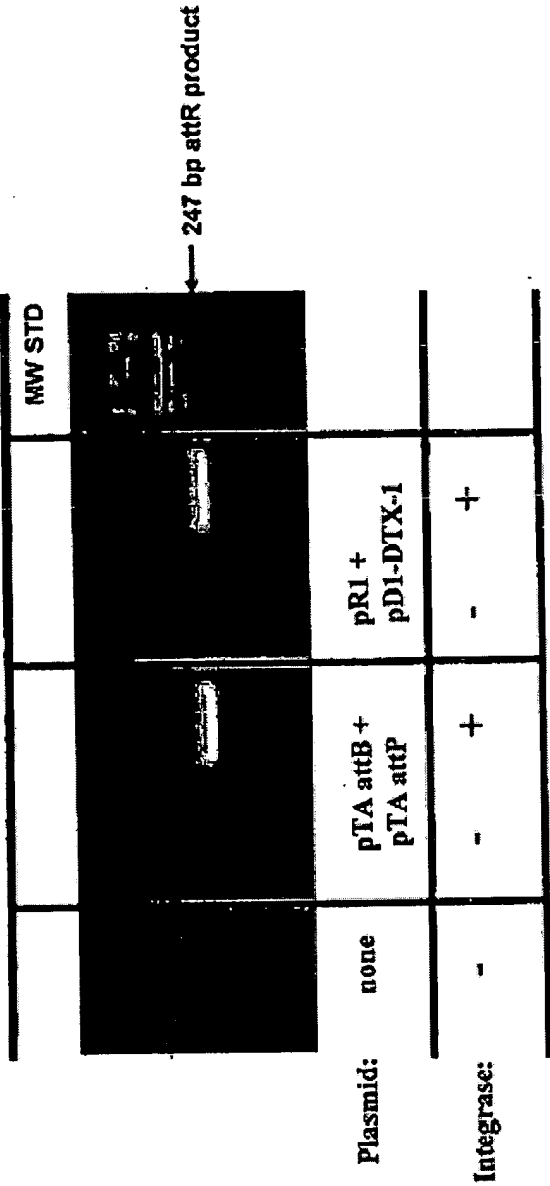


FIG. 22

<u>Cells</u>	<u>Sequence at Junction</u>
DG44	R4 attL: cccatgaccatgccGAAGCAGTGGTA (SEQ ID NO:79) Observed: cccatgaccatgc. tgggcaagatt (SEQ ID NO:80) SINE-like element
DG44	R4 attL: cccatgaccatgccGAAGCAGTGGTA (SEQ ID NO:81) Observed: cccatgaccatgc. tgac (SEQ ID NO:82) unidentified sequence

FIG. 23

A) ϕ C31 att L site from a DG44 target vector clone,
transfected with donor expression and ϕ C31 integrate plasmids

```

<-----pSV40 region----->
      (from donor expression vector)
Predicted: ...attccagaagtagtgaggaggttttttgaggcctagggttttgcaaaaagctccc
Observed: ..attccagaagtagtgaggaggttttttgaggcctagggttttgcaaaaagctccc

<-----attL----->
<----- attB-derived -----> <--
Predicted: gtgcacgaaataggtcacggtctcgaaagcgcggtgcggtgcagggcgtgccccTTgag
Observed: gtgcacgaaataggtcacggtctcgaaagcgcggtgcggtgcagggcgtgccccTTgag
              crossover point/

-----attL----->
-----attP-derived-----><-----puromycin resistance
              (from target vector)
Predicted: ttctctcagttggggcgtaggtgcgccacATGaccgagtagataagccccacggtgcg...
Observed: ttctctcagttggggcgtaggtgcgccacATGaccgagtagataagccccacggtgcg...
              Start codon/

Predicted: SEQ ID NO:83
Observed: SEQ ID NO:84

B)  $\phi$ C31 att R site from a DG44 target vector clone,
transfected with donor expression and  $\phi$ C31 integrate plasmids

<-----attR-----> <-----pCMV----->
<-attP-derived-> <-- attB-derived --> (from donor vector)
      (from target vector)
Predicted: ccaactggggtaacctTTgggctccccggcggtact...aattgcatgaagaatc
Observed: ccaactggggtaacctTTgggctccccggcggtact...aattgcatgaagaatc
              crossover point/

Predicted: SEQ ID NO:85
Observed: SEQ ID NO:86

```

FIG. 24

A) ϕ C31 att L site from a PER.C6 target vector clone, transfected with donor expression and ϕ C31 integrase plasmids

```

<-----pSV40 region----->
      (from donor expression vector)
Predicted: ...attccagaagtagtgaggagcttttttggaggcccttaggcttttgcataaaagctccc
Observed: ...attccagaagtagtgaggagcttttttggaggcccttaggcttttgcataaaagctccc

<-----attL----->
<----- attB-derived -----> <--
Predicted: gtccagcaaataggtcacggctctcgaagccgcggtgcgggtgccagggcgctgccctTtgag
Observed: gtccagcaaataggtcacggctctcgaagccgcggtgcgggtgccagggcgctgccctTtgag
              crossover point/

-----attL----->
-----attP-derived-----><-----puromycin resistance--
              (from target vector)
Predicted: ttcttcagttggggcgtagggtcgccaccATGaccgagtagacaagcccaagcgtgcg...
Observed: ttcttcagttggggcgtagggtcgccaccATGaccgagtagacaagcccaagcgtgcg...
              Start codon/

Predicted: SEQ ID NO: 87
Observed: SEQ ID NO: 88

```

B) ϕ C31 att R site from a PER.C6 target vector clone transfected with donor expression and ϕ C31 integrase plasmids

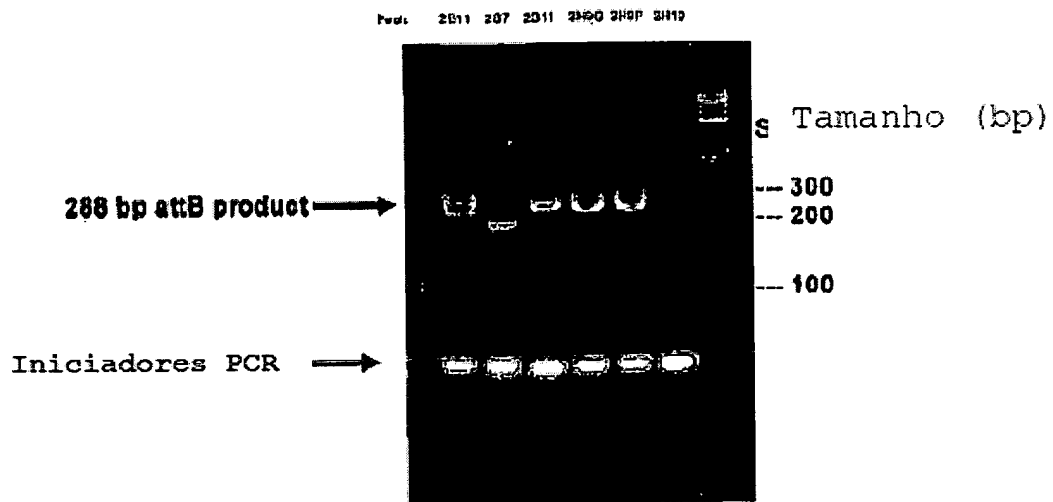
```

<----- attR ----->
<--- attP-derived ----> <----- attB-derived ----->
      (from target vector)      (from donor vector)
Predicted: gtagtgcaccaactggggtaaccTTtgggtcccccggcgctactccaccatctct
Observed: gtagtgcaccaactggggtaaccTTtgggtcccccggcgctactccaccatctct
              crossover point/

Predicted: SEQ ID NO: 89
Observed: SEQ ID NO: 90

```

FIG. 25



B) ϕ C31 att R PCR

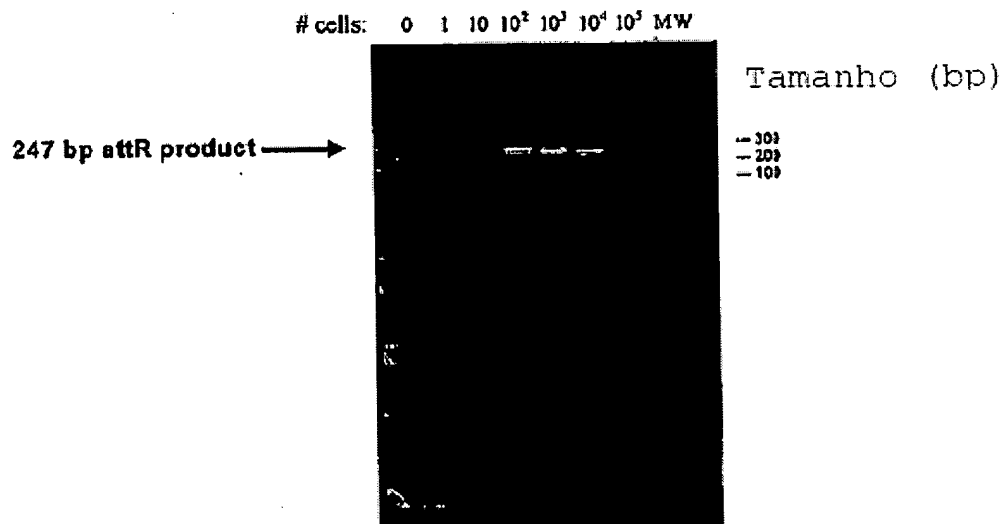


FIG. 26

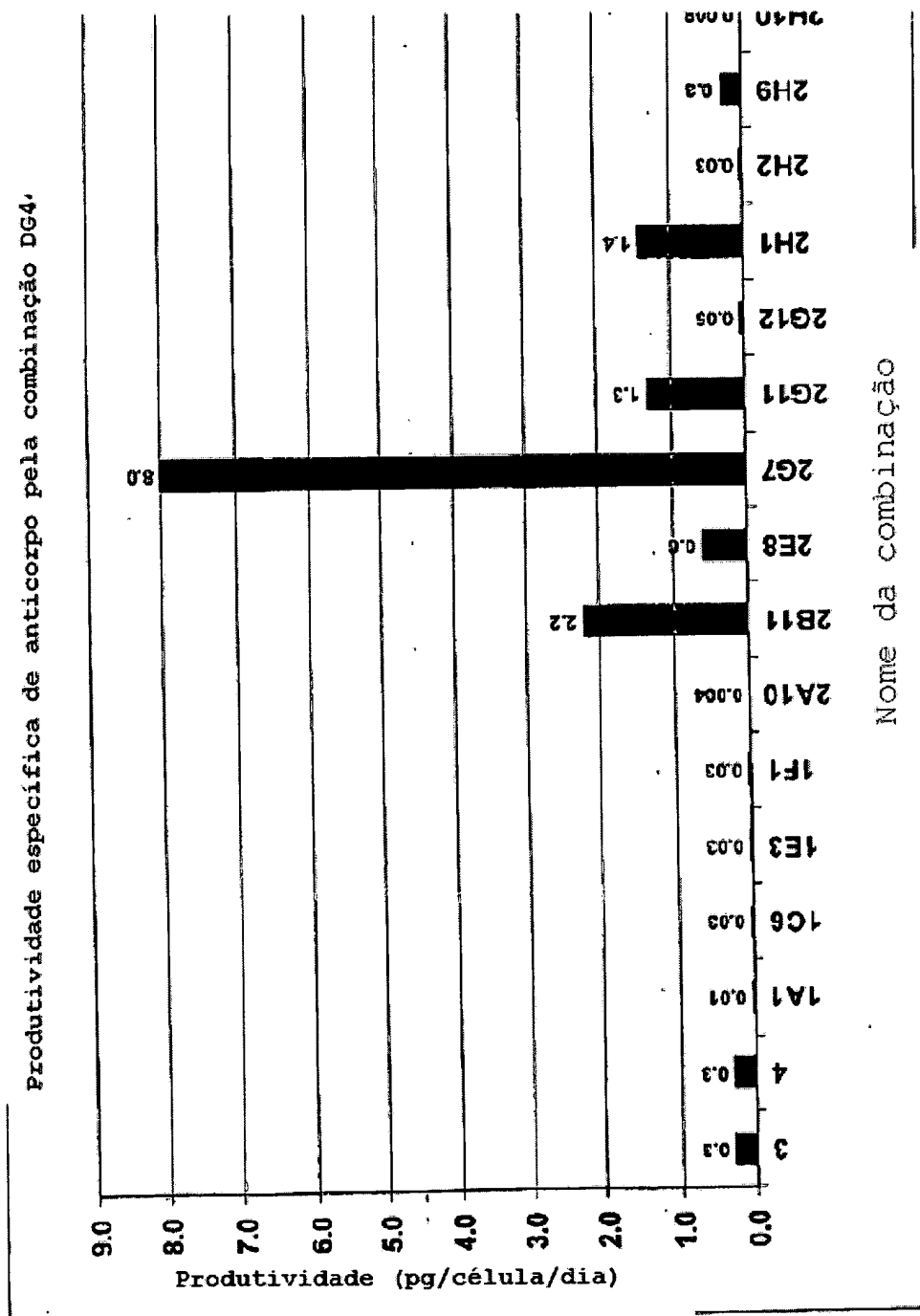


FIG. 27A

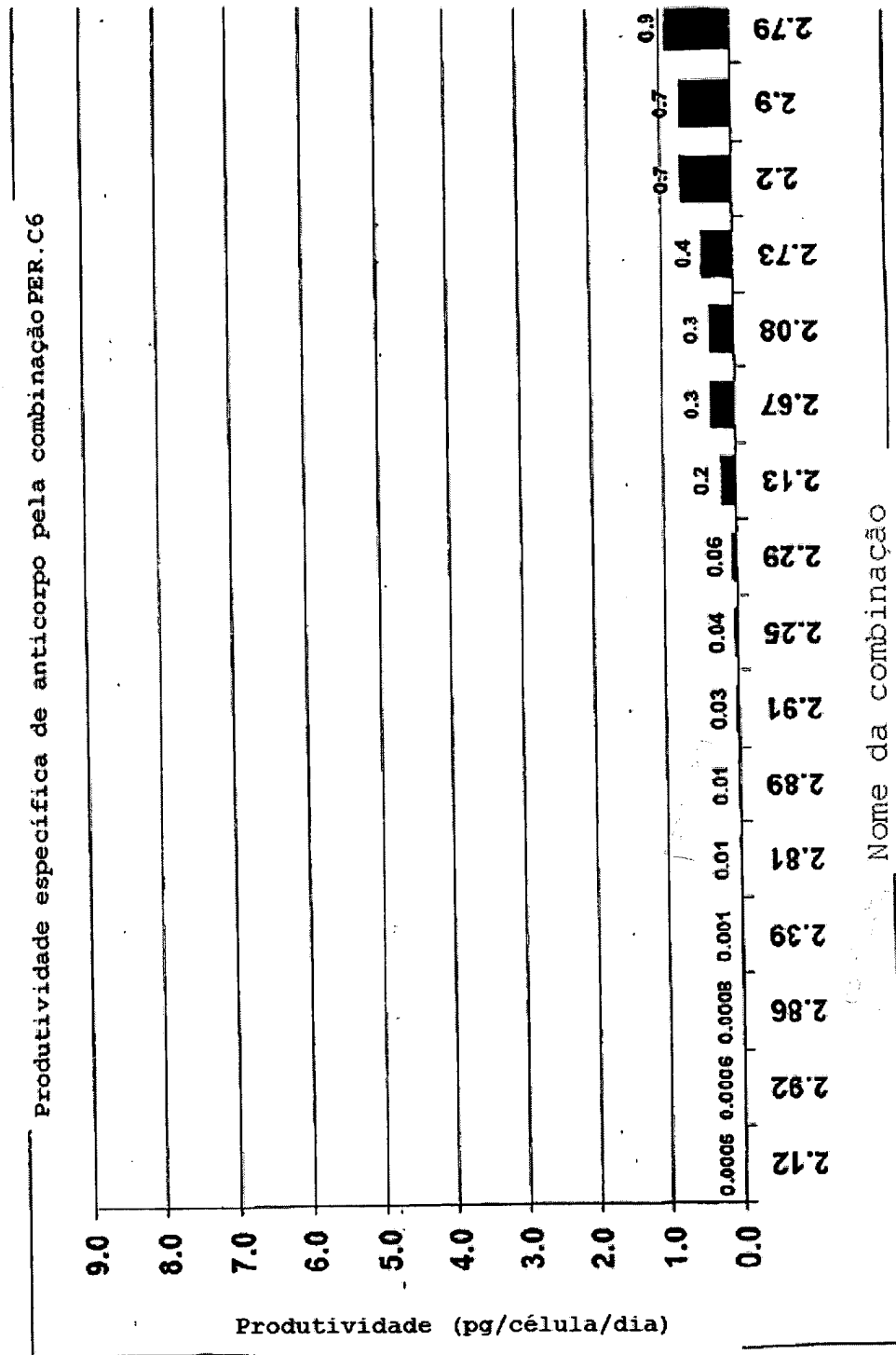


FIG. 27B

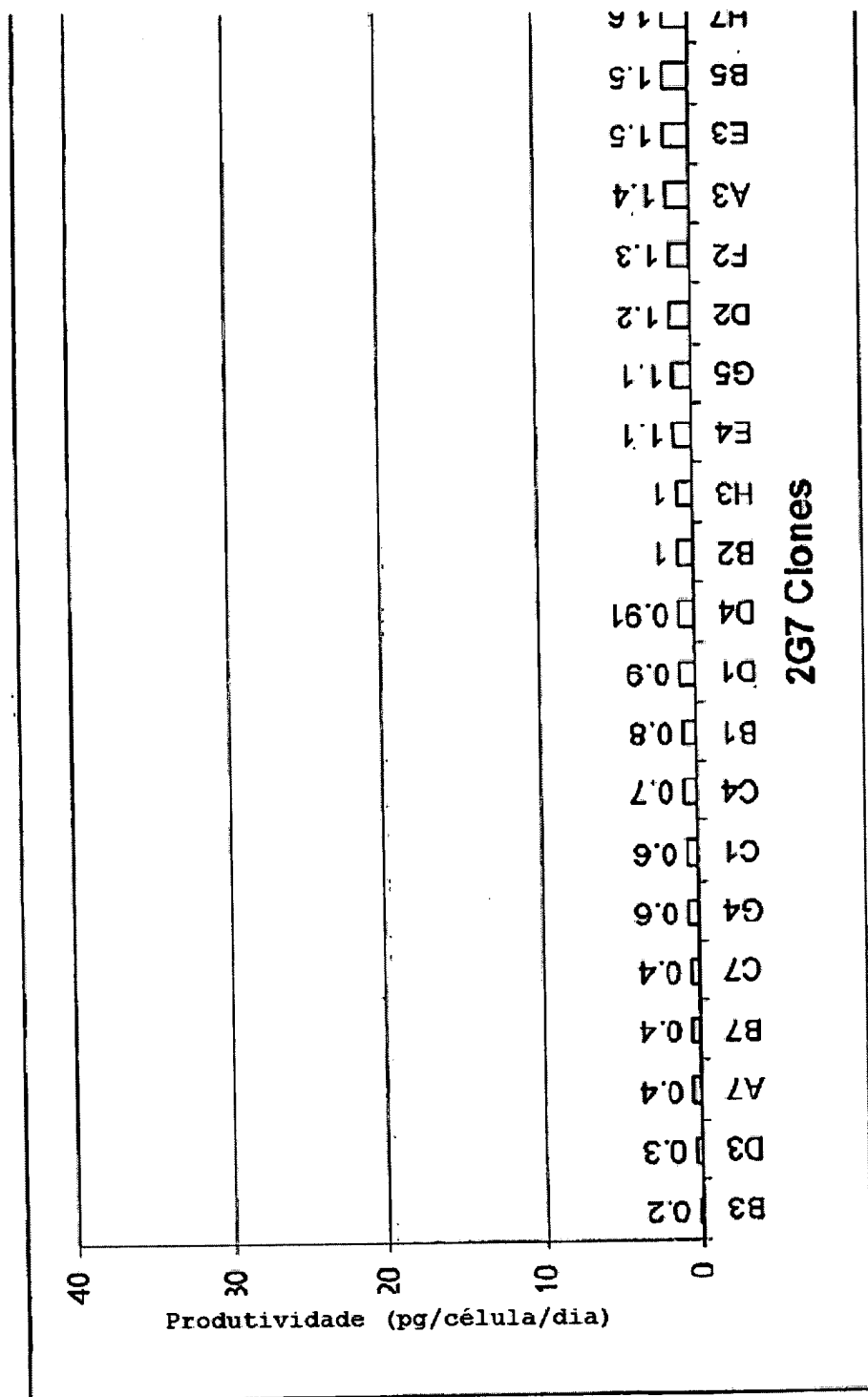


FIG. 28A

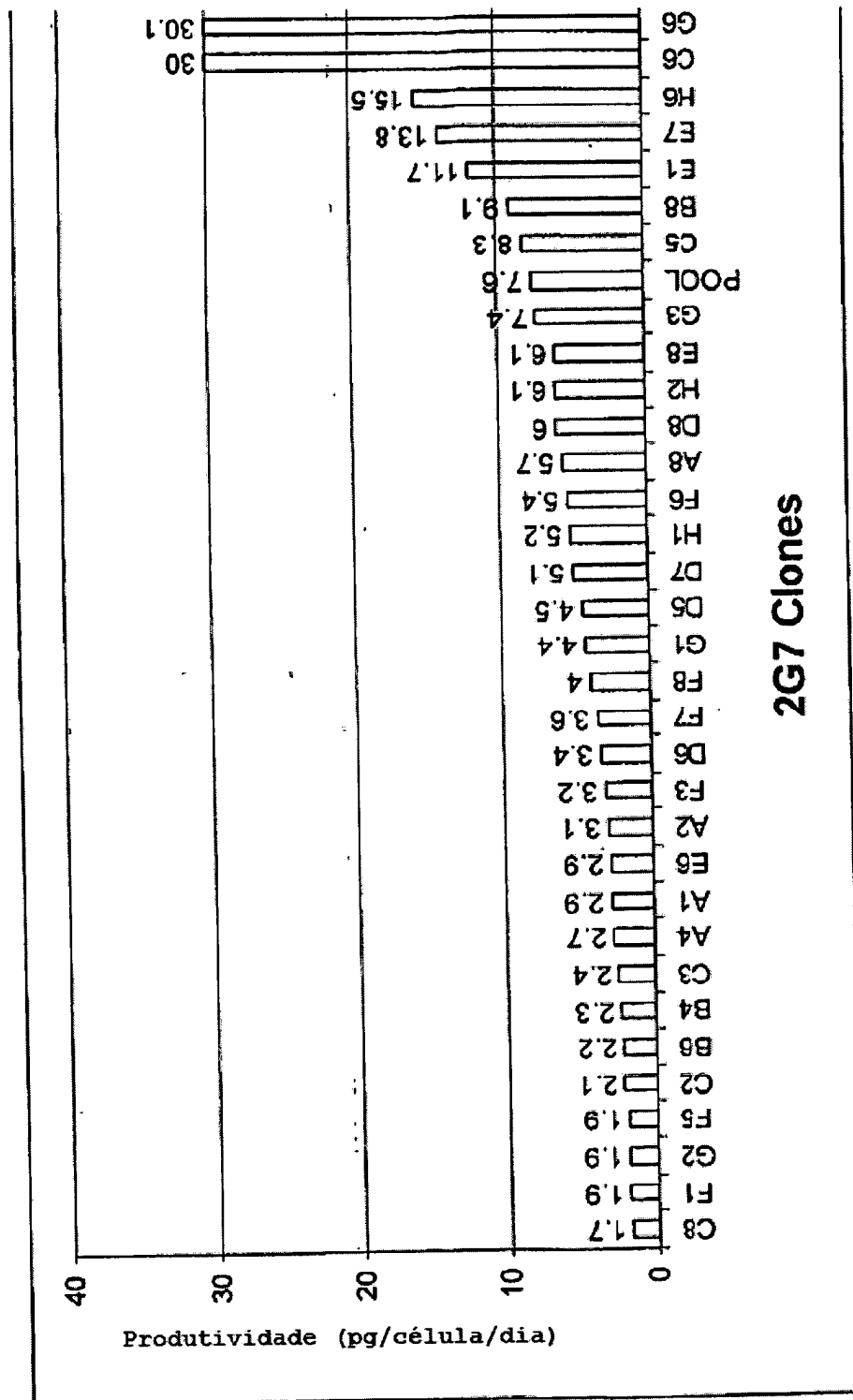


FIG. 28B

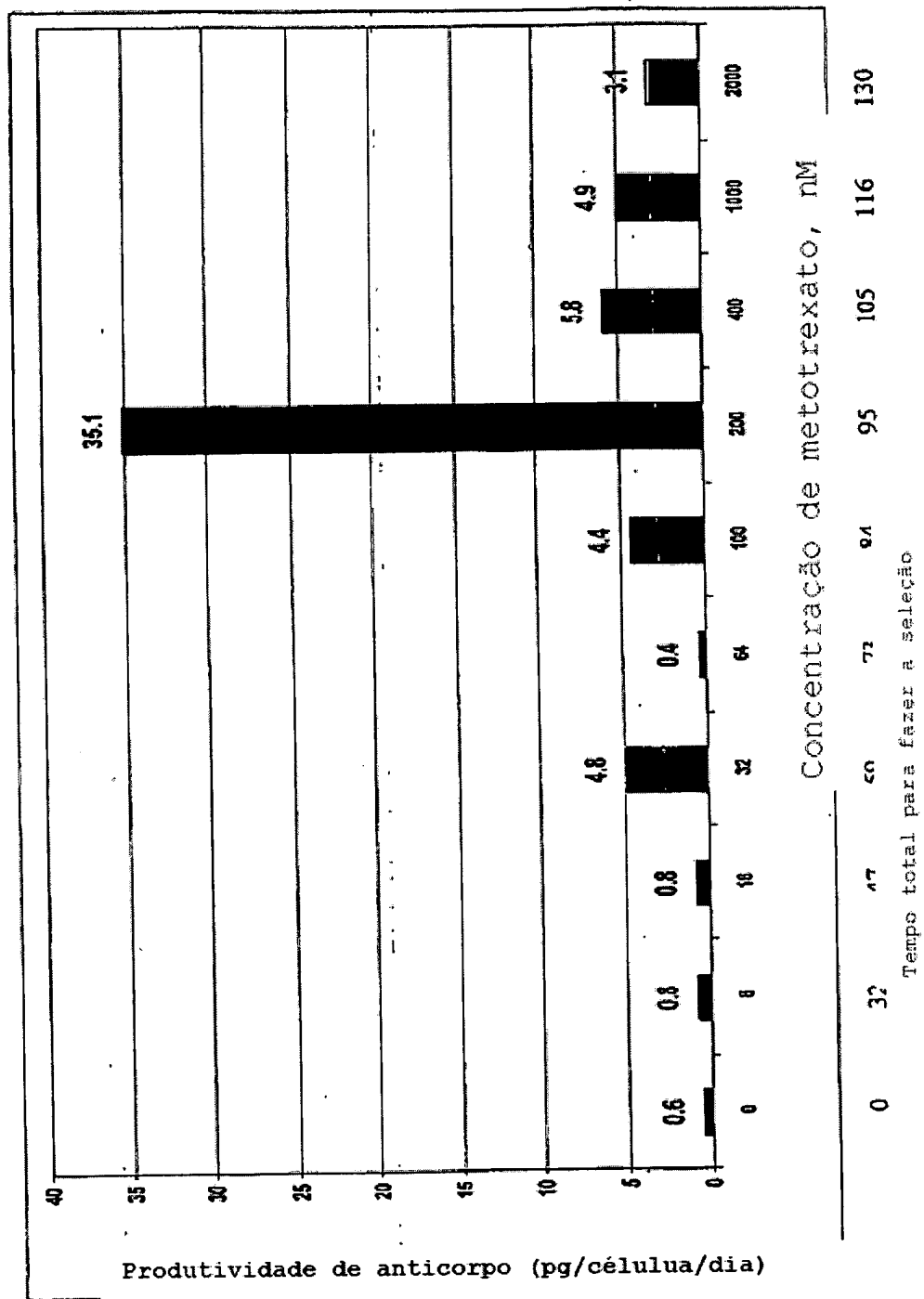


FIG. 29

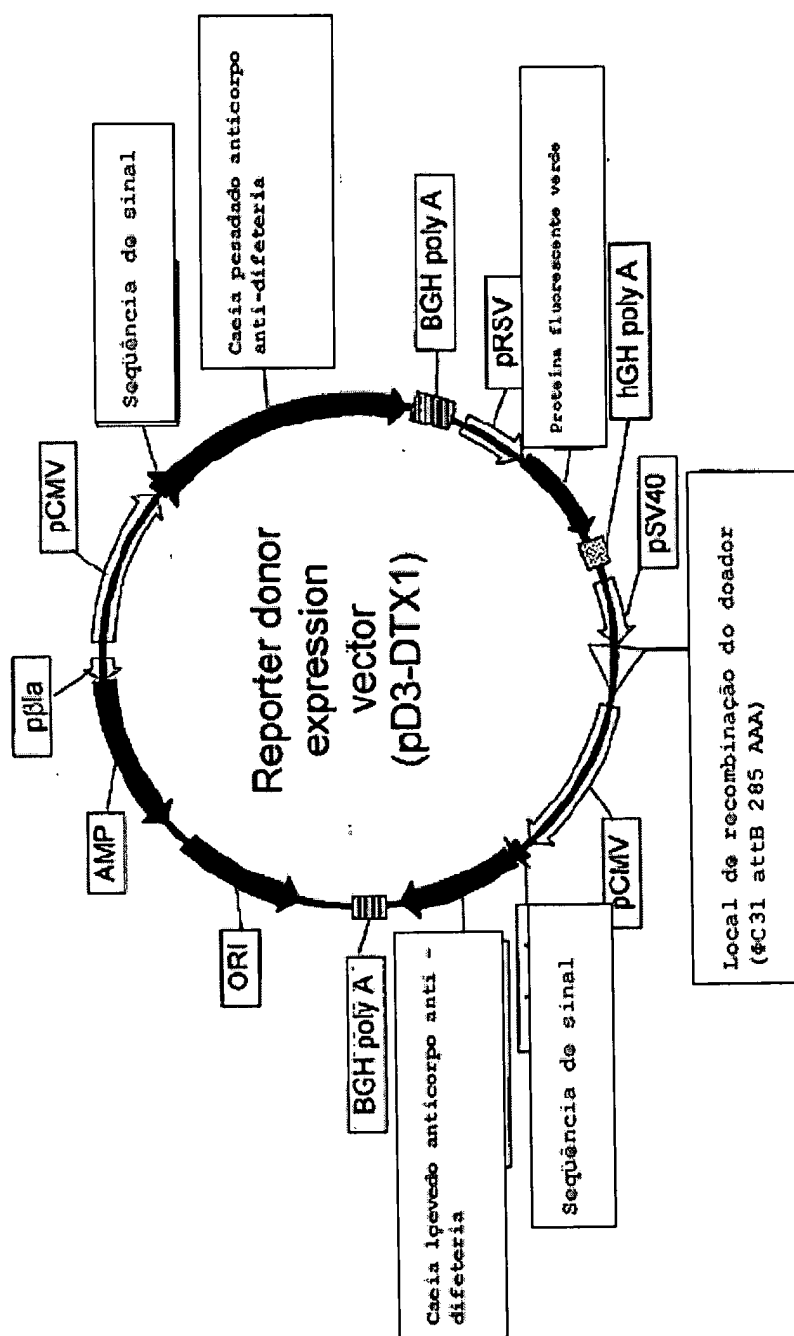
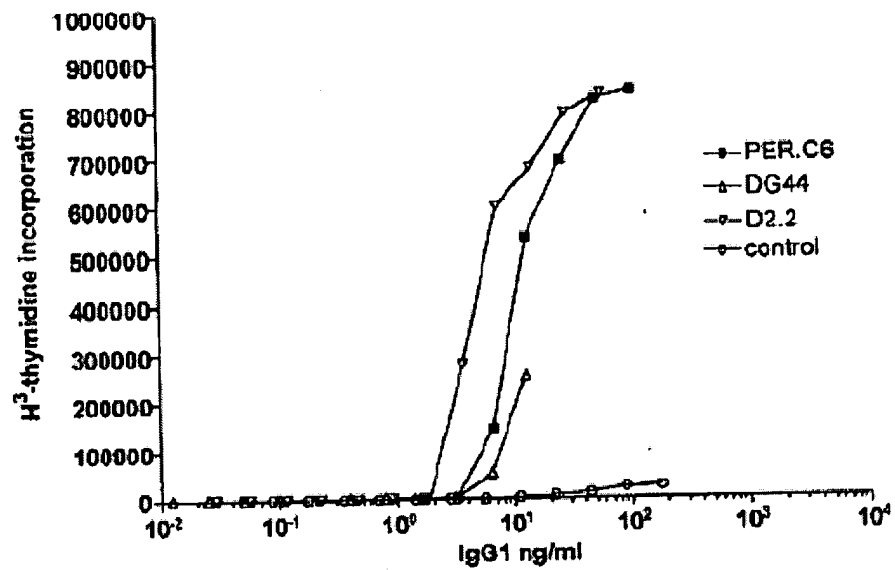


FIG. 30

Célula:	<u>DG44</u>	<u>PER.C6</u>
Total IgG (mg/L):	0.3	1.0
	↓	↓
Toxina IgG anti-difitérica ; (IU/L):	12	45
	↓	↓
Atividade específica / (IU/mg):	40	45

FIG. 31

Atividade de neutralização de anticorpos toxina anti-diftér



Source	EC50 (ng/ml)
PER.C6	11
DG44	8
D2.2	5

FIG. 32

FIG. 33A

pR1 (5437 bp)

CCTTTCGTCTTCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGC
 AGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTTATA
 AGCTGCAATAAACAAAGTTAACAAACAATTCATTCATTTTATGTTTCAGGTTTCAGGGGGA
 GGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGCTGATTATGATC
 CTCTAGAGTCGGTGGGCCCTCGGGGGCGGGTGCGGGGTTCGGCGGGGCCGCCCCGGGTGGCTTC
 GGTTCGGAGCCATGGGGTCGTGCGCTCCTTTCGGTTCGGGCGCTGCGGGTTCGTGGGGCGGGCGT
 CAGGCACCGGGCTTGCGGTCATGCACAGGTGCAGCGTTCCTTCGGGCACCTCGACGTCCGC
 GGTGACGGTGAAGCCGAGCCGCTCGTAGAAGGGGAGGTTGCGGGGCGCGAGGTCTCCAGGA
 AGGCGGGCACCCCGGCGCGCTCGGCCGCTCCACTCCGGGGAGCACGACGGCGCTGCCCAGA
 CCTTTCCTGCTGGTGGTTCGGGCGAGACGCCGACGGTGGCCAGGAACCAACGCGGGCTCCTTGGG
 CCGGTGCGGCGCCAGGAGGCCCTTCATCTGTTGCTGCGCGGCCAGCCGGGAACCGCTCAACT
 CGGCCATGCGCGGGCCGATCTCGGCGAACACCGCCCCCGCTTCGACGCTCTCCGGCGTGGTC
 CAGACCGCCACCGCGCGCGCTCGTCCGCGACCCACACCTTGCCGATGTCGAGCCCGACCGG
 CGTGAGGAAGAGTTCTTGACGCTGGGTGACCCGCTCGATGTGGCGGTTCGGGGTTCGACGGTGT
 GCGCGTGGCGGGGTAGTCGGCGAACCGCGCGCGAGGGTGCCTACGGCCCCGGGGGACGTCCG
 TCGCGGGTGGCGAGGCGCACCGTGGGCTTGTAATCGGTTCATGGTGGCGACCCCTACGCCCCCA
 ACTGAGAGAACTCAAAGGTTACCCAGTTGGGGCACTACTCCCGAAAACCGCTTCTGACCTG
 GGAAAACGTGAAGCCCCGGGGCAAAGGGCGAATTCTGACAGATAAATTAGGCAAAGGAATTCC
 TCGACCTGCAGCCCCAAGCTAATTGCCCCCTCGTGGGGACGCCGTACAGGGACGTGCACCTCT
 CCCGCTGCACCGCTCCAGCGTCGCGCGCGGCTCGAAGGACGGGGCCGGGATGACGATGCAG
 GCGGCGTGGGAGGTGGCGCCCAAGTTGCCCATGACCATGCCGAAGCAGTGGTAGAAGGGCAC
 CGGCAGACACACCGGTCTGCTCCGTGTAGCCGACCGTGCAGGCCACCCAGTAGCCGTGT
 TGAGGATGTTGTGGTGGGAGAGCGTGGCGCCCTTGGGGAAGCCGGTGGTGCAGGAGGTGTAC
 TGGATGTTGACCGGGAAGGGCGAATTAGCTTGGCACTGGCGCCAGAAATCCGCGCGGTGGTT
 TTTGGGGGTTCGGGGTGTGTTGGCAGCCACAGACGCCCGGTGTTCGTGTGCGCCAGTACATG
 CGGTCCATGCCAGGCCATCCAAAACCATGGGTCTGTCTGCTCAGTCCAGTCTGGACCTG
 ACCCCACGCAACGCCCAAATAATAACCCCAAGAACCATAAACCATTCCCCATGGGGGACC
 CCGTCCCTAACCCACGGGGCCAGTGGCTATGGCAGGGCCTGCCGCCCGACGTTGGCTGCGA
 GCCCTGGGCCCTTCACCCGAACCTGGGGGTGGGGTGGGGGAAAAGGAAGAAACGCGGGCGTAT
 TGGCCCCAATGGGGTCTCGGTGGGGTATCGACAGAGTGCCAGCCCTGGGACCGAACCCCGCG
 TTTATGAACAAACGACCCAAACACCCGTGCGTTTATTCTGTCTTTTATTGCGGTCATAGCG
 CGGGTTCCCTTCGGTATTGTCTCCTTCGGTGTTCAGTTAGCCTCCCCCATCTCCCGATCCC
 CACGAGTGTGGGGCGTGGTTCCTACTATCGGCGAGTACTTCTACACAGCCATCGGTCCAG
 ACGGCCGCGCTTCTGCGGGCGATTGTGTACGCCCGACAGTCCCGGCTCCGGATCCGACGAT
 TGGTTCGATCGACCCCTGCGCCCAAGCTGCATCATCGAAATTGCCGTCAACCAAGCTCTGAT
 AGAGTTGGTCAAGACCAATGCGGAGCATATACGCCCGGAGCCGCGGCGATCCTGCAAGCTCC
 GGATGCCTCCGCTCGAAGTAGCGGTCTGCTGCTCCATACAAGCCAACACGGGCTCCAGAA
 GAAGATGTTGGCGACCTCGTATTGGGAATCCCGAACATCGCCTCGCTCCAGTCAATGACCG
 CTGTTATGCGGCCATTGTCCGTGAGGACATGTTGGAGCCGAAATCCGCGTGCACGAGGTGC
 CGGACTTCGGGGCAGTCCCTCGGCCCAAAGCATCAGCTCATCGAGAGCCTGCGCGACGGACGC
 ACTGACGGTGTGCTCCATCACAGTTTGCCAGTGATACACATGGGGATCAGCAATCCGCGCAT
 TGAAATCACGCCATGTAGTGTATTGACCGATTCTTTCGGTCCGAATGGGCCGAACCCGCTC
 GTCTGGCTAAGATCGGCCGAGCGATCGCATCCATGGCCTCCGCGACCGGCTGCAGAACAGC
 GGGCAGTTTCGGTTTCAGGCAGGTCTTGCAACGTGACACCTGTGCACGGCGGGAGATGCAAT

FIG. 33B

AGGTCAGGCTCTCGCTGAATTCCCAATGTCAAGCACTTCCGGAATCGGGAGCGCGGCCGAT
 GCAAAGTGCCGATAAACATAACGATCTTTGTAGAAACCATCGGCGCAGCTATTTACCCGAG
 GACATATCCACGCCCTCCTACATCGAAGCTGAAAGCACGAGATTCTTCGCCCTCCGAGAGCT
 GCATCAGGTCCGAGACGCTGTGAACTTTTCGATCAGAACTTCTCGACAGACGTCGCGGTG
 AGTTCAGGCTTTTTCATATCAAGCTGATCTTGCGGCACGCTGTTGACGCTGTTAAGCGGGTC
 GCTGCAGGGTCGCTCGGTGTTGAGGCCACACGCGTCACCTTAATATGCGAAGTGGACCTGG
 GACCGCGCCGCCCGGACTGCATCTGCGTGTTCGAATTCGCCAATGACAAGACGCTGGGCGGG
 GTTTGTGTATCATAGAACTAAAGACATGCAATATATTTCTTCGGGGACACCGCCAGCAA
 ACGCGAGCAACGGGCCACGGGGATGAAGCAGGGCGGCACCTCGCTAACGGATTCACTACTCC
 AAGAATTGGAGCCAATCAATCTTGCGGAGAACTGTGAATGCGCAAACCAACCTTGGCAGA
 ACATATCCATCGCGTCCGCCATCTCCAGCAGCCGACGCGGCGCATCTCGGGGCGGACGCGC
 TGGGCTACGTCTTGCTGGCGTTCCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTT
 CTCGCTTCGGCGGCATCGGGATGCCCGCGTTGACGGCCATGCTGTCCAGGCAGGTAGATGA
 CGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCGCCAGCAAAGGCCAG
 GAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATC
 ACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCG
 TTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCCTGCCGCTTACCGGATACCT
 GTCGCGCTTTCTCCCTTCGGGAAGCGTGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCA
 GTTCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTAGCCCCGAC
 CGCTCGCGCTTATCCGGTAACCTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCC
 ACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGT
 TCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTCGCGCTCTG
 CTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGC
 TGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAG
 AAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAACGAAAACCTCACGTTAAGGG
 ATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAG
 TTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCA
 GTGAGGCACCTATCTCAGCGATCTGTCTATTTCTGTTTATCCATAGTTGCCTGACTCCCCGTC
 GTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCG
 AGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACAGCCAGCCGGAAGGGCCGAGC
 GCAGAAGTGGTCTTGCAACTTTATCCGCCCTCCATCCAGTCTATTAATTGTTGCCGGAAGCT
 AGAGTAAGTAGTTCGCCAGTTAATAGTTTGGCGAACGTTGTTGCCATTGCTGCAGGCATCGT
 GGTGTACGCTCGTCTGTTTGGTATGGCTTCATTACAGCTCCGGTTCCTAACGATCAAGGCCGAG
 TTACATGATCCCCCATGTTGTGCAAAAAAGCGTTAGCTCCTTCGGTCTCCGATCGTTGTC
 AGAAGTAAGTTGGCCGAGTGTATCACTCATGGTTATGGCAGCACTGCATAATTTCTTTAC
 TGTATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAG
 AATAGTGATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGATAATACCGCGCCA
 CATAGCAGAACTTTAAAAGTGCTCATCATTTGGAAACGTTCTTCGGGGCGAAAACCTCTCAAG
 GATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCAACCAACTGATCTTCAG
 CATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAA
 AAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTCAATATTATTG
 AAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATA
 AACAAATAGGGGTTCCGCGCACATTTCCCGGAAAAGTGCCACCTGACGTCCTAAGAAACCAT
 ATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGC

(SEQ ID NO:91)

FIG. 34A

pD1-DTX-1 (8110 bp)

GACGGATCGGGAGATCTCCCGATCCCCCTATGGTGGACTCTCAGTACAATCTGCTCTGATGCC
 GCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGGAG
 CAAAATTTAAGCTACAACAAGGCAAGCCTTGACCGACAATTGCATGAAGAATCTGCTTAGGG
 TTAGGCGTTTTGCGCTGCTTCGCTAGGTGGTCAATATTGGCCATTAGCCATATTATTATTG
 GTTATATAGCATAAATCAATATTGGCTATTGGCCATTGCATACGTTGTATCCATATCATAAT
 ATGTACATTTATATTGGCTCATGTCCAACATTACCGCCATGTTGACATTGATTATTGACTAG
 TTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCCGCGTTA
 CATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCCGCCCATTGACGTCA
 ATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGA
 GTATTTACGGTAAACTGCCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCC
 CTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGG
 GACTTTCTACTTGGCAGTACATCTACGTATTAGTTCATCGCTATTACCATGGTGTATGCGGTT
 TTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCGAAGTCTCCACC
 CCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCTG
 AACAACTCCGCCCCATTGACGCAATGGGCGGTAGGCGGTACGGTGGGAGGTCTATATAAG
 CAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCC
 ATAGAAGACACCGGGACCGATCCAGCCTCCGCGGCCGGGAACGGTGCATTGGAAGCTTGGA
 CCGAGCTCGGATCCGCCACCATGGCATGCCCTGGCTTCCTGTGGGCACTTGTGATCTCCACG
 TGTCTTGAATTTCCATGGCTGAAGTGCAGCTGGTGGAGTCTGGGGGAGGCTTGGTCAAGCC
 TGGAGGGTCCCTGAGACTCTCCTGTGAAGCCTCTGGATTCTCTTCACTGACTACTACATGA
 GCTGGATCCGCCAGGCTCCAGGGAAGGGGCTGGAATGGATTTTCATACATTAGTCTTAGTGGT
 AGTACCCTATACTACGCAGACTCTATGAGGGGGCCGATTACCATCTCCAGGGACAACGCCAA
 GAACTCACTGTATCTGCAAATGAACAGCCTGAGAGTCGAGGACACGGCCGTGTATTTCTGTG
 CGAGAGAGTACCCCACTTCTAAAGTCGCTATTACCCCGAACTGGTTCGACCTCTGGGGC
 CAGGGAACCCCTGGTACCGTCTCGAGCGCGAGCACCAAGGGCCCATCGGTCTTCCCCCTGGC
 ACCCTCCTCAAGAGCACCTCTGGGGGCACAGCGGCCCTGGGCTGCCTGGTCAAGGACTACT
 TCCCCGAACCGGTGACGGTGTCTGTGAACCTCAGGCGCCCTGACCAGCGGCGTGCACACCTTC
 CCGGCTGTCTTACAGTCTCAGGACTCTACTCCCTCAGCAGCGTGGTGACCGTGCCTCCAG
 CAGCTTGGGCACCCAGACCTACATCTGCAACGTGAATCACAAGCCAGCAACCAAGGTGG
 ACAAGAGAGTTGAGCCCAAATCTTGTGACAAAATCACACATGCCACCGTGCACACCTTC
 GAACTCCTGGGGGACCGTCAGTCTTCTCTTCCCCCCTCAAGGACACCTCATGAT
 CTCCCGGACCCCTGAGGTACATGCGTGGTGGTGGACGTGAGCCACGAAGACCTGAGGTCA
 AGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGAGGAG
 CAGTACAACAGCACGTACCGTGTGGTCAGCGTCTCACCGTCTGACACCAGGACTGGCTGAA
 TGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCAGCCCCCATCGAGAAAACCA
 TCTCCAAGCCAAAGGGCAGCCCCGAGAACCACAGGTGTACACCCTGCCCCCATCCCGGGAT
 GAGCTGACCAAGAACCAGGTGAGCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACAT
 CGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCTCCCGTGC
 TGGACTCCGACGGCTCCTTCTTCTCTACAGCAAGCTCACCGTGGACAAAGAGCAGGTGGCAG
 CAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGCACAACCACTACACGCAGAA
 GAGCCTCTCCCTGTCTCCGGGTAAATAATTAATTAATAAATAAATAAATAAATAAATAAATAA
 GCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
 AAGTTTAAACCGCTGATCAGCCTCGACTGTGCCCTCTAGTTGCCAGCCATCTGTTGTTTGCC
 CCTCCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTCTCTAATAAATAA

FIG. 34B

GAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCA
 GGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTA
 TGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCCCCCTGTAGC
 GCGCATTAAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGC
 CCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCTTTCCCC
 GTCAGCTCTAAATCGGGGCATCCCTTTAGGGTTCGGATTAGTGCTTTACGGCACCTCGAC
 CCCAAAAAATTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTT
 TCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCAAACTGGAACAA
 CACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGGGGATTTTCGGCCTAT
 TGGTTAAAAAATGAGCTGATTTAAACAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGT
 CAGTTAGGGTGTGGAAGTCCCAGGCTCCCAGGCAGGCAGAAAGTATGCAAAGCATGCATC
 TCAATTAGTCAGCAACCAGGTGTGGAAGTCCCAGGCTCCCAGCAGGCAGAAAGTATGCAA
 AGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACCCGCCATCCCGCCCCCT
 AACTCCGCCCAGTTCCGCCCATTTCTCCGCCCATGGCTGACTAATTTTTTTTATTTATGCAG
 AGGCCGAGGCCGCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGC
 CTAGGCTTTTGCAAAAAGCTCCCGTCGACGAAATAGGTACGGTCTCGAAGCCGCGGTGCGG
 GTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCACCTCACCCTCTGGTCCATCA
 TGATGAACGGGTGAGGTGGCGGTAGTTGATCCCGGCGAACGCGCGGCGCACCGGGAAGCCC
 TCGCCCTCGAAACCGCTGGGCGCGGTGGTACGGTGAGCACGGGACGTGCGACGGCGTCCGC
 GGGTGCGGATACGCGGGGCAGCGTCAGCGGGTCTCGACGGTCACGGCGGGCATGTGACGCT
 ATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAA
 TTGTTATCCGCTCACAATTCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGG
 GTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCG
 GGAAACCTGTCTGTGCCAGAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCT
 TCGCTAGGTGGTCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAATCAA
 TATTGGCTATTGGCCATTGCATACGTTGTATCCATATCATAATATGTACATTTATATTGGCT
 CATGTCCAACATTACCGCCATGTGACATTGATTATTGACTAGTTATTAATAGTAATCAATT
 ACGGGGTCAATTAGTTCATAGCCCATATATGGAGTTCGCGGTTACATAACTTACGGTAAATGG
 CCCGCTGGCTGACCGCCCAACGACCCCCCGCCATTGACGTCAATAATGACGTATGTTCCCA
 TAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAATGCC
 CACTTGGCAGTACATCAAGTGATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGG
 TAAATGGCCCGCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCTACTTTGGCAGT
 ACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGG
 CGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAG
 TTTGTTTTGGCACCAAATCAACGGGACTTTCCAAATGTGTAACAACCTCCGCCCCATTGA
 CGCAAATGGGCGGTAGGCGGTGACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAAC
 CGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACGGGACCG
 ATCCAGCCTCCGCGGCCGGAACGGTGCAATTGGAAGCTTGGTACCGGTGAATTGCGCGCGCC
 ACCATGGCATGCCCTGGCTTCCTGTGGGCACCTTGTGATCTCCACCTGCCTCGAGTTTTCCAT
 GGCTGAAACGACACTCACGCAGTCTCCAGCCACCCTGTCTTTGTCTCCAGGGCAAAGAGCCA
 CCTCTCCTGCAGGGCCAGTCAGAGTGTTAGCACCTTCTTAGCCTGGTACCAACAGAAACCT
 GGCCAGGCTCCCAGGCTCCTCATCTATGATGCATCCAACAGGGCCACTGGCATCCCAGCCAG
 GTTCAGTGGCAGTGGGTCTGGGACAGACTTCACTCTACCATCAGCAGCCTAGAGCCTGAAG
 ATTTTGCAGTTTATTACTGTCTAGCAGCGAAACATCTGGCCCTCTTTTCGGCGGAGGGACCAA
 GGGATATCAAACGTACGGTGGCTGCACCATCTGTATTTCATCTTCCGCCATCTGATGAGCA
 GTTGAAATCTGGAATGCCCTCTGTTGTGTGCTGCTGAATAACTTCTATCCCAGAGAGGCCA

FIG. 34C

AAGTACAGTGGAAAGGTGGATAACGCCCTCCAATCGGGTAACTCCCAGGAGAGTGTACAGAG
 CAGGACAGCAAGGACAGCACCTACAGCCTCAGCAGCACCCCTGACGCTGAGCAAAGCAGACTA
 CGAGAAACACAAAGTCTACGCCTGCGAAGTCACCCATCAGGGCCTGAGCTCGCCCGTCACAA
 AGAGCTTCAACAGGGGAGAGTGTAAAGCGGCCGCAATTGCTAGCGTTAACGGATCGATCCG
 AGCTCGGTACCAAGCTTAAGTTTAAACCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCA
 GCCATCTGTTGTTTGGCCCTCCCCCGTGCCCTTCTTGACCCTGGAAGGTGCCACTCCCCTG
 TCCTTTCTAATAAAATGAGGAAATTGCCATCGCATTTGTCTGAGTAGGTGTATTCTATTCTG
 GGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGG
 GGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGCATTAATGAATCGGCCA
 ACGCGCGGGGAGAGGCGGTTTGGCTATTGGGCGCTCTTCCGCTTCTCGCTCACTGACTCGC
 TGCGCTCGGTGCTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAAATACGGTTA
 TCCACAGAATCAGGGGATAACGCAGGAAGAACAATGTGAGCAAAGGCCAGCAAAGGCCAG
 GAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATC
 ACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCG
 TTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCTGTTCCGACCCCTGCCGCTTACCGGATACCT
 GTCCGCTTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCAGCTGTAGGTATCTCA
 GTTCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTACGCCCGAC
 CGCTGCGCTTATCCGGTAACCTATCGTCTTGAGTCCAACCCGTAAGACACGACTTATCGCC
 ACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGT
 TCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTG
 CTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGC
 TGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAG
 AAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAACGAAACTCACGTTAAGGG
 ATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAG
 TTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCA
 GTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTC
 GTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCG
 AGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGC
 GCAGAAAGTGGTCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCT
 AGAGTAAGTAGTTCCGCCAGTTAATAGTTTGGCGAACGTTGTTGCCATGCTACAGGCATCGT
 GGTGTCACGCTCGTCTGTTTGGTATGGCTTCATTACGCTCCGGTTCCCAACGATCAAGGCGAG
 TTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCTCTCGATCGTTGTC
 AGAAGTAAGTTGGCCGAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTAC
 TGTATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAG
 AATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCA
 CATAGCAGAACTTTAAAGTGCTCATCATTTGGAACCGTTCTTCGGGGCGAAAACTCTCAAG
 GATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGACCCCACTGATCTTCAG
 CATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAAAACAGGAAGGCAAAATGCCGCAAAA
 AAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCTTTTCAATATTATTG
 AAGCATTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATA
 AACAAATAGGGGTTCGCGCGACATTTCCCCGAAAAGTGCCACCTGACGTC

FIG. 35A

pR1-DHFR (7093 bp)

CCTTTCGTCTTCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGC
 AGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAAACCATTTATA
 AGCTGCAATAAACAAGTTAACAACAACAATTGCATTTCATTTTATGTTTCAGGTTTCAGGGGA
 GGTGTGGGAGGTFTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGCTGATTATGATC
 CTCTAGAGTCGGTGGGCCTCGGGGGCGGGTGCGGGGTCCGGCGGGGCCGCCCGGGTGGCTTC
 GGTCCGAGCCATGGGGTCGTGCGCTCCTTTCCGGTCCGGGCGCTGCGGGTCTGGGGCGGGCGT
 CAGGCACCGGGCTTGGGGTCATGCACCAGGTGCGCGGTCTTCCGGGCACCTCGACGTCGGC
 GGTGACGGTGAAGCCGAGCCGCTCGTAGAAGGGGAGGTTGCGGGGCGCGAGGTCTCCAGGA
 AGGCGGGCACCCCGGCGCGCTCGGCCGCTCCACTCCGGGGAGCACGACGGCGCTGCCCAGA
 CCTTTCCTTGGTGGTGGGGCGAGACGCCGACGGTGGCCAGGAACACGCGGGCTCCTTGGG
 CCGGTGCGGCGCCAGGAGGCCTTCCATCTGTTGCTGCGCGGCCAGCCGGGAACCGCTCAACT
 CGGCCATGCGCGGGCCGATCTCGGCGAACACCGCCCCCGCTTCGACGCTCTCCGGCGTGGTC
 CAGACCGCCACCGCGCGCCGTCGTCCGCGACCCACACCTTGCCGATGTGAGCCCGACGCG
 CGTGAGGAAGAGTTCTTGACGCTCGGTGACCCGCTCGATGTGGCGGTCCGGGTGACGGTGT
 GCGCGCTGGCGGGGTAGTCCGGCGAACCGCGCGCGGAGGGTGCCTACGGCCCCGGGGGACGTG
 TCGCGGGTGGCGAGGCGCACCGTGGGCTTGTACTCGGTTCATGGTGGCGACCCCTACGCCCCCA
 ACTGAGAGAACTCAAAGGTTACCCCAAGTTGGGGCACTACTCCGAAAACCGCTTCTGACCTG
 GGAACCGTGAAGCCCCGGGGCAAGGGCGAATTCTGCGAGATAAATTAGGCAAAGGAATTCC
 TCGACCTGCAGCCCAAGCTAATTCGCCCTTCGTGGGGACGCCGTACAGGGACGTGCACCTCT
 CCCGCTGCACCGCCTCCAGCGTCGCCGCCGGCTCGAAGGACGGGGCCGGGATGACGATGCAG
 GCGGCGTGGGAGGTGGCGCCCAAGTTGCCCATGACCATGCCGAAGCAGTGGTAGAAGGGCAC
 CGGCAGACACACCCGGTCTGCTCCGTGTAGCCGACCGTCCGGCCACCCAGTAGCCGTTGT
 TGAGGATGTTGTGGTGGGAGAGCGTGGCGCCCTTGGGGAAGCCGGTGGTGCAGGAGGTGTAC
 TGGATGTTGACCGGAAGGGCGAATTAGCTTGGCACTGGCGCCAGAAATCCCGCGGGTGGTT
 TTTGGGGGTGGGGGTGTTTGGCAGCCACAGACGCCCGGTGTTCTGTGTCGCCCAGTACATG
 CGGTCCATGCCCAGGCCATCCAAAACCATGGGTCTGTCTGCTCAGTCCAGTCCGTGGACCTG
 ACCCCACGCAACGCCCAAATAAATACCCCAACGAACCATAAACCAATCCCATGGGGGACC
 CCGTCCCTAACCCACGGGGCAGTGGCTATGGCAGGGCCTGCCGCCCGACGTTGGCTGCGA
 GCCCTGGGCCTTACCCGAACTTGGGGGTGGGGTGGGGAAAAGGAAGAAACGCGGGCGTAT
 TGGCCCCAATGGGGTCTCGGTGGGGTATCGACAGAGTGCCAGCCCTGGGACCGAACCCCGG
 TTTATGAACAAACGACCCAAACACCCGTGCGTTTTATTCTGTCTTTTATTGCGGTATAGCG
 CGGGTTCCTTCCGGTATTGTCTCCTTCCGTGTTTCAGTTAGCCTCCCCCATCTCCCGATCCC
 CACGAGTGCTGGGGCGTCCGTTTCCACTATCGGCGAGTACTTCTACACAGCCATCGGTCCAG
 ACGGCCGCGCTTCTGCGGGCGATTGTGTACGCCCGACAGTCCCGGTCCGGATCGGACGAT
 TGGCTCGCATCGACCCCTGCGCCCAAGCTGCATCATCGAAATTGCCGTCAACCAAGCTCTGAT
 AGAGTTGGTCAAGACCAATGCGGAGCATATACGCCCGGAGCCGCGGCGATCCTGCAAGCTCC
 CGATGCCCTCCGCTCGAAGTAGCGCGTCTGCTGCTCCATACAAGCCAACACGGCCTCCAGAA
 GAAGATGTTGGCGACCTCGTATTGGGAATCCCCGAACATCGCCTCGCTCCAGTCAATGACCG
 CTGTTATGCGGCCATTGTCCGTGAGGACATTGTTGGAGCCGAAATCCGCGTGACGAGGTGC
 CGGACTTCGGGGCAGTCTCGGCCCAAAGCATCAGCTCATCGAGAGCCTGCGCGACGGACGC
 ACTGACGGTGTGTCATCACAGTTTGCCAGTGATACACATGGGGATCAGCAATCGCGCATA
 TGAAATCACGCCATGTAGTGTATTGACCGATTCTTGGCGTCCGAATGGGCGAACCGCTC
 GTCTGGCTAAGATCGGCCGAGCGATCGCATCCATGGCCTCCGCGACCGGCTGCAGAACAGC
 GGCAGTTCCGTTTCAGGCAGGTCTTGCAACGTGACACCCTGTGCACGGCGGGAGATGCAAT

FIG. 35B

AGGTCAGGCTCTCGCTGAATTCCTCCCAATGTCAAGCACTTCCGGAATCGGGAGCGCGGCCGAT
 GCAAAGTGCCGATAAAACATAACGATCTTTGTAGAAACCATCGGCGCAGCTATTTACCCGCAG
 GACATATCCACGCCCTCCTACATCGAAGCTGAAAGCACGAGATTCTTCGCCCTCCGAGAGCT
 GCATCAGGTCCGAGACGCTGTGGAACCTTTTCGATCAGAAACTTCTCGACAGACGTCGCGGTG
 AGTTCAGGCTTTTTTCATATCAAGCTGATCTTGCGGCACGCTGTTGACGCTGTTAAGCGGGTC
 GCTGCAGGGTCGCTCGGTGTTGAGGCCACACGCGTCACCTTAATATGCGAAGTGGACCTGG
 GACCGCGCCGCCCCGACTGCATCTGCGTGTTCGAATTCGCCAATGACAAGACGCTGGGGCGG
 GTTTGTGTCTATCATAGAACTAAAGACATGCAAAATATATTTCTTCGGGGACACCGCCAGCAA
 ACGCGAGCAACGGGCCACGCGGATGAAGCAGGGCGGCACCTCGCTAACGGATTCACTACTCC
 AAGAAGTGCAACCGGTGTTTAATTCGCCCTTAAAAACACCGGTGAAAGTTTAAACAAACCTG
 CAGGAATGAAAGACCCCGCTGACGGGTAGTCAATCACTCAGAGGAGACCTCCCAAGGAAC
 AGCGAGACCACAAGTCGATGCAACTGCAAGAGGGTTTATTGGATACACGGGTACCCGGGCG
 ACTCAGTCAATCGGAGGACTGGCGCCCCGAGTGAGGGGTTGTGGGCTCTTTTATTGAGCTCG
 GGGAGCAGAAGCGCGCGAAGCAGAAGCGAGAAGCGAACTGATTGGTTAGTTCAAATAAGGCAC
 AGGGTCATTTAGGTCCTTGGGGCACCTTGAAACATCTGATGGTCTCTAGAAACTGCTGA
 GGGCTGGACCGCATCTGGGGACCATCTGTTCTTGGCCCTGAGCCGGGGCAGGAACTGCTTAC
 CACAGATATCCTGTTTGGCCCATATTAGCTGTTCATCTGTTCTTGGCCCTGAGCCGGGGC
 AGGAACTGCTTACCACAGATATCCTGTTTGGCCCATATTAGCTGTTCATCTGTTCTCTGAC
 CTTGATCTGAACTTCTCTATTCTCAGTTATGTATTTTTCCATGCCTTGCAAAATGGCGTTAC
 TTAAGCTAGCTTGCCAAACCTACAGGTGGGGTCTTTTCAATTTAATTAAGGGCGAATTAACCT
 TAATTAAGATCTAAAGCCAGCAAAAGTCCCATGGTCTTATAAAATGCATAGCTTTAGGAG
 GGGAGCAGAGAACTTGAAAGCATCTTCCTGTTAGTCTTTCTTCTCGTAGACTTCAAACCTTAT
 ACTTGATGCCTTTTCTCCTGGACCTCAGAGAGGACGCTGGGTATTCTGGGAGAAGTTTA
 TATTTCCCAAATCAATTTCTGGGAAAAACGTGTCACTTTCAAATTCCTGCATGATCCTTGT
 CACAAAGAGTCTGAGGTGGCCCTGCTTGATTCTGCTTCTGTAACAGAACTGCCCTCCGA
 CTATCCAAACCATGTCTACTTTACTTGCCAAATTCGGTTGTTCAATAAGTCTTAAGGCATCA
 TCCAAACTTTTGGCAAGAAAATGAGCTCCTCGTGGTGGTTCTTTGAGTTCTCTACTGAGAAC
 TATATTAATTCGTCTCTTTAAAGGTCGATTCTTCTCAGGAATGGAGAACCAGTTTTCTAC
 CCATAATCACCAGATTCTGTTTACCTTCCACTGAAGAGGTTGTGGTCAATCTTTGGAAGTAC
 TTGAACTCGTTCTTGAGCGGAGGCCAGGGTAGGTCTCGTTCTTGCCAATCCCATATTTTG
 GGACACGCGGACGATGCAGTTCAATGGTTCGAACCATGATGGCAGCGGGGATAAAGCTTTTG
 CAAAAGCCTAGGCCTCCAAAAAGCCTCCTCACTACTTCTGGAATAGCTCAGAGGCCGAGGC
 GGCTTCGGCCTCTGCATAAAATAAAAAAATTAGTCAGCCATGGGGCGGAGAATGGGCAGAAC
 TGGGCGGAGTTAGGGGCGGGATGGGCGGAGTTAGGGGCGGACTATGGTTGCTGACTAATTG
 AGATGCATGCTTTGCATACTTCTGCCTGCTGGGGAGCCTGGGGACTTTCCACACCTGTTGC
 TGACTAATTGAGATGCATGCTTTGCATACTTCTGCCTGCTGGGGAGCCTGGGGACTTTCCAC
 ACCCTAAGTACACACATTCCACAGGGCGCGCCCCCTTGGCAGAACATATCCATCGCGTCCG
 CCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGGCCGACGCGCTGGGCTACGTCTTGCTGG
 CGTTGCGGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATC
 GGGATGCCCGGCTTGACAGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCT
 TCAAGGATCGCTCGCGGCTCTTACCAGCGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGC
 GTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAA
 GTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCC
 CTCGTGCGCTCTCCTGTTCCGACCTGCGGCTTACCGGATACCTGTCGCGCTTTCTCCCTTC
 GGGAAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTC
 GCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTACGCCGACCGCTGCGCCTTATCCGGT

FIG. 35C

AACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGG
 TAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTA
 ACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTC
 GGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAAACAAACCAACCGCTGGTACCGCTCCTTTTTT
 TGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTT
 CTACGGGGTCTGACGCTCAGTGGAAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTA
 TCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTAAATCAATCTAAAG
 TATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAG
 CGATCTGTCTATTTTCGTTTATCCATAGTTGCTGACTCCCCGTCGTGTAGATAACTACGATA
 CGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGC
 TCCAGATTTATCAGCAATAAACCCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAA
 CTTTATCCGCCTCCATCCAGTCTATTAAFTGTTGCCGGGAAGCTAGAGTAAGTAGTTGCGCA
 GTTAATAGTTTGCGCAACGTTGTTGCCATTGCTGCAGGCATCGTGGTGTACGCTCGTCGTT
 TGGTATGGCTTCATTAGCTCCGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGT
 TGTGCAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCA
 GTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTGTCATGCCATCCGTAAG
 ATGCTTTTCTGTGACTGCTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGAC
 CGAGTTGCTCTTGCCCGGCGTCAACACGSGATAATACCGCGCCACATAGCAGAACTTTAAAA
 GTGCTCATCATTGGAAAAAGTTCCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAG
 ATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTTACCA
 GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACA
 CGGAAATGTTGAATACTCATACTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTA
 TTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGC
 GCACATTTCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTTATTATCATGACATTAACC
 TATAAAAATAGGCGTATCACGAGGC

FIG. 36A

pDI-DTX1-G418 (9455 bp)

GACGGATCGGGAGATCCACGCGTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCA
 GGCTCCCCAGGCAGGCAGGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTG
 GAAAGTCCCCAGGCTCCCCAGCAGGCAGGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCA
 ACCATAGTCCCGCCCCCTAACTCCGCCCCATCCCGCCCCCTAACTCCGCCCCAGTTCCGCCCCATT
 TCCGCCCCCATGGCTGACTAATTTTATTTTATTTATGTCAGAGGCCGAGGCCGCCCTCTGCCTCTG
 AGCTATTCCAGAAGTAGTGAGGAGGCTTTTGGAGGCCCTAGGCTTTTGCAAAAAGCTCCCG
 GGAGCTTGGATATCCATTTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTTCGCATGAT
 TGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCCGGCTATG
 ACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGG
 CGCCCGGTTCTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGC
 AGCGCGGCTATCGTGGCTGGCCACGACGGCGTTCTTGCGCAGCTGTGCTCGACGTTGTCA
 CTGAAGCGGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCACT
 CACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCT
 TGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTAATC
 GGATGGAAGCCGGTCTTGTGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCA
 GCCGAACCTGTTCCGCGAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCA
 TGGCGATGCCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTTCATCGACT
 GTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCT
 GAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCGTGCTTTACGGTATCGCCGCTCCCGA
 TTCCGACGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTT
 CGGTGCTACGAGATTTTCGATTCCACCGCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTT
 TTCCGGGACGCCGCTGGATGATCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCCA
 CCCCACCTTGTTTATTGTCAGCTTATAATGGTTACAAATAAAGCAATAGCATCAAAATTTCA
 CAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCT
 TATCATGTCTGTCTAGAGAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCT
 TCGCTAGGTGGTCAATATTGGCCATTAGCCATATTATTTCATTGGTTATATAGCATAAATCAA
 TATTGGCTATTGGCCATTGCATACGTTGTATCCATATCATAATATGTACATTTATATTGGCT
 CATGTCCAACATTACCGCCATGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATT
 ACGGGGTCAATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAAATTACGGTAAATGG
 CCGGCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCA
 TAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCC
 CACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGG
 TAAATGGCCCGCCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCTTACTTGGCAGT
 ACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGG
 CGTGGATAGCGGTTTGACTCACGGGGATTTCAGTCTCCACCCCATTGACGTCAATGGGAG
 TTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTGTAACAACCTCCGCCCCATTGA
 CGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAAC
 CGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTGGACCTCCATAGAAGACACCGGGACCG
 ATCCAGCCTCCGCGGCCGGGAACGGTGCAATTGGAAGCTTGGTACCGAGCTCGGATCCGCCAC
 CATGGCATGCCCTGGCTTCTGTGGGCACCTGTGATCTCCACGTGTCTTGAATTTTCCATGG
 CTGAAGTGCAGCTGGTGGAGTCTGGGGGAGGCTTGGTCAAGCCTGGAGGGTCCCTGAGACTC
 TCCTGTGAAGCCTCTGGATTTCATCTTCAGTGACTACTACATGAGCTGGATCCGCCAGGCTCC
 AGGGAAGGGGCTGGAATGGATTTTATACATTAGTCCCTAGTGGTAGTACCCTATACTACGCAG
 ACTCTATGAGGGGCCGATTCAACATCTCCAGGGACAACGCCAAGAAGTCACTGTATCTGCAA

FIG. 36B

ATGAACAGCCTGAGAGTCGAGGACACGGCCGTGTATTTCTGTGCGAGAGAGTACCCACAAAC
 TTCTAAAGTCGCTATTACCCCGAACTGGTTTCGACCTCTGGGGCCAGGGAACCTGGTCACCG
 TCTCGAGCGCGAGCACCAAGGGCCCATCGGTCTTCCCCCTGGCACCCCTCCTCCAAGAGCACC
 TCTGGGGGACAGCGGCCCTGGGCTGCCTGGTCAAGGACTACTTCCCCGAACCGGTGACGGT
 GTCGTGGAACCTCAGGCGCCCTGACCAGCGGCGTGACACCTTCCCGGCTGTCTACAGTCCT
 CAGGACTCTACTCCCTCAGCAGCGTGGTGACCGTGCCCTCCAGCAGCTTGGGCACCCAGACC
 TACATCTGCAACGTGAATCACAAAGCCAGCAACACCAAGGTGGACAAGAGAGTTGAGCCCAA
 ATCTTGTGACAAAACCTCACACATGCCACCGTGCCACAGCACCTGAACTCCTGGGGGGACCGT
 CAGTCTTCTCTTCCCCCCTAAAGGACACCCCTCATGATCTCCCGGACCCCTGAGGTC
 ACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCCTGAGGTCAAGTTCAACTGGTACGTGGA
 CGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAGCACGTACC
 GTGTGGTCAGCGTCTTACCGTCTTGACACAGGACTGGCTGAATGGCAAGGAGTACAAGTGC
 AAGGTCTCCAACAAAGCCCTCCAGCCCCATCGAGAAAACCATCTCCAAGCCAAAGGGCA
 GCCCCGAGAACCACAGGTGTACACCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGG
 TCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCAGCGACATCGCCGTGGAGTGGGAGAGC
 AATGGGCAGCCGGAGAACTACAAGACCAAGCCCTCCCGTGTGGACTCCGACGGCTCCTT
 CTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCAT
 GCTCCGTGATGCATGAGGCTCTGCACAAACCTACACGCAGAAGAGCCTCTCCCTGTCTCCG
 GGTAAATAATTAATTAATAAAAAAACACCGGCGAAAAAGCGATCGCAAAAAACCAAGTGTGGTG
 GAATTCCTGCAGATAACGCTAGCGAATTCACCGGTACCAAGCTTAAGTTTAAACCGCTGATCA
 GCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCCTTCCTT
 GACCCTGGAAGGTGCCACTCCCACTGTCTTTCCTAATAAAATGAGGAAATTGCATCGCATT
 GTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGAT
 TGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAG
 AACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGG
 GTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTC
 GCTTTCCTTCCCTTCCTTCTCGCCACGTTCCGCGGCTTTCCTCCGTCAAGCTCTAAATCGGGG
 CATCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGG
 GTGATGGTTACGTAAGTGGGCCATCGCCCTGATAGACGGTTTTTTCGCCCTTTGACGTTGGAG
 TCCACGTTCTTTAATAGTGGACTCTTGTTCAAACTGGAACAACACTCAACCTATCTCGGT
 CTATTCTTTTGATTTATAAGGGATTTTGGGGATTTCCGGCTATTGGTTAAAAAATGAGCTGA
 TTTAACAAAAATTTAACCGGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGT
 CCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAG
 GTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGT
 CAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCCGCCCTAACTCCGCCCAGTTCCGCC
 CATTCTCCGCCCATGGCTGACTAATTTTTTTTATTTATGTCAGAGGCCGAGGCCGCTCTGC
 CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGC
 TCCCGTCGACGAAATAGGTCACGGTCTCGAAGCCGCGGTGCGGGTGCCAGGGCGTGCCCTTG
 GGCTCCCCGGGCGCGTACTCCACCTCACCCATCTGGTCCATCATGATGAACGGGTGAGGTG
 GCGGTAGTTGATCCCGGCGAACGCGCGGCGCACCGGGAAGCCCTCGCCCTCGAAACCGCTGG
 GCGCGGTGGTCACGGTGAGCACGGGACGTGCGACGGCGTCGGCGGGTGGGATACGCGGGG
 AGCGTCAGCGGGTCTCGACGGTCACGGCGGGCATGTGACGTATACCGTCGACCTCTAGCT
 AGAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTTGTTATCCGCTCACAATT
 CCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCCTGGGGTGCCTAATGAGTGAGCTA
 ACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCAGTCGGGAAACCTGTGCGTCCAGA
 ATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCTAGGTGGTCAATATT

FIG. 36C

GGCCATTAGCCATATTATTTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATTG
 CATACGTTGTATCCATATCATAATATGTACATTTATATTGGCTCATGTCCAACATTACCGCC
 ATGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATA
 GCCCATATATGGAGTTCGCGGTTACATAAATTACGGTAAATGGCCCGCTGGCTGACCGCCC
 AACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGAC
 TTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAG
 TGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCAT
 TATGCCCAGTACATGACCTTATGGGACTTTCTTACTTGGCAGTACATCTACGTATTAGTCAT
 CGCTATTACCATGGTGATGCGGTTTGGCAGTACATCAATGGGCGTGATAGCGGTTTGACT
 CACGGGGATTTCCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTTTTGGCACCAAAAT
 CAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCCATTGACGCAAAATGGGCGGTAGGCG
 TGTACGGTGGGAGGTCTATAAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGAC
 GCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGCGGCCGG
 GAACGGTGCATTGGAAGCTTGGTACCGGTGAATTCGGCGCGCCACCATGGCATGCCCTGGCT
 TCCTGTGGGCACCTTGTGATCTCCACCTGCCTCGAGTTTTCCATGGCTGAAACGACACTCACG
 CAGTCTCCAGCCACCCTGTCTTTGTCTCCAGGGGAAAGAGCCACCCTCTCCTGCAGGGCCAG
 TCAGAGTGTTAGCACCTTCTTAGCCTGGTACCAACAGAAAACCTGGCCAGGCTCCAGGCTCC
 TCATCTATGATGCATCCAACAGGGCCACTGGCATCCAGCCAGGTTCAAGTGGCAGTGGGTCT
 GGGACAGACTTCACTCTCACCATCAGCAGCCTAGAGCCTGAAGATTTTGAGTTTATTACTG
 TCAGCAGCGAAACATCTGGCCCTCTTTCGGCGGAGGGACCAAAGTGGATATCAAACGTACGG
 TGGCTGCACCATCTGTATTCTCTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACCTGCC
 TCTGTGTGTGCTGCTGAATAACTTCTATCCCAGAGAGGCCAAAGTACAGTGGAAAGGTGGA
 TAAAGCCCTCCAATCGGGTAACTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCA
 CCTACAGCCTCAGCAGCACCCCTGACGCTGAGCAAAGCAGACTACGAGAAACACAAAGTCTAC
 GCCTGCGAAGTCACCCATCAGGGCCTGAGCTCGCCCGTCACAAAGAGCTTCAACAGGGGAGA
 GTGTTAAGCGGCCGCAATTCGCTAGCGTTAACGGATCGATCCGAGCTCGGTACCAAGCTTAA
 GTTTAAACCGCTGATCAGCCTCGACTGTGCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCC
 TCCCCCGTGCTTCTTGACCTGGAAGGTGCCACTCCCACTGTCCTTTCTTAATAAAATGA
 GGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGG
 ACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATG
 GCTTCTGAGGCGGAAAGAACCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGT
 TTGCGTATTGGGCGCTCTTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTCTGCTCGGCT
 GCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATA
 ACGCAGGAAAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAAAGGCCCGG
 TTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAG
 TCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCC
 TCGTGGCTCTCCTGTTCCGACCCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCG
 GGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCTGTCG
 CTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTAGCCCGACCGCTGCGCTTATCCGGTA
 ACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGT
 AACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAA
 CTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCG
 GAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTT
 GTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTC
 TACGGGGTCTGACGCTCAGTGGAAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTAT
 CAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTTAAATCAATCTAAAGT

FIG. 36D

ATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGC
GATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATAC
GGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCT
CCAGATTTATCAGCAATAAACAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAAC
TTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAGTTCCGCCAG
TTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACAGCTCGTCTGTTT
GGTATGGCTTCATTTCAGCTCCGGTTCCTCAACGATCAAGGCGAGTTACATGATCCCCATGTT
GTGCAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAG
TGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTTCATGCCATCCGTAAGA
TGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACC
GAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAG
TGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGA
TCCAGTTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTTACCAG
CGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACAC
GGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTATCAGGGTTAT
TGTCTCATGAGCGGATACATATTGAATGTATTTAGAAAAATAACAAATAGGGGTTCCGCG
CACATTTCCCGAAAAGTGCCACCTGACGTC

FIG. 37A

pD3-DTX1 (9753 bp)

GACGGATCGGGAGATCTCCCGATCCCCTATGGTCTGACTCTCAGTACAATCTGCTCTGATGCC
 GCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTGCTGAGTAGTGCGCGAG
 CAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGG
 TTAGGCGTTTTGCGCTGCTTCGCTAGGTGGTCAATATTGGCCATTAGCCATATTATTCATTG
 GTTATATAGCATAAATCAATATTGGCTATTGGCCATTGCATACGTTGTATCCATATCATAAT
 ATGTACATTTATATTGGCTCATGTCCAACATTACCGCCATGTTGACATTGATTATTGACTAG
 TTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCGCGTTA
 CATAACTTACGGTAAATGGCCCCGCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCA
 ATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGA
 GTATTTACGGTAAACTGCCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCC
 CTATTGACGTCAATGACGGTAAATGGCCCCGCTGGCATTATGCCCAGTACATGACCTTATGG
 GACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTT
 TTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACC
 CCATTGACGTCAATGGGAGTFTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTGCT
 AACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAG
 CAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCC
 ATAGAAGACACCGGGACCGATCCAGCCTCCGCGGCCGGGAACGGTGCATTGGAAGCTTGGTA
 CCGAGCTCGGATCCGCCACCATGGCATGCCCTGGCTTCCTGTGGGCACTTGTGATCTCCACG
 TGTCTTGAATTTTCCATGGCTGAAGTGCAGCTGGTGGAGTCTGGGGAGGCTTGGTCAAGCC
 TGGAGGGTCCCTGAGACTCTCCTGTGAAGCCTCTGGATTCACTTCAGTGACTACTACATGA
 GCTGGATCCGCCAGGCTCCAGGGAAGGGGCTGGAATGGATTTTCATACATTAGTCCTAGTGGT
 AGTACCCTATACTACGCAGACTCTATGAGGGGCCGATTACCATCTCCAGGGACAACGCCAA
 GAACCTCACTGTATCTGCAAATGAACAGCCTGAGAGTTCGAGGACACGGCCGTGTATTTCTGTG
 CGAGAGAGTACCCCACAACTTCTAAAGTCGCTATTACCCCGAACTGGTTTCGACCTCTGGGGC
 CAGGGAACCCCTGGTCAACGCTCTGAGCGCGAGCACCAAGGGCCCATCGGTCTTCCCCCTGGC
 ACCCTCCTCCAAGAGCACCTCTGGGGGCACAGCGGCCCTGGGCTGCCTGGTCAAGGACTACT
 TCCCCGAACCGGTGACGCTGTGCTGGAACCTCAGGCGCCCTGACCAGCGCGTGCACACCTTC
 CCGGCTGTCTACAGTCTCAGGACTCTACTCCCTCAGCAGCGTGGTGACCGTGCCCTCCAG
 CAGCTTGGGCACCCAGACCTACATCTGCAACGTGAATCACAAGCCCAGCAACCAAGGTGG
 ACAAGAGAGTTGAGCCCAAATCTTGTGACAAAACCTCACACATGCCACCGTGCCCAGCACCT
 GAACTCCTGGGGGGGACCGTCAGTCTTCCTCTTCCCCCAAAACCCAAGGACACCCCTCATGAT
 CTCCCGGACCCCTGAGGTACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCA
 AGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAG
 CAGTACAACAGCACGTACCGTGTGGTCAGCGTCTTCAACGCTCCTGCACCAGGACTGGCTGAA
 TGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCAGCCCCCATCGAGAAAACCA
 TCTCCAAGCCAAAGGGCAGCCCCGAGAACCAAGGTGTACACCCTGCCCCCATCCCGGGAT
 GAGCTGACCAAGAACCAGGTGAGCTGACCTGCCTGGTCAAAGGCTTCTATCCAGCGACAT
 CGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAGACCAGCCTCCCGTGC
 TGGACTCCGACGGCTCCTTCTTCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAG
 CAGGGGAACGTCTTCTCATGCTCGGTGATGCATGAGGCTCTGCACAACCACTACACGCAGAA
 GAGCCTCTCCCTGTCTCCGGGTAAATAATTAATTAATAAAAAAACACCGGCGAAAAAAGCGATC
 GCAAAAAACAGTGTGGTGGAAATCTGCAGATAACGCTAGCGAATTCACCGGTACCAAGCTT
 AAGTTTAAACCGCTGATCAGCCTGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTGCC
 CCTCCCCCGTGCCTTCTTGACCTGGAAGGTGCCACTCCCACTGTCTTCTTAATAAAAT

FIG. 37B

GAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCA
 GGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTA
 TGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGC
 GGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGC
 CCTAGCGCCCGCTCCTTTTCGCTTCTTCCCTTCTTTCTCGCCACGTTTCGCGGGCTTTCCCC
 GTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGAC
 CCCAAAAAATTGATTAGGGTGATGGTTACGCTAGTGGGCCATCGCCCTGATAGACGGTTTTT
 TCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAA
 CACTCAACCCCTATCTCGGTCTATTCTTTTGATTAGAGAGAGCTAGCATTAAATAAGGACA
 GGGAAAGGGAGCAGTGGTTACGCGCTGTAATCCCAGCAATTTGGGAGGCCAAGGTGGGTAGAT
 CACCTGAGATTAGGAGTTGGAGACCAGCCTGGCCAATATGGTGAAACCCCGTCTCTACCAAA
 AAAACAAAAATTAGCTGAGCCTGGTCATGCATGCCTGGAATCCCAACAACCTCGGGAGGGCTGA
 GGCAGGAGAATCGCTTGAACCCAGGAGGCGGAGATTGCAGTGAGCCAAGATTGTGCCACTGC
 ACTCCAGCTTGGTTCCCAATAGACCCCGCAGGCCCTACAGGTTGTCTTCCCAACTTGCCCTT
 TGCTCCATAACACCCCCCTCCACCCCATATAATTATAGAAGGACACCTAGTCAGACAAATG
 ATGCAACTTAATTTTATTAGGACAAGGCTGGTGGGCACTGGAGTGGCAACTTCCAGGGCCAG
 GAGAGGCACTGGGGAGGGGTACAGGGATGCCACCCGTAGATCTCTCGAGCTATTACACCCA
 CTCGTGCAGGCTGCCCAGGGGCTTGCCAGGCTGGTCAGCTGGGCGATGGCGGTCTCGTGCT
 GCTCCACGAAGCCCGCTCTCCACGTAGGTCTTCTCCAGGCGGTGCTGGATGAAGTGGTAC
 TCGGGGAAGTCCTTCACCACGCCCTTGCTCTTCATCAGGGTGCGCATGTGGCAGCTGTAGAA
 CTTGCCGCTGTTTCAGGCGGTACACCAGGATCACCTGGCCCACCAGCACGCCGTGCTTCATGT
 ACACCACCTCGAAGCTGGGCTGCAGGCCGGTGATGGTCTTCTTCATCAGGGGGCGTCTGTTG
 GGGAAAGTTGCGGCCCTTGTAATCCACGCGGTACACGAACATCTCCTCGATCAGGTTGATGTC
 GCTGCGGATCTCCACCAAGCCCGCTCTCTGATAGCGCAGGGTGCGCTCGTACACGAAGCCGG
 CGGGGAAGCTCTGGATGAAGAAGTCGCTGATGTCTCGGGGTACTTGGTGAAGGTGCGGTTG
 CCGTACTGGAAGGCGGGGCTCAGGATGTGGAAGGCGAAGGGCAGGGGGGCGCCCTTGGTGAC
 GCGGATCTGCACCAGCTGGTTGCCGAACAGGATGTTGCCCTTGCCGAGCCCTCCATGGTGA
 ACACGTGGTTGTTTACCACGCCCTCCAGGTTACCTTGAAGCTCATGATCTCCTGCAGGCCG
 GTGTTCTTCAGGATCTGCTTGCTCACCATGGTGAATTCATCGATGTTTGAATCCCAATTCT
 TTGCCAAAGTGATGGGCCAGCACACAGACCAGCACGTTGCCCAGGAGTTGGAGGTGCACACC
 AATGTGGTGAATGGTCAAATGGCGTTTGCTGTATCGAGCTAGGCACTTAAATACAATATCTC
 TGCAATGCGGAATTCAGTGGTTCTGCTCAATCCATGTCAGACCCGTCTGTTGCCTTCTTAATA
 AGGCAOGATCGTACCACCTTACTTCCACCAATCGGCATGCACGGTGCTTTTCTCTCCTTGT
 AAGGCATGTTGCTAACTCATCGTTACCATGTTGCAAGACTACAAGAGTATTGCATAAGACTA
 CATTTCCCCCTCCCTATGCAAAAGCGAAACTACTATATCCTGAGGGGATGACAATTGTGAA
 TGCATAAGGGATTTTGGGGATTTGGGCTATTGGTTAAAAAATGAGCTGATTTAACAAAAAT
 TTAACGCGAATTAATTTCTGTGGAATGTGTGTGAGTTAGGGTGTGGAAAGTCCCAGGCTCCC
 CAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTC
 CCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAG
 TCCCGCCCCCTAACTCCGCCCCATCCGCCCCCTAACTCCGCCCCAGTTCCGCCCCATTCTCCGCCC
 CATGGCTGACTAATTTTTTTTATTTATGAGAGGCCGAGGCCGCTCTGCCTCTGAGCTATT
 CCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCCTAGGCTTTTGCAAAAAGCTCCCGTCGACGA
 AATAGGTCACGGTCTCGAAGCCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGC
 GCGTACTCCACCTCACCCATCTGGTCCATCATGATGAACGGGTGAGGTGGCGGTAGTTGAT
 CCCGGCGAAGCGCGCGGCACCGGGAAGCCCTCGCCCTCGAAACCGCTGGGCGCGGTGGTCA
 CCGTGAGCACGGGACGTGCGACGGCGTGGCGGGTGCGGATACGCGGGGCAGCGTCAGCGGG

FIG. 37C

TTCTCGACGGTCACGGCGGGCATGTCGACGTATACCGTCGACCTCTAGCTAGAGCTTGGCGT
 AATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCGCTCACAATCCACACAACATA
 CGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAAT
 TGGCTTGGCGCTCACTGCCCGCTTCCAGTCGGGAAACCTGTGCTGCCAGAATTGCATGAAGA
 ATCTGCTTAGGGTTAGGCGTTTTCGCTGCTTCGCTAGGTGGTCAATATTGGCCATTAGCCA
 TATTATTCAATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATTGCATACGTTGTAT
 CCATATCATAATATGTACATTTATATTGGCTCATGTCCAACATTACCGCCATGTTGACATTG
 ATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGG
 AGTTCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGC
 CCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTCCATTGACG
 TCAATGGGTGGAGTATTTACGGTAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGC
 CAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCAGTAC
 ATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCAT
 GGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC
 CAAGTCTCCACCCCATGACGTCAATGGGAGTTTGTTCCTGGCACCAAATCAACGGGACTTT
 CCAAATGTGTAACAACCTCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGA
 GGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCT
 GTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCGCGGCCGGGAACGGTGCATT
 GGAAGCTTGGTACCGGTGAATTGGGCGCGCCACCATGGCATGCCCTGGCTTCCTGTGGGCAC
 TTGTGATCTCCACCTGCCTCGAGTTTTCATGGCTGAAACGACACTCACGCAGTCTCCAGCC
 ACCCTGTCTTTGTCTCCAGGGGAAAGAGCCACCCTCTCTGCGAGGGCCAGTCAGAGTGTAG
 CACCTTCTTAGCCTGGTACCAACAGAAACCTGGCCAGGCTCCAGGCTCCTCATCTATGATG
 CATCCAACAGGGCCACTGGCATCCAGCCAGGTTCACTGGCAGTGGGTCTGGGACAGACTTC
 ACTCTCACCATCAGCAGCCTAGAGCCTGAAGATTTTGAGTTTATTAAGTGTGAGCAGCGAAA
 CATCTGGCCCTCTTTCGGCGGAGGGACCAAAGTGGATATCAAACGTACGGTGGCTGCACCAT
 CTGTATTCTCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAAGTGCCTCTGTTGTGTGC
 CTGCTGAATAACTTCTATCCAGAGAGGCCAAAGTACAGTGGAAAGGTGGATAACGCCCTCCA
 ATCGGGTAACTCCAGGAGAGTGTACAGAGCAGGACAGCAAGGACAGCACCTACAGCCTCA
 GCAGCACCTGACGCTGAGCAAAGCAGACTACGAGAAACAACAAGTCTACGCCCTGCGAAGTC
 ACCCATCAGGGCCTGAGCTCGCCCGTCACAAAGAGCTTCAACAGGGGAGAGTGTAAAGCGGC
 CGCAATTGCTAGCGTTAACGGATCGATCGAGCTCGGTACCAAGCTTAAGTTTAAACCGCT
 GATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTGCCCTCCCCCGTGCCT
 TCCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTCTAATAAAATGAGGAAATTGCATC
 GCATTGTCTGAGTAGGTGTCTATTCTGTTGGGGGTGGGGTGGGGCAGGACAGCAAGGGGG
 AGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCG
 GAAAGAACCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGG
 CGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTGCTTCGGCTGCGGCGAGCGGT
 ATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGA
 ACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGCTGGCGTTT
 TCCATAGGCTCCGCCCCCTGAAGGAGCATCAAAAAATCGACGCTCAAGTCAGAGGTGGCG
 AAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCTCTC
 CTGTTCCGACCTGCGCTTACCGGATACCTGTCCGCTTCTCCCTTCGGGAAGCGTGGCG
 CTTTCTCAATGCTCAGCTGTAGGTATCTCAGTTCCGTGTAGGTGCTTCGCTCCAAGCTGGG
 CTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCTTATCCGTAACATATCGTCTTG
 AGTCCAACCCGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGC
 AGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACAC

FIG. 37D

TAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTG
GTAGCTCTTGATCCGGCAAACAAACCACGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAG
CAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGA
CGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCT
TCACCTAGATCCTTTTAAATTAAAAATGAAGTTTAAATCAATCTAAAGTATATATGAGTAA
ACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATT
TCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTAC
CATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCA
GCAATAAACCCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTC
CATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTGCCAGTTAATAGTTTGC
GCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCTTGGTATGGCTTCA
TTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGC
GGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCA
TGGTTATGGCAGCACTGCATAATTCTCTTACTGTTCATGCCATCCGTAAGATGCTTTTCTGTG
ACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTG
CCCGGCGTCAATACGGGATAATAACCGGCCACATAGCAGAACTTTAAAAGTGCTCATCATTG
GAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATG
TAACCCACTCGTGCACCCACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGGTG
AGCAAAAACAGGAAGCAAAATGCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAA
TACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGC
GGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCG
AAAAGTGCCACCTGACGTC

RESUMO

VETOR ALVO PARA INTEGRAÇÃO EM SÍTIOS-ESPECÍFICOS,
MÉTODO DE SÍTIO-ESPECÍFICO QUE INTEGRA UM POLINUCLEOTÍDEO
CODIFICANDO UMA PROTEÍNA DE INTERESSE EM UM GENOMA DE UMA
5 CÉLULA EUCARIÓTICA, CÉLULA EUCARIÓTICA ISOLADA, KIT PARA USO
NO SÍTIO-ESPECÍFICO QUE INTEGRA UM POLINUCLEOTÍDEO DENTRO DE
UM GENOMA DE UMA CÉLULA IN VITRO E KIT PARA USO EM PRODUZIR
UMA PROTEÍNA EM UMA CÉLULA

Esta invenção fornece um sistema de integração
10 local-específico e métodos para a geração de linhagens de
células eucarióticas pra produção de proteína. O sistema
fornecido inclui um primeiro vetor alvo integrando o local-
específico e um segundo vetor doador integrando o local-
específico compreendendo o gene de interesse. Também são
15 fornecidos linhagem de células mamíferas produzidas pelos
métodos e sistemas revelados, bem como kits que incluem o
sistema revelado.