

LUIS ALBERTO QUISPE POMATAY

Mapeamento de regiões cromossômicas com alto potencial de expressão gênica como alvo para seu uso na produção de proteínas recombinantes

Dissertação apresentada ao Programa de Pós-Graduação Interunidades em Biotecnologia da Universidade de São Paulo, Instituto Butantan e Instituto de Pesquisas Tecnológicas para obtenção do Título de Mestre em Biotecnologia.

São Paulo
2019

LUIS ALBERTO QUISPE POMATAY

**Mapeamento de regiões cromossômicas com alto potencial de expressão gênica como
alvo para seu uso na produção de proteínas recombinantes**

Dissertação apresentada ao Programa de Pós-Graduação Interunidades em Biotecnologia da Universidade de São Paulo, Instituto Butantan e Instituto de Pesquisas Tecnológicas para obtenção do Título de Mestre em Biotecnologia.

Área de concentração: Biotecnologia

Orientador: Dr. Marcos Angelo Almeida Demasi

Versão corrigida. A versão original eletrônica, encontra-se disponível tanto na Biblioteca no ICB quanto na Biblioteca Digital de Teses e Dissertações da USP (BDTD).

São Paulo
2019

CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
Serviço de Biblioteca e informação Biomédica
do Instituto de Ciências Biomédicas da Universidade de São Paulo

Ficha Catalográfica elaborada pelo(a) autor(a)

Quispe Pomatay, Luis Alberto
Mapeamento de regiões cromossômicas com alto
potencial de expressão gênica como alvo para seu
uso na produção de proteínas recombinantes / Luis
Alberto Quispe Pomatay; orientador Marcos Angelo
Almeida Demasi. -- São Paulo, 2019.
99 p.

Dissertação (Mestrado)) -- Universidade de São
Paulo, Instituto de Ciências Biomédicas.

1. Biofármacos. 2. vetores lentivirais . 3. CHO.
4. inserção sítio-dirigida. 5. Proteínas
recombinantes. I. Almeida Demasi, Marcos Angelo ,
orientador. II. Título.

COMISSÃO NACIONAL DE
ÉTICA EM PESQUISA



PARECER CONSUBSTANCIADO DA CONEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Reconhecimento de regiões com alto potencial de expressão gênica como alvo para seu uso na produção de proteínas recombinantes

Pesquisador: Marcos Angelo Almeida Demasi

Área Temática: Projetos de pesquisa que envolvam organismos geneticamente modificados (OGM), células-tronco embrionárias e organismos que representem alto risco coletivo, incluindo organismos relacionados a eles, nos âmbitos de: experimentação, construção, cultivo, manipulação, transporte, transferência, importação, exportação, armazenamento, liberação no meio ambiente e descarte;

Versão: 1

CAAE: 16605619.9.0000.0065

Instituição Proponente: Faculdade de Medicina da Universidade de São Paulo

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 3.559.866

Apresentação do Projeto:

As informações contidas nos campos "Apresentação do Projeto", "Objetivo da Pesquisa" e "Avaliação dos Riscos e Benefícios" foram obtidas dos documentos contendo as Informações Básicas da Pesquisa (PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1361898.pdf de 28/06/2019) e do Projeto Detalhado.

INTRODUÇÃO

A alta potência e especificidade das proteínas recombinantes terapêuticas como ferramentas para o tratamento de doenças têm permitido um crescimento constante da indústria biofarmacêutica. Em parte, este crescimento se sustenta no aumento da capacidade produtiva do setor, e que está relacionado com melhorias no processo de produção, como, por exemplo, a otimização de meios de cultura e de estratégias de operação dos biorreatores, e o estabelecimento de novas plataformas de produção. Um dos fatores limitantes para o estabelecimento de uma plataforma de produção baseada em células de mamífero é o tempo prolongado e os custos associados ao processo de obtenção de clones super-produtores. Esta limitação está intimamente relacionada às metodologias comumente empregadas para a geração de transfectantes celulares estáveis, que

Endereço: SRTVN 701, Via W 5 Norte, lote D - Edifício PO 700, 3º andar

Bairro: Asa Norte

CEP: 70.719-040

UF: DF

Município: BRASÍLIA

Telefone: (61) 3315-5877

E-mail: conep@saude.gov.br

COMISSÃO NACIONAL DE ÉTICA EM PESQUISA



Continuação do Parecer: 3.558.865

envolvem eventos aleatórios de integração genômica dos vetores de expressão e o emprego de protocolos de amplificação gênica, os quais estão associados a eventos de inserção em sítios cromossômicos com atividade transcricional instável e eventos drásticos de rearranjos genéticos, respectivamente. Estes processos resultam numa grande heterogeneidade dos transfectantes estáveis obtidos quanto aos níveis e à estabilidade de expressão do produto de interesse, reduzindo drasticamente a probabilidade de se isolar um clone super-produtor. Uma alternativa que vem sendo explorada nos últimos anos se baseia na inserção sítio-dirigida do cassete de expressão em regiões do genoma da célula hospedeira previamente associadas à uma expressão gênica estável e em altos níveis. A indústria biofarmacêutica tem mostrado um crescente interesse em metodologias de geração de clones superprodutores empregando a integração genômica sítio-específica em células CHO visando à produção de proteínas recombinantes terapêuticas. Desta forma, existe uma demanda no sentido de se acumular conhecimento sobre regiões genômicas em linhagens produtoras como as células CHO que suportam uma expressão gênica estável e em altos níveis. A nossa proposta envolve a utilização de um vetor lentiviral bicistrônico de expressão como ferramenta de varredura de sítios genômicos nas células CHO-DG44 associados a uma expressão elevada e estável da proteína de interesse. Ao final do Projeto, espera-se contribuir com o mapeamento de alguns destes sítios genômicos nas células CHO-DG44, uma das plataformas mais utilizadas para a produção de biofármacos, e os quais poderiam ser utilizados para a inserção sítio-dirigida de vetores de expressão de interesse pela metodologia de CRISPR-Cas9.

HIPÓTESE

O racional da proposta de varredura utilizando um vetor lentiviral se baseia no fato de várias evidências indicarem que vetores lentivirais baseados no vírus HIV-1 (Human Immunodeficiency Virus-1) se integram em um subconjunto de genes com alta atividade transcricional no genoma da célula hospedeira[33, 34]. Ao final do Projeto, espera-se contribuir com o mapeamento de alguns destes sítios genômicos nas células CHO-DG44, uma das plataformas mais utilizadas para a produção de biofármacos, e os quais poderiam ser utilizados para a inserção sítio-dirigida de vetores de expressão de interesse pela metodologia de CRISPR-Cas9.

METODOLOGIA

- Construção e caracterização do vetor lentiviral bicistrônico de expressão de uma variante secretada da proteína EYFP (pLV-spEYFP-Dhfr);
- Transdução das células CHO-DG44 com o vetor pLV-spEYFP-Dhfr;

Endereço: SRTVN 701, Via W 5 Norte, lote D - Edifício PO 700, 3º andar
Bairro: Águas Norte CEP: 70.715-040
UF: DF Município: BRASÍLIA
Telefone: (61)3315-5877 E-mail: consp@saude.gov.br

COMISSÃO NACIONAL DE ÉTICA EM PESQUISA



Continuação do Parecer: 3.559.855

- Isolamento e clonagem das células CHO-DG44 super-produtoras da proteína spEYFP por separação de células ("cell sorting") por citometria de fluxo;
- Avaliação do número de cópias do vetor pLV-spEYFP-DHFR integradas no genoma dos clones isolados através da metodologia de PCR em tempo-real;
- Avaliação da estabilidade de expressão da proteína spEYFP ao longo de 50 passagens por fluorimetria do meio condicionado pelos clones isolados;
- Mapeamento do sítio de inserção genômica do vetor pLV-spEYFP nos clones super-produtores e estáveis através da metodologia de sequenciamento de DNA por "DNA walking";
- Transfecção estável das células CHO-DG44 com o vetor pLV-spEYFP-Dhfr;
- Amplificação gênica das células CHO-DG44 transfectadas estávelmente com o vetor pLV-spEYFP-Dhfr;
- Isolamento e clonagem das células CHO-DG44 super-produtoras da proteína spEYFP por separação de células ("cell sorting") por citometria de fluxo (estratégia transfecção estável);
- Avaliação do número de cópias do vetor pLV-spEYFP-DHFR integradas no genoma dos clones isolados através da metodologia de PCR em tempo-real (estratégia transfecção estável).

Objetivo da Pesquisa:

OBJETIVO PRIMÁRIO

Mapear sítios genômicos nas células CHO-DG44 associados a uma expressão elevada e estável da proteína de interesse.

OBJETIVO SECUNDÁRIO

Comparar a performance desta metodologia de obtenção de clones super-produtores frente a estratégia convencionalmente utilizada envolvendo a inserção aleatória no genoma do vetor de expressão através da transfecção estável, seguida de amplificação gênica e isolamento de clones superprodutores.

Avaliação dos Riscos e Benefícios:

RISCOS

1) As duas linhagens celulares necessárias para a realização desse trabalho são as células CHO-DG44, derivadas de ovário de hamster chinês, e as células HEK293T, derivadas de rim embrionário humano. O nível de biossegurança associado à manipulação das células CHO-DG44 é o 1 (NB1). Quanto às células HEK293T, o nível de biossegurança recomendado é o 2 (NB2) devido a estas células apresentarem regiões oncogênicas do adenovírus e do SV40 integradas em seu genoma.

Endereço: SRTVN 701, Via W 5 Norte, lote D - Edifício PO 700, 3º andar

Bairro: Asa Norte

CEP: 70.713-040

UF: DF

Município: BRASILIA

Telefone: (61)3215-5877

E-mail: conep@saude.gov.br

COMISSÃO NACIONAL DE ÉTICA EM PESQUISA



Continuação do Parecer: 3.555/055

Adicionalmente, as células HEK293T expressam a proteína oncogênica grande T (large T) derivada do vírus SV40.

2) O sistema de expressão lentiviral utilizado no laboratório é de terceira geração, e o seu uso está associada a um aumento da margem de biossegurança para os operadores e o público geral quando comparado a versões anteriores destes sistemas. Para sistemas de expressão lentivirais de terceira geração, recomenda-se que a produção de partículas lentivirais recombinantes e transdução das células-alvo sejam realizadas em laboratórios de nível de biossegurança 2 (NB2) com algumas recomendações especiais relativas à utilização de material perfurocortante, e o uso de vestimentas de proteção individual de forma a reduzir o potencial de exposição de mucosas às partículas lentivirais. É esperado que esse nível de biossegurança seja adequado para volumes de produção semanal de até 100 ml de meio de cultivo condicionado contendo partículas lentivirais recombinantes com títulos de 10^6 a 10^7 UFC/ml.

BENEFÍCIOS

A indústria biofarmacêutica tem mostrado um crescente interesse em metodologias de geração de clones superprodutores empregando a integração genômica sítio-específica em células CHO visando à produção de proteínas recombinantes terapêuticas. Desta forma, existe uma demanda no sentido de se acumular conhecimento sobre regiões genômicas em linhagens produtoras como as células CHO que suportem uma expressão gênica estável e em altos níveis. A implementação de estratégias que permitam direcionar a inserção sítio-específica de cassetes de expressão de interesse em regiões genômicas com estas propriedades permitiria que se melhorasse a eficiência do processo de obtenção de clones celulares superprodutores, possibilitando uma redução tanto do prazo como dos custos deste processo.

Comentários e Considerações sobre a Pesquisa:

Estudo experimental *in vitro*, com células comercializáveis depositadas em Biorrepositório.

As duas linhagens que serão utilizadas são:

Células CHO-DG44, derivadas de ovário de hamster chinês;

Células HEK293T, derivadas de rim embrionário humano (comercializáveis, guardadas em biorrepositório).

O estudo envolve metodologias de transgenia (OGM) e edição gênica e, portanto, deverá seguir o que rege a Lei 11.105 de 24 de março de 2005 (Lei de Biossegurança).

O coordenador do centro de pesquisa manifesta compromisso em requerer o Certificado de Biossegurança junto a Comissão Interna de Biossegurança em Organismos Geneticamente

Endereço: SRTVN 701, Via W 5 Norte, lote D - Edifício PO 700, 2º andar

Bairro: Asa Norte

CEP: 70.719-040

UF: DF

Município: BRASÍLIA

Telefone: (61)3315-5877

E-mail: cnrepi@ans.gov.br

COMISSÃO NACIONAL DE ÉTICA EM PESQUISA



Contribuição do Parecer: 3.559.005

Modificados (CTNBIO), de acordo com a Lei 11.105 de Biossegurança vigente.

Por se tratar de pesquisa IN VITRO com células COMERCIALIZÁVEIS, que não envolve o recrutamento de participantes de pesquisa nem a utilização de material biológico proveniente de participantes de pesquisas anteriores, o presente estudo não necessita da aprovação do comitê de ética em pesquisa com seres humanos (CEP/Conep) para ser executado.

Considerações sobre os Termos de apresentação obrigatória:

Vide item "Conclusões ou Pendências e Lista de Inadequações".

Conclusões ou Pendências e Lista de Inadequações:

Diante do exposto, a Comissão Nacional de Ética em Pesquisa - Conep, de acordo com as atribuições definidas na Resolução CNS nº 466 de 2012 e na Norma Operacional CNS nº 001 de 2013, manifesta-se pela retirada do estudo, uma vez que o protocolo de pesquisa não necessita de aprovação pelo sistema CEP/Conep para ser executado. Desta forma, a Conep não procederá com a análise do protocolo de pesquisa.

Considerações Finais a critério da CONEP:

A Comissão Nacional de Ética em Pesquisa (Conep) delibera por encerrar a tramitação do projeto, razão pela qual não emitirá parecer de análise ética sobre a presente proposição.

Situação do protocolo: Retirado

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PE_INFORMACOES_BASICAS_DO_PROJETO_1361898.pdf	28/06/2019 10:15:43		Aceito
Outros	Formulario_Cadastro_de_Pesquisa_assinado.pdf	28/06/2019 10:14:24	Marcos Angelo Almeida Demasi	Aceito
Outros	carta_anuencas_coordenador_CIMTRA.pdf	25/06/2019 10:41:13	Marcos Angelo Almeida Demasi	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_de_Pesquisa_Luis_Pomatay.pdf	25/06/2019 09:59:03	Marcos Angelo Almeida Demasi	Aceito
Folha de Rosto	FOLHA_DE_ROSTO.pdf	25/06/2019 09:56:21	Marcos Angelo Almeida Demasi	Aceito

Situação do Parecer:

Endereço: SRTVN 701, Via W 5 Norte, lote D - Edifício PO 700, 3º andar
Bairro: Águas Norte CEP: 70.719-040
UF: DF Município: BRASÍLIA
Telefone: (61) 3215-5877 E-mail: conep@saude.gov.br

COMISSÃO NACIONAL DE
ÉTICA EM PESQUISA



Continuação do Parecer: 3.559.005

Retirado

BRASILIA, 06 de Setembro de 2019

Assinado por:
Jorge Alves de Almeida Venancio
(Coordenador(a))

Endereço: SRTVN 701, Via W 5 Norte, lote D - Edifício PO 700, 3º andar
Bairro: Asa Norte CEP: 70.719-040
UF: DF Município: BRASILIA
Telefone: (61)3375-5877 E-mail: conep@saude.gov.br

registro de sa

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Reconhecimento de regiões com alto potencial de expressão gênica como alvo para seu uso na produção de proteínas recombinantes

Pesquisador: Marcos Angelo Almeida Demazi

Área Temática: Projetos de pesquisa que envolvam organismos geneticamente modificados (OGM), células-tronco embrionárias e organismos que representem alto risco coletivo, incluindo organismos relacionados a eles, nos âmbitos de: experimentação, construção, cultivo, manipulação, transporte, transferência, importação, exportação, armazenamento, liberação no meio ambiente e descarte;

Versão: 1

CAAE: 16605619.9.0000.0065

Instituição Proponente: Faculdade de Medicina da Universidade de São Paulo

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 3.524.868

Apresentação do Projeto:

Trata-se de um projeto de pesquisa com fins de formação acadêmica (mestrado) com objetivo de mapear regiões cromossômicas associadas à expressão estável e em altos níveis de proteínas recombinantes de interesse biotecnológico, empregando experimentos *in vitro* com linhagens celulares já estabelecidas e envolvendo organismos geneticamente modificados.

Objetivo da Pesquisa:

O principal objetivo do projeto de pesquisa é o mapeamento de regiões cromossômicas associadas à expressão estável e em altos níveis de proteínas recombinantes de interesse biotecnológico; e o objetivos secundário seria comparar o desempenho desta metodologia de obtenção de clones super-produtores frente a estratégia convencionalmente utilizada envolvendo a inserção aleatória no genoma do vetor de expressão através da transfecção estável, seguida de amplificação gênica e isolamento de clones superprodutores.

Avaliação dos Riscos e Benefícios:

Trata-se de um estudo experimental *in vitro*. As duas linhagens celulares necessárias para a realização desse trabalho são as células CHO-DG44, derivadas de ovário de hamster chinês, e as

Endereço: DOUTOR ARNALDO 251 21º andar sala 36
Bairro: PACAEMBU CEP: 01.246-903
UF: SP Município: SÃO PAULO
Telefone: (11) 3893-4401 E-mail: cep.fm@usp.br

Continuação do Parecer: 1.524.858

células HEK293T, derivadas de rim embrionário humano. O nível de biossegurança associado à manipulação das células CHO-DG44 é 1 (NB1). Quanto às células HEK293T, o nível de biossegurança recomendado é 2 (NB2) devido a estas células apresentarem regiões oncogênicas do adenovírus e do SV40 integradas em seu genoma. Adicionalmente, as células HEK293T expressam uma proteína oncogênica derivada do vírus SV40.

O sistema de expressão lentiviral utilizado no laboratório é de terceira geração (NB3), e o seu uso está associada a um aumento da margem de biossegurança para os operadores e o público geral quando comparado a versões anteriores destes sistemas; assim, é esperado que esse nível de biossegurança seja adequado para volumes de produção semanal de até 100 ml de meio de cultivo condicionado contendo partículas lentivirais recombinantes com títulos de 1.000.000 a 10.000.000 UFC/ml.

Comentários e Considerações sobre a Pesquisa:

O projeto é factível de ser desenvolvido, será supervisionado por pesquisador com experiência no tema e será desenvolvido em laboratório com pré-requisitos de biossegurança necessários.

Considerações sobre os Termos de apresentação obrigatória:

O proponente solicita liberação do TCLE por se tratar de um estudo experimental in vitro empregando somente cultura de linhagens celulares já estabelecidas em laboratório.

Tem anuência do coordenador do Centro de Inovação e Medicina Translacional, que informa sua submissão a Comissão Interna de Biossegurança em Organismos Geneticamente Modificados.

Recomendações:

Não há recomendações.

Conclusões ou Pendências e Lista de Inadequações:

Não foi verificada inadequações e/ou pendências para este projeto.

Considerações Finais a critério do CEP:

O presente projeto, ceguiu nesta data para análise da CONEP e só tem o seu Início autorizado após a aprovação pela mesma.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PE_INFORMACOES_BASICAS_DO_PROJETO_1361898.pdf	28/06/2019 10:15:43		Aceito
Outros	Formulario_Cadastro_de_Pesquisa_assinado.pdf	28/06/2019 10:14:24	Marcos Angelo Almeida Damasi	Aceito

Endereço: DOUTOR ARNALDO 251 21ª andar sala 36
Bairro: PAGAEMBU CEP: 01.246-900
UF: SP Município: SAO PAULO
Telefone: (11)3850-4401 E-mail: cep.fm@usp.br

USP - FACULDADE DE
MEDICINA DA UNIVERSIDADE
DE SÃO PAULO - FMUSP



Continuação do Parecer: 3.524.858

Outros	carta_anuencia_coordenador_CIMTRA.pdf	25/06/2019 10:41:13	Marcos Angelo Almeida Demasi	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_de_Pesquisa_Luis_Pomatay.pdf	25/06/2019 09:59:03	Marcos Angelo Almeida Demasi	Aceito
Folha de Rosto	FOLHA_DE_ROSTO.pdf	25/06/2019 09:56:21	Marcos Angelo Almeida Demasi	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Sim

SÃO PAULO, 22 de Agosto de 2019

Assinado por:

Marta Aparecida Azevedo Kolke Folguera
(Coordenador(a))

Endereço: DOUTOR ARNALDO 251 21º andar sala 36
Bairro: PACAEMBU CEP: 01.246-903
UF: SP Município: SÃO PAULO
Telefone: (11) 3893-4401 E-mail: csp.fm@usp.br

AGRADECIMENTOS

Gostaria de agradecer nestas linhas à ajuda que muitas pessoas me deram durante o desenvolvimento deste trabalho.

A minha amada mãe, Alejandra, e o meu amado pai, Pedro, por sempre ter incentivado o meu amor pelo conhecimento e pela descoberta, sem sua ajuda incondicional, eu não teria chegado até aqui.

Os meus irmãos, Lidia, Magdalena e Alfredo, que ao igual que meus pais, tem estado sempre comigo, mesmo não estando fisicamente por perto, pouco adianta eu tentar expressar o tamanho da minha gratidão a vocês.

Ao Dr. Marcos Almeida Demasi, meu orientador e principal colaborador ao longo de tudo este processo, que com sua direção, conhecimento, ensino e colaboração tem permitido o sucesso no desenvolvimento deste trabalho.

À Profa Dra. Mari Cleide Sogayar, que apostou por mim desde um começo e me deu a oportunidade de me aventurar pelo fantástico mundo da ciência.

Aos meus colegas do laboratório, que me acompanharam no dia-a-dia do trabalho científico dividindo frustrações, alegrias, desesperos e sucessos.

Aos técnicos do laboratório, zizi e alan, que tem estado sempre me ajudando nas necessidades dos meus experimentos.

Aos membros da secretaria da pos-graduação, Eliane e Fabia, que sempre me tem socorrido em todo tramite e documento que é precisado.

À Universidade de São Paulo, por ser a sede de tudo o conhecimento adquirido nesses anos.

Ao Programa Interunidades em biotecnologia por sua excelência na formação de mestrados e doutorados

O presente trabalho foi realizado com apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil. Processo N° 134579/2016-0

RESUMO

Pomatay LAQ. **Mapeamento de regiões cromossômicas com alto potencial de expressão gênica como alvo para seu uso na produção de proteínas recombinantes.** [Dissertação (Mestrado em Biotecnologia) – Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2019.

Nas últimas décadas, observou-se uma melhoria de várias etapas do processo de fabricação de produtos biofarmacêuticos, que acabaram se traduzindo em aumento da capacidade de produção do setor. No entanto, a produção de produtos biofarmacêuticos envolve processos complexos e de baixo rendimento. Entre os principais fatores limitantes relacionados ao estabelecimento de uma plataforma de produção baseada em células de mamíferos estão o tempo prolongado e altos custos associados ao processo de obtenção de clones celulares de alta produtividade. Essas limitações estão intimamente relacionadas às metodologias comumente empregadas para a geração de transfectantes celulares estáveis, que envolvem eventos aleatórios de integração genômica de vetores de expressão, e o uso de protocolos de amplificação gênica, que estão associados, respectivamente, a eventos de inserção em sítios cromossômicos com atividade transcricional instável e reordenamentos genéticos drásticos. Uma alternativa que tem sido explorada nos últimos anos é baseada na inserção sítio-dirigida do cassete de expressão de interesse em regiões do genoma da célula hospedeira previamente associadas à uma atividade transcricional elevada e estável. A indústria biofarmacêutica tem mostrado um crescente interesse por este tipo de alternativa para a geração de clones superprodutores em células CHO, uma das plataformas de expressão mais utilizadas para a produção de biofármacos. Nesse sentido, existe uma demanda para se acumular conhecimento sobre regiões genômicas que suportam níveis estáveis e elevados de expressão gênica, e o objetivo principal deste trabalho consistiu na varredura desses “hotspots” genômicos na linhagem celular CHO-DG44. Nossa estratégia experimental envolveu: a) a construção e o uso de um vetor lentiviral bicistrônico de expressão, o pLV-spEYFP-DHFR, para a varredura de sítios genômicos em células CHO-DG44 associadas a uma expressão elevada e estável da proteína de interesse, e b) o uso de uma variante secretada da EYFP, a spEYFP, como proteína repórter, com o intuito de evitar um possível efeito deletério associado ao acúmulo citoplasmático da EYFP. Alternativamente, as células CHO-DG44 foram transfectadas com o mesmo vetor de expressão com o objetivo de se isolar clones celulares para servir como referência de níveis de expressão da spEYFP associados a números mais elevados de cópias genômicas do vetor de expressão. Explorando-se a transdução das células CHO-DG44 com o

vetor pLV-spEYFP-DHFR, foram isolados e caracterizados oito clones celulares associados a eventos de integração genômica únicos do vetor de expressão. Estes clones celulares apresentaram níveis de expressão estáveis da proteína repórter por até dois meses, e comparáveis aos observados em clones celulares contendo de 35 a 45 cópias do mesmo vetor de expressão aleatoriamente integrado no genoma das células CHO-DG44. As nossas tentativas de mapeamento dos sítios genômicos de inserção através da técnica de DNA “walking” foram dificultadas por várias limitações associadas a fatores que limitaram a sensibilidade da técnica em detectar eventos únicos de inserção. Este trabalho foi a primeira tentativa de mapear sítios genômicos transcricionalmente ativos e estáveis em células CHO-DG44, e buscou, desta forma, complementar os achados envolvendo outras linhagens de células CHO importantes, a CHO-K1 e a CHO-S.

Palavras chaves: Biofármacos, CHO, inserção sítio-dirigida, vetores lentivirais

ABSTRACT

Pomatay LAQ. **Mapping of chromosomal regions with high potential for gene expression as a target for their use in the production of recombinant proteins.** Master thesis[Biotechnology] - Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2019

In recent decades, there has been an improvement of several stages of the manufacturing process of biopharmaceutical products, which eventually resulted in increased production capacity in the sector. However, the production of biopharmaceuticals involves complex and low-yield processes. Among the main limiting factors related to the establishment of a production platform based on mammalian cells are the prolonged time and high costs associated with the process of obtaining highly producing cell clones. These limitations are closely related to the methodologies commonly used for the generation of stable cellular transfectants, which involve random genomic integration events of expression vectors, and the use of gene amplification protocols, which are associated, respectively, with insertion events in chromosomal sites with unstable transcriptional activity and drastic genetic rearrangements. An alternative that has been explored in recent years is based on the site-directed insertion of the expression cassette of interest in regions of the host cell genome previously associated with a high and stable transcriptional activity. The biopharmaceutical industry has shown a growing interest in this type of alternative for the generation of super-producing clones in CHO cells, one of the most used expression platforms for the production of biopharmaceuticals. In this sense, there is a demand to accumulate knowledge about genomic regions that support stable and high levels of gene expression, and the main objective of this work consisted of screening for these genomic hotspots in the CHO-DG44 cell line. Our experimental strategy involved: a) the construction and use of a bicistronic lentiviral expression vector, the pLV-spEYFP-DHFR, for screening for genomic sites in CHO-DG44 cells associated with a high and stable expression of the protein of interest, and b) the use of a secreted variant of EYFP, spEYFP, as a reporter protein, in order to avoid a possible deleterious effect associated with cytoplasmic accumulation of EYFP. Alternatively, CHO-DG44 cells were transfected with the same expression vector in order to isolate cellular clones to serve as a reference for expression levels of spEYFP associated with higher numbers of genomic copies of the expression vector. Exploring the transduction of CHO-DG44 cells with the pLV-spEYFP-DHFR vector, we isolated and characterized eight cell clones associated with single genomic integration events of the expression vector. These cell clones presented stable levels of expression of the reporter protein for up to two months, and

comparable to those observed in cell clones containing 35 to 45 copies of the same expression vector integrated into the CHO-DG44 cell genome. Our attempts to map the lentiviral integration sites through DNA "walking" were hampered by various limitations associated with factors which limited the sensitivity of the technique to detect single genomic insertion events. This work was the first attempt to identify genomic sites which are transcriptionally active and stable in CHO-DG44 cells, and it sought to complement findings involving other important CHO cell lines, such as CHO-K1 and CHO-S.

Key words: Biopharmaceutical, CHO, site-directed insertion, lentiviral vector

LISTA DE FIGURAS

Figura 1: Representação esquemática destacando a interação entre a cromatina e a lâmina nuclear, estrutura que confere estabilidade mecânica ao envelope nuclear.	30
Figura 2: Mapa genético do vector lentiviral de terceira geração p156RRLsinPPTCMVIRESPRE (pLV_GFP).....	49
Figura 3: Representação esquemática dos principais elementos do vector lentiviral bicistrônico de expressão pLV-spEYFP-DHFR..	61
Figura 4: Análise por eletroforese em gel de agarose 1% m/v do perfil de restrição do clone C6H do vector recombinante pLV-spEYFP-DHFR com as enzimas de restrição indicadas..	622
Figura 5: Avaliação da expressão da proteína spEYFP por microscopia de fluorescência.	633
Figura 6: Avaliação da secreção da proteína spEYFP (~27 KDa) por SDS-PAGE 10% em condições não redutoras, seguida por detecção de fluorescência por escaneamento a laser.....	644
Figura 7: Avaliação da funcionalidade das partículas lentivirais recombinantes produzidas a partir do clone P3C do vector pLV-spEYFP-DHFR por microscopia de fluorescência.	666
Figura 8 : Enriquecimento por FACS das células spEYFP-positivas a partir das células CHO-DG44 transduzidas (A) ou transfectadas estavelmente (B) com o vector lentiviral bicistrônico de expressão pLV-spEYFP-DHFR.	677
Figura 9: Clones derivados das células CHO-DG44 produtores da proteína spEYFP isolados a partir das estratégias de transdução ou transfecção estável por lipofecção do clone P3C do vector pLV-spEYFP-DHFR.....	699
Figura 10: Avaliação da secreção da proteína spEYFP (~27KDa) por SDS-PAGE em condições não redutoras, seguida por detecção de fluorescência por escaneamento a laser.	70
Figura 11: Obtenção da curva-padrão para a quantificação do número de cópias da região codificadora da DHFR por qPCR.	71
Figura 12: Quantificação absoluta por qPCR do número de cópias do elemento bicistrônico spEYFP-IRES-Dhfr integrado no genoma das células CHO-DG44.....	733
Figura 13: Análise dos níveis de expressão relativa da spEYFP entre os clones isolados a partir das células CHO-DG44 transduzidas ou transfectadas estavelmente com o vector pLV-spEYFP-DHFR.	744
Figura 14: Avaliação da estabilidade da secreção da proteína spEYFP ao longo de 40 gerações.....	766
Figura 15: Geração das bibliotecas de fragmentos de gDNA dos clones SV2, SV4, SV5 e SV7.	777
Figura 16: Etapas de “Nested” PCR para o reconhecimento dos sítios de integração do vector de expressão pLV-spEYFP-DHFR nas células CHO-DG44.	788

Figura 17: Digestão do vetor pLV-spEYFP-DHFR com as enzimas de restrição SspI, HpaI e EcoRV.	799
Figura 18: Avaliação da eficiência do protocolo de purificação dos fragmentos de DNA baseado em adsorção do DNA à sílica.....	80
Figura 19: Detecção por “Nested” PCR da ligação dos oligonucleotídeos adaptadores aos fragmentos de DNA oriundos da digestão do vetor pLV-spEYFP-DHFR com as enzimas de restrição HpaI, SspI e EcoRV.	81
Figura 20: Amplificação dos sítios genômicos de integração do vetor de expressão pLV-spEYFP-DHFR nos clones SV4 e SV5.	822

LISTA DE ABREVIATURAS E SIGLAS

BHK: *Baby Hamster Kidney* (rim de hamster bebé)

cDNA: *complementary DNA* (ADN complementar)

CHO: *chinese hamster ovary* (ovario de hamster Chines)

COSMC: *Core 1 beta3-Gal-T-Specific Molecular Chaperone* (Chaperona molecular específica do núcleo 1 beta3-Gal-T)

cHS4: *chicken β -globin 5C hypersensitive site 4* (sítio 4 da região hipersensível do gene da β -globina de galinha)

CRISPR: *Clustered Regularly Interspaced Short Palindromic Repeats* (Repetições Palindrômicas Curtas Agrupadas e Regularmente Interespaçadas)

DNA: *Deoxyribonucleic Acid* (ácido desoxirribonucleico – ADN)

DHFR: dihidrofolato-redutase

DMSO: dimetilsulfóxido

DMEM: *Dulbecco's Modified Eagle Medium* (meio Dulbecco's Modified Eagle)

EGFP: *Enhanced green fluorescent protein* (proteína verde fluorescente melhorada)

EYFP: *Enhanced Yellow Fluorescent Protein* (proteína amarela fluorescente melhorada)

EDTA: *Ethylenediaminetetraacetic Acid* (ácido etilenodiaminotriacético)

FRT: *Flipase Recognition Target* (alvo de reconhecimento da flipase)

FITC: *fluorescein isothiocyanate* (isotiocianato de fluoresceína)

FUT8: *Alpha-(1,6)-fucosyltransferase* (Alpha-(1,6)-fucosiltransferase)

Gcg: glucagon

Gly: glicina

GS: glutamine synthetase (glutamina sintetase)

Grik: *Glutamate receptor, ionotropic, kainate 1* (Receptor de glutamato ionotrópico, cainato 1)

HIV-1: *Human Immunodeficiency Virus-1* (vírus de imunodeficiência humana 1)

Hprt: *Hypoxanthine-guanine PhosphoRibosylTransferase* (Hipoxantina-guanina fosfo-ribossil-transferase)

IRES: *Internal Ribosome Entry Site* (elemento interno de ligação ao ribossomo)

kDa: quilodalton

LB: Luria-Bertani; *Lysogeny Broth* (meio lisogênico)

LCR: *Locus Control Region* (regiões de controle de locus)

M: molar

MTX: *methotrexate* (metotrexato)

MSX: *Methionine Sulfoximine* (metionina sulfoxinamina)

MOI: *Multiplicity of infection* (multiplicidade de infecção)

NAP: Núcleo de Apoio à Pesquisa

Nm: nanômetro

NSO: *non-secreting murine myeloma* (mieloma murino)

NUCEL: Núcleo de Terapia Molecular e Celular

Pb: pares de bases

PBS: *Phosphate Buffered Saline* (solução salina tamponada com fosfato)

PCR: *Polymerase Chain Reaction* (reação em cadeia da polimerase)

RCF: *relative centrifugal force* (força centrífuga relativa)

RE: Retículo endoplasmático

RNA: *Rybonucleic Acid* (ácido ribonucleico – ARN)

RMCE: *Recombinase mediated cassette Exchange* (troca de cassete mediada por recombinase)

RT-PCR: *reverse transcription polymerase chain reaction* (transcrição reversa seguida pela reação em cadeia da polimerase)

SDS: *Sodium Dodecyl Sulfate* (dodecil sulfato de sódio)

SDS-PAGE: *Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis* (eletroforese em gel de poliacrilamida contendo SDS)

UCOE: *Ubiquitous Chromatin Opening Elements* (elementos de abertura de cromatina ubíquos)

STAR: *Stabilizing and Antirepressor*

SFB: Soro Fetal Bovino

S/MARs: *Scaffold/Matrix Attachment Regions* (regiões de ligação na matriz)

T_m: temperatura de *melting*

TALENs: *Transcription Activator-like Effector Nucleases* (Nucleases com efetores do tipo ativador transcricional)

U: Unidade

Ufc: unidade formadora de colônias

USP: Universidade de São Paulo

SUMÁRIO

1	INTRODUÇÃO	27
1.1	<i>Indústria biofarmacêutica e a produção de proteínas terapêuticas</i>	27
1.2	<i>Células de mamíferos como plataformas de expressão</i>	28
1.3	<i>Estrutura genômica e o controle da expressão gênica</i>	29
1.4	<i>Emprego de regiões moduladoras da estrutura de cromatina em vetores de expressão</i>	31
1.5	<i>O uso das células CHO como plataforma de expressão de biofármacos</i>	32
1.6	<i>Genoma das células CHO e a instabilidade da expressão</i>	34
1.7	<i>O fenômeno de amplificação gênica nas células CHO</i>	35
1.8	<i>Expressão e síntese de proteínas</i>	36
1.9	<i>Recombinação homóloga e a inserção sítio-dirigida</i>	37
1.10	<i>Vetores de expressão lentivirais como ferramenta de varredura de sítios genômicos com atividade transcricional elevada e estável</i>	41
1.11	<i>A proposta do trabalho</i>	42
2	OBJETIVOS	44
2.1	<i>Objetivo geral</i>	44
2.2	<i>Objetivos experimentais específicos</i>	44
3	MATERIAIS E MÉTODOS	45
3.1	<i>Linhagens celulares e manutenção</i>	45
3.2	<i>Construção e caracterização do vetor lentiviral bicistrônico de expressão de uma variante secretada da proteína EYFP (pLV-spEYFP-DHFR)</i>	46
3.3	<i>Transfecção transitória das células HEK 293T com o vetor pLV-spEYFP-DHFR</i>	50
3.4	<i>Produção das partículas lentivirais recombinantes em células HEK 293T</i>	51

3.5	<i>Transdução estável das células CHO-DG44 cultivadas em aderência com as partículas virais recombinantes</i>	51
3.6	<i>Transfecção estável das células CHO-DG44 com o vetor pLV-spEYFP-DHFR</i>	52
3.7	<i>Enriquecimento das populações celulares mistas das células CHO-DG44 produtoras da proteína spEYFP através de separação de células (“cell sorting”) por citometria de fluxo</i>	52
3.8	<i>Isolamento das células CHO-DG44 superprodutoras da proteína spEYFP por diluição limitante</i>	53
3.9	<i>Seleção das células CHO-DG44 transfectadas estavelmente com o vetor pLV-spEYFP-DHFR com MTX</i>	54
3.10	<i>Extração de DNA genômico (gDNA)</i>	54
3.11	<i>Determinação, através de PCR em tempo-real (qPCR), do número de cópias do elemento spEYFP-IRES-DHFR integradas no genoma dos clones celulares gerados</i>	54
3.12	<i>Avaliação dos níveis de secreção da proteína spEYFP</i>	56
3.13	<i>Avaliação da estabilidade dos níveis de secreção da proteína spEYFP</i>	56
3.14	<i>Determinação do sitio de integração baseado na estratégia de “DNA walking”</i>	57
3.14.1	<i>DNA walking</i>	57
3.14.2	<i>Geração das bibliotecas de gDNA</i>	58
3.14.3	<i>Reação de ligação da sequência adaptadora às bibliotecas de gDNA</i>	58
3.14.4	<i>“Nested” PCR</i>	59
3.15	<i>Análise estatística</i>	60
4	RESULTADOS	61
4.1	<i>Construção e caracterização do vetor lentiviral bicistrônico de expressão de uma variante secretada da proteína EYFP (pLV-spEYFP-DHFR)</i>	61
4.2	<i>Teste de expressão e secreção da proteína spEYFP</i>	63
4.3	<i>Avaliação da produção e funcionalidade das partículas lentivirais recombinantes geradas a partir do clone P3C do vetor pLV-spEYFP-Dhfr</i>	65
4.4	<i>Geração de células CHO-DG44 com expressão estável da proteína spEYFP</i>	66
4.5	<i>Isolamento e caracterização de clones celulares a partir das populações celulares mistas enriquecidas CHO-DG44 spEYFP-DHFR SV e CHO-DG44-spEYFP-DHFR ST</i>	68

4.6	<i>Determinação do número de cópias do elemento spEYFP-IRES-DHFR integradas no genoma dos clones celulares gerados</i>	70
4.7	<i>Quantificação relativa e análise da estabilidade de expressão da spEYFP</i>	73
5.	DISCUSSÃO	84
6.	CONCLUSÕES	91
	BIBLIOGRAFIA	92

1 INTRODUÇÃO

O projeto que originou esta dissertação foi desenvolvido no Núcleo de Apoio à Pesquisa (NAP) NUCEL/NETCEM (Núcleo de Terapia Molecular e Celular), vinculado à Faculdade de Medicina da USP. O centro de pesquisa apresenta como uma de suas quarto linhas de pesquisa o desenvolvimento de linhagens celulares produtoras de proteínas recombinantes de interesse terapêutico e biotecnológico. Então, a procura por novas estratégia que permitam o melhoramento ou otimização na geração de clones produtores de proteínas recombinantes é um interesse constante de nosso grupo. Foi nesse contexto que este projeto foi elaborado.

1.1 Indústria biofarmacêutica e a produção de proteínas terapêuticas

Em 1986 foi aprovada a primeira aplicação terapêutica de uma proteína recombinante produzida em células de mamífero, o ativador de plasminogênio tecidual humano (Genentech, EUA), revelando as vantagens desta plataforma de expressão em comparação às outras da época, e impulsionando a sua utilização na indústria biofarmacêutica¹. Algumas características das proteínas recombinantes para o tratamento de diversas doenças, como a sua alta potência e especificidade², têm acelerado a expansão destes produtos³, atingindo vendas de \$188 bilhões em 2017 (La Merie Business Intelligence, 2018)⁴.

Ante este crescimento, o aumento gradual nos níveis de produção dos biofármacos é um tema de interesse ao longo dos últimos anos⁵. Desta forma, surgiu a necessidade de se estabelecer plataformas de produção de biofármacos associadas a uma maior produtividade. Estratégias para otimizar a produtividade têm sido implementadas, como, por exemplo, a otimização da composição dos meios de cultivo, o controle mais rígido dos parâmetros de cultivo, a utilização de promotores/acentuadores específicos da transcrição gênica associados a níveis de expressão mais elevados e estáveis do produto de interesse, e o emprego de protocolos de amplificação gênica. Entretanto, aliar alta produtividade e manutenção das características estruturais e funcionais da proteína terapêutica é um desafio constante no desenvolvimento de uma plataforma celular de expressão de biofármacos⁵.

1.2 Células de mamíferos como plataformas de expressão

A determinação da plataforma de expressão é uma etapa crucial no estabelecimento do processo de fabricação de biofármacos⁶, e esta deve ser compatível com a complexidade e as diferentes características estruturais de um determinado biofármaco⁷. Dentre as diversas plataformas de expressão existentes, as células de mamíferos são comumente usadas como sistemas de expressão heteróloga para a produção de proteínas complexas, pois, diferentemente das bactérias ou fungos, apresentam organelas especializadas e sistemas enzimáticos específicos que permitem a ocorrência do dobramento tridimensional correto e a introdução de modificações pós-traducionais específicas necessárias para a atividade terapêutica de muitas destas proteínas⁸.

Nestes últimos anos, cerca de 80% das proteínas terapêuticas em fase de desenvolvimento foram produzidas em células de mamíferos⁹. Dentre as culturas de células de mamíferos mais utilizadas na indústria biofarmacêutica, pode-se destacar: a) a linhagem celular proveniente das células de ovário de hamster chinês (CHO – Chinese Hamster Ovary), e linhagens mutantes derivadas a partir desta linhagem, deficientes para a expressão da enzima Dihidrofolato redutase (DHFR)¹⁰, sendo este o sistema de expressão mais empregado para a produção de produtos recombinantes para aplicações terapêuticas²; b) as células derivadas de mieloma murino NS0, também comumente utilizadas como hospedeiras para a produção de proteínas recombinantes³; e c) as células derivadas de rim embrionário humano HEK-293 (Human Embryonic Kidney), geralmente usadas em abordagens de expressão transitória em grande escala¹¹.

Classicamente, o estabelecimento de uma plataforma de produção de proteínas recombinantes envolve a modificação genética da linhagem celular de interesse através da inserção, em seu genoma, de vetores de expressão contendo o gene recombinante de interesse, associado a elementos regulatórios da expressão gênica, e marcadores de seleção apropriados para a obtenção de clones celulares que possuam uma produção elevada e estável da proteína recombinante de interesse⁷. Os marcadores de seleção mais comumente utilizados são as regiões codificadoras dos genes codificadores da enzima DHFR e glutamina sintetase (GS), os quais permitem que se explore o fenômeno de amplificação gênica¹². Através de protocolos de seleção envolvendo a utilização de concentrações crescentes dos inibidores metotrexato (MTX), no caso da DHFR, e metionina sulfoximina (MSX), no caso da GS, é possível obter a amplificação do número de cópias do gene recombinante, associada a aumentos expressivos nos níveis de produção da proteína recombinante de interesse¹³.

A produção de proteínas recombinantes em linhagens celulares de mamíferos pode ser conseguida com sistemas de expressão transitória ou estável¹⁴. Os sistemas de expressão transitória possuem como principal vantagem a rapidez na obtenção de grandes quantidades do produto. Hoje em dia é predominantemente feita em linhagens celulares derivadas de HEK293 e CHO adaptadas para cultivo em suspensão¹³. Dentro das limitações do sistema estão a baixa taxa de crescimento celular específico, a baixa concentração celular, e o alto consumo das principais fontes de carbono e energia (geralmente glicose e glutamina)².

Por outro lado, os sistemas de expressão estável proporcionam a possibilidade de escalonamento da produção, além de permitir uma maior reprodutibilidade, consistência na qualidade do produto e a exploração por períodos mais prolongados do processo de produção¹⁴. Existem algumas dificuldades com o sistema de expressão estável, destacando-se o silenciamento da expressão gênica, e a perda do transgene recombinante associada a instabilidades genômicas no sítio de inserção⁵. Entretanto, a utilização de diversos elementos funcionais em vetores de expressão, como, por exemplo, regiões acentuadoras da transcrição, ou isoladores do contexto estrutural da região de inserção genômica, tem se mostrado benéficas para a manutenção da estabilidade da expressão da proteína de interesse¹⁴, conforme será discutido mais adiante.

1.3 Estrutura genômica e o controle da expressão gênica

A estrutura genômica dos eucariontes muda constantemente durante, por exemplo, o desenvolvimento ou em resposta a estímulos ambientais, alterando, desta forma, a expressão e posição no espaço nuclear dos genes¹⁵. Esta organização estrutural dinâmica desempenha uma função central no controle da expressão gênica¹⁶.

Quanto à atividade transcricional, o genoma é classicamente segregado em dois domínios, sendo um deles caracterizado por ser um ambiente altamente repressivo à transcrição gênica, e situado na lâmina nuclear, chamado de heterocromatina¹⁷. O outro, localizado no interior do núcleo, num compartimento associado à alta atividade transcricional, é conhecido como eucromatina¹⁸ (Figura 1).

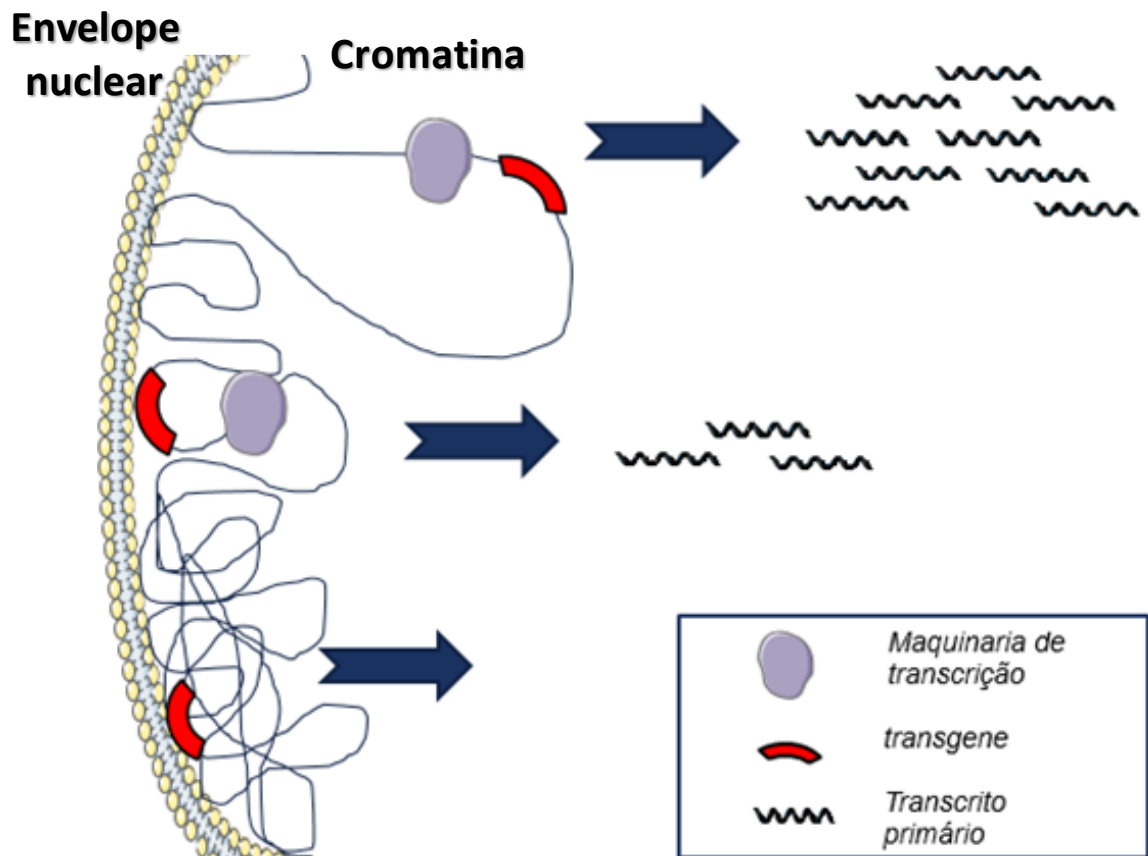


Figura 1: Representação esquemática destacando a interação entre a cromatina e a lâmina nuclear, estrutura que confere estabilidade mecânica ao envelope nuclear. O nível de interação da cromatina e a lâmina nuclear está associado ao grau de condensamento da cromatina, o qual, por sua vez, tem um impacto direto sobre a atividade transcricional de regiões genômicas específicas.

A heterocromatina apresenta uma estrutura topologicamente compactada que impede a acessibilidade dos genes pelos complexos de transcrição¹⁹. A heterocromatina pode se formar transitoriamente, como parte de um ciclo dinâmico entre configurações "abertas" e "fechadas", ou assumir uma estrutura compacta permanente¹⁶.

A eucromatina é caracterizada por regiões ricas em nucleotídeos AT, e é onde se situam as origens de replicação do DNA, os sítios de ligação para a DNA topoisomerase II, e as regiões de ligação à matriz nuclear, conhecidas pela sigla S/MARs (Scaffold/Matrix Attachment Regions)²⁰. Outra característica estrutural marcante da eucromatina, e que está relacionada com o controle da transcrição gênica, são regiões conhecidas como "fábricas de transcrição"^{15,21}, que são sítios nucleares específicos, de aproximadamente 80nm de diâmetro, onde ocorre a maior parte da transcrição mediada pela RNA polimerase II, e onde os genes são recrutados de forma dinâmica, resultando na ativação da sua transcrição²².

O controle da organização estrutural da cromatina é um processo dinâmico conhecido como remodelamento de cromatina. O remodelamento da cromatina pode ser mediado por alterações estruturais dos nucleossomos, ou podem ser mediadas por alterações químicas no DNA. Os nucleossomos estão intimamente relacionados com o controle do grau de compactação da cromatina, e, conseqüentemente, com a atividade transcricional de regiões do genoma. Basicamente, os nucleossomos podem influenciar a estrutura da cromatina através da alteração da estrutura do próprio nucleossomo, ou através de modificações químicas das histonas, ou ainda, através do reposicionamento dos nucleossomos²³. Como mencionado acima, outro mecanismo de regulação da atividade transcricional envolve a modificação química do DNA, mais especificamente a adição ou remoção de grupos metila, sendo este processo conhecido como metilação²⁴.

1.4 Emprego de regiões moduladoras da estrutura de cromatina em vetores de expressão

Dentro dos elementos regulatórios de ação cis mais utilizados em vetores de expressão em células de mamíferos, destacam-se regiões promotoras e/ou acentuadoras da transcrição, as regiões de ligação de matriz nuclear (S/MARs)²⁵, as regiões isoladoras de cromatina²⁶, regiões de controle de locus (LCRs)²⁷, e elementos de abertura de cromatina ubíquos (UCOEs)²⁸. Vários destes elementos permitem a ligação de proteínas que interagem com a cromatina, e, desta forma, potencializam ou inibem os efeitos da estrutura da cromatina adjacente sobre a atividade transcricional dos minigenes recombinantes integrados no genoma²⁸. De forma geral, tem-se demonstrado que estes elementos podem ser combinados com diferentes promotores encontrados em vários vetores de expressão comerciais, sendo esta estratégia efetiva em linhagens celulares utilizadas para a produção de biofármacos, conferindo uma expressão estável do gene recombinante e em níveis elevados²⁹.

Um dos elementos utilizados são as S/MARs, que medeiam a ligação de “loops” de cromatina às proteínas estruturais nucleares. As S/MARs estão associadas a regiões do genoma com alta atividade transcricional, já que, ao atuar como pontos de ligação, facilitam a ancoragem da cromatina à matriz nuclear durante a interfase³⁰. Desta forma, as S/MARs proporcionam o recrutamento de fatores de transcrição nesta região, o que restringe, ao mesmo tempo, a difusão da heterocromatina através da desmetilação do DNA e a acetilação de histonas¹⁷, aumentando assim as taxas de transcrição e diminuindo a probabilidade de

silenciamento da transcrição^{31,32}. Esta característica das S/MARs é essencial para a ocorrência de processos como o processamento e transporte de RNA, transdução de sinal, e na replicação e transcrição do DNA³³. As S/MARs exibem um conjunto de atividades que têm sido difíceis de atribuir a qualquer motivo específico de DNA, sendo muito difícil a predição destes elementos a partir de dados de sequências primárias de DNA^{17,34}. Ainda assim, todas estas características têm incentivado a utilização destes elementos em vetores de expressão em células de mamífero, ou através de co-transfecções dos vetores de expressão e vetores contendo elementos S/MARs, para aumentar significativamente os níveis de expressão³⁵. Entretanto, existem relatos de que altas densidades de elementos S/MARs em um único locus cromossômico podem saturar atividades de ligação das S/MARs em áreas específicas do núcleo, pois nem todos sítios presentes podem ser capazes de se ligar à matriz nuclear simultaneamente³⁴.

Outro elemento modulador da estrutura de cromatina utilizado foi originalmente derivado do gene da β -globina humana, sendo conhecido como região de controle de locus (LCR – Locus Control Region) ²⁷. Ainda outra classe destes elementos é conhecida como elementos insuladores, sendo o mais explorado o insulador derivado do sítio 4 da região hipersensível do gene da β -globina de galinha (cHS4 – *chicken β -globin 5C hypersensitive site 4*), o qual bloqueia uma possível influência do contexto do sítio de integração sobre a atividade promotora do vetor de expressão³⁶. Outros elementos isoladores utilizados são os elementos derivados de genes expressos ubiquamente (UCOEs – Ubiquitous Chromatin Opening Elements)²⁸, e os elementos STAR (*Stabilizing and Antirepressor*) ³⁷. Adicionalmente, tem-se empregado elementos genômicos que permitam uma expressão estável do transgene derivados da linhagem célula CHO. Um exemplo é o elemento regulatório designado E77, isolado a partir da linhagem celular CHO-K1, tendo sido demonstrado sua aplicabilidade na obtenção de níveis estáveis de expressão quando empregado como acentuador da transcrição gênica e combinado à região promotora derivada do citomegalovírus humano (CMV)³⁸.

1.5 O uso das células CHO como plataforma de expressão de biofármacos

As células CHO foram originadas no laboratório do Dr. Theodore Puck no final da década de 1950³⁹. As duas linhagens derivadas da CHO mais utilizadas para a expressão heteróloga de proteínas recombinantes, conhecidas como CHO-DUKX-B11 e CHO-DG44, contêm mutações

no gene da enzima dihidrofolato redutase, e originaram-se a partir da mutagênese *in vitro* da linhagem CHO parental⁴⁰.

Nestes últimos anos, a diversidade de linhagens e otimização de bioprocessos na cultura das células CHO têm permitido um aumento dramático na produtividade de proteínas terapêuticas, tornando as células CHO a plataforma de expressão mais utilizada para a produção biofarmacêutica, sendo responsável por cerca de 70% da produção das proteínas recombinantes utilizadas comercialmente⁴¹.

A popularidade das células CHO como plataforma para produção de biofármacos pode ser atribuída às seguintes razões:

- A experiência acumulada durante as últimas quatro décadas com a utilização clínica de diversas proteínas terapêuticas produzidas nas células CHO as consolidaram como uma plataforma segura para a produção de biofármacos, permitindo uma maior facilidade para obter a aprovação das agências reguladoras, como a FDA (*Food and Drug Administration*), para a comercialização de proteínas terapêuticas¹;
- A capacidade das células CHO de se adaptarem e crescer em altas densidades em sistemas de cultivos em suspensão, característica esta ideal para a produção de biofármacos a nível industrial¹³;
- A capacidade das células CHO de serem cultivadas em meios de cultivo livre de soro e quimicamente definidos, garantindo uma maior reprodutibilidade entre diferentes lotes de cultura de células, uma característica muito valorizada para cultura em larga escala em biorreactores⁴²;
- As células CHO possuem maquinaria para a promoção de modificações pós-traducionais complexas em diversas classes de proteínas recombinantes, tornando-as biocompatíveis e bioativas em seres humanos¹. Especificamente, um exemplo é a glicosilação de glicoproteínas com a ausência de epítipo de α -galactose, que é imunogênica⁴³;
- Vários sistemas de amplificação gênica foram estabelecidos nas células CHO como estratégia para aumentar os rendimentos de produção das proteínas recombinantes⁴¹, permitindo que se atinja níveis de produção de biofármacos de até 5-10 g/L em cultura em suspensão das células CHO⁴⁴.

1.6 Genoma das células CHO e a instabilidade da expressão

O hamster chinês (*Cricetulus griseus*) apresenta células primárias diploides, com um cariótipo de 22 cromossomos (20 autossômicos e dois sexuais)⁴⁵. As células CHO não são totalmente diploides no sentido funcional, já que alguns genes estão presentes no estado hemizigoto, o que tem facilitado a geração de sublinhagens mutantes derivadas destas células^{46,47}. Apesar de muitas das variações de sequência serem compartilhadas entre membros da mesma família de células CHO, como, por exemplo, a CHO-K1, CHO-DG44 ou CHO-S, a distribuição não uniforme de mutações em cada linhagem celular parece ter alguma relevância fenotípica, particularmente na rota biossintética envolvida com a glicosilação, ou quanto a efetividade de execução do processo de apoptose⁴⁸. Adicionalmente, o próprio processo de seleção de clones produtores a partir de transfectantes estáveis da CHO e linhagens derivadas está associado a uma importante instabilidade genômica, relacionada a rearranjos cromossômicos decorrentes de eventos de translocações e recombinações homólogas, especialmente em resposta a procedimentos de amplificação genica, gerando a acumulação de várias mutações ao longo do processo de seleção/isolamento do clone produtor⁸.

Recentemente, sequenciou-se o genoma do hamster chinês (*Cricetulus griseus*)⁴⁸, encontrando-se um total de 24.044 genes codificadores de proteína. Posteriormente, em 2011, ocorreu a publicação da sequência do genoma de uma das linhagens derivada das células CHO, as células CHO-K1⁴⁹. Descreveu-se que a CHO-K1 mantinha o mesmo número de genes do hamster chinês, mas distribuídos em 21 cromossomos estruturalmente diferentes⁴⁹. Desde então, uma série de genomas de diferentes linhagens celulares derivadas da CHO foi publicada, como a sublinhagem CHO-DG44⁴⁸, gerada por mutagênese química a partir de uma sublinhagem da CHO deficiente em prolina (CHOpro3-), seguida de exposição à irradiação gama, com o intuito de se deletar os alelos do gene *Dhfr*⁵⁰. No genoma da CHO-DG44 foi observado um total de 20 cromossomos em vez dos 22 encontradas nas células de hamster chinês, sendo apenas 7 deles normais, sendo, aparentemente, os cromossomos um e quatro os mais estáveis⁷. Aliás, no cromossomo dois, onde o gene *Dhfr* de hamster está originalmente situado, verificou-se que 89 genes foram reduzidos de diploides a haploides, redução esta também observada nos cromossomos 9/10 e X, nos quais observou-se uma redução de 70% no número de genes⁵¹.

A disponibilidade pública de sequências dos genomas de várias sublinhagens derivadas da CHO tem complementado os dados gerados a partir de outras abordagens para a caracterização das células CHO, como a transcriptômica e a proteômica⁵¹. Estes recursos genômicos forneceram uma grande quantidade de oportunidades em biotecnologia, permitindo, por exemplo, o desenvolvimento de abordagens visando a engenharia das células CHO para o melhoramento da produtividade e a qualidade das proteínas terapêuticas produzidas⁵².

1.7 O fenômeno de amplificação gênica nas células CHO

A amplificação gênica é um fenômeno amplamente difundido, reportado em bactérias, leveduras, insetos e mamíferos⁴⁶. O fenômeno consiste na replicação repetida de algum gene ou genes sem o proporcional incremento de outros genes no genoma⁵³. Atualmente, as tecnologias de desenvolvimento de linhagens celulares utilizadas pela maioria das empresas biofarmacêuticas são baseadas na tecnologia de amplificação gênica associada ao locus de inserção genômica do cassete de expressão da DHFR⁵⁴, ou o sistema baseado na glutamina sintetase (GS) desenvolvido pela empresa Lonza⁵⁵. Ambos sistemas utilizam um fármaco específico para inibição da atividade enzimática selecionável e essencial para o metabolismo celular. Um destes inibidores é o MTX, um análogo da dihidrofolato, que inibe a DHFR, interferindo com a síntese *de novo* de purinas e pirimidinas, e, desta forma, com o processo de síntese de ácidos nucleicos no sistema de amplificação gênica baseado na DHFR⁵⁶. No caso do sistema baseado na enzima GS, o inibidor utilizado é a metionina sulfoximina (MSX)⁴².

O processo de amplificação gênica pode ser iniciado por eventos de quebras cromossômicas, e pode levar à formação de uma variedade de estruturas cromossômicas, incluindo regiões cromossômicas expandidas, elementos extra-cromossômicos, e cromossomos dicêntricos⁵⁷. Este processo é estimulado por incrementos graduais na concentração de seleção de MTX ou MSX, e que conduzem a um maior número de genes amplificados nas regiões teloméricas em comparação a outras regiões do genoma⁵⁸. No caso de se empregar linhagens mutantes das células CHO que não contenham cópias funcionais do gene *Dhfr*, como, por exemplo, as linhagens CHO-DUXB11 e a CHO-DG44, é necessária a suplementação do meio de cultivo das células parentais com uma fonte de purina, timidina e glicina para contornar a deficiência de DHFR, já que a via biossintética de síntese de purinas e pirimidinas de salvamento é funcional nestas linhagens. A deficiência bioquímica na via de síntese *de novo* de

purinas e pirimidinas é inclusive utilizada para selecionar eventos de integração genômica estáveis nas células CHO-DG44 e CHO-DUXB11⁴⁶.

Uma problemática da metodologia de amplificação gênica é que os genes recombinantes amplificados, e incorporados no genoma das células hospedeiras, nem sempre são estáveis. Por um lado, a manutenção artificial de níveis elevados de expressão gênica de uma determinada proteína de interesse, estranha às células CHO, pode constituir um fator de estresse celular, seja por um possível efeito citotóxico relacionado a altos níveis da proteína recombinante, ou à saturação da capacidade metabólica e biossintética das células¹³. Uma das alternativas das células lidar com esse estresse associado à super-expressão de proteínas recombinantes é através da regulação da expressão gênica, o que pode eventualmente resultar na instabilidade da produção de proteína recombinante¹¹. Por outro lado, o próprio processo de amplificação gênica está associado à eventos drásticos de rearranjos genéticos e ao surgimento de estruturas genômicas aberrantes, comprometendo a estabilidade de expressão da proteína recombinante e inclusive a integridade de genes endógenos⁸.

1.8 Expressão e síntese de proteínas

A expressão e síntese de proteínas é um processo complexo que envolve várias etapas, incluindo a transcrição, a tradução, o processamento pós-traducional e secreção, no caso de proteínas secretadas. Os níveis de expressão celular de uma determinada proteína são determinados pela eficiência de cada um desses pontos de controle individual⁸. No entanto, em células eucarióticas, os níveis finais de expressão proteica são extremamente dependentes da taxa de transcrição gênica, e, conseqüentemente, da acessibilidade de trechos específicos do DNA à maquinaria de transcrição. Como mencionado anteriormente, trechos do DNA genômico se encontram altamente condensados nas estruturas de cromatina, e esta condensação muitas vezes dificulta a acessibilidade do DNA aos complexos de transcrição⁵⁹. A ativação da transcrição requer o rearranjo da estrutura da cromatina, sendo esta etapa importante para o controle da transcrição, podendo a cromatina sofrer modificações estruturais devido a uma variedade de modificações químicas, como a metilação, acetilação, fosforilação e ubiquitinação⁶⁰. A acetilação das histonas é uma das modificações estruturais associada ao remodelamento da cromatina mais bem caracterizada, relacionada à ativação transcricional, embora modificações como a fosforilação, ubiquitinação e hipometilação das histonas

desempenhem um papel relevante na ativação gênica⁶¹. Por outro lado, a metilação do DNA é vista como um mecanismo de bloqueio que garante a manutenção do estado inativo⁶².

Como mencionado acima, a transcrição é considerada como o fator dominante que controla os níveis finais de expressão proteica, principalmente através do controle da taxa de transcrição e dos eventos de iniciação da transcrição⁶³. Entretanto, eventos pós-transcricionais também podem impactar nos níveis finais de expressão proteica. Eventos envolvidos com a maturação do mRNA, como a adição enzimática da 7-metilguanossina na extremidade 5' do mRNA, ou ainda, a poliadenilação no extremo 3', ou ainda o processo de “splicing” para remoção das sequências intrônicas, e até mesmo a estrutura secundária assumida pelo mRNA⁶⁴, desempenham um impacto significativo sobre os rendimentos de expressão das proteínas, seja pelo aumento da estabilidade do mRNA, ou pelo aumento da taxa de exportação nuclear do mRNA, ou ainda devido ao aumento da taxa de tradução⁶⁵.

Para as proteínas que são destinadas à secreção, além dos controles transcricionais e pós-transcricionais mencionados acima, existem fatores que regulam o fluxo de proteínas em vários estágios da via secretória, e que podem impactar no controle geral da produção⁶⁶. Nas células de mamíferos, esses fatores podem incluir processamentos efetuados no retículo endoplasmático (ER) e no complexo de Golgi, os quais desempenham uma função-chave nos processos de glicosilação e outras modificações pós-traducionais, como também na segregação das proteínas, dependendo do seu destino final⁶⁷.

1.9 *Recombinação homóloga e a inserção sítio-dirigida*

Em eucariotos superiores, a recombinação homóloga é um evento essencial, que faz parte de processos celulares como o reparo do DNA e a troca de material genético entre as cromátides irmãs durante os eventos de mitose e meiose⁶⁸.

Baseando-se na ocorrência deste fenômeno, foram desenvolvidas estratégias alternativas para a obtenção de clones produtores estáveis de proteínas recombinantes⁶⁹, constituindo uma alternativa para superar o silenciamento dos genes recombinantes associado ao “efeito de posição”¹¹.

Para sistemas de expressão baseados em células de mamíferos, foi desenvolvido o sistema de troca de cassete mediada por recombinase (RMCE – *Recombinase mediated cassette*

exchange), o qual foi introduzido em 1994 como uma estratégia que permitia o uso repetido de *loci* genômicos determinados, extensamente caracterizados e com propriedades de expressão conhecidas⁷⁰. Para se efetuar a substituição de um cassete de expressão contendo qualquer transgene ou construção de DNA desejada, é necessária a ocorrência de um evento de recombinação recíproca dupla, que envolve a utilização de duas sequências altamente homólogas flanqueando tanto o cassete genômico alvo como a região codificadora a ser inserida⁷¹. Dentre os sistemas de recombinação mais utilizados estão o sistema Cre/loxP, derivado de bactérias, e o sistema Flp/FRT, derivado de levedura, que reconhecem, respectivamente, as sequências loxP e FRT, e catalisam o evento de quebra da fita de DNA mediada por um resíduo específico de tirosina situado no sítio ativo destas enzimas⁷¹. Resumidamente, o processo de RMCE envolve uma primeira etapa de integração genômica aleatória, na linhagem celular de interesse, de um vetor de expressão de um gene repórter flanqueado por duas sequências acceptoras heterotípicas e incompatíveis, criando-se um alelo acceptor. Na sequência, são selecionadas as células contendo os eventos de inserção genômica estável associados a uma alta expressão do gene repórter⁴². A partir destes clones celulares é feita a recombinação homóloga duplamente recíproca mediada por uma recombinase, empregando-se um segundo vetor de expressão contendo o gene de interesse flanqueado pelas regiões de homologia presentes no primeiro vetor⁷².

O sistema Cre/lox do bacteriófago P1 foi desenvolvido como uma ferramenta versátil devido à sua simplicidade e atividade em sistemas heterólogos⁷³. Requer uma única proteína recombinase, Cre, que não necessita de cofatores adicionais para mediar a recombinação recíproca entre dois dos seus locais de reconhecimento no DNA, conhecido como loxP, de 34 pb, que permanecerá no DNA das células hospedeiras após a integração⁷⁴. Os monômeros Cre ligam-se a sítios loxP em duas repetições invertidas de 13 pb, que são separadas por uma região espaçadora de 8 pb, o que fornece a direcionalidade ao local alvo⁶⁹. No entanto, esta reação é reversível no sítio loxP. Apesar desta desvantagem, o sistema Cre/loxP é capaz de gerar eventos de recombinação com alta especificidade, em cerca de 80% dos eventos de recombinação⁷⁵.

O sistema Flp/FRT, derivado de levedura, apresenta um conjunto de locais-alvo de reconhecimento da recombinase Flipase (Flp), conhecidos como FRT (*Flipase Recognition Target*), sendo atualmente o sistema com maior especificidade de integração, e menor reatividade cruzada dos sítios-alvo de recombinação, em comparação a outras recombinases¹⁴. A construção de variantes das sequências FRT requer que se mantenha intacta as duas repetições invertidas de 13 pb separadas por um espaçador assimétrico de 8 pb. Sua assimetria

determina a direção do alinhamento do local de integração no evento de recombinação, o que, consequentemente, levará a inversão ou excisão⁶⁸.

Outra classe de ferramentas de edição de genes para mutagênese sítio-dirigida ou integração de transgene são as meganucleases⁴². As meganucleases são endonucleases específicas que reconhecem sítios específicos no DNA compreendidos por 12 ou mais pares de bases⁷⁶. As meganucleases têm sido utilizadas para o desenvolvimento de estratégias de integração direcionada de transgenes visando a melhorar a eficiência na obtenção de linhagens celulares com expressão estável⁷⁷.

É importante mencionar que alguns relatos na literatura indicam algumas desvantagens quanto à utilização destas recombinases, como, por exemplo, a possibilidade de ocorrência de aberrações cromossômicas associadas à presença de pseudo-sítios de recombinação, sua eficiência ser limitada à 37°C⁷⁸, ou uma diminuição da frequência de recombinação homóloga a medida que o tamanho da cassette inserido aumenta¹³, o que limita a quantidade de DNA heterólogo que pode ser inserido em uma única integração³⁵.

Conceitualmente, estas estratégias envolvendo a inserção sítio-dirigida do cassette de expressão se mostram muito vantajosas frente às dificuldades associadas à instabilidade clonal decorrente da inserção aleatória do cassette de expressão no genoma, e da instabilidade genômica associada aos protocolos de amplificação gênica. Desta forma, o processo de varredura clonal, necessário para o isolamento de clones com uma expressão estável e em altos níveis da proteína recombinante de interesse, potencialmente se torna mais rápido e eficiente, reduzindo-se os custos e o tempo necessários para o desenvolvimento de uma linhagem celular produtora. Adicionalmente, existe, ainda, a possibilidade de se reutilizar o mesmo sítio genômico para se expressar diferentes proteínas recombinantes, minimizando, desta forma, os tempos para geração e isolamento de novos clones celulares.

O potencial desta alternativa foi explorado para a produção do anticorpo humano policlonal anti-RhD através da integração de 25 cassetes individuais de expressão do anticorpo em um sítio de integração definido, flanqueado pela sequência FRT, através de recombinação homóloga mediada pela recombinase flipase¹²². Este sítio particular faz parte de uma plataforma de expressão baseada em células CHO comercializada pela empresa Invitrogen (Flp-InTM cell line). Este sistema foi utilizado para a geração de células CHO Lec3.2.8.1 com a capacidade de produzir complexos multiproteicos através da integração de cassetes de troca contendo dois

transgenes vinculados por um elemento interno de entrada do ribossomo IRES (*Internal Ribosome Entry Site*) para sua co-expressão¹²².

Mais recentemente, com a finalização do sequenciamento do genoma das células CHO⁴⁹, tornou-se possível a promoção de inserções genômicas sítio-específicas independentemente da inserção prévia de sequências de reconhecimento para recombinação homóloga, como no caso do sistema Flp-FRT. Assim, há descrições de inserções sítio-específicas no genoma bem sucedidas, utilizando-se, como ferramentas de edição genômica, as nucleases efetoras com atividade similares a ativadores transcricionais (TALENs – Transcription Activator-like Effector Nucleases), ou sistema baseado nas nucleases guiadas por RNA-guia, CRISPR-Cas9 (*Clustered Regularly Interspaced Palindromic Repeats-CRISPR associated protein 9*)⁸⁰. O primeiro relato de edição precisa do genoma usando a ferramenta CRISPR/Cas9 em células CHO envolveu a edição genômica de dois genes envolvidos nas vias de O- e N-glicosilação, COSMC (*Core 1 beta3-Gal-T-Specific Molecular Chaperone*) e FUT8 (*Alpha-(1,6)-fucosyltransferase*)⁸¹.

Embora a maior parte dos exemplos descritos na literatura explorem a utilização das metodologias de inserção genômica sítio-específicas para obtenção de clones celulares contendo uma única ou um pequeno número de cópias do vetor de expressão, é possível o mapeamento de sítios propícios para a exploração da amplificação gênica a partir destes sítios únicos de integração. A determinação de sítios-alvo para serem explorados em metodologias empregando a inserção genômica sítio-específica pode ser realizada a partir de regiões genômicas com propriedades previamente caracterizadas e definidas, ou, alternativamente, a partir de varreduras aleatórias por regiões genômicas associadas a uma expressão gênica duradoura e elevada.

Esta estratégia já foi usada para o desenvolvimento de linhagens celulares CHO produtoras de proteínas recombinantes através da técnica de CRISPR/Cas9 baseados em elementos genômicos extensamente caracterizados, como a região telomérica do cromossomo 8 onde se localiza o gene *C12orf35*⁷⁸. Outro exemplo é o locus associado ao gene *Hprt* (Hypoxanthine-guanine PhosphoRibosylTransferase), que foi relatado como um sítio ideal de integração capaz de ser explorado na geração de clones celulares, derivados das células CHO-K1, estáveis e de alta expressão para um fragmento de cadeia única da porção variável de um anticorpo recombinante anti-prion (scFv-Fc - single-chain Fv fused with Fc)⁸³. Ainda outro exemplo, é o locus associado ao gene *Grik1* (*Glutamate receptor, ionotropic, kainate 1*), um sítio descoberto

recentemente por esforços de identificação e redirecionamento de pontos de acesso transcricionais no genoma humano⁸⁴. Outro exemplo bem conhecido é o locus murino conhecido como Rosa26, utilizado amplamente na engenharia genética em camundongos e células humanas por ser um local ótimo para expressão de genes heterólogos⁸⁵.

A indústria biofarmacêutica tem mostrado um crescente interesse em metodologias de geração de clones superprodutores empregando a integração genômica sítio-específica em células CHO visando à produção de proteínas recombinantes terapêuticas. Desta forma, existe uma demanda no sentido de se acumular conhecimento sobre regiões genômicas em linhagens produtoras, como as células CHO, que suportem uma expressão gênica estável e em altos níveis.

1.10 Vetores de expressão lentivirais como ferramenta de varredura de sítios genômicos com atividade transcricional elevada e estável

Os lentivírus constituem uma subclasse dos retrovírus que infecta diversos organismos, e que foram adaptados para serem utilizados como ferramenta de transferência gênica devido à alta eficiência de integração do genoma viral no genoma da célula hospedeira. Após o evento de infecção e internalização da partícula viral, o material genético do lentivírus é convertido em DNA através da atividade de uma enzima codificada pelo genoma viral, a transcriptase reversa, formando então um complexo de pré-integração, constituído pela molécula de DNA viral linear, a integrase e proteínas do envelope viral. Uma característica muito importante quanto ao evento de integração genômica, e que será explorada no presente trabalho, é que os sítios onde ocorrem a integração genômica não são aleatórios, ocorrendo preferencialmente em regiões transcricionalmente ativas localizadas em sítios cromossômicos próximos aos poros nucleares, e tendo a integrase viral um papel fundamental na seleção destes sítios⁸⁶. Aparentemente, esta é uma estratégia do vírus causador da imunodeficiência humana (HIV – *Human Immunodeficiency Virus*), um dos lentivírus a partir do qual se derivou a maioria dos vetores lentivirais, e empregada para a otimização da expressão dos genes provirais, mediante a integração destes genes em unidades de transcrição ativas no genoma da célula hospedeira⁸⁷. Esta estratégia parece ser fundamental no contexto da infecção pelo HIV, já que as células infectadas por este vírus persistem apenas por um ou dois dias, e, sendo assim, seria importante que a integração genômica ocorresse em uma localização ideal para uma rápida produção das partículas virais nesse curto período de tempo⁸⁸. Portanto, essa estratégia evitaria os efeitos

negativos na expressão dos genes virais que por ventura fossem inseridos em certas regiões transcricionalmente inativas do genoma da célula hospedeira⁸⁸.

A alta eficiência de integração genômica do material genético viral nas células hospedeiras, inclusive naquelas que não estão se dividindo, motivou o desenvolvimento de estratégias de transferência gênica empregando os lentivírus como uma alternativa à transfecção⁸⁹. Desta forma, foram desenhados vetores de expressão derivados do genoma do HIV, o que envolveu a deleção ou mutação de regiões patogênicas do vírus, mas mantendo a sua capacidade de transduzir vários tipos celulares, inclusive aqueles difíceis de transfectar por outras metodologias, e de inserir o material genético em regiões de alta atividade transcricional⁹⁰. Apesar destas qualidades, os vetores de expressão lentivirais apresentam algumas desvantagens que restringem a sua utilização para a produção de proteínas recombinantes terapêuticas, como a necessidade de uma contenção de nível 2 de biossegurança no momento da produção e manipulação das partículas lentivirais recombinantes, e a probabilidade de mutação do transgene devido ao processo de transcrição reversa do genoma viral após o evento de transdução⁹¹. Por estas razões, o uso de vetores lentivirais é mais restrito em relação a outros métodos de transferência gênica mais tradicionais, embora existam alguns relatos envolvendo a exploração destes vetores para a geração de linhagens de células recombinantes derivadas das células CHO, BHK (Baby Hamster Kidney) e HEK293⁹².

1.11 A proposta do trabalho

A integração genômica aleatória do vetor de expressão do gene de interesse obtida através do emprego das metodologias clássicas de transfecção celular estável evita prever o sítio cromossômico de inserção, o número de cópias integradas, e o nível de expressão das células geradas⁸⁸. Além disso, o sítio de inserção cromossômica possui um grande impacto sobre a regulação da expressão gênica, podendo causar, em casos extremos, o eventual silenciamento da expressão do gene de interesse⁹³. Adicionalmente, o emprego da metodologia de transfecção estável está associado a uma baixa probabilidade de geração de clones celulares altamente produtores, sendo necessário a utilização de abordagens alternativas para superar esta desvantagem, como o aumento artificial do número de cópias do minigene recombinante através de processos envolvendo o fenômeno de amplificação gênica¹. Entretanto, o próprio emprego da metodologia de amplificação gênica é um fator importante de variabilidade clonal,

estando associada à ocorrência de eventos drásticos de rearranjos genéticos, inclusive podendo levar à perda de cópias do cassete de expressão, ou, ainda, ao surgimento de subpopulações celulares resistentes à ação do MTX e MSX⁹⁴. Conseqüentemente, faz-se necessário a varredura de um grande número de clones após a execução dos protocolos de amplificação gênica, de tal forma que se aumente a probabilidade de se isolar clones que expressem estavelmente quantidades adequadas da proteína terapêutica de interesse. Este é um obstáculo significativo para o estabelecimento rápido de linhagens celulares de alta produção, e que demanda uma melhor compreensão dos sítios genômicos e dos fenótipos celulares associados a uma alta produtividade, o que poderia reduzir substancialmente o tempo de varredura por clones produtores.

Nossa proposta envolve a utilização de um vetor lentiviral bicistrônico de expressão como ferramenta de varredura de sítios genômicos nas células CHO-DG44 associados a uma expressão elevada e estável da proteína de interesse. Adicionalmente, é nosso objetivo, também, a comparação da *performance* desta metodologia de obtenção de clones superprodutores frente à estratégia convencionalmente utilizada, envolvendo a inserção aleatória no genoma do vetor de expressão, através da transfecção estável, seguida de amplificação gênica e isolamento de clones superprodutores. O vetor lentiviral bicistrônico de expressão explorado em ambas as estratégias possui como primeiro cistron um gene repórter, uma forma secretada da proteína fluorescente amarela EYFP (Enhanced Yellow Fluorescent Protein), e como segundo cistron a região codificadora da enzima DHFR. O racional da proposta de varredura utilizando um vetor lentiviral se baseia no fato de várias evidências indicarem que vetores lentivirais baseados no vírus HIV-1 se integram em um subconjunto de genes com alta atividade transcricional no genoma da célula hospedeira⁶⁹. Desta forma, espera-se que o mapeamento de sítios genômicos associados a uma produção elevada e duradoura da proteína recombinante de interesse possa contribuir com o estabelecimento mais eficiente de plataformas de expressão de alta produtividade.

2 OBJETIVOS

2.1 *Objetivo geral*

- Varredura de sítios genômicos com atividade transcricional estável e elevada nas células CHO-DG44 visando o estabelecimento mais eficiente de plataformas de expressão de proteínas recombinantes baseada em células de mamíferos.

2.2 *Objetivos experimentais específicos*

- Construção e caracterização do vetor lentiviral bicistrônico de expressão de uma variante secretada da proteína EYFP (pLV-spEYFP-DHFR);
- Transdução e transfecção estável das células CHO-DG44 com o vetor pLV-spEYFP-DHFR;
- Isolamento e clonagem das células CHO-DG44 superprodutoras da proteína spEYFP por separação de células (“cell sorting”) por citometria de fluxo;
- Amplificação gênica das células CHO-DG44 transfectadas estavelmente com o vetor pLV-spEYFP-DHFR;
- Avaliação do número de cópias do vetor pLV-spEYFP-DHFR integradas no genoma dos clones isolados através da metodologia de PCR em tempo-real;
- Avaliação da estabilidade de expressão da proteína spEYFP ao longo de 40 passagens por fluorimetria do meio condicionado pelos clones isolados;
- Mapeamento do sítio de inserção genômica do vetor pLV-spEYFP-DHFR nos clones superprodutores e estáveis através da metodologia de sequenciamento de DNA por “DNA walking”.

3 MATERIAIS E MÉTODOS

As subseções a seguir descrevem os materiais, os métodos e os equipamentos utilizados para o desenvolvimento das linhagens produtoras de uma variante secretada da proteína EYFP, do isolamento de populações monoclonais e da avaliação dos níveis de produção da proteína recombinante

3.1 *Linhagens celulares e manutenção*

Duas linhagens celulares foram necessárias para a realização desse trabalho, as células CHO-DG44³⁹ para a produção estável de uma variante secretada da proteína EYFP, e as células HEK 293T, derivadas de rim embrionário humano, para a avaliação funcional do vetor de expressão pLV-spEYFP-Dhfr e produção das partículas lentivirais recombinantes a partir deste mesmo vetor, através de experimentos de transfecção transitória⁹⁵.

A linhagem CHO-DG44 é amplamente utilizada para a produção industrial de biofármacos, e foi derivada a partir das células CHO, sendo deficientes para o gene *Dhfr*⁹⁶, o qual é responsável pela produção da enzima dihidrofolato-redutase (DHFR). A linhagem HEK 293T foi derivada de células de rim embrionário humano, posteriormente transformadas com a proteína “large” T de SV40, e originalmente referidas como 293tsA1609neo⁹⁵. Ambas as linhagens estão depositadas no biorrepositório do Núcleo de Apoio à Pesquisa NUCEL (Núcleo de Terapia Molecular e Celular), vinculado à Faculdade de Medicina da USP, no qual o projeto de pesquisa foi desenvolvido.

A linhagem celular CHO-DG44 foi cedida ao NUCEL em 2006 pela Pesquisadora Científica do Instituto Butantan, Doutora Ana Maria Moro. As células HEK293T foram transferidas para o laboratório na década de 90, tendo sido fornecidas pela Profa. Dra. Eugênia Costanzi-Strauss, do Depto. de Biologia Celular, Instituto de Ciências Biomédicas da Universidade de São Paulo – USP.

As células HEK 293T foram mantidas em meio D-MEM suplementado com 10% de soro fetal bovino (SFB). Para o cultivo das células CHO-DG44 foi utilizado o meio HAM's

F12 suplementado com 10% de SFB. As populações celulares derivadas a partir da transfecção estável das células CHO-DG44 com o vetor de expressão pLV-spEYFP-DHFR foram mantidas em meio Alfa-MEM sem nucleosídeos e deoxinucleosídeos (Invitrogen) suplementado com 7% de SFB dialisado.

As células foram mantidas em frascos de cultivo T descartáveis, contendo o meio adequado para cada linhagem, e armazenadas em atmosfera umidificada com 2% de CO₂ à 37°C. As células foram sub-cultivadas quando atingiram 80% da confluência, conforme descrito em Freshney⁹⁷. Os estoques celulares foram mantidos em meio de cultivo contendo 10% de SFB e 10% de DMSO (dimetilsulfóxido) em reservatório contendo nitrogênio líquido à -190°C.

3.2 *Construção e caracterização do vetor lentiviral bicistrônico de expressão de uma variante secretada da proteína EYFP (pLV-spEYFP-DHFR)*

A estratégia adotada para a obtenção do vetor de expressão lentiviral bicistrônico da proteína de fusão entre o peptídeo sinal do Fator VIII humano (sp) (nucleotídeos 172 a 228 da sequência de referência do GenBank NM_000132.3) e a proteína fluorescente EYFP, pLV-spEYFP-DHFR, envolveu dois ciclos de PCR. O primeiro ciclo consistiu na utilização do vetor de expressão de uma forma variante da cadeia leve do Fator VIII humano recombinante pIQID-B-LC⁹⁸ como molde para a síntese do iniciador senso spEYFP (5'CCCGGGATCCACCGGTCCGCCACCATGCAAATAGAGCTCTCCACCTGCTTCTTTC TGTGCCTTTTGCGATTCTGCTTTAGTATGGTGAGCAAGGGCGAGGA3'). Para este primeiro ciclo de PCR, o meio reacional consistiu do vetor pCXNIID-Lch (400 ng), tampão Phusion HF Buffer 1X (Invitrogen), MgCl₂ 2 mM, dNTPs 0,2 mM, iniciadores senso spLCE (5'CCCGGGATCCACCGGTCCGCCACCATGCAAATAGAGCTCTCCAC3') e spLCR antisense (5' TCCTCGCCCTTGCTCACCATACTAAAGCAGAATCGCAAAAAGG3') na concentração de 0,5 µM, Taq polimerase (Invitrogen) 1 U e água milliQ[®] suficiente para um volume final de 50 µl. As condições de ciclagem foram 98°C por 30s, 98°C por 10 s, 55°C por 30 s, 72°C por 30 s, repetição de 35 ciclos, e um período de extensão final à 72°C por 10 minutos.

O fragmento de DNA correspondente ao iniciador senso spEYFP, de aproximadamente 100 pb e contendo a sequência codificadora do peptídeo sinal do Fator VIII humano fusionada ao início da região codificadora da proteína fluorescente EYFP, foi purificado através de uma eletroforese em gel de agarose 2%, utilizando o QIAquick Gel Extraction Kit (Quiagen), seguindo as instruções do fabricante. O iniciador senso spEYFP purificado foi concentrado por um fator de cinco vezes no rotoevaporador a vácuo Speed-Vac[®], e usado no segundo ciclo de PCR para a síntese completa do elemento spEYFP-IRES-DHFR. Para este segundo ciclo de PCR, o meio reacional consistiu do vetor pNU1-EYFP-DHFR (400 ng), tampão Phusion HF Buffer 1X (Invitrogen), MgCl₂ 2 mM, dNTPs 0,2 mM, iniciador senso spEYFP (8 ng), iniciador antisense complementar a um trecho anterior à sequência sinalizadora de poliadenilação do vetor pNU1-EYFP-DHFR (2222R) (5'CCCGTCGACCCAGCCACCACCTTCTGATA3') na concentração de 0,5 μM, Taq polimerase (Invitrogen) 1 U e água milliQ[®] suficiente para um volume final de 20 μl. O vetor bicistrônico de expressão pNU1 foi obtido a partir da inserção de um sítio de múltipla clonagem, e remoção do cassete de resistência ao antibiótico Neomicina (Neo), do elemento IVS (Intervening Sequence) e do cassete de recombinação attP1-attP2 do vetor de expressão pIQID⁹⁸. O vetor pNU1-EYFP foi obtido a partir da subclonagem da sequência codificadora da EYFP, clonada no vetor comercial pEYFP-N1 (Clontech), no vetor pNU1, explorando-se os sítios reconhecidos pelas enzimas de restrição EcoRI e NotI. As condições de ciclagem foram 98°C por 30 s, 98°C por 10 s, 55°C por 30 s, 72°C por 120 s, repetição de 30 ciclos, e um período de extensão final à 72°C por 10 minutos. O produto resultante desta reação de PCR, o fragmento de DNA spEYFP-IRES-DHFR, de aproximadamente 2140 pb, foi purificado a partir de eletroforese em gel de agarose 1%, e concentrado como descrito anteriormente. Na sequência, o fragmento spEYFP-IRES-DHFR foi co-digerido com as enzimas de restrição Sall e BamHI, de acordo com orientação do fabricante (ThermoScientific). Em paralelo, digeriu-se o vetor lentiviral bicistrônico de terceira geração p156RRLsinPPTCMVIRESPRE (pLV-EGFP)(Figura 2)⁹⁹ com as mesmas enzimas para eliminação do elemento IRES-EGFP. O fragmento de DNA correspondente ao vetor pLV foi purificado através de uma eletroforese em gel de agarose 1%, e purificado como descrito anteriormente. A partir do elemento spEYFP-IRES-DHFR e da porção do vetor pLV co-digeridos com as enzimas Sall e BamHI, realizou-se a reação de ligação destes fragmentos, utilizando a enzima T4 DNA Ligase (Invitrogen), seguindo-se as indicações do fabricante. Uma alíquota da reação de ligação foi utilizada para a transformação por eletroporação (cubetas de 2 mm, e um pulso de 2500 V/25 μF/200 Ω -

Eletroporador Gene Pulser XCELL – Biorad) das bactérias *Escherichia coli*, cepa XL10 gold, eletrocompetentes. As bactérias eletroporadas foram inoculadas em meio LB à 37°C sob agitação por uma hora, plaqueadas em meio LB-Agar contendo ampicilina (100 µg/ml), e incubadas por 16 horas à 37°C, o que permitiu a seleção de clones bacterianos que receberam o vetor de expressão. A varredura das colônias bacterianas recombinantes foi realizada por PCR a partir dos inóculos bacterianos, empregando-se três pares distintos de iniciadores: a) spLC F (5'TACCGCGGGCCCGGGATCCAGTCGACCACCATGCAAATAG3'/ 2222R), com amplicon esperado de 2140 pb; b) pNU1 seq5F (5'AGTTGTGGAAAGAGTCAAATGG3')/ WPRE R (5'GCCACAACCTCCTCATAAAGAGACAG3'), com amplicon esperado de aproximadamente 1 Kpb, pNU1 seq6F (5'GAACTCAAAGAACCACCACG3')/ WPRE R, com amplicon esperado de aproximadamente 550 pb. Por meio desta estratégia de varredura, os clones bacterianos associados a detecção dos fragmentos de DNA esperados nas três reações de PCR foram inoculados em meio LB+Ampicilina (100 µg/ml) à 37°C por 16 horas, para posterior recuperação e purificação dos plasmídeos recombinantes de interesse, através de uma minipreparação de DNA plasmidial, usando o Plasmid Miniprep Kit (Qiagen) segundo orientações do fabricante. A partir das minipreparações de DNA plasmidial obtidas, avaliou-se a integridade dos vetores isolados através da reação de digestão pelas enzimas XhoI (fragmentos de DNA esperados de 820 pb e 7,9 Kpb), BamHI/SalI (fragmentos de DNA esperados 2140 pb e 6,6 Kpb), PstI (fragmentos esperados de 300 pb e 8,6 Kpb) e EcorV (fragmentos esperados de 830 pb e 7,9 Kpb), sendo que somente aqueles clones bacterianos com o perfil de restrição esperado foram avaliados por sequenciamento de DNA. As análises da integridade da sequência do elemento spEYFP-IRES-DHFR foram realizadas por sequenciamento de DNA (método de Sanger) através da utilização do kit BigDye (ThermoScientific), empregando-se os iniciadores flanqueando o sítio de clonagem e segundo as orientações do fabricante. As reações de sequenciamento foram analisadas em um sequenciador de DNA modelo ABI 3700, no laboratório GATELab, do Instituto de Biologia da USP. As montagens das sequências consenso foram realizadas através do programa Seqman (DNASTAR), e os alinhamentos com a sequência teórica do trecho spEYFP-IRES-DHFR foram realizados através da ferramenta BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

3.3 *Transfecção transitória das células HEK 293T com o vetor pLV-spEYFP-DHFR*

A funcionalidade do cassete de expressão do vetor pLV-spEYFP-DHFR purificado de cada um dos clones bacterianos isolados conforme descrito no item anterior foi avaliada através de experimentos de transfecção transitória em células HEK 293T. Paralelamente, foi avaliado se a proteína spEYFP estava sendo de fato secretada. Os experimentos de transfecção transitória das células HEK 293T foram realizados mediante a técnica de lipofecção, e para tanto utilizou-se o reagente Lipofectamina 2000[®], Invitrogen, conforme a orientação do fabricante. A funcionalidade do cassete de expressão do vetor foi avaliada na monocamada celular 48 horas após o momento da transfecção, através da observação da fluorescência das células em microscópio de fluorescência Nikon, modelo TE-300, com o conjunto de filtros de excitação/emissão e espelho dicróico para o fluoróforo isotiocianato de fluoresceína (FITC) (excitação 460 nm – 510 nm e emissão 515 nm – 565 nm). A confirmação da secreção da proteína spEYFP foi realizada através da análise por fluorimetria do meio condicionado por 48 horas após o momento da transfecção no aparelho SpectraMax Paradigm Multi-Mode Microplate, utilizando-se o meio condicionado por 48 h pelas células HEK 293T transfectadas com o vetor pLV-EGFP como controle negativo. Para análise por fluorimetria, utilizou-se os comprimentos de onda de excitação em 505 nm, e de emissão em 530 nm. Alternativamente, utilizaram-se as amostras de meio de cultivo condicionado para avaliação da secreção da proteína spEYFP através de eletroforese em gel de poliacrilamida contendo SDS (SDS-PAGE 10%), seguida de escaneamento à laser do gel para detecção de fluorescência. Para tanto, primeiramente as amostras de meio condicionado foram concentradas 10X em colunas de Ultrafiltração Amicon Ultra de 10 K, seguindo as indicações do fabricante. Posteriormente, as amostras concentradas foram fracionadas por SDS-PAGE 10% em condições não redutoras (ausência de β -Mercaptoetanol) e na ausência de desnaturação térmica. Ao término da eletroforese, os géis foram escaneados a laser no aparelho Typhoon Trio, utilizando como comprimento de onda de excitação um laser de 532 nm. Posteriormente, os géis foram fixados e corados com Coomassie blue G-250 coloidal.

3.4 Produção das partículas lentivirais recombinantes em células HEK 293T

Para a produção das partículas lentivirais recombinantes, aproximadamente 1×10^6 células HEK 293T foram transfectadas transitoriamente pelo método de lipofecção, segundo orientações do fabricante, (Lipofectamine 2000 - Invitrogen), com uma mistura dos DNAs plasmidiais correspondentes ao vetor de transferência (pLV-spEYFP-DHFR), e aos vetores de expressão de elementos estruturais do lentivírus (pLP1/pMDL, RevpLP2/pREV e pVSVg) na proporção indicada em Tiscornia *et al.*¹⁰⁰. Os meios de cultura condicionados contendo as partículas lentivirais recombinantes foram coletados de 24 em 24 horas até 72 h após a transfecção, clarificados por centrifugação ($10.000 \times g$ por 10 min à 4°C), e estocados à -70°C . A titulação das preparações lentivirais foi realizada através da transdução das células HEK 293-T, conforme descrito em Tiscornia *et al.*¹⁰⁰. Cerca de 48 h após a transdução viral, avaliou-se as monocamadas por microscopia de fluorescência, e os títulos virais foram calculados a partir da determinação da porcentagem de células fluorescentes com auxílio do programa ImageJ a partir de fotos tiradas de 10 campos aleatórios da monocamada celular, sendo os valores representados a média desses 10 campos, conforme descrito por Tiscornia e colaboradores¹⁰⁰.

3.5 Transdução estável das células CHO-DG44 cultivadas em aderência com as partículas virais recombinantes

Para a transdução das células CHO-DG44 com o vetor pLV-spEYFP-IRES-DHFR, cerca de 2×10^4 células suspensas em meio HAM's F12 suplementado com 10% de SFB foram transferidas para tubos de microcentrífuga de modo que o volume final em cada tubo fosse de 200 μl . Com as células ainda em suspensão, foi adicionado um volume do meio condicionado por 48 horas contendo as partículas lentivirais, de tal forma que se obteve uma multiplicidade de infecção (MOI) de 0,5. Após a adição dos volumes de meio condicionado contendo as partículas virais, as células foram plaqueadas em poços de uma placa de 48 poços. No dia seguinte, após lavagem da monocamada com PBSA, foi feita a troca de meio para o meio de cultivo de seleção (α -MEM contendo 7% SFB dialisado) baseado na expressão da DHFR. O meio de cultivo foi renovado periodicamente, e a população celular mista resultante (CHO-DG44 spEYFP-DHFR SV) foi mantida em cultura até que atingisse um número de células suficiente (cerca de 4×10^6 células) para congelamento. Cerca de uma semana após a transdução, as células foram observadas ao microscópio de fluorescência Nikon modelo TE-300, com o

conjunto de filtros e espelho dicróico para o fluoróforo isotiocianato de fluoresceína (FITC) (excitação 460 nm – 510 nm e emissão 515 nm – 565 nm) para se estimar a porcentagem de células transduzidas.

3.6 *Transfecção estável das células CHO-DG44 com o vetor pLV-spEYFP-DHFR*

Para a transfecção estável das células CHO-DG44 com o vetor pLV-spEYFP-DHFR, plaqueou-se cerca de 1×10^5 células em um poço de placa de 24 poços em meio HAM's F12 suplementado com 10% de SFB. A transfecção foi realizada por lipofecção (Lipofectamina 2000[®] - Invitrogen), segundo orientações do fabricante, empregando-se 500 ng do vetor plasmidial. Cerca de 48 horas após a transfecção, a monocamada celular foi tripsinizada, as células foram subdivididas para quatro poços de uma placa de seis poços, e realizou-se a troca de meio para o meio de cultivo de seleção (α -MEM contendo 7% SFB dialisado) baseado na expressão da DHFR. O meio de cultivo foi renovado periodicamente, e a população celular mista resultante (CHO-DG44 spEYFP-DHFR ST) esta foi mantida em cultura até que atingisse um número de células suficiente (cerca de 4×10^6 células) para congelamento. A população mista foi observada ao microscópio de fluorescência Nikon modelo TE-300, com o conjunto de filtros e espelho dicróico para o fluoróforo isotiocianato de fluoresceína (FITC) (excitação 460 nm – 510 nm e emissão 515 nm – 565 nm) para se estimar a porcentagem de células positivas para a expressão da spEYFP.

3.7 *Enriquecimento das populações celulares mistas das células CHO-DG44 produtoras da proteína spEYFP através de separação de células (“cell sorting”) por citometria de fluxo*

O enriquecimento celular a partir das populações celulares mistas CHO-DG44 spEYFP-DHFR ST e CHO-DG44 spEYFP-DHFR SV foi realizado através da separação de células ativada por fluorescência. Resumidamente, culturas confluentes das células CHO-DG44 spEYFP-DHFR, mantidas em garrafas T25 em meio α -MEM contendo 7% SFB dialisado, foram tripsinizadas e ressuspensas em meio regular de cultivo (volume final 2,5 ml, equivalente a cerca de 5×10^5 células/ml) suplementado com cinco vezes a concentração usual de ampicilina (500 μ g/ml) e estreptomicina (500 μ g/ml). O enriquecimento das populações celulares mistas para as células mais fluorescentes foi realizado no equipamento FACS ARIA-

Ilu Special Order System (BD Biosciences), utilizando o bico injetor ou “nozzle” de 70 μm , e o laser de 488 nm. Cerca de 10 % das células, proveniente das populações celulares mistas iniciais que apresentaram as maiores intensidades de fluorescência, foram coletadas em 5 mL de meio regular de cultivo suplementado com cinco vezes a concentração usual de ampicilina e estreptomicina, e retornadas para a cultura. O meio de cultivo foi renovado periodicamente, e as populações celulares mistas enriquecidas resultantes (CHO-DG44 spEYFP-DHFR ST e CHO-DG44-spEYFP-DHFR SV) foram mantidas em cultura até que atingissem um número suficiente (cerca de 4×10^6 células) para congelamento.

3.8 Isolamento de clones a partir das células CHO-DG44 superprodutoras da proteína spEYFP por diluição limitante

O isolamento dos clones superprodutores foi feito através do experimento de clonagem por diluição limitante em placas de 96 poços. A partir de suspensões das populações enriquecidas CHO-DG44 spEYFP-DHFR ST e CHO-DG44 spEYFP-DHFR SV com uma concentração celular total acima de 1×10^6 células/mL e uma porcentagem de viabilidade maior que 90%, avaliada através da utilização do corante Trypan Blue a 0,4% m/v, realizou-se diluições seriadas sucessivas até a obtenção de uma concentração final de 3,75 células viáveis/mL. Plaqueou-se 200 μL desta suspensão em cada poço da placa de 96 poços, de tal forma que se obtivesse 0,75 células/poço. Após o plaqueamento, cada um dos poços foi inspecionado no microscópio óptico para confirmação de quais poços realmente continham uma única célula por poço. As células foram mantidas em meio regular de cultivo até que se atingisse uma confluência de 80 % no poço, para então, se proceder com a transferência para poços de placas de 24 poços e, posteriormente, quando as monocamadas atingiram de novo essa porcentagem de confluência, as células foram transferidas para uma garrafa T25. Os clones celulares foram mantidos até se atingir um número suficiente (cerca de 4×10^6 células) para congelamento. Cada um destes clones isolados foi avaliado quanto aos níveis de produção da proteína EYFP através da análise do meio de cultivo condicionado coletado, previamente concentrado 10X por ultrafiltração em microcolunas Nikon de corte de 10KDa, por SDS-PAGE 10%, seguida de escaneamento do gel para detecção de fluorescência, conforme descrito anteriormente.

3.9 Seleção das células CHO-DG44 transfectadas estavelmente com o vetor pLV-spEYFP-DHFR com MTX

O clone celular ST1, isolado a partir da população mista CHO DG44 spEYFP-DHFR ST, foi submetido à seleção com o agente MTX. . O processo de seleção foi iniciado a partir de uma monocamada confluyente das células ST1 mantidas em garrafa de cultivo T75. Neste momento, adicionou-se ao meio de cultura 30 nM de MTX, e renovou-se o meio de cultura contendo MTX a cada 3 dias. Após a formação de pequenos agrupamentos de células, estas foram tripsinizadas, e transferidas para garrafas T25. Considerou-se as células resistentes ao MTX após três subcultivos consecutivos na presença do inibidor, obtendo-se, desta forma, as células ST1 30nM MTX.

3.10 Extração de DNA genômico (gDNA)

Para extração do gDNA das diferentes populações e clones celulares gerados, monocamadas celulares confluentes mantidos em garrafas de cultivo T75 foram tripsinizadas e ressuspensas em meio de cultivo regular, e as suspensões celulares resultantes foram posteriormente centrifugadas à 200xg por 5 min para a obtenção dos sedimentos celulares. Sedimentos celulares correspondentes a cerca de 5×10^6 células foram utilizados para a extração do gDNA, utilizando o kit Illustra tissue and cells genomic prep Mini Spin mini 250 kit (GE Healthcare), seguindo as orientações do fabricante. A avaliação da pureza e a quantificação das amostras de gDNA foram realizados por espectrofotometria nos comprimentos de onda de 260 nm e 280 nm, empregando-se o equipamento NanoDrop. As amostras de gDNA obtidas foram armazenadas à -20°C até o momento das análises.

3.11 Determinação, através de PCR em tempo-real (qPCR), do número de cópias do elemento spEYFP-IRES-DHFR integradas no genoma dos clones celulares gerados

Realizou-se, através de qPCR, a quantificação absoluta do número de cópias do elemento spEYFP-IRES-DHFR integradas no genoma de cada um dos clones isolados. Para a quantificação absoluta do número de cópias integradas, obteve-se uma curva-padrão empregando-se diluições sucessivas (fator de diluição de 5) do vetor pLV-spEYFP-DHFR,

compreendendo o intervalo de $1,5 \times 10^7$ a $4,8 \times 10^3$ cópias do vetor. As amostras de gDNA obtidas foram diluídas em água milliQ (fator de diluição de 2), para concentrações finais na faixa de 60 ng a 7,5 ng. Empregou-se 1,5 μ l das diluições das amostras de gDNA, que foram transferidas para placas apropriadas para a realização da qPCR (MicroAmp Fast Optical 96-Well Reaction Plate with barcode), juntamente com 3,0 μ l de Fast SYBR[®] Green PCR Master Mix (Applied Biosystems), e os oligonucleotídeos iniciadores do gene de interesse na concentração final de 0,1 μ M, em um volume final de 6,0 μ l. Cada amostra foi analisada em triplicata. Foram utilizados oligonucleotídeos iniciadores específicos para a região codificadora da DHFR (DHFR “*forward*”: 5’ TTGAACTGCATCGTCGCCGTGTCCCAA 3’ e “*reverse*”: 5’ TCGTTCCTGAGCGGAGGCCAGGGTAG 3’). Como controle do procedimento de avaliação do número de cópias integradas no genoma por qPCR, determinou-se o número de cópias de um gene endógeno das células CHO-DG44, o *Glucagon (Gcg)*. Dados de sequenciamento genômico das células CHO-DG44 indicam que o *Gcg* é diploide nestas células¹⁰¹. Assim, foi avaliado o número de cópias do gene *Gcg* na linhagem parental CHO DG44 seguindo a mesma metodologia descrita acima. Para a obtenção da curva-padrão, empregou-se o vetor pUC18-gluc que contém um trecho de 256 pb do *Gcg* de hamster (*Cricetulus griseus*) correspondendo aos nucleotídeos 348 a 603 da entrada XM_027440922.1 do Genbank. Foram utilizados os oligonucleotídeos iniciadores específicos Gluc forward (5’ ATTGCCAAACGCCACGAT 3’) e “reverse” (5’ CCAAGCAATGAATTCCTTTGC 3’). As reações de qPCR foram realizadas no equipamento ViiA[™] 7 Real Time PCR System nas condições-padrão de ativação da enzima AmpliTaq Fast DNA Polymerase à 95°C por 20 segundos, e 40 ciclos de desnaturação (95°C por 3 segundos) e anelamento e extensão (60°C por 20 segundos). Os valores de ciclo limiar (Ct – Cycle threshold) foram determinados de forma automática pelo equipamento, e foram utilizados para a obtenção das curvas-padrão do número de cópias dos vetores pLV-spEYFP-DHFR e pUC18-Gluc, correlacionando os valores médios de Ct e o número de cópias em escala logarítmica. A eficiência de amplificação foi determinada para cada curva-padrão, e apenas valores entre 90% e 110%, com um valor de $R^2 > 0.90$ foram considerados. O número de cópias do elemento spEYFP-IRES-DHFR e do gene *Gcg* foram determinados usando as curvas-padrão obtidas e assumindo-se que o conteúdo de gDNA de uma célula CHO na fase G0 do ciclo celular é 5,4pg¹⁰²

3.12 Avaliação dos níveis de secreção da proteína spEYFP

A determinação dos níveis de secreção da proteína spEYFP foi realizada em cada um dos clones produtores estáveis isolados por ambas metodologias de transferência gênica exploradas. Avaliou-se a quantidade acumulada de spEYFP nos meios de cultivo condicionados por 72h pelos clones celulares avaliados. A coleta das amostras de meio condicionado foi realizada a partir de cultivos em placas de seis poços, sendo que foram plaqueadas 1×10^5 células/ml em um volume final de 2 ml, e as monocamadas foram incubadas a 37°C em um ambiente umidificado e atmosfera contendo 2% de CO₂. Após 72 h, os meios condicionados foram coletados, clarificados por centrifugação a 200xg por 5 min, e armazenados a -20°C até o momento das análises. Para a avaliação dos níveis de spEYFP, primeiramente as amostras de meio condicionado foram concentradas por um fator de 10X por ultrafiltração em microcolunas Amicon de corte de 10KDa (10KDa Amicon Ultra Ultrafiltration columns). Na sequência, as amostras de meio condicionado concentradas foram analisadas por SDS-PAGE 10%, seguida de escaneamento do gel para detecção de fluorescência, conforme descrito anteriormente (Item 4.3). Todos estes experimentos foram feitos em triplicata. A partir das imagens digitalizadas obtidas, realizou-se a densitometria das duas bandas proteicas correspondentes à spEYFP, com auxílio do programa ImageQuant (ImageQuant software release 7 - Molecular Devices/GE Healthcare). Para a comparação relativa dos níveis de secreção da spEYFP entre os diferentes clones avaliados, os valores de volumes das bandas proteicas obtidos foram normalizados pelo número de células viáveis determinado no momento da coleta do meio condicionado.

3.13 Avaliação da estabilidade dos níveis de secreção da proteína spEYFP

Avaliou-se a estabilidade dos níveis de secreção da proteína spEYFP ao longo de aproximadamente 40 gerações (cerca de dois meses) em cultivo para cada um dos clones produtores isolados a partir de ambas estratégias de transferência genica exploradas. Plaqueou-se 5×10^4 células/ml, com uma viabilidade superior a 90%, em poços de placas de seis poços em um volume final de 2 ml. As monocamadas foram incubadas à 37°C em um ambiente umidificado e atmosfera contendo 2% de CO₂. Os subcultivos foram realizados em intervalos de 72 h, e imediatamente antes do subcultivo, o meio condicionado por 72 h foi coletado ou descartado, e as células foram removidas da superfície de cultivo por tripsinização. Determinou-

se o número de células viáveis, e realizou-se um novo plaqueamento nas mesmas condições descritas acima. A coleta de amostras de meio condicionado para análise dos níveis de expressão da spEYFP foi feita em intervalos aproximados de 10 gerações. As amostras de meio condicionado para análise dos níveis de spEYFP foram clarificadas por centrifugação (200xg, 5 min), e armazenadas à -20°C até o momento da análise. Este procedimento foi realizado sucessivamente até se atingir aproximadamente 40 gerações. O número de gerações foi determinado empregando-se a seguinte fórmula:

$$\text{Número de gerações} = \frac{\ln\left[\frac{\text{número de células finais}}{\text{número de células iniciais}}\right]}{\ln 2}$$

As quantidades relativas da spEYFP foram determinadas em duplicatas através da análise das amostras coletadas do meio de cultivo condicionado por 72 h, previamente concentradas 10X por ultrafiltração em microcolunas Amicon, de corte de 10KDa. As amostras de meio condicionado concentradas foram fracionadas por SDS-PAGE 10%, seguida de escaneamento do gel para detecção de fluorescência, e análise densitométrica das imagens digitalizadas, conforme descrito nos itens 4.3 e 4.12. Para a determinação da estabilidade dos níveis de secreção da spEYFP de cada clone analisado, obteve-se uma curva de regressão linear dos valores de produtividade relativa obtidos ao longo das 40 gerações. O critério de estabilidade dos níveis de secreção da spEYFP utilizado foi a obtenção de inclinações da curva de regressão inferiores a 30%.

3.14 Determinação do sitio de integração baseado na estratégia de “DNA walking”

Os sítios genômicos de integração do vetor lentiviral pLV-spEYFP-DHFR foram identificados pela estratégia de “DNA walking”,

3.14.1 DNA walking

A identificação de sequências nucleotídicas desconhecidas que flanqueiam as regiões de DNA já caracterizadas pode ser realizada por vários métodos diferentes baseados em PCR, comumente conhecidos como DNA Walking.¹¹⁹ Esta metodologia consiste no uso de uma sequência conhecida para o desenho de um dos iniciadores, que se anela em uma região próxima

ao sítio de inserção genômica. As amostras de gDNA são submetidas a digestão com enzimas de restrição e posteriormente se faz a ligação com oligonucleotídeos sintéticos, chamados de adaptadores, nas extremidades dos fragmentos de DNA digeridos. Posteriormente, realizam-se dois ciclos de amplificação por PCR de um fragmento de DNA da sequência desconhecida, empregando-se o iniciador da sequência conhecida, e um iniciador que se liga à sequência do adaptador.¹²³

3.14.2 Geração das bibliotecas de gDNA

As amostras de gDNA purificado de cada um dos clones produtores isolados a partir da estratégia de transferência gênica empregando vetores lentivirais foram fragmentadas através da digestão com enzimas de restrição para a formação das bibliotecas de gDNA, os fragmentos gerados foram ligados à uma sequência adaptadora específica, os quais foram, posteriormente, submetidos à PCR (configuração “*nested*” PCR), conforme detalhado a seguir.

Para a geração das bibliotecas de fragmentos de gDNA representativas de cada um dos clones produtores avaliados, foi feita a digestão das amostras de gDNA com as enzimas de restrição SspI, EcoRV e HpaI. As reações de digestão consistiram de 2,5 µg de gDNA, 1U de cada uma das enzimas (ThermoScientific), 2X do tampão Tango (ThermoScientific), em um volume final de 50 µl. As reações foram inicialmente incubadas por 2h à 37°C, homogeneizadas no vórtex, e incubadas por mais 16 h à 37°C. As enzimas foram inativadas à 65°C por 10 minutos. Por fim, as reações de digestão foram concentradas por um fator de cinco vezes no rotoevaporador a vácuo Speed-Vac®.

3.14.3 Reação de ligação da sequência adaptadora às bibliotecas de gDNA

Inicialmente, procedeu-se com o preparo do oligonucleotídeo dupla-fita utilizado como adaptador, que consistiu no anelamento de dois oligonucleotídeos sintéticos, caracterizados pelas seguintes sequências: GenomaW_A 5' GTAATACGACTCACTATAGGGCACGCGTGGTTCGACGGCCCGGGCTGGT 3' e Gw_A 5' PO₄-ACCAGCCC-N₂H 3'. A reação de anelamento consistiu de 40 µM de cada um dos

oligonucleotídeos acima, 1X do tampão T4 Ligation Buffer (Fermentas / Thermo Fischer) e 1U de T4 Polinucleotide Kinase (Fermentas / Thermo Fischer), a qual foi incubada inicialmente a 37°C por 30 minutos, seguida por incubação a 95°C por 5 minutos, e redução da temperatura para 25°C, com a taxa de decréscimo de 5°C/min. O adaptador foi ligado aos fragmentos de gDNA, através da incubação de 3 µl do adaptador (12,5 µM), 5 µl das bibliotecas de fragmentos de gDNA, 1 µl do tampão de ligação 10X (Fermentas/Thermo Fischer), e 1µl da enzima T4 DNA ligase (10 U/µl, Fermentas/Thermo Fischer) à 16°C por cerca de 16 h. Ao término da incubação, as reações foram interrompidas através de incubação à 70°C por 5 min.

3.14.4 “Nested” PCR

Os sítios genômicos de integração do vetor pLV-spEYFP-DHFR foram amplificados através de dois ciclos sucessivos de PCR (“*nested*” PCR). No primeiro ciclo, o meio reacional consistiu de 1 µl do meio reacional de ligação da sequência adaptadora aos fragmentos de gDNA, tampão Taq DNA Polymerase PCR Buffer 1X (Invitrogen), MgCl₂ 2 mM, dNTPs 0,2 mM, iniciador senso LSP1 (5’GCTTCAGCAAGCCGAGTCCTGCGTTCGAG3’) e iniciador antisenso AP1 (5’GTAATACGACTCACTATAGGGC3’) na concentração de 0,5 µM, Taq polimerase (Invitrogen) 1U e água milliQ[®] suficiente para um volume final de 20µl. As condições de ciclagem foram 95°C por 5 minutos, 95°C por 30 s, 57,5°C por 30 s, 72°C por 5 minutos, repetição de 35 ciclos, e um período de extensão final à 72°C por 10 minutos. Para o segundo ciclo da “*nested*” PCR foi usado 1 µl do meio reacional do primeiro ciclo, diluído previamente por um fator de 10 em água deionizada, o mesmo meio reacional descrito para o primeiro ciclo, alterando-se os iniciadores pelo iniciador senso LSP2 (5’GCTCCTCTGGTTTCCCTTTCGCTTTCAA3’) e o iniciador antisenso AP2 (5’ACTATAGGGCACGCGTGGT3’). As condições de ciclagem foram as mesmas utilizadas para o primeiro ciclo da “*nested*” PCR. Os amplicons obtidos foram analisados através de uma eletroforese em gel de agarose 1,5% m/v, e foram purificados utilizando o kit QIAquick Gel Extraction (Quiagen), seguindo as instruções do fabricante. Os fragmentos de DNA purificados foram analisados por sequenciamento de DNA (método de Sanger) através da utilização do kit BigDye (ThermoScientific), empregando-se os iniciadores LSP2 e AP2, segundo as orientações do fabricante. As reações de sequenciamento foram analisadas em um sequenciador de DNA modelo ABI 3700, no laboratório GATELab, do Instituto de Biologia da USP. Os dados

resultantes do sequenciamento de DNA foram analisados para determinar o sítio genômico de inserção do vetor lentiviral. A porção da sequência que compõe a sequência genômica flanqueando o sítio de inserção foi alinhada contra sequências genômicas das células CHO, usando-se a ferramenta BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

3.15 Análise estatística

A significância estatística dos resultados foi obtida através de análise de variância (Anova) de uma via, e posteriormente o teste Tukey de múltipla comparação, utilizando-se o software Prism (GraphPad Software, Inc). Os dados serão considerados significantes estatisticamente em $P < 0,05$.

4 RESULTADOS

O principal objetivo deste trabalho envolveu o reconhecimento de sítios genômicos associados a níveis de expressão altos e estáveis de proteínas recombinantes nas células CHO-DG44. Nossa abordagem experimental envolveu: a) o emprego de um vetor lentiviral bicistrônico de expressão, pLV-spEYFP-DHFR (Figura 3), de forma a explorar o tropismo natural do provírus derivado dos lentivírus para a inserção em sítios genômicos com alta atividade transcricional da célula hospedeira⁸⁷; e b) o uso de uma forma variante secretada da EYFP (spEYFP) como proteína repórter, evitando um possível limite ou teto nos níveis de expressão da proteína repórter relacionado à toxicidade celular atribuída ao acúmulo citoplasmático de proteínas fluorescentes, e permitindo a possibilidade de seleção de subfenótipos celulares com uma alta capacidade de secreção¹⁰³.



Figura 3: Representação esquemática dos principais elementos do vetor lentiviral bicistrônico de expressão pLV-spEYFP-DHFR. Uma forma variante secretada da EYFP (spEYFP) foi obtida a partir da fusão entre as sequências codificadoras do peptídeo sinal do Fator VIII humano da coagulação (spFVIII) e da EYFP. LTR: Repetição Terminal Longa derivada do vírus HIV-1; CMV: Região promotora derivada do Citomegalovírus humano; IRES: Sítio interno de entrada do ribossomo; Dhfr: região codificadora da enzima DHFR de camundongo.

4.1 Construção e caracterização do vetor lentiviral bicistrônico de expressão de uma variante secretada da proteína EYFP (pLV-spEYFP-DHFR)

A construção do vetor de expressão lentiviral bicistrônico da proteína de fusão entre o peptídeo sinal do Fator VIII (sp) e a proteína fluorescente EYFP, envolveu dois ciclos de PCR, conforme detalhado na sessão 4.2 de Materiais e Métodos. O produto obtido a partir da segunda reação de PCR, spEYFP-IRES-DHFR, foi purificado a partir de eletroforese em gel de agarose, posteriormente co-digerido com as enzimas de restrição SalI e BamHI, e ligado ao vetor lentiviral bicistrônico pLV-EGFP digerido com as mesmas enzimas. Uma alíquota da reação de ligação foi utilizada para a transformação por eletroporação da cepa XL10 Gold das bactérias *Escherichia coli*. As colônias recombinantes obtidas foram inoculadas em poços de uma placa de 96 poços contendo meio LB e ampicilina na concentração de 100 µg/ml, e a varredura quanto a obtenção do vetor pLV-spEYFP-DHFR foi realizada por PCR, empregando-se os pares de

iniciadores spLC F/ 2222R, pNU1 seq5F/WPRE R, e pNU1 seq6F/WPRE R, conforme descrito na seção 4.2 de Materiais e Métodos. Todos os clones bacterianos (P3C, P3E, P8D, C6H, C1B e C4C) associados à detecção dos fragmentos de DNA esperados nas três reações de PCR mencionadas acima foram selecionados para a extração do DNA plasmidial, e determinação dos respectivos perfis de restrição com as enzimas XhoI, BamHI/SalI, PstI e EcoRV para avaliação da integridade dos vetores isolados (Figura 4).

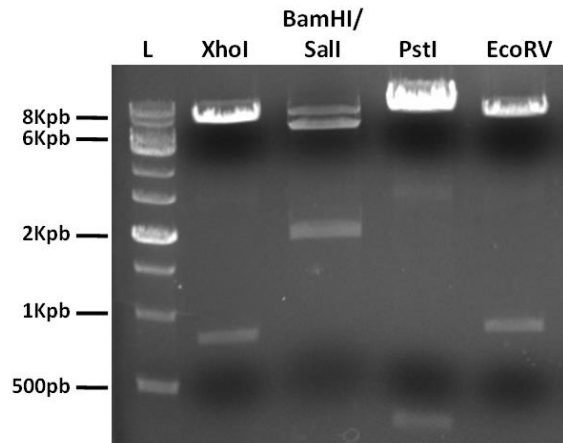


Figura 4: Análise por eletroforese em gel de agarose 1% m/v do perfil de restrição do clone C6H do vetor recombinante pLV-spEYFP-DHFR com as enzimas de restrição indicadas. Os fragmentos de DNA esperados eram 7,8 Kpb e 800 pb (XhoI), 6,3 Kpb e 2,3 Kpb (BamHI/SalI), 8,4 Kpb e 200 pb (PstI) e 7,9 Kpb e 900 pb (EcoRV). L: Padrão de peso molecular.

A Figura 4 representa a análise do perfil de restrição de um dos clones obtidos do vetor pLV-spEYFP-DHFR, o C6H. Como pode ser observado, o perfil de fragmentos de DNA obtido está de acordo com àquele esperado, indicando a integridade estrutural do vetor de expressão pLV-spEYFP-DHFR (clone C6H). Da mesma forma, foram avaliados os outros clones bacterianos (clones P3C, P3E, P8D, C1B e C4C), os quais também apresentaram o perfil de restrição esperado (dados não mostrados). Todos os clones positivos confirmados por PCR e perfil de restrição foram sequenciados com o objetivo de se avaliar a identidade e a integridade das regiões codificadoras da proteína de fusão spEYFP e DHFR, e do elemento IRES, conforme descrito na seção 4.2 de Materiais e Métodos. Os alinhamentos entre as sequências de DNA consenso dos clones P3C, P3E, P8D, C6H, C1B e C4B e a sequência teórica do elemento spEYFP-IRES-DHFR, realizados através da ferramenta BLAST, mostraram, na maioria dos clones avaliados, deleções de diferentes trechos do cassete de expressão, como, por exemplo, a deleção de um segmento da extremidade 3' da região codificadora da spEYFP no clone P8D, ou na extremidade 5' da região codificadora da DHFR no clone C1B (dados não mostrados). De todos os clones avaliados, apenas a sequência consenso do clone P3C apresentou similaridade

completa com a sequência teórica no trecho avaliado (dado não mostrado), complementando os dados obtidos na caracterização parcial por PCR e pelo perfil de restrição. Seguiu-se com a confirmação das funcionalidades de elementos-chave do vetor pLV-spEYFP-DHFR (clone P3C), a saber: a) o cassete de expressão do elemento bicistrônico spEYFP-DHFR; b) a secreção do repórter spEYFP; e c) a produção de partículas lentivirais recombinantes funcionais.

4.2 *Teste de expressão e secreção da proteína spEYFP*

Com o objetivo de se avaliar a expressão e secreção da proteína spEYFP, realizaram-se experimentos de transfecção transitória das células HEK-293T, pelo método de lipofecção, com o clone P3C do vetor pLV-spEYFP-DHFR (Figuras 5 e 6).

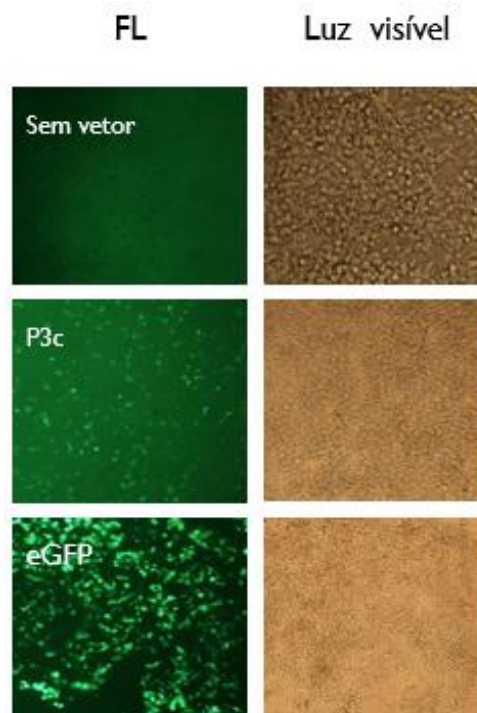


Figura 5: Avaliação da expressão da proteína spEYFP por microscopia de fluorescência. Células HEK-293T foram transfectadas transitoriamente por lipofecção com o vetor pLV-spEYFP-DHFR (clone P3C), e avaliadas por microscopia de campo claro (luz visível) e fluorescência (FL) 48 horas após a transfecção. Como controle positivo da detecção da fluorescência da proteína fluorescente EYFP, transfectou-se as células HEK-293T com o vetor de expressão plasmidial pLV-EGFP. Como controle negativo, empregou-se as células HEK-293T não submetidas ao procedimento de transfecção.

Na avaliação das monocamadas celulares por microscopia de fluorescência (Figura 5, FL), pode-se observar que o clone P3C do vetor pLV-spEYFP-DHFR é funcional quanto à expressão da variante secretável da proteína fluorescente EYFP (spEYFP), em experimentos de transfecção transitória das células HEK-293T. Neste experimento detectou-se a emissão de fluorescência intracelular da spEYFP, possivelmente em trânsito pela via secretória. As células HEK-293T não submetidas ao procedimento de transfecção não apresentaram emissão de fluorescência nas condições de detecção utilizadas, demonstrando que este tipo celular não possui componentes celulares que emitem fluorescência nas condições de detecção empregadas. Nas células HEK-293T transfectadas com o vetor de expressão pNU1-EYFP, utilizadas como controle positivo, observou-se a presença de fluorescência indicativa de acúmulo intracelular da proteína EYFP, como o esperado.

Seguiu-se com a confirmação da secreção da spEYFP por fluorimetria, conforme descrito na seção 4.3 de Materiais e Métodos. Entretanto, não houve detecção de fluorescência nos sobrenadantes dos meios de cultura coletados após 48 horas da transfecção transitória das células HEK-293T com o vetor pLV-spEYFP-DHFR (dado não mostrado). Alternativamente, avaliou-se a presença da proteína spEYFP no meio de cultura condicionado pelas células HEK-293T transfectadas transitoriamente com o vetor pLV-spEYFP-DHFR através de SDS-PAGE 10%, seguido de escaneamento a laser do gel para detecção da fluorescência (Figura 6).

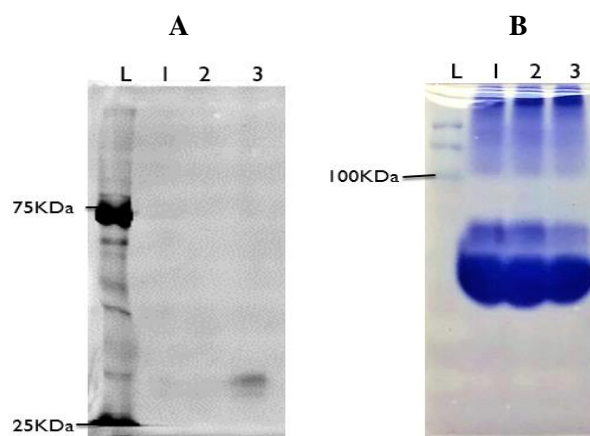


Figura 6: Avaliação da secreção da proteína spEYFP (~27 KDa) por SDS-PAGE 10% em condições não redutoras, seguida por detecção de fluorescência por escaneamento a laser. (A) Amostras de meio condicionado por 48 horas pelas células HEK-293T transfectadas transitoriamente por lipofecção com o vetor pLV-spEYFP-DHFR (clone P3C – canaleta 3) foram concentradas (10x) por ultrafiltração e fracionadas por SDS-PAGE 10% em condições não redutoras, seguida por detecção da emissão de fluorescência a 540 nm por escaneamento a laser (530 nm). Como controle negativo, empregou-se meio condicionado por 48h pelas células HEK-293T não submetidas ao procedimento de transfecção (canaleta 1), ou pelas células HEK-293T transfectadas com o vetor de expressão da proteína fluorescente não secretada EGFP, pLV-EGFP (canaleta 2). (B) Coloração com Coomassie blue coloidal do gel de acrilamida após o escaneamento a laser. L: Padrão de peso molecular.

Pode-se observar na figura acima que foi possível constatar que a proteína spEYFP é secretada para o meio de cultivo das células HEK-293T transfectadas transitoriamente com o clone P3C do vetor pLV-spEYFP-DHFR a partir da detecção de um duplete de bandas proteicas fluorescentes com tamanhos correspondentes à proteína EYFP (~27KDa), entretanto a razão da existência de duas bandas referente a spEYFP é desconhecida. A possibilidade da detecção da proteína spEYFP no meio de cultivo ser devido ao extravasamento do conteúdo intracelular associado a eventos de morte celular foi descartado frente ao resultado obtido com um dos controles negativos utilizados, o meio condicionado pelas células HEK-293T transfectadas com o vetor de expressão da proteína fluorescente EGFP, pLV-EGFP, a qual, por não apresentar um peptídeo sinal, é acumulada no meio intracelular (Figura 6A, canaleta 2). Estes experimentos demonstraram a funcionalidade e secreção da proteína spEYFP, expressa a partir do clone P3C do vetor pLV-spEYFP-DHFR.

4.3 Avaliação da produção e funcionalidade das partículas lentivirais recombinantes geradas a partir do clone P3C do vetor pLV-spEYFP-Dhfr

A produção das partículas lentivirais recombinantes foi realizada a partir da co-transfecção transitória das células HEK-293T com o vetor de transferência (clone P3C do vetor pLV-spEYFP-DHFR), e dos vetores de expressão de elementos estruturais do lentivírus pLP1/pMDL, RevpLP2/pREV e pVSVg, conforme descrito na seção 4.4 de Materiais e Métodos. Da mesma forma, foram geradas as partículas lentivirais com o vetor de transferência pLV-EGFP como controle do procedimento. Os meios condicionados contendo as partículas lentivirais recombinantes foram titulados através da transdução viral em células HEK-293T. Ao se avaliar as monocamadas por microscopia de fluorescência, observou-se células fluorescentes nas transduções envolvendo o vetor controle pLV-EGFP, e do vetor teste pLV-spEYFP-DHFR, demonstrando-se que as partículas lentivirais recombinantes produzidas a partir do clone P3C do vetor pLV-spEYFP-DHFR são funcionais (Figura 7).

Os títulos das partículas lentivirais recombinantes produzidos a partir dos vetores lentivirais pLV-EGFP e pLV-spEYFP-DHFR (clone P3C) foram determinados através da determinação da porcentagem de células fluorescentes nas monocamadas de células, conforme descrito na seção 4.4 de Materiais e Métodos. Através dessa metodologia determinou-se que o

vector pLV-EGFP e pLV-spEYFP-DHFR (clone P3C) apresentaram um título viral de 8×10^5 e 3×10^5 unidades formadoras de colônias por mililitro (UFC/mL), respectivamente

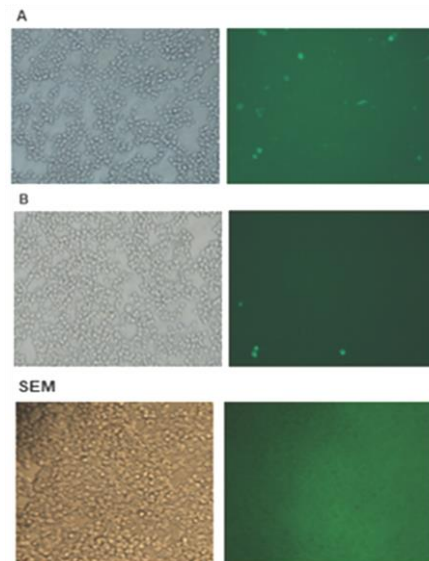


Figura 7: Avaliação da funcionalidade das partículas lentivirais recombinantes produzidas a partir do clone P3C do vetor pLV-spEYFP-DHFR por microscopia de fluorescência. Células HEK-293T foram transduzidas com as partículas lentivirais recombinantes do clone P3C do vetor pLV-spEYFP-DHFR (B) e do vetor pLV-EGFP (A), e avaliadas por microscopia de campo claro (luz visível) e fluorescência (FL) 48 horas após a transdução. Como controle positivo da capacidade de empacotamento e transdução do vetor lentiviral, as células HEK-293T foram transduzidas com o vetor pLV-EGFP. Como controle negativo, empregou-se as células HEK-293T não submetidas ao procedimento de transdução (SEM).

4.4 Geração de células CHO-DG44 com expressão estável da proteína spEYFP

Após a confirmação da funcionalidade das partículas lentivirais recombinantes produzidas a partir do clone P3C do vetor pLV-spEYFP-DHFR, estas foram utilizadas para a transdução estável do vetor pLV-spEYFP-DHFR nas células CHO-DG44, conforme descrito na seção 3.5 de Materiais e Métodos. Como há uma relação entre o MOI e o número de eventos de integração genômica de vetores de expressão lentivirais, optamos por empregar um MOI relativamente baixo, de 0,5, com o objetivo de aumentar nossas chances de resgatar clones celulares nas células CHO-DG44 associados à eventos únicos de inserção genômica. Alternativamente, as células CHO-DG44 foram transfectadas estavelmente com o mesmo vetor de expressão, conforme descrito na seção 3.6 de Materiais e Métodos. O objetivo foi isolar clones celulares que seriam utilizados como referência de níveis de expressão da spEYFP associados a números mais elevados de cópias do vetor de expressão pLV-spEYFP-DHFR integradas no genoma das células CHO-DG44.

A partir das populações mistas obtidas, procedeu-se à seleção das células com maiores níveis de expressão da proteína spEYFP, denominada de etapa de enriquecimento. Para esta finalidade, foi realizada a separação das células ativada por fluorescência (“cell sorting”) por citometria de fluxo no equipamento FACS ARIA II, conforme descrito na seção 4.7 de Materiais e Métodos. Primeiramente, realizou-se uma varredura, empregando-se cerca de 10.000 eventos, das populações mistas para a obtenção do perfil de distribuição da intensidade de fluorescência em cada população. Tomando-se como referência as distribuições de intensidade de fluorescência obtidas, isolou-se o último decil de cada população (10% da população celular com as maiores intensidades de fluorescência da proteína spEYFP) (Figura 8). As populações celulares mistas obtidas foram avaliadas por microscopia de fluorescência antes e após a seleção de células ativada por fluorescência (Figura 8).

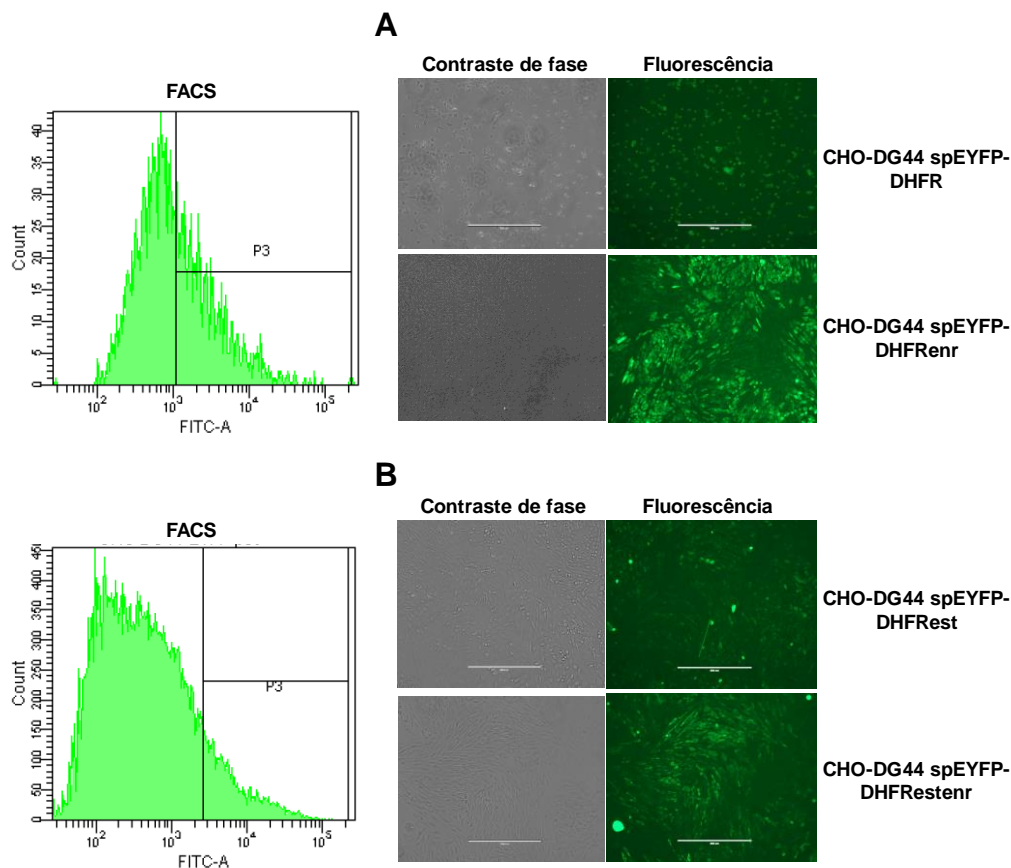


Figura 8: Enriquecimento por FACS das células spEYFP-positivas a partir das células CHO-DG44 transduzidas (A) ou transfetadas estavelmente (B) com o vetor lentiviral bicistrônico de expressão pLV-spEYFP-DHFR. As janelas (“gates”), aplicadas para o isolamento de 10% das células spEYFP-positivas com as maiores intensidades de fluorescência, são mostradas nos histogramas de intensidade relativa de fluorescência representados na parte esquerda das figuras (A) e (B). As monocamadas celulares das populações mistas CHO-DG44 spEYFP-DHFR e CHO-DG44 spEYFP-DHFRest foram avaliadas por microscopia de fluorescência antes e depois do enriquecimento por FACS, como representado na parte direita das figuras (A) e (B), demonstrando a eficácia do procedimento de enriquecimento.

Observa-se na Figura 8 que as populações mistas de células CHO-DG44, geradas a partir da transdução ou transfecção estável do vetor pLV-spEYFP-DHFR, apresentaram-se heterogêneas em relação à intensidade de fluorescência e à proporção do número de células fluorescentes. Nota-se também a maior proporção de células fluorescentes associada a metodologia da transdução (Figura 8A). Na parte inferior à direita das Figuras 8A e 8B, observa-se o aumento da proporção de células fluorescentes e da intensidade de fluorescência nas populações celulares enriquecidas, demonstrando a eficácia do procedimento de enriquecimento celular. Desta forma, foram geradas as populações celulares mistas enriquecidas CHO-DG44 spEYFP-DHFR ST e CHO-DG44-spEYFP-DHFR SV.

4.5 Isolamento e caracterização de clones celulares a partir das populações celulares mistas enriquecidas CHO-DG44 spEYFP-DHFR SV e CHO-DG44-spEYFP-DHFR ST

Após a recuperação das populações celulares enriquecidas a partir dos dois procedimentos de transferência gênica explorados, transdução e transfecção estáveis, procedeu-se ao isolamento de clones celulares produtores da spEYFP por clonagem por diluição limitante em placas de 96 poços, conforme descrito na seção 4.4 de Materiais e Métodos. Através deste método, isolaram-se clones produtores da spEYFP a partir das populações mistas enriquecidas de células CHO-DG44 transduzidas com as partículas lentivirais recombinantes do clone P3C do vetor pLV-spEYFP-DHFR (clones SV1, SV2, SV3, SV4, SV5, SV6, SV7 e SV8), ou transfectadas estavelmente com o mesmo vetor (clones ST1, ST2, ST3, ST4, ST5 e ST6). A Figura 9 representa a análise por microscopia de fluorescência de cada um dos clones isolados, confirmando a produção da spEYFP.

Cada um destes clones isolados foi avaliado quanto a capacidade de secreção da proteína spEYFP através do fracionamento dos meios de cultivo condicionados por 72 h, previamente concentrados por um fator de 10X mediante ultracentrifugação em microcolunas Amicon de corte de 10KDa, por SDS-PAGE 10%, seguida de escaneamento do gel a laser para detecção de fluorescência (Figura 10).

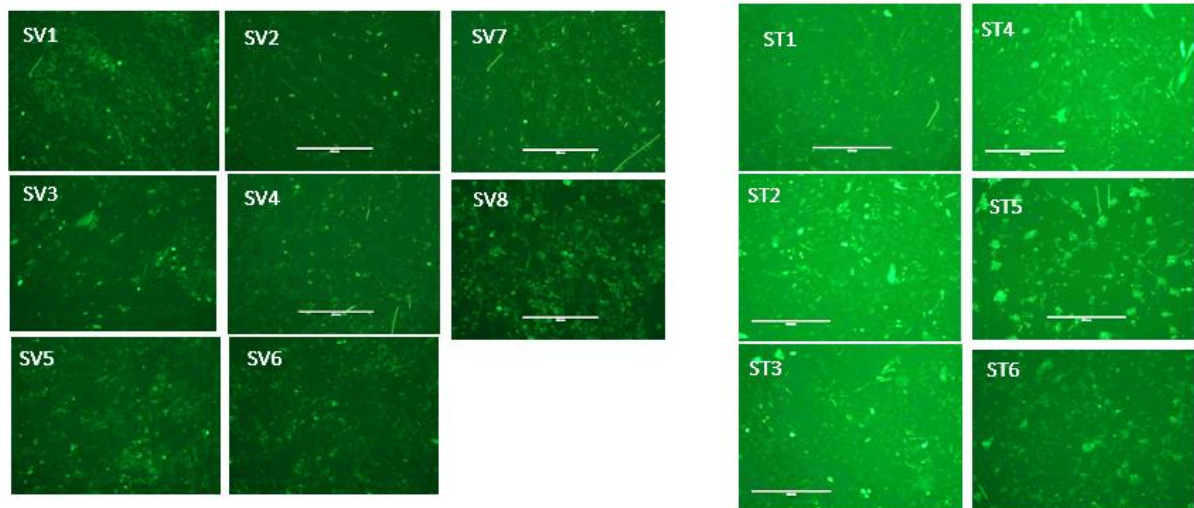


Figura 9: Clones derivados das células CHO-DG44 produtores da proteína spEYFP isolados a partir das estratégias de transdução ou transfecção estável por lipofecção do clone P3C do vetor pLV-spEYFP-DHFR. Avaliação por microscopia de fluorescência dos clones isolados. SV: clones produtores da proteína spEYFP associados à estratégia de transdução. ST: clones produtores da proteína spEYFP associados a estratégia de transfecção estável pelo método de lipofecção.

Pode-se observar na Figura 10 que todos os clones produtores da spEYFP e oriundos das células CHO-DG44 transduzidas (clones SV) ou transfectadas estavelmente (clones ST) com o vetor pLV-spEYFP-DHFR P3C secretam a proteína spEYFP, e que no caso dos meios condicionados pelas linhagens parentais sem modificação, células HEK 293T e CHO DG44, não foi possível detectar bandas proteicas fluorescentes no comprimento de onda avaliado.

Selecionou-se um dos clones celulares como representante da estratégia de transfecção estável (ST1), o qual foi submetido a uma rodada de seleção com 30nM de MTX, na tentativa de selecionar sub-populações celulares com um maior número de cópias genômicas do vetor de expressão pLV-spEYFP-DHFR. Gerou-se, desta forma, a população celular ST1-30nM.

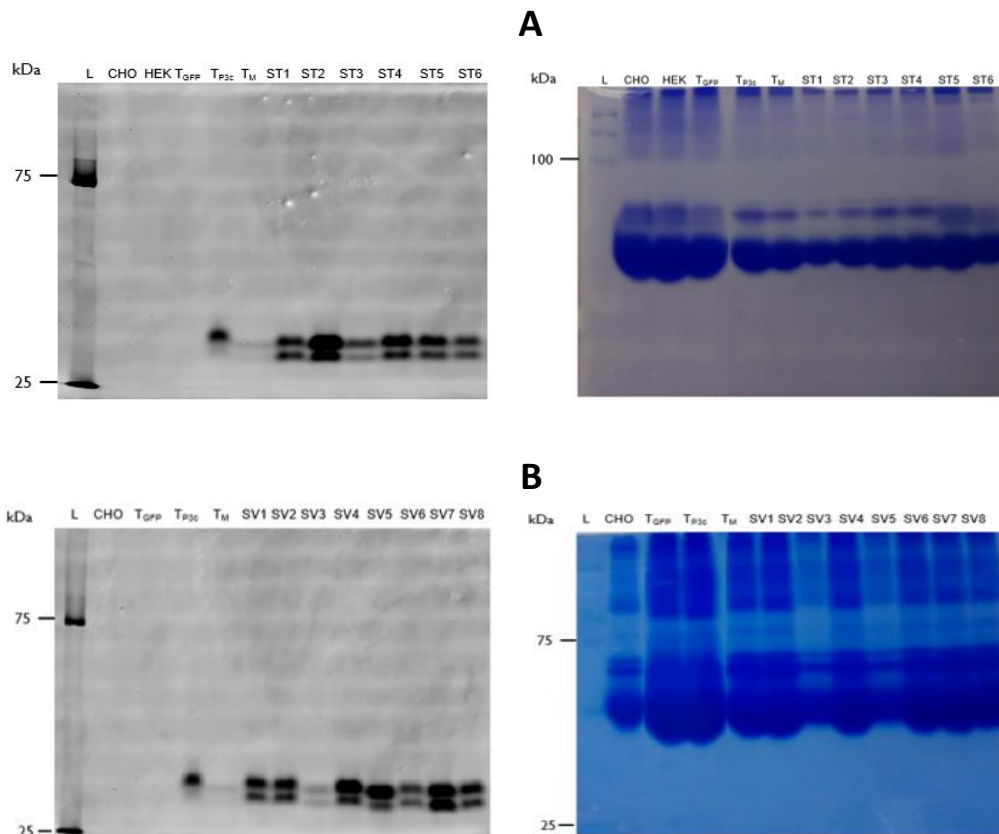


Figura 10: Avaliação da secreção da proteína spEYFP (~27KDa) por SDS-PAGE em condições não redutoras, seguida por detecção de fluorescência por escaneamento a laser. Foram avaliados os clones celulares representativos das estratégias de transfecção estável (A) ou transdução (B) das células CHO-DG44 com o vetor pLV-spEYFP-DHFR. Amostras de 5 μ l de meio condicionado por 72 horas pelos clones produtores da spEYFP isolados a partir das populações celulares mistas enriquecidas das células CHO-DG44 transduzidas com as partículas lentivirais do vetor pLV-spEYFP-Dhfr (clone P3C) (Figura 9B, canaletas SV1 a SV8), ou transfectadas estavelmente pelo método de lipofecção com o mesmo vetor (Figura 9A, canaletas ST1 a ST6), foram concentradas (10x) por ultrafiltração e fracionadas por SDS-PAGE 10% em condições não redutoras, seguida por detecção da emissão de fluorescência a 540nm por escaneamento a laser (530nm) (figuras à esquerda), e coloração com Coomassie blue coloidal (figuras à direita). Como controle positivo, empregou-se 5 μ l de meio condicionado por 48 horas das células HEK293T transfectadas transitoriamente com o vetor pLV-spEYFP-DHFR e concentrado por um fator de 10x por ultrafiltração (canaletas T_{P3C}). Como controles negativos da detecção da spEYFP, utilizou-se alíquotas de 5 μ l de meio de cultivo α -MEM (canaletas M), 5 μ l de meio condicionado por 96h pelas células CHO-DG44 e concentrado por um fator de 10x por ultrafiltração (canaletas CHO), e 5 μ l de meio condicionado por 48h pelas células HEK293T transfectadas transitoriamente com o vetor pLV-EGFP e concentrado por um fator de 10x por ultrafiltração (canaletas T_{GFP}). L: Padrão de peso molecular.

4.6 Determinação do número de cópias do elemento spEYFP-IRES-DHFR integradas no genoma dos clones celulares gerados

Com a demonstração de que os clones isolados por ambas metodologias de transferência genica produzem e secretam a proteína-modelo utilizada, a spEYFP, a próxima etapa consistiu

em determinar o número de cópias integradas do elemento spEYFP-IRES-DHFR no genoma dos clones isolados a partir das células CHO DG44. Avaliou-se o número de cópias do elemento spEYFP-IRES-DHFR integradas no genoma dos clones SV1-SV8, ST1 e ST130nM. Empregou-se um protocolo de contagem do número de cópias baseado na quantificação absoluta por qPCR. A avaliação por qPCR foi baseada na utilização de iniciadores específicos para a detecção de um trecho da sequência codificadora da DHFR, presente no cassete de expressão do vetor pLV-spEYFP-DHFR, e que foi integrado no genoma das células CHO-DG44. Para a quantificação absoluta do número de cópias integradas no genoma dos clones celulares gerados, gerou-se uma curva-padrão empregando quantidades definidas do vetor de expressão pLV-spEYFP-DHFR (Figura 11).

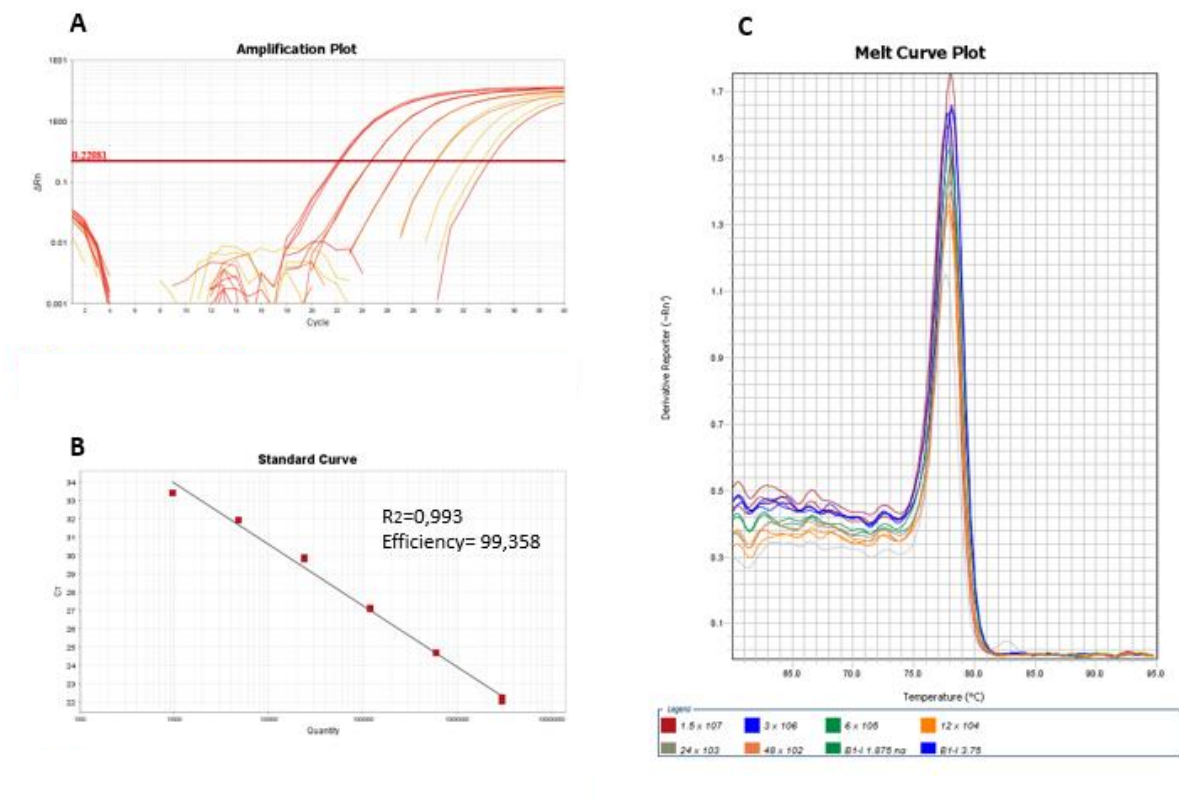


Figura 11: Obtenção da curva-padrão para a quantificação do número de cópias da região codificadora da DHFR por qPCR. A) Curvas de amplificação de um trecho da região codificadora da DHFR a partir do NTC (“No Template Control”) e das diferentes diluições do vetor pLV-spEYFP-DHFR que compõem os pontos da curva-padrão. B) Curva-padrão obtida com as diluições do vetor pLV-spEYFP-DHFR. Estão representados os valores de eficiência de amplificação (“Efficiency”) e o coeficiente de determinação (R^2). C) Curvas de dissociação dos amplicons obtidos empregando-se as diferentes diluições do vetor pLV-spEYFP-DHFR.

Na Figura 11A, pode-se observar as curvas de amplificação obtidas a partir de cada uma das diluições do vetor pLV-spEYFP-DHFR, as quais demonstram o acúmulo exponencial dos

amplicons. Adicionalmente, observa-se uma diferença de Ct constante entre as diferentes diluições exploradas, possibilitando a obtenção de uma curva-padrão (Figura 11B) com um R^2 de 0,99, indicando uma forte correlação entre os dados do Ct e o número de cópias do vetor pLV-spEYFP-DHFR, e de uma eficiência de amplificação de aproximadamente 99%, indicando um perfil de amplificação bem próximo do teórico. Pode-se observar nas curvas de amplificação (Figura 11A) a formação de um produto inespecífico, associado a um valor mais elevado de Ct, em uma das amostras do NTC (“No Template Control”). Entretanto, a amplificação deste produto inespecífico provavelmente não interferiu com a utilização da curva-padrão para se realizar a avaliação do número de cópias integradas, pois a sua amplificação ocorreu somente nos últimos ciclos de amplificação. As curvas de dissociação obtidas (Figura 11C) evidenciaram a obtenção de um único amplicon durante os ciclos de amplificação das diferentes diluições do vetor pLV-spEYFP-DHFR exploradas, sendo provavelmente um produto específico de amplificação da região codificadora da DHFR, não tendo sido observado nenhum outro pico relativo à amplificação de algum produto adicional.

Após ter sido verificada a especificidade e linearidade da curva-padrão gerada, esta foi utilizada para a quantificação do número de cópias do elemento spEYFP-IRES-DHFR integradas no genoma dos clones isolados pelo método de transdução (clones SV1-SV8), além do clone ST1 e da população celular ST1-30nM, associados à estratégia de transfecção celular (Figura 12).

Como mostrado na Figura 12, todos os clones celulares associados a estratégia de transdução apresentaram aproximadamente um único evento de integração do elemento spEYFP-IRES-DHFR no genoma das células CHO-DG44. Em contraste, o clone usado como referência do método convencional de transfecção celular estável, clone ST1, apresentou aproximadamente 34 cópias do elemento spEYFP-IRES-DHFR integradas no genoma das células CHO-DG44, e, após este clone ter sido submetido à seleção com 30nM de MTX, o número de cópias/genoma aumentou para aproximadamente 47. Como controle negativo, realizou-se esta mesma avaliação com a amostra de gDNA da linhagem parental CHO DG44, não tendo sido observada amplificação de produto. Adicionalmente, como controle da metodologia, realizou-se a determinação do número de cópias/genoma de um gene endógeno, o glucagon (*Gcg*). Segundo dados da literatura, o *Gcg* apresenta duas cópias na linhagem parental CHO-DG44¹⁰¹, e o valor obtido foi muito similar ao publicado.

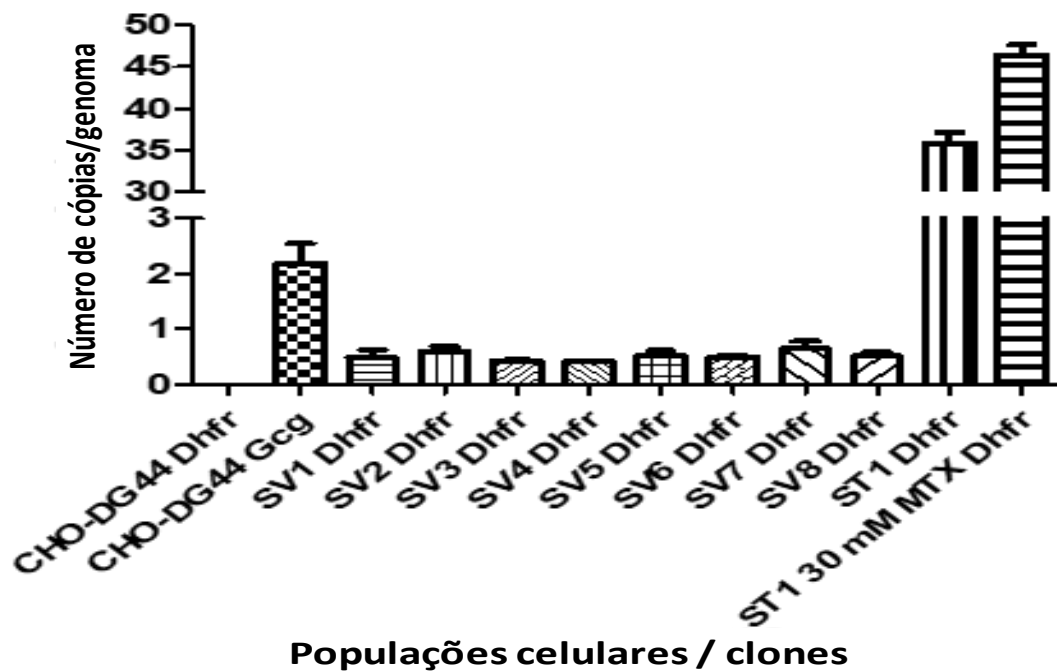


Figura 12: Quantificação absoluta por qPCR do número de cópias do elemento bicistrônico spEYFP-IRES-Dhfr integrado no genoma das células CHO-DG44. Empregou-se iniciadores específicos para a DHFR, e diluições seriadas (fator de diluição de duas vezes) das amostras de gDNA obtidas de clones isolados das populações celulares mistas enriquecidas CHO-DG44 spEYFP-DHFRenr (indicados como SV1 a SV8) ou CHO-DG44 spEYFP-DHFRestenr (indicado como ST1). A mesma análise foi realizada com o gDNA do clone ST1 submetido à seleção com 30nM MTX (indicado como ST1 30nM MTX) e as células CHO-DG44 parentais. O número de cópias/genoma do gene Gcg foi determinado na amostra de gDNA das células CHO-DG44 parentais como controle da análise do número de cópias genômicas por qPCR. As barras representam as determinações médias do número de cópias/genoma e a barra de erro o desvio padrão considerando as réplicas técnicas (três) de cada ponto (quatro no total) da diluição seriada das amostras de gDNA.

4.7 Quantificação relativa e análise da estabilidade de expressão da spEYFP

Em seguida, comparamos os níveis relativos de spEYFP acumulados nos meios de cultivo condicionados por 72 h pelos clones celulares e populações celulares avaliados. Amostras de meio de cultivo condicionados foram analisadas por 10% SDS-PAGE, e a spEYFP foi detectada pelo escaneamento a laser dos géis. As bandas proteicas da spEYFP, de 27 KDa e 30 KDa, foram quantificadas por densitometria digital, e os volumes de bandas proteicas obtidos foram normalizados pelo número de células no momento da coleta de meio condicionado. O valor de expressão relativa da spEYFP associado à população original de células mistas obtidas a partir

da transdução das células CHO-DG44 (CHO-DG44 spEYFP-DHFR) foi definida como 1,0, e foi utilizado como referência para se determinar a expressão relativa da spEYFP entre os diferentes clones e populações celulares avaliados (Figura 13).

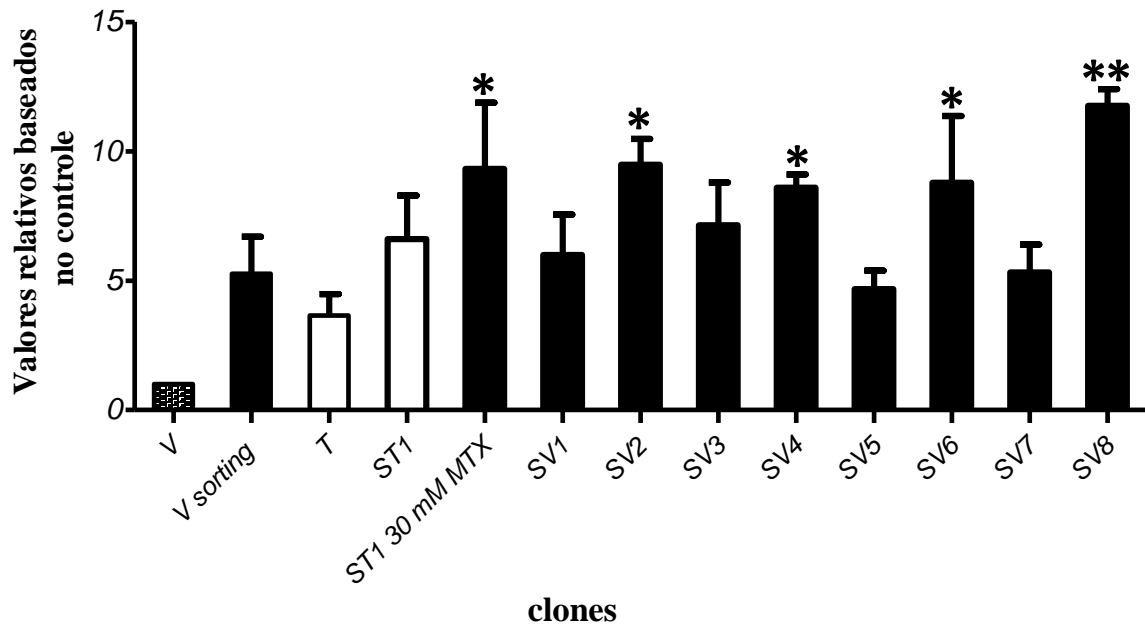


Figura 13: Análise dos níveis de expressão relativa da spEYFP entre os clones isolados a partir das células CHO-DG44 transduzidas ou transfectadas estavelmente com o vetor pLV-spEYFP-DHFR. Os valores dos volumes densitométricos das bandas proteicas (27 kDa e 30 kDa) correspondentes à proteína spEYFP, detectadas por escaneamento a laser após os experimentos de SDS-PAGE, e relativos aos clones e populações celulares indicadas, foram normalizados pelo número de células viáveis no momento das coletas de meio condicionado por 72 horas. Os valores de expressão relativa normalizada da spEYFP da população mista CHO-DG44 spEYFP-DHFR (V) foram utilizados como referência, sendo definida como 1,0. V sorting e T: Populações enriquecidas associadas as estratégias de transdução ou transfecção das células CHO-DG44, respectivamente SV1-SV8: Clones isolados a partir da população de células mistas enriquecidas das células CHO-DG44 transduzidas com o vetor pLV-spEYFP-DHFR; ST1: Clone isolado a partir da população de células mistas enriquecidas das células CHO-DG44 transfectadas estavelmente com o vetor pLV-spEYFP-DHFR; ST1 30nM MTX: População de células geradas a partir da seleção do clone ST1 com 30 nM de MTX. As barras indicam as médias de expressão, e as barras de erro o desvio-padrão (n=3). Significância estatística: *p<0,05 ou **p<0,001, considerando a amostra spEYFP-LV como referência (ANOVA seguida de Tukey).

Os valores de secreção normalizados mostraram que a população gerada pela estratégia convencional de transfecção apresentou níveis de secreção 3,6X maiores da proteína spEYFP quando comparada à população mista gerada pela estratégia de transdução (Figura 13), embora esta diferença não tenha sido estatisticamente significativa. Um dos clones isolados a partir desta população mista, o clone ST1, acumulou no meio condicionado por 72 h níveis relativos da spEYFP 6,6X maiores que àqueles observados na população mista gerada pela estratégia de

transdução (V). O clone ST1, após ter sido submetido à seleção com uma concentração de 30nM de MTX (ST1 30 nM MTX), passou a acumular cerca de nove vezes mais spEYFP em relação à população mista gerada pela estratégia de transdução (Figura 13). Quanto à estratégia de transdução, a população mista enriquecida mostrou níveis de secreção cerca de 5X maiores que o da população mista de transdução inicial, embora esta diferença não tenha sido significativa (Figura 13). Para os clones celulares associados à estratégia de transdução, destacam-se os clones SV2, SV4, SV6, e particularmente o clone SV8, que apresentou um acúmulo relativo da spEYFP cerca de 12X maior do que àquele observado na população mista gerada pela estratégia de transdução (Figura 13). É importante ressaltar que todos os clones derivados da população de células mistas CHO-DG44 spEYFP-DHFR SV (clones SV1-SV8) apresentaram níveis de expressão semelhantes ao observado para as células ST1 e ST1 30nM MTX (Figura 13). Os dados de expressão da spEYFP e a análise de número de cópias do elemento spEYFP-IRES-DHFR integradas no genoma (Figura 12) sugerem que os sítios únicos de integração genômica do elemento spEYFP-IRES-DHFR associados aos clones SV1 a SV8 possui uma elevada atividade transcricional, uma vez que apenas uma cópia desse elemento está associada a níveis de expressão da spEYFP semelhantes aos associados a 35 a 45 cópias do mesmo elemento integradas aleatoriamente no genoma.

Após avaliação dos níveis relativos de secreção da spEYFP associados a cada um dos clones isolados, avaliou-se se esses níveis de secreção eram mantidos ao longo do tempo. Assim, acompanhou-se a estabilidade dos níveis de secreção da spEYFP ao longo de 40 gerações, ou aproximadamente dois meses em cultivo (Figura 14).

A avaliação da estabilidade dos níveis de secreção da proteína spEYFP (Figura 14) indicou que todos os clones isolados, tanto da estratégia de transdução como de transfecção estável, apresentaram níveis estáveis de secreção da proteína spEYFP ao longo das 40 gerações, de acordo com os critérios de estabilidade descritos na seção 4.13 de Materiais e Métodos.

De uma forma geral, a metodologia empregada de varredura de sítios genômicos baseados na utilização de partículas lentivirais recombinantes permitiu o isolamento de clones celulares associados a eventos únicos de inserção genômica, os quais permitem uma produtividade relativa da proteína repórter estável e em níveis similares aos encontrados em clones que possuem dezenas de cópias do mesmo cassete de expressão. Estes resultados indicam que estes sítios únicos de integração genômica do elemento spEYFP-IRES-DHFR apresentam um

contexto genômico que permite uma expressão estável e de alta capacidade, comparável àquela observada em clones celulares contendo dezenas de cópias do mesmo elemento de expressão.

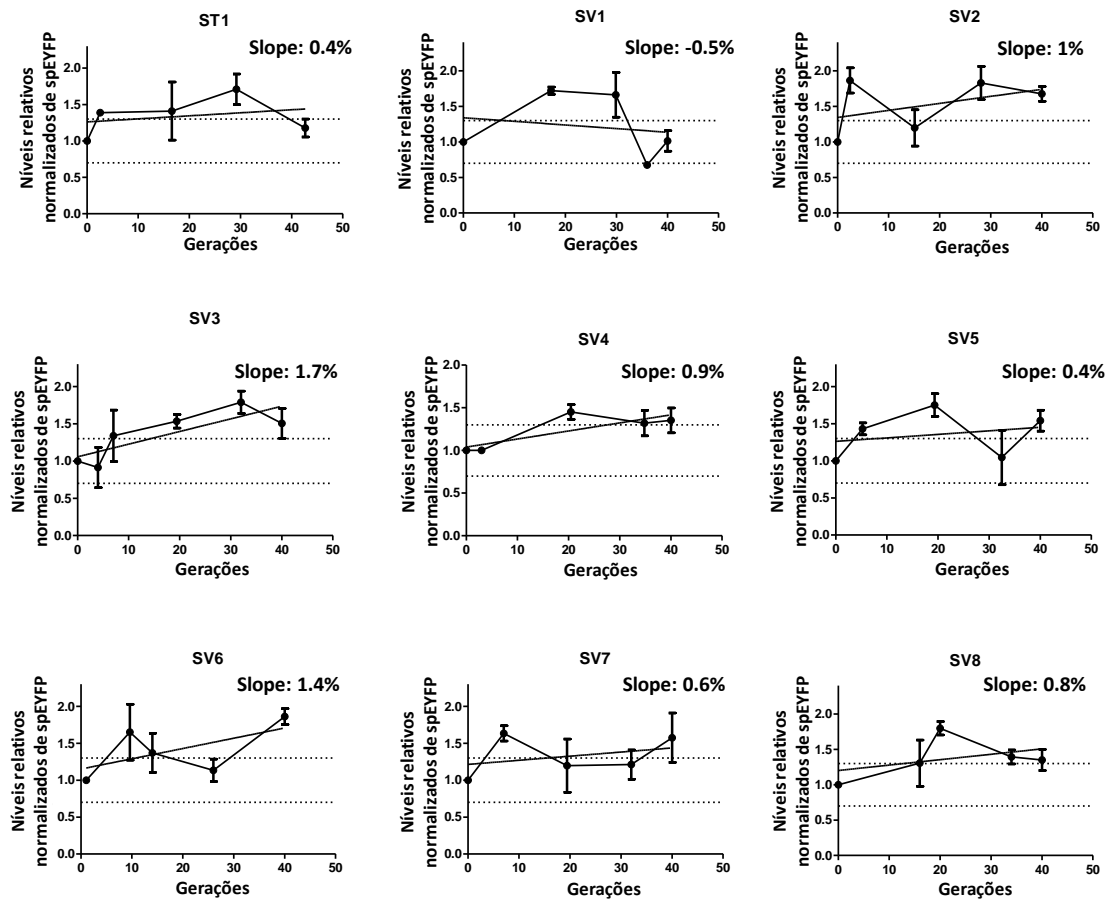


Figura 14: Avaliação da estabilidade da secreção da proteína spEYFP ao longo de 40 gerações. Avaliou-se os clones associados às estratégias de transfecção estável (clone ST1) ou transdução das células CHO-DG44 (clones SV1-SV8) com o vetor pLV-spEYFP-DHFR. Os níveis relativos de secreção da spEYFP de cada um dos clones isolados foram avaliados no tempo 0 e em intervalos de 10 gerações, ao longo de 40 gerações (cerca de dois meses). As amostras de meios de cultivo condicionados por 72 h foram avaliadas por 10% SDS-PAGE em condições não redutoras, seguida por detecção de fluorescência por escaneamento a laser. Os valores de volume densitométricos das bandas proteicas correspondentes à spEYFP, normalizados pelo número de células viáveis no momento da coleta, foram utilizados para se obter uma curva de regressão linear dos valores de produtividade relativa obtidos ao longo das 40 gerações. O critério de estabilidade dos níveis de secreção da spEYFP utilizado foi a obtenção de inclinações da curva de regressão inferiores a 30%. Na parte superior à direita de cada gráfico estão representados os valores de inclinação (“slope”) das curvas.

4.14 Determinação do sítio de integração baseado na estratégia de “DNA walking”

Com o isolamento de clones celulares contendo eventos únicos de inserção genômica, potencialmente associados à uma atividade estável e elevada de transcrição gênica, a etapa final

do trabalho consistiu na identificação destas regiões genômicas de inserção através da técnica de DNA “walking”, conforme descrito na seção 4.14.1 de Materiais e Métodos. A primeira etapa consistiu na fragmentação das amostras de gDNA através da digestão com as enzimas de restrição HpaI e EcoRV para a geração das bibliotecas de fragmentos de DNA, conforme indicado na Figura 15.

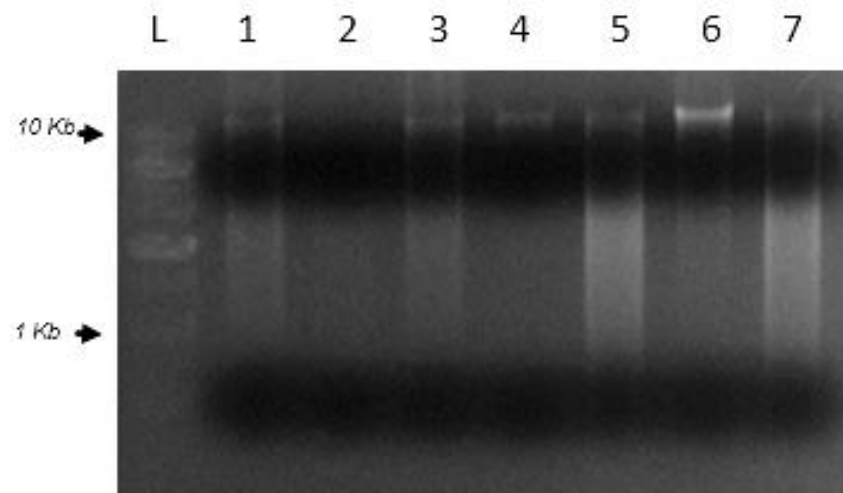


Figura 15: Geração das bibliotecas de fragmentos de gDNA dos clones SV2, SV4, SV5 e SV7. Análise por eletroforese em gel de agarose 1% m/v das amostras de gDNA (200ng) dos clones SV2, SV4, SV5 e SV7 fragmentados através da digestão com as enzimas de restrição HpaI ou EcoRV por 16 h à 37°C. L: marcador de peso molecular de 1 Kpb; 1) Amostra de gDNA do clone SV2 digerido com HpaI; 2) Amostra de gDNA do clone SV2 digerido com EcoRV; 3) Amostra de gDNA do clone SV4 digerido com HpaI; 4) Amostra de gDNA do clone SV4 digerido com EcoRV; 5) Amostra de gDNA do clone SV5 digerido com HpaI; 6) Amostra de gDNA do clone SV5 digerido com EcoRV; 7) Amostra de gDNA do clone SV7 digerido com HpaI.

Como pode ser observado na Figura 15, as amostras de gDNA dos clones SV2, SV4, SV5 e SV7 foram digeridas pelas enzimas de restrição EcoRV e HpaI, evidenciado pela presença de fragmentos de gDNA de diversos tamanhos. As amostras de gDNA não digeridas apresentam um perfil de fragmentos com tamanhos ao redor de 20 Kpb (dados não mostrados).

Após a geração das bibliotecas de fragmentos de gDNA, seguiu-se com a etapa de ligação dos oligonucleotídeos adaptadores, conforme descrito na seção 4.14.2 de Materiais e Métodos. Na sequência, uma alíquota das reações de ligação (1 µl) foi utilizada para a realização dos dois ciclos de PCR do “Nested” PCR (Figura 16).

Como pode ser observado na Figura 16, não houve a amplificação de produtos específicos após os dois ciclos de PCR.

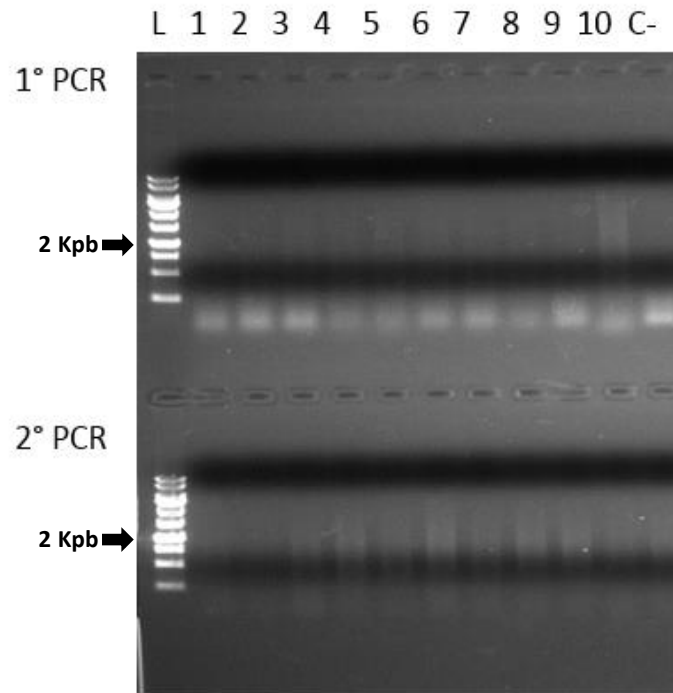


Figura 16: Etapas de “Nested” PCR para o reconhecimento dos sítios de integração do vetor de expressão pLV-spEYFP-DHFR nas células CHO-DG44. Análise por eletroforese em gel de agarose 1% m/v dos dois ciclos de PCR do “Nested” PCR, etapa final da metodologia empregada para o reconhecimento dos sítios de integração baseada em DNA “walking”. L: marcador de peso molecular de 1 Kpb. 1) Amostra de gDNA do clone SV2 digerido com a enzima de restrição HpaI; 2) Amostra de gDNA do clone SV2 digerido com EcoRV; 3) Amostra de gDNA do clone SV4 digerido com HpaI; 4) Amostra de gDNA do clone SV4 digerido com EcoRV; 5) Amostra de gDNA do clone SV5 digerido com HpaI; 6) Amostra de gDNA do clone SV5 digerido com EcoRV; 7) Amostra de gDNA do clone SV7 digerido com HpaI; 8) Amostra de gDNA do clone SV7 digerido com EcoRV; 9) Amostra de gDNA do clone SV8 digerido com HpaI; 10) Amostra de gDNA do clone SV8 digerido com EcoRV; C-) amostra sem DNA.

Como a nossa tentativa inicial de amplificação dos sítios de inserção pela técnica de DNA “walking” resultou na amplificação de produtos inespecíficos e sem a formação de nenhuma banda de DNA definida, buscou-se descobrir as etapas que limitavam o sucesso desta metodologia. Para isso, realizou-se o protocolo empregando-se o vetor pLV-spEYFP-DHFR como um controle de cada etapa do procedimento, basendo-se na sua sequência nucleotídica teórica. Seguiu-se o procedimento previamente descrito, empregando-se 1 µg do vetor, que foi

digerido individualmente ou simultaneamente, com as enzimas de restrição SspI, HpaI e EcoRV (Figura 17).

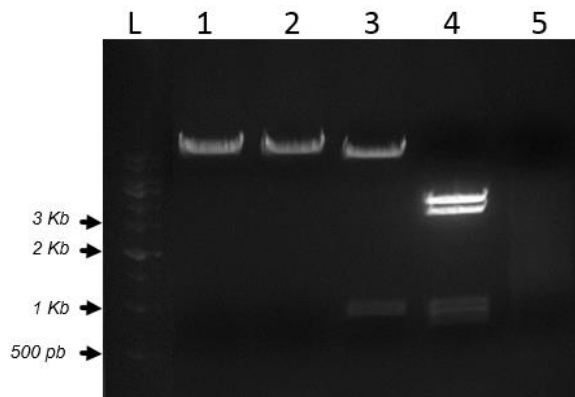


Figura 17: Digestão do vetor pLV-spEYFP-DHFR com as enzimas de restrição SspI, HpaI e EcoRV. Análise por eletroforese em gel de agarose 1% m/v do vetor pLV-spEYFP-DHFR (1 µg) após ter sido digerido com as enzimas de restrição SspI (fragmentos esperados 8,7 Kpb e 133 pb), HpaI (fragmento esperado 8,85 Kpb) ou EcoRV (fragmentos esperados 8 Kpb e 849 pb) por 16 h à 37°C. L: marcador de peso molecular de 1 Kpb; 1) Amostra do pLV-spEYFP-DHFR digerido com HpaI; 2) Amostra de pLV-spEYFP-DHFR digerido com SspI; 3) Amostra de pLV-spEYFP-DHFR digerido com EcoRV; 4) Amostra de pLV-spEYFP-DHFR digerido com HpaI, SspI e EcoRV (fragmentos esperados 3,9 Kpb, 3,2 Kpb, 850 pb, 731pb e 133pb); 5) Amostra de gDNA da linhagem CHO-DG44 digerido com HpaI, SspI e EcoRV.

Como pode ser observado na Figura 17, a digestão do vetor pLV-spEYFP-DHFR individualmente ou simultaneamente com as enzimas de restrição SspI, HpaI e EcoRV gerou fragmentos de DNA de tamanhos que se aproximaram dos valores teóricos esperados. Uma das nossas dúvidas era se alguma das enzimas de restrição utilizadas para fragmentação das amostras de gDNA clivaria algum sítio nas regiões 5'LTR e 3'LTR do vetor, o que comprometeria a continuidade da região compreendida entre os sítios de anelamento dos iniciadores LSP1 e LSP2, localizados na região U5 da região 5'LTR, e a extremidade do elemento 5' do provírus. A digestão do vetor com as três enzimas de restrição simultaneamente demonstrou que o mesmo não é clivado em alguma região próxima à região na qual se liga os iniciadores LSP1 e LSP2, o que, caso acontecesse, não permitiria a formação de um amplicon específico relativo ao sítio genômico de integração.

De acordo com o protocolo original, purificou-se os fragmentos de DNA gerados pela digestão individual ou simultânea com as três enzimas de restrição avaliadas, através de adsorção do DNA em colunas de sílica. Em uma das alternativas avaliadas, misturou-se os fragmentos de DNA (500ng) do vetor pLV-spEYFP-DHFR digerido, individualmente ou simultaneamente, com as enzimas EcoRV, HpaI e SspI ao gDNA (1 µg) das células CHO-

DG44 digerido com as mesmas enzimas, e seguiu-se com a purificação dos fragmentos através de adsorção do DNA em colunas de sílica. Na sequência, avaliou-se a eficiência de recuperação dos fragmentos de DNA gerados pela digestão com as enzimas de restrição e purificados, comparando-se a situação na qual o material genético não foi purificado. A título de ilustração, a Figura 18 representa os dados obtidos com a digestão realizada apenas com a enzima SspI. Dados similares foram obtidos nas demais configurações exploradas.

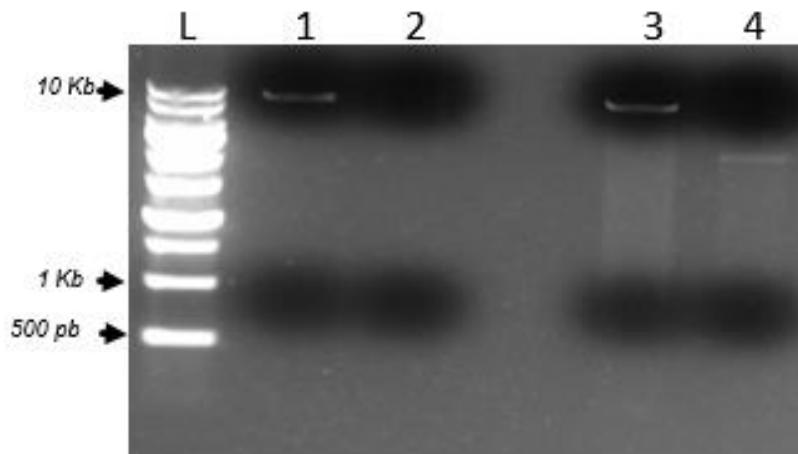


Figura 18: Avaliação da eficiência do protocolo de purificação dos fragmentos de DNA baseado em adsorção do DNA à sílica. Análise por eletroforese em gel de agarose 1% m/v de uma das configurações avaliadas, envolvendo a digestão com a enzima de restrição SspI por 16 h à 37°C. L: marcador de peso molecular de 1 Kpb; 1) Amostra (500ng) do vetor pLV-spEYFP-DHFR digerido com SspI; 2) Amostra (500ng) do vetor pLV-spEYFP-DHFR digerido com SspI e purificação dos fragmentos de DNA resultantes por adsorção do DNA à sílica; 3) Amostra (500ng) do vetor pLV-spEYFP-DHFR digerido com SspI e misturado à uma amostra (1 µg) de gDNA de CHO-DG44 digerida com SspI; 4) Amostra (500ng) do vetor pLV-spEYFP-DHFR digerido com SspI e misturado à uma amostra (1 µg) de gDNA de CHO-DG44 digerida com SspI e purificação dos fragmentos de DNA resultantes por adsorção do DNA à sílica.

Pode-se observar na Figura 18 (canaletas 1 vs. 2, e 3 vs 4) que houve uma perda significativa dos fragmentos de DNA oriundos tanto da amostra de vetor pLV-spEYFP-DHFR como da amostra de gDNA das células CHO-DG44 digeridas com a enzima de restrição SspI, após o processo de purificação por adsorção do DNA à sílica. Concluiu-se que a etapa de purificação limita a quantidade de material disponível para a realização da etapa de ligação dos fragmentos de DNA ao adaptador, o que poderia resultar em uma redução da sensibilidade da etapa de “Nested” PCR. Devido a esta possível limitação, decidiu-se seguir o protocolo com os fragmentos de DNA não submetidos à etapa de purificação. Assim, utilizaram-se os meios reacionais das reações de digestão com as enzimas de restrição diretamente na etapa de ligação dos oligonucleotídeos adaptadores, e os meios reacionais das reações de ligação do adaptador

foram usados diretamente para a primeira reação de PCR da etapa de “Nested” PCR (Figura 19).

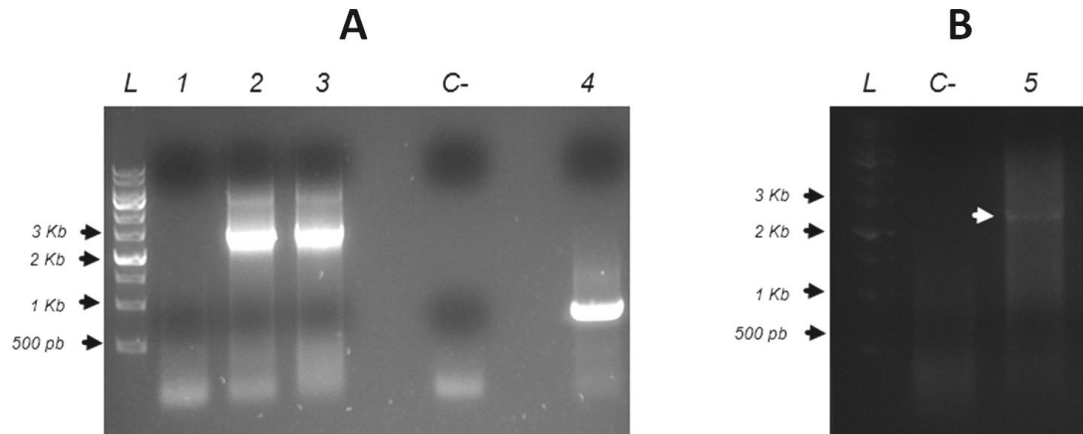


Figura 19: Detecção por “Nested” PCR da ligação dos oligonucleotídeos adaptadores aos fragmentos de DNA oriundos da digestão do vetor pLV-spEYFP-DHFR com as enzimas de restrição HpaI, SspI e EcoRV. Análise por eletroforese em gel de agarose 1% m/v dos dois ciclos de “Nested” PCR. As Figuras (A) e (B) representam os primeiros e segundo ciclos de “Nested” PCR, respectivamente. L: marcador de peso molecular de 1 Kpb; 1) Amostra de gDNA do clone SV2 digerido com a enzima de restrição SspI e purificada por adsorção do DNA à sílica; 2) Amostra do vetor pLV-spEYFP-DHFR digerido com HpaI, SspI e EcoRV; 3) Amostra do vetor pLV-spEYFP-DHFR digerido com as enzimas HpaI, SspI e EcoRV misturada à amostra de gDNA de CHO-DG44 digerida com as mesmas enzimas; 4) Amostra do vetor pLV-spEYFP-DHFR digerido com HpaI, SspI e EcoRV submetida à PCR empregando iniciadores que reconhecem a região codificadora do elemento IRES (controle positivo da presença de fragmentos de DNA oriundos do vetor após os procedimentos de fragmentação, purificação e ligação do adaptador); 5) Amostra da PCR descrita em 3), diluída 10 vezes com água milliQ. C-) Amostra sem DNA (controle negativo).

Baseando-se na sequência nucleotídica teórica do vetor pLV-spEYFP-DHFR, ao digerir-se o vetor simultaneamente com as enzimas de restrição SspI, HpaI e EcoRV um dos fragmentos gerados possuiria a região de reconhecimento do iniciador reverso LSP1, e duas extremidades passíveis de ligação ao adaptador, uma delas gerada pela digestão com a enzima SspI, e a outra pela digestão com a enzima HpaI. Caso a ligação deste fragmento de DNA ao adaptador fosse bem sucedida, um amplicon de 2.586 pb, relativo a distância entre os sítios de anelamento do iniciador LSP1 e de reconhecimento da enzima de restrição SspI, deveria ser gerado após a execução da etapa de “Nested” PCR, como de fato ocorreu (Figura 19A, canaletas 2 e 3). Este resultado positivo com os fragmentos de DNA do vetor indicaram que uma provável limitação para a obtenção de amplicons a partir de fragmentos de gDNA seria a baixa abundância relativa do alvo (sítio de inserção), já que novamente não se observou amplicons quando se empregou uma amostra de gDNA extraída do clone SV2 como material de partida, o qual foi digerido

com a enzima SspI, purificado por adsorção do DNA à sílica, ligado ao adaptador, e usado na primeira PCR (Figura 19A, canaleta 1). É importante ressaltar que em ambos os casos explorados, ou seja, na amostra que possuía apenas fragmentos do vetor e aquela na qual se efetuou a mistura de fragmentos de DNA do vetor à fragmentos de gDNA da linhagem parental CHO-DG44, observaram-se amplicons específicos e do tamanho esperado, indicando a ausência de elementos inibitórios da reação de PCR. Para a segunda PCR do protocolo de “Nested” PCR (Figura 19B, canaleta 5), a PCR foi realizada diluindo-se uma alíquota proveniente da primeira PCR (Figura 19A, canaleta 3) previamente 10 vezes com água milliQ. Contrastando com a primeira PCR (Figura 19A, canaleta 3), observou-se a presença de alguns produtos inespecíficos juntamente com uma banda de DNA definida do tamanho esperado (aproximadamente 2 500 pb) (Figura 19B, canaleta 5).

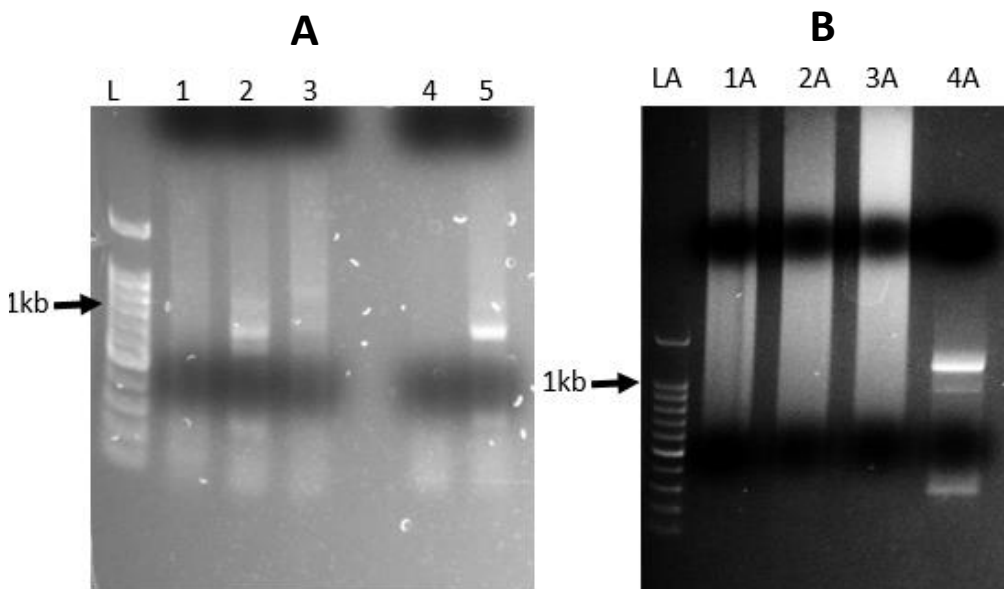


Figura 20: Amplificação dos sítios genômicos de integração do vetor de expressão pLV-spEYFP-DHFR nos clones SV4 e SV5. Análise por eletroforese em gel de agarose 1% m/v dos primeiros (A) e segundo (B) ciclos de “Nested” PCR, etapa final da metodologia empregada para o reconhecimento do sítio genômico de integração baseada na estratégia de “DNA walking”. L: marcador de peso molecular de 100 pb. 1) Primeira PCR realizada com amostra da reação de ligação entre o adaptador e o gDNA do clone SV5 digerido com as enzimas de restrição EcoRV e HpaI; 2) Primeira PCR realizada com amostra da reação de ligação entre o adaptador e o gDNA do clone SV4 digerido com EcoRV e HpaI; 3 e 5) Primeira PCR realizada com amostra da reação de ligação do adaptador e o gDNA do clone SV5 digerido com EcoRV e HpaI, empregando-se iniciadores para amplificação de parte do elemento IRES e da região codificadora da spEYFP (fragmento esperado 800pb – controle positivo da presença do vetor); 4) Primeira PCR realizada sem adição de DNA (controle negativo). LA: marcador de peso molecular de 1 Kpb; 1A) Segunda PCR realizada a partir da amostra da primeira PCR referente ao clone SV4 diluída 100 vezes, e empregando-se os mesmos iniciadores (AP1 e LSP1) da primeira PCR; 2A) Segunda PCR realizada a partir da amostra da primeira PCR referente ao clone SV4, empregando-se os iniciadores AP2 e LSP1; 3A) Segunda PCR realizada a partir da amostra da primeira PCR referente ao clone SV4 sem diluição; 4A) Segunda PCR realizada a partir da amostra da primeira PCR referente ao clone SV4 e diluída por um fator de 10 vezes, e empregando-se os iniciadores AP2 e LSP2.

Frente a estes resultados, resolveu-se realizar uma nova tentativa de amplificar o sítio genômico de inserção do vetor pLV-spEYFP-DHFR com dois dos clones isolados (clones SV4 e SV5) seguindo as novas condições estabelecidas, conforme descrito acima. A figura 19 representa os resultados obtidos após a realização das duas etapas de PCR.

Como pode ser observado na Figura 20, após a realização da primeira PCR (Figura 20A, canaletas 1, 2 e 3), pode-se visualizar amplicons de vários tamanhos nos dois clones avaliados (SV4 e SV5). Como controle negativo, realizou-se uma PCR sem adição de DNA (Figura 20A, canaleta 4). Adicionalmente, como controle positivo da presença do vetor na amostra de gDNA, realizou-se duas PCRs para a detecção do cassete de expressão na amostra de gDNA do clone SV5, empregando-se iniciadores específicos para amplificação de um amplicon de 800pb correspondente a um trecho do elemento IRES e da região codificadora da proteína spEYFP (Figura 20, canaletas 3 e 5). Nas canaletas de 1 a 4 da Figura 20B, pode-se observar os resultados referentes a configurações alternativas da segunda PCR, realizadas a partir da primeira reação de PCR, referente ao clone SV4. Observa-se que em uma das condições exploradas foi possível obter um amplicon de aproximadamente 1200 pb (Figura 20 – canaleta 4A). Nas demais condições exploradas foram observadas apenas produtos inespecíficos. Estão em andamento os experimentos para a determinação da sequência deste fragmento de DNA de 1200 pb.

5. DISCUSSÃO

Uma das etapas mais cruciais para o estabelecimento de uma plataforma de produção de proteínas recombinantes terapêuticas baseada em células de mamíferos é a obtenção de clones celulares expressando estavelmente grandes quantidades do produto de interesse, sendo fundamental que esta alta produtividade esteja vinculada a manutenção das características estruturais, bioquímicas e funcionais da proteína recombinante. A estratégia de obtenção de clones celulares superprodutores envolve a manipulação genética de determinadas linhagens celulares, mais especificamente, a inserção genômica estável da região codificadora da proteína de interesse ligada funcionalmente a elementos regulatórios da transcrição gênica, e associada à marcadores de seleção que possam ser explorados para o isolamento de clones produtores¹.

Os desafios associados a estas estratégias envolvem procedimentos laboriosos e prolongados de varredura tanto de eventos aleatórios e favoráveis de inserção genômica, bem como de fenótipos celulares compatíveis com a manutenção das características estruturais e funcionais da proteína de interesse quando produzida em grandes quantidades. Estes procedimentos de seleção e varredura diminuem grandemente a eficiência de obtenção de clones produtores. Neste sentido, o estabelecimento de novas metodologias que permitam estabelecer, em menos tempo, clones celulares superprodutores estáveis, torna-se relevante. Avanços provenientes de estudos “ômicos” e do campo da engenharia de células de mamíferos têm permitido a otimização dos processos de criação de linhagens celulares de mamífero através de múltiplas modificações genéticas para aumentar a estabilidade e os níveis de expressão de proteínas recombinantes⁴². Estudos genômicos permitiram a identificação de trechos nos genomas de várias sublinhagens de interesse das células CHO com potencial para a exploração de novas estratégias para o melhoramento na produção de proteínas recombinantes¹¹. Da mesma forma, a aplicação de ferramentas de biologia de sistemas tem se mostrado muito relevante para a compreensão das complexas alterações fenotípicas sofridas pelas células CHO durante o processo de obtenção de clones superprodutores¹⁰⁴. Ressalta-se ainda a recente disponibilidade de genomas de referência de várias sublinhagens das células CHO, fornecendo não só um mapa genético para a aplicação de técnicas de manipulação genética às células CHO, mas também um catálogo de alterações genéticas presentes nestas células, permitindo um melhor entendimento da sua biologia⁴⁸. Estas pesquisas evidenciaram a plasticidade do genoma das células CHO, e deixaram claros os desafios associados à aplicação deste conhecimento para a manipulação genética controlada das células CHO, já que são inúmeros as mutações pontuais,

os rearranjos cromossômicos e as variações do número de cópias observadas nas suas diferentes sublinhagens⁵². Há ainda amplas alterações epigenéticas que podem reconfigurar sua maquinaria celular, como, por exemplo, a metilação do DNA e as modificações de histonas¹⁰⁵.

Frente a estes desafios, a nossa proposta envolveu o reconhecimento de regiões genômicas com estável e elevada atividade transcricional como possíveis alvos para inserção sítio-dirigida de genes de interesse, visando otimizar o processo de desenvolvimento de plataforma de expressão de proteínas recombinantes em células de mamíferos. No presente trabalho, nos propusemos a avaliar a viabilidade de combinar o uso de uma proteína repórter fluorescente secretável (spEYFP), partículas lentivirais recombinantes como sistema de entrega gênica, e o enriquecimento das células expressando níveis mais elevados da spEYFP por FACS para mapeamento de regiões genômicas com atividade transcricional elevada e estável nas células CHO-DG44. Além disso, também empregamos clones celulares isolados de células CHO-DG44 transfectadas estavelmente com o mesmo vetor lentiviral como referência dos níveis de expressão spEYFP associados a um maior número de cópias do vetor de expressão integrado no genoma das células CHO-DG44. Empregou-se um vetor lentiviral bicistrônico de terceira geração derivado do vírus HIV. A escolha desta abordagem está fundamentada numa característica bem estudada destes vírus, que é a capacidade de inserção preferencial do provírus em um subconjunto de genes transcricionalmente ativos⁶⁹.

Uma das características mais importantes relacionada ao emprego das metodologias de transferência gênica é a heterogeneidade clonal quanto aos níveis de produção da proteína de interesse, e que são observados quando se emprega protocolos de transfecção estável. Os vetores de expressão em células de mamíferos frequentemente possuem elementos regulatórios destinados a promover uma expressão constitutiva, e, idealmente, independente do contexto de inserção genômica do gene de interesse. Assim, é comum, por exemplo, a utilização de promotores virais fortes para conduzir a expressão de genes recombinantes em células CHO, como, por exemplo, o promotor do citomegalovírus humano (CMV), a região promotora derivada da repetição terminal longa (LTR) do vírus do sarcoma mieloproliferativo, e promotores derivados do vírus do sarcoma de Rous e do vírus Símio 40 (SV40E)¹⁰⁶. Entretanto, no contexto genômico, a função destes elementos depende de componentes da maquinaria de transcrição celular e de elementos genéticos endógenos próximos ao locus cromossômicos de inserção, os quais podem afetar negativamente a expressão do transgene¹⁹. Sabe-se que o nível de expressão relativo de um determinado transgene de interesse é variado entre os diferentes subclones que compõem uma determinada população celular mista¹⁰⁶. Quando se considera as

estratégias envolvendo metodologias de transfecção estável, a variação nos níveis de expressão do transgene está relacionada a padrões específicos de integração de transgenes, como, por exemplo, inserções de múltiplas cópias do cassete de expressão, muitas vezes organizadas como *loci* complexos, ou a rearranjos genômicos dos sítios de inserção do transgene¹⁰⁸, ou ainda à localização genômica dos eventos de integração, o denominado “efeito de posição”, o qual possui um efeito importante na taxa de transcrição do transgene¹⁰⁹. Estes efeitos posicionais, aliados a variabilidade da capacidade biossintética e secretória associada a sub-fenótipos celulares¹¹⁰ são as principais razões da variedade dos níveis de expressão e a instabilidade da produtividade celular durante o cultivo a longo prazo em clones celulares isolados a partir das estratégias clássicas de manipulação genética envolvendo transfecção celular¹⁰⁸. Neste trabalho, não foi observada uma variabilidade estatisticamente significativa nos níveis de expressão da proteína repórter utilizada, a spEYFP, nos diferentes clones avaliados. Embora tenha se avaliado um número muito limitado de clones neste trabalho, apenas oito no caso da estratégia de transdução viral, e apenas um no caso da estratégia envolvendo a transfecção plasmidial estável, é possível que esta baixa variabilidade clonal quanto aos níveis de expressão da spEYFP reflita o tropismo das sequências provirais derivadas dos lentivírus por regiões genômicas ativas transcricionalmente⁸⁷. Outro fator que pode ter contribuído para a baixa heterogeneidade quanto aos níveis de expressão da spEYFP entre os diferentes clones analisados foi a etapa de seleção por FACS, indicando a efetividade desta abordagem.

A produtividade e estabilidade são dois fatores cruciais na etapa de desenvolvimento de uma plataforma de expressão de biofármacos⁸². Por isso, boa parte do esforço e tempo despendido na fase de estabelecimento de uma plataforma de expressão estão relacionados na geração de populações celulares clonais com perfil de produção reprodutível e confiável¹¹². Os marcadores de seleção mais comumente empregados durante o processo de obtenção de um clone celular produtor são as regiões codificadoras da enzimas DHFR e Glutamina Sintetase (GS), que permitem que se explore o fenômeno da amplificação gênica⁴⁶. Protocolos de amplificação gênica envolvendo o uso de concentrações crescentes de metotrexato (MTX), inibidor do DHFR, e a metionina sulfoximina (MSX), inibidora da GS, tornam possível a obtenção da amplificação do número de cópias do gene recombinante para centenas a milhares de cópias/genoma, com aumentos expressivos nos níveis de produção da proteína recombinante de interesse⁴⁶. No presente trabalho, selecionamos um número limitado de clones celulares, sendo que foi relativamente simples isolar clones celulares contendo uma única cópia do cassete de expressão bicistrônico da spEYFP integrada no genoma das células CHO-DG44, e associada

a níveis de expressão da spEYFP estáveis e similares aos observados em um clone celular contendo aproximadamente 45 cópias do mesmo elemento de expressão. É razoável especular que a triagem de um número mais elevado de clones permitiria a identificação de sítios genômicos associados a níveis de expressão da proteína recombinante de interesse ainda mais elevados. No entanto, é improvável que um único sítio genômico suporte níveis de expressão de uma proteína recombinante determinada equivalentes àqueles que poderiam ser obtidos com centenas ou milhares de cópias/genoma comumente obtidas quando se emprega estratégias de amplificação gênica. A exploração de múltiplos “hotspots” genômicos provavelmente será necessário para se alcançar tais níveis de expressão. Nesse sentido, Guidukov e seus colegas demonstraram o aumento linear dos níveis de expressão de um anticorpo monoclonal recombinante associado ao uso de três “hotspots” genômicos diferentes em combinação com o uso de várias cópias transgênicas, destacando a importância do mapeamento empírico ou teórico de sítios genômicos adicionais¹¹³.

Como repórter para o mapeamento destes sítios genômicos ativos transcricionalmente, empregou-se uma variante artificial secretada da proteína fluorescente EYFP. A EYFP é uma proteína intracelular monomérica e de baixo peso molecular. A EYFP foi derivada a partir de uma forma mutante da proteína GFP, e apresenta uma maturação mais rápida em comparação com a proteína de origem, assim como um deslocamento no espectro de emissão em direção ao vermelho, com pico em 528 nm, e uma maior intensidade do brilho em células vivas¹¹⁴. A EYFP é relativamente sensível ao pH ácido, e degrada-se em processo induzido pelo íon cloreto (Cl⁻), sendo esses problemas exacerbados quando é expressa à 37°C e/ou direcionada a certas organelas¹¹⁵. Ao ser sintetizada, precisa amadurecer adequadamente antes de emitir fluorescência, passando por duas etapas para a sua maturação: dobramento e formação do cromóforo¹¹⁵. Como a proteína EYFP é intracelular, fato este que dificultaria a avaliação mais rigorosa dos seus níveis de expressão, optou-se por modificá-la para que fosse secretada no meio de cultura, e desta forma impedir grandes acúmulos desta proteína no meio intracelular, já que existem relatos indicando a toxicidade celular associada a superexpressão e acúmulo intracelular da EGFP^{103,115}. A fusão da região codificadora da proteína EYFP com o peptídeo sinal da cadeia leve do fator VIII humano foi feita através de mutação dirigida por PCR. Curiosamente, não foi possível a detecção por fluorimetria da proteína spEYFP nos meios de cultivos condicionados pelas populações celulares mistas e clones celulares obtidos a partir da transfecção ou transdução das células CHO-DG44. Desconhece-se os motivos que impediram a detecção da fluorescência da spEYFP diretamente no meio de cultivo condicionado, mas

suspeita-se que seja devido à uma provável supressão (“quenching”) do sinal fluorescente da spEYFP por compostos presentes no meio de cultivo ou, talvez, por uma incapacidade da técnica de se detectar baixas concentrações da proteína spEYFP presente nas amostras de meios condicionados. A detecção da secreção da proteína fluorescente spEYFP se tornou possível após o fracionamento das amostras de meio condicionado por SDS-PAGE 10%, seguido por detecção da fluorescência por escaneamento a laser do gel. Com o emprego desta metodologia, constatou-se que a proteína spEYFP é secretada para o meio de cultivo das células HEK-293T transfectadas transitoriamente, ou nas populações mistas das células CHO transfectadas e transduzidas estavelmente com o vetor pLV-spEYFP-Dhfr (Figura 9).

Neste trabalho, não se avaliou de forma sistemática uma possível relação entre os níveis intracelulares de fluorescência e a quantidade acumulada de spEYFP nos meios de cultivos condicionados pelos clones avaliados. Como a fluorescência intracelular detectada através de microscopia de fluorescência estaria associada à detecção da forma madura da spEYFP, a expectativa era de que altos níveis de fluorescência intracelular estariam associados a uma maior capacidade de expressão e secreção da spEYFP. Foi nessa premissa, inclusive, que nos baseamos para incluir a etapa de enriquecimento de clones super-produtores por FACS. Entretanto, observou-se que parte dos clones selecionados apresentavam, aparentemente, níveis mais elevados de intensidade de fluorescência intracelular, sendo que não havia uma correspondência evidente com os níveis acumulados de spEYFP no meio condicionado. O processo de secreção de proteínas pelas células de mamíferos é complexo, e envolve a translocação das cadeias nascentes de polipeptídeos do citosol para o lúmen do retículo endoplasmático (ER – Endoplasmic Reticulum), através da sua ligação com a partícula de reconhecimento de sinal (SRP – Signal Recognition Particle), e é no interior do ER onde ocorre o enovelamento das proteínas que serão direcionadas através da via secretória¹¹⁶. Adicionalmente, cada sequência de peptídeo sinal promove a secreção de proteínas de forma variável, possivelmente porque um determinado peptídeo sinal nativo pode não funcionar corretamente fora do seu contexto nativo, ou, ainda, devido a diferenças ligadas à célula hospedeira ou ao próprio processo de secreção¹¹⁶. Por exemplo, existem relatos relacionando a disfunção da via secretória associada a deficiências no funcionamento do complexo SRP, acarretando na retenção do peptídeo sinal, e gerando assim um acúmulo dos polipeptídeos na forma de corpos de inclusões precipitados no ER, e, conseqüentemente, a baixos níveis de secreção¹¹⁷. Desta forma, é possível que altos níveis de fluorescência intracelular não estejam associados necessariamente a altos níveis de secreção da proteína repórter fluorescente,

possivelmente por uma alteração ou deficiência da via de secreção. Por outro lado, em parte dos clones, a baixa intensidade de fluorescência intracelular aparenta ser um indicativo de altos níveis de expressão e secreção da proteína repórter. Um destes exemplos seria o clone ST1, que apresentou baixos níveis aparentes de fluorescência intracelular acompanhados de bons níveis de secreção da spEYFP, indicando um trânsito eficiente da spEYFP pela via secretória.

Quanto à determinação dos sítios genômicos de integração associados aos clones isolados pela estratégia de transdução estável, experimentamos várias dificuldades na metodologia empregada (DNA “Walking”), possivelmente devido à limitada capacidade da técnica de detectar eventos relativamente raros, associado a um baixo número de cópias, ou ainda a dificuldades técnicas associadas aos reagentes utilizados, como, por exemplo, a processividade e atividade da enzimas Taq polimerase ou da Ligase utilizadas¹¹⁸. A identificação de sequências de nucleotídeos desconhecidas flanqueadas por regiões de DNA já caracterizadas pode ser realizada por vários métodos baseados em PCR, comumente conhecidos como “Genome Walking” (GW)^{119,120}. Uma variante destas abordagens, denominada de método R-GW (Restriction-based GW), requer a digestão preliminar do DNA genômico por enzimas de restrição adequadas, cujos sítios devem estar localizados a uma distância adequada da região onde ocorre a junção entre as sequências conhecidas e desconhecidas, sendo que esses fragmentos de DNA são posteriormente auto-circularizados ou ligados a adaptadores projetados especificamente¹¹⁹. Esta estratégia tem sido empregada em um grande número de investigações para identificar e caracterizar os sítios de inserção no genoma humano de sequências provirais retrovirais ou provenientes de vetores derivados de retrovirus¹¹⁹. Optamos pela estratégia de “Genome” ou DNA “Walking” para o mapeamento das regiões de inserção genômica do cassete de expressão associadas aos clones isolados, sendo que obtivemos, na nossa primeira tentativa, apenas produtos inespecíficos após as reações de “Nested” PCR (Figura 15). Problemas similares, associados à uma baixa sensibilidade e eficiência da metodologia de GW foram relatados na literatura¹²⁰. Nas tentativas seguintes de se padronizar a metodologia de GW, empregamos o próprio vetor pLV-spEYFP-DHFR, que apresenta em sua sequência sítios únicos ou duplos de reconhecimento para as enzimas de restrição utilizadas, e aliado ao fato de dispormos da sua sequência nucleotídica teórica completa, permitiu ter um controle mais rígido de cada uma das etapas da estratégia de GW. Inicialmente, foram usados 500 ng do vetor para a execução do protocolo de GW, sendo que experimentamos o mesmo problema de inespecificidade nas reações de PCR (dados não mostrados). Desta forma, a quantidade do vetor foi aumentada para 1 µg, o que nos permitiu a obtenção do resultado esperado (Figura 18). O

uso do vetor pLV-spEYFP-DHFR possibilitou a identificação da etapa de purificação dos fragmentos de DNA, gerados pela digestão com as enzimas de restrição, por adsorção do DNA à sílica como o principal gargalo da técnica (Figura 17). A retirada da etapa de purificação dos fragmentos de DNA por adsorção à sílica permitiu que se aumentasse a disponibilidade de fragmentos de DNA para ligação ao adaptador, uma etapa inerentemente ineficiente¹¹⁸. De uma forma geral, a perda dos fragmentos de DNA ao longo do procedimento explicaria o nosso insucesso nas tentativas iniciais. Adicionalmente, um complicante adicional seria a própria abundância relativa dos sítios genômicos de inserção, pois, considerando que os clones isolados possuem uma única cópia integrada genômica do vetor, 1 µg de gDNA dos clones, em teoria, apresentaria cerca de 5×10^6 cópias do sítio de inserção, caso a eficiência de purificação dos fragmentos de gDNA fosse 100%. A título de comparação, nas condições que tivemos sucesso, empregando-se 1 µg do vetor, esta quantidade representaria em teoria cerca de 10^{13} cópias da sequência alvo, indicando claramente que a sensibilidade da metodologia também foi um fator limitante importante. Nós aparentemente obtivemos êxito quando os meios reacionais das reações de digestão com as enzimas de restrição foram concentradas por um fator de 10 vezes antes de serem empregada nas próximas etapas, o que nos permitiu a obtenção de uma banda de DNA definida após a etapa de “Nested” PCR no caso do clone SV4 (Figura 19).

Finalmente, a implementação de novas estratégias de geração de plataformas de expressão de biofármacos, seja através da modificação de vetores de expressão, ou pela engenharia de vetores e bioprocessos, ou ainda através do emprego de abordagens envolvendo a metodologia de amplificação gênica ou técnicas moleculares de integração sítio-dirigida, tem se traduzido em aumentos expressivos nos rendimentos de produção de proteínas recombinantes terapêuticas¹²¹. A nossa proposta de utilizar vetores lentivirais para o mapeamento de sítios genômicos nas células CHO-DG44 associados a níveis de produção estáveis e elevados para seu posterior uso como plataformas de expressão, associado a técnicas de edição genômica, como, por exemplo CRISPR/Cas9, vem a se juntar a estas correntes destinadas a aumentar a produtividade e o rendimento das plataformas de produção de biofármacos baseadas em células de mamíferos. Concluindo, nós descrevemos a primeira tentativa de mapeamento de sítios genômicos transcricionalmente ativos e estáveis em células CHO-DG44, e que poderiam ser explorados como possíveis alvos para a inserção dirigida de transgenes de interesse, complementando esforços semelhantes em outras linhagens celulares CHO comercialmente importantes, como CHO-K1REF e CHO-SREF. A crescente disponibilidade de sítios genômicos de integração favoráveis e bem caracterizados constituirá

as bases para abordagens novas e inovadoras que visam o desenvolvimento e engenharia de linhagens celulares para a produção de proteínas recombinantes.

6. CONCLUSÕES

- Isolou-se clones celulares contendo uma única cópia do vetor de expressão bicistrônico da spEYFP integrada no genoma das células CHO-DG44, e associada a níveis de expressão da spEYFP estáveis e similares aos observados em um clone celular contendo aproximadamente 45 cópias do mesmo elemento de expressão.

- Obteve-se o vetor lentiviral bicistrônico de expressão (pLV-spEYFP-DDHFR) de uma forma variante secretada da EYFP (spEYFP). As análises realizadas por transfecção transitória da linhagem celular HEK293T demonstraram a funcionalidade do vetor quanto a expressão e secreção da spEYFP (análises de microscopia de fluorescência e SDS-PAGE seguida de escaneamento a laser) e produção de partículas lentivirais recombinantes (experimentos de transdução viral e microscopia de fluorescência);

- A transdução das células CHO-DG44 com partículas lentivirais empregando-se um MOI de 0.5 permitiu a obtenção eficiente de clones celulares associados a eventos únicos de integração do vetor de expressão no genoma das células CHO-DG44.

BIBLIOGRAFIA*

1. Kim JY, Kim Y-GG, Lee GM. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl Microbiol Biotechnol.* 2012;93(3):917–30. doi: 10.1007/s00253-011-3758-5.
2. Altamirano C, Berrios J, Vergara M, Becerra S. Advances in improving mammalian cells metabolism for recombinant protein production. *Electron J Biotechnol.* 2013;16(3). doi: 10.2225/vol16-issue3-fulltext-2.
3. Andersen DC, Krummen L. Recombinant protein expression for therapeutic applications. *Curr Opin Biotechnol.* 2002;13(2):117–23. doi: 10.1016/s0958-1669(02)00300-2 .
4. La Merie Business Intelligence. Blockbuster Biologics 2018: Sales of Recombinant Therapeutic Antibodies & Proteins[Internet]. 2019 mar. [citado em 30 set. 2019]. Disponível em: <https://lamerie.com/report/blockbuster-biologics-2018-sales-of-recombinant-therapeutic-antibodies-proteins/>.
5. Wurm F, Wurm M. Cloning of CHO Cells, Productivity and Genetic Stability—A Discussion. *Processes.* 2017;5(4):20. doi:10.3390/pr5020020.
6. Chen C, Le H, Goudar CT. Integration of systems biology in cell line and process development for biopharmaceutical manufacturing. *Biochem Eng J.* 2016;107:11–7. doi: 10.1016/j.bej.2015.11.013.
7. Derouazi M, Martinet D, Besuchet Schmutz N, Flaction R, Wicht M, Bertschinger M, et al. Genetic characterization of CHO production host DG44 and derivative recombinant cell lines. *Biochem Biophys Res Commun.* 2006;340(4):1069–77. doi: 10.1016/j.bbrc.2005.12.111.
8. Barnes LM, Bentley CM, Dickson AJ. Stability of protein production from recombinant mammalian cells. *Biotechnol Bioeng.* 2003;81(6):631–9. doi: 10.1002/bit.10517.
9. Walsh G. Biopharmaceutical benchmarks 2018. *Nat Biotechnol.* 2018;36(12):1136–45. doi: 10.1038/nbt.4305.
10. Hacker DL, Balasubramanian S. Recombinant protein production from stable mammalian cell lines and pools. *Curr Opin Struct Biol.* 2016;38(1):129–36. doi: 10.1016/j.sbi.2016.06.005.
11. Kim Y-G. Omics-Based CHO Cell Engineering – Entrance into Post-Genomic Era. *Adv*

*De acordo com: International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to Biomedical Journal: sample references. [Cited 2011 Jul 15]. Available from: <http://www.icmje.org>

- Genet Eng Biotechnol. 2012;01(01). doi: 10.4172/2324-9021.1000e103.
12. Fomina-Yadlin D, Mujacic M, Maggiora K, Quesnell G, Saleem R, McGrew JT. Transcriptome analysis of a CHO cell line expressing a recombinant therapeutic protein treated with inducers of protein expression. *J Biotechnol.* 2015;212:106–15. doi: 10.1016/j.jbiotec.2015.08.025.
 13. Bandaranayake AD, Almo SC. Recent advances in mammalian protein production. *FEBS Letters.* 2014. 588:253–60. doi: 10.1016/j.febslet.2013.11.035.
 14. Baser B, Spehr J, Büssow K, van den Heuvel J. A method for specifically targeting two independent genomic integration sites for co-expression of genes in CHO cells. *Methods.* 2015;95:3–12. doi: 10.1016/j.ymeth.2015.11.022.
 15. Papantonis A, Cook PR. Genome architecture and the role of transcription. *Curr Opin Cell Biol.* 2010;22(3):271–6. doi: 10.1016/j.ceb.2010.03.004.
 16. Chakalova L, Debrand E, Mitchell JA, Osborne CS, Fraser P. Replication and transcription: Shaping the landscape of the genome. *Nat Rev Genet.* 2005;6(9):669–77. doi: 10.1038/nrg1673 .
 17. Bode J, Winkelmann S, Götze S, Spiker S, Tsutsui K, Bi C, et al. Correlations between Scaffold/Matrix Attachment Region (S/MAR) Binding Activity and DNA Duplex Destabilization Energy. *J Mol Biol.* 2006;358(2):597–613. doi: 10.1016/j.jmb.2005.11.073.
 18. Cremer M, Hase J V., Volm T, Brero A, Kreth G, Walter J, et al. Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosom Res.* 2001;9(7):541–67. doi: 10.1023/A:1012495201697 .
 19. Brown AJ, James DC. Precision control of recombinant gene transcription for CHO cell synthetic biology. *Biotechnology Advances.* 2016;34:492–503. doi: 10.1016/j.biotechadv.2015.12.012.
 20. Platts AE, Quayle AK, Krawetz SA. In-silico prediction and observations of nuclear matrix attachment. *Cell Mol Biol Lett.* 2006;11(2):191–213. doi: 10.2478/s11658-006-0016-4.
 21. Jackson DA, Iborra FJ, Manders EMM, Cook PR. Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol Biol Cell.* 1998;9(6):1523–36. doi: 10.1091/mbc.9.6.1523.
 22. Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, Debrand E, et al. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet.* 2004;36(10):1065–71. doi: 10.1038/ng1423 .
 23. Neusser M, Schubel V, Koch A, Cremer T, Müller S. Evolutionarily conserved, cell type and species-specific higher order chromatin arrangements in interphase nuclei of primates. *Chromosoma.* 2007;116(3):307–20. doi: 10.1007/s00412-007-0099-3.

24. Schneider R, Grosschedl R. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes Dev.* 2007;21(23):3027–43. doi: 10.1101/gad.1604607.
25. Phi-Van L, Strätling WH. Dissection of the ability of the chicken lysozyme gene 5' matrix attachment region to stimulate transgene expression and to dampen position effects. *Biochemistry.* 1996;35(33):10735–42. doi: 10.1021/bi9603783.
26. Wang Y, DeMayo FJ, Tsai SY, O'Malley BW. Ligand-inducible and liver-specific target gene expression in transgenic mice. *Nat Biotechnol.* 1997;15(3):239–43. doi: 10.1038/nbt0397-239.
27. Alami R. beta-globin YAC transgenes exhibit uniform expression levels but position effect variegation in mice. *Hum Mol Genet.* 2000;9(4):631–6. doi: 10.1093/hmg/9.4.631.
28. Barnes LM, Dickson AJ. Mammalian cell factories for efficient and stable protein expression. *Curr Opin Biotechnol.* 2006;17(4):381–6. doi: 10.1016/j.copbio.2006.06.005.
29. O'Gorman S, Fox DT, Wahl GM. Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science.* 1991;251(4999):1351 LP – 1355. doi: 10.1126/science.1900642.
30. Mirkovitch J, Mirault M-E, Laemmli UK. Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell.* 1984;39(1):223–32. doi: 10.1016/0092-8674(84)90208-3.
31. Wang T-Y, Yang R, Qin C, Wang L, Yang X-J. Enhanced expression of transgene in CHO cells using matrix attachment region. *Cell Biol Int.* 2008;32(10):1279–83. doi: 10.1016/j.cellbi.2008.07.014.
32. Wang T-Y, Zhang J-H, Jing C-Q, Yang X-J, Lin J-T. Positional effects of the matrix attachment region on transgene expression in stably transfected CHO cells. *Cell Biol Int.* 2010;34(2):141–5. doi: 10.1042/CBI20090017.
33. Martelli AM, Bortul R, Fackelmayer FO, Tazzari PL, Bareggi R, Narducci P, et al. Biochemical and morphological characterization of the nuclear matrix from apoptotic HL-60 cells. *J Cell Biochem.* 1999;72(1):35–46. doi: 10.1002/(SICI)1097-4644(19990101)72:1<35::AID-JCB5>3.0.CO;2-S.
34. Girod P-A, Nguyen D-Q, Calabrese D, Puttini S, Grandjean M, Martinet D, et al. Genome-wide prediction of matrix attachment regions that increase gene expression in mammalian cells. *Nat Methods.* 2007;4(9):747–53. doi: 10.1038/nmeth1076.
35. Zahn-Zabal M, Kobr M, Girod PA *et al.* Development of stable cell lines for production or regulated expression using matrix attachment regions. *J. Biotechnol.* 2001;87: 29–42. doi: 10.1016/S0168-1656(00)00423-5.
36. Arumugam PI, Scholes J, Perelman N, Xia P, Yee J-K, Malik P. Improved Human β -

- globin Expression from Self-inactivating Lentiviral Vectors Carrying the Chicken Hypersensitive Site-4 (cHS4) Insulator Element. *Mol Ther.* 2007;15(10):1863–71. doi: 10.1038/sj.mt.6300259.
37. Majocchi S, Aritonovska E, Mermoud N. Epigenetic regulatory elements associate with specific histone modifications to prevent silencing of telomeric genes. *Nucleic Acids Res.* 2014;42(1):193–204. doi: 10.1093/nar/gkt880.
 38. Kang SY, Kim YG, Kang S, Lee HW, Lee EG. A novel regulatory element (E77) isolated from CHO-K1 genomic DNA enhances stable gene expression in Chinese hamster ovary cells. *Biotechnol J.* 2016;11(5):633–41. doi: 10.1002/biot.201500464.
 39. Puck TT. Development of the Chinese Hamster Ovary (CHO) Cell for Use in Somatic Cell Genetics. In *Molecular Cell Genetics*; Gottesman, M.M., Ed.; JohnWiley and Sons: New York, NY, USA, 1985:37–64.
 40. Urlaub G, Chasin LA. Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. *Proc Natl Acad Sci U S A.* 1980;77(7):4216–20. doi: 10.1073/pnas.77.7.4216.
 41. Birch JR, Onakunle Y. Biopharmaceutical Proteins: Opportunities and Challenges. In: Smales CM, James DC, editors. *Therapeutic Proteins*. New Jersey: Humana Press; 2005:001–16. doi: 10.1385/1-59259-922-2:001.
 42. Lai T, Yang Y, Ng SK. Advances in mammalian cell line development technologies for recombinant protein production. *Pharmaceuticals.* 2013;6(5):579–603. doi: 10.3390/ph6050579.
 43. Ghaderi D, Zhang M, Hurtado-Ziola N, Varki A. Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation. *Biotechnol Genet Eng Rev.* 2012;28:147–75. PMID: 22616486.
 44. Lim Y, Wong NSC, Lee YY, Ku SCY, Wong DCF, Yap MGS. Engineering mammalian cells in bioprocessing – current achievements and future perspectives. *Biotechnol Appl Biochem.* 2010;55(4):175–89. doi: 10.1042/BA20090363.
 45. Ray M, Mohandas T. Proposed banding nomenclature for the Chinese hamster chromosomes (*Cricetulus griseus*). *Cytogenet Cell Genet.* 1976;16(1–5):83–91. doi: 10.1159/000130559.
 46. Omasa T. Gene amplification and its application in cell and tissue engineering. *J Biosci Bioeng.* 2002;94(6):600–5. doi: 10.1016/S1389-1723(02)80201-8.
 47. Siminovitch L. On the nature of heritable variation in cultured somatic cells. *Cell.* 1976;7(1):1–11. doi: 10.1016/0092-8674(76)90249-X.
 48. Lewis NE, Liu X, Li Y, Nagarajan H, Yerganian G, O'Brien E, et al. Genomic landscapes of Chinese hamster ovary cell lines as revealed by the *Cricetulus griseus* draft genome. *Nat Biotechnol.* 2013;31(8). doi: 10.1038/nbt.2624.

49. Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, et al. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat Biotechnol.* 2011;29(8):735–41. doi: 10.1038/nbt.1932.
50. Urlaub G, Käs E, Carothers AM, Chasin LA. Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells. *Cell.* 1983;33(2):405–12. doi: 10.1016/0092-8674(83)90422-1.
51. Kaas C, Kristensen C, Betenbaugh MJ, Andersen M. Sequencing the CHO DXB11 genome reveals regional variations in genomic stability and haploidy. *BMC Genomics.* 2015;16(1):160. doi: 10.1186/s12864-015-1391-x.
52. Wuest DM, Harcum SW, Lee KH. Genomics in mammalian cell culture bioprocessing. *Biotechnol Adv.* 2012;30(3):629–38. doi: 10.1016/j.biotechadv.2011.10.010.
53. Cammack R, Atwood T, Campbell P, Parish H, Smith A, Vella F, et al. *Oxford Dictionary of Biochemistry and Molecular Biology.* Oxford University Press; Oxford, 1997.
54. Kaufman RJ. Selection and coamplification of heterologous genes in mammalian cells. In: *Methods in enzymology.* United States; 1990:537–66. doi:10.1016/0076-6879(90)85044-O.
55. Sautter K & Enenkel B. Selection of high-producing CHO cells using NPT selection marker with reduced enzyme activity. *Biotechnol. Bioeng.* 2005;89: 530–538. doi: 10.1002/bit.20374 .
56. Kaufman RJ. Selection and coamplification of heterologous genes in mammalian cells. In: *Methods in enzymology.* United States; 1990:537–66. doi:10.1016/0076-6879(90)85044-O.
57. Davies SL, Lovelady CS, Grainger RK, Racher AJ, Young RJ, James DC. Functional heterogeneity and heritability in CHO cell populations. *Biotechnol Bioeng.* 2013;110(1):260–74. doi: 10.1002/bit.24621.
58. Yoshikawa T, Nakanishi F, Ogura Y, Oi D, Omasa T, Katakura Y, et al. Amplified Gene Location in Chromosomal DNA Affected Recombinant Protein Production and Stability of Amplified Genes. *Biotechnol Prog.* 2000;16(5):710–5. doi: 10.1021/bp000114e.
59. Woodcock CL, Dimitrov S. Higher-order structure of chromatin and chromosomes. *Curr Opin Genet Dev.* 2001;11(2):130–5. doi: 10.1016/S0959-437X(00)00169-6
60. Workman JL, Kingston RE. Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu Rev Biochem.* 1998;67(1):545–79. doi: 10.1146/annurev.biochem.67.1.545.
61. Berger SL. An embarrassment of niches: the many covalent modifications of histones in transcriptional regulation. *Oncogene.* 2001;20(24):3007–13. doi: 10.1038/sj.onc.1204324.

62. Siegfried Z, Simon I. DNA methylation and gene expression. *Wiley Interdiscip Rev Syst Biol Med.* 2010;2(3):362–71. doi: 10.1002/wsbm.64.
63. Gross G, Hauser H. Heterologous expression as a tool for gene identification and analysis. *J Biotechnol.* 1995;41(2–3):91–110. doi: 10.1016/0168-1656(95)00070-7.
64. Fabre E, Hurt EC. Nuclear transport. *Curr Opin Cell Biol.* 1994;6(3):335–42. doi: 10.1016/0955-0674(94)90023-X.
65. Le Hir H, Nott A, Moore MJ. How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci.* 2003;28(4):215–20. doi: 10.1016/S0968-0004(03)00052-5.
66. Harter C, Wieland F. The secretory pathway: mechanisms of protein sorting and transport. *Biochim Biophys Acta - Rev Biomembr.* 1996;1286(2):75–93. doi: 10.1016/0304-4157(96)00003-2.
67. Abeijon C, Hirschberg CB. Topography of glycosylation reactions in the endoplasmic reticulum. *Trends Biochem Sci.* 1992;17(1):32–6. doi: 10.1016/0968-0004(92)90424-8
68. Schlake T, Bode J. Use of Mutated FLP Recognition Target (FRT) Sites for the Exchange of Expression Cassettes at Defined Chromosomal Loci. *Biochemistry.* 1994;33(43):12746–51. doi: 10.1021/bi00209a003.
69. Louwerse JD, van Lier MCM, van der Steen DM, de Vlaam CMT, Hooykaas PJJ, Vergunst AC. Stable Recombinase-Mediated Cassette Exchange in Arabidopsis Using *Agrobacterium tumefaciens*. *Plant Physiol.* 2007;145(4):1282–93. doi: 10.1104/pp.107.108092.
70. Turan S, Kuehle J, Schambach A, Baum C, Bode J. Multiplexing RMCE: Versatile Extensions of the FLP-Recombinase-Mediated Cassette-Exchange Technology. *J Mol Biol.* 2010;402(1):52–69. doi: 10.1016/j.jmb.2010.07.015.
71. Lauth M. Stable and efficient cassette exchange under non-selectable conditions by combined use of two site-specific recombinases. *Nucleic Acids Res.* 2002;30(21):115e – 115. doi: 10.1093/nar/gnf114.
72. Baer A, Bode J. Coping with kinetic and thermodynamic barriers: RMCE, an efficient strategy for the targeted integration of transgenes. *Curr Opin Biotechnol.* 2001;12(5):473–80. doi: 10.1016/S0958-1669(00)00248-2.
73. Ow DW. Recombinase-directed plant transformation for the post-genomic era. *Plant Mol Biol.* 2002;48(1–2):183–200. PMID: 11860209.
74. Ahmadi M, Damavandi N, Eidgahi MRA, Davami F. Utilization of site-specific recombination in biopharmaceutical production. *Iran Biomed J.* 2016;20(2):68–76. doi: 10.7508/ibj.2016.02.001.
75. Colcher D, Milenic D, Roselli M, Raubitschek A, Whittle N, Bodmer M, et al. Characterization and Biodistribution of Recombinant and Recombinant / Chimeric Constructs of Monoclonal Antibody B72 . 3. 1989;3(22):1738–45.

76. Chevalier BS, Stoddard BL. Homing endonucleases : structural and functional insight into the catalysts of intron / intein mobility. *Nucleic Acids Research* . 2001;29(18):3757–74. doi: 10.1093/nar/29.18.3757.
77. Datta P, Linhardt RJ, Sharfstein ST. An 'omics approach towards CHO cell engineering. *Biotechnol Bioeng*. 2013;110(5):1255–71. doi: 10.1002/bit.24841.
78. Zhu J, Park CW, Sjeklocha L, Kren BT, Steer CJ. High-Level Genomic Integration, Epigenetic Changes, and Expression of Sleeping Beauty Transgene. *Biochemistry*. 2010;49(7):1507–21. doi: 10.1021/bi9016846.
79. Baser B, Spehr J, Büsow K, van den Heuvel J. A method for specifically targeting two independent genomic integration sites for co-expression of genes in CHO cells. *Methods*. 2015;95:3–12. doi: 10.1016/j.ymeth.2015.11.022.
80. Ma D, Liu F. Genome Editing and Its Applications in Model Organisms. *Genomics Proteomics Bioinformatics*. 2016;13(6):336–44. doi: 10.1016/j.gpb.2015.12.001.
81. Ronda C, Pedersen LE, Hansen HG, Kallehauge TB, Betenbaugh MJ, Nielsen AT, et al. Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. *Biotechnol Bioeng*. 2014;111(8):1604–16. doi: 10.1002/bit.25233.
82. Zhao M, Wang J, Luo M, Luo H, Zhao M, Han L, et al. Rapid development of stable transgene CHO cell lines by CRISPR/Cas9-mediated site-specific integration into C12orf35. *Appl Microbiol Biotechnol*. 2018;102(14):6105–17. doi: 10.1007/s00253-018-9021-6.
83. Kawabe Y, Komatsu S, Komatsu S, Murakami M, Ito A, Sakuma T, et al. Targeted knock-in of an scFv-Fc antibody gene into the hprt locus of Chinese hamster ovary cells using CRISPR/Cas9 and CRIS-PITCh systems. *J Biosci Bioeng*. 2018;125(5):599–605. doi: 10.1016/j.jbiosc.2017.12.003.
84. Cheng JK, Lewis AM, Kim DS, Dyess T, Alper HS. Identifying and retargeting transcriptional hot spots in the human genome. *Biotechnol J*. 2016;11(8):1100–9. doi: 10.1002/biot.201600015.
85. Irion S, Luche H, Gadue P, Fehling HJ, Kennedy M, Keller G. Identification and targeting of the ROSA26 locus in human embryonic stem cells. *Nat Biotechnol*. 2007;25(12):1477–82. doi: 10.1038/nbt1362.
86. Schröder ARW, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. *Cell*. 2002;110(4):521–9. doi: 10.1016/S0092-8674(02)00864-4.
87. Lewinski MK, Bisgrove D, Shinn P, Chen H, Hoffmann C, Hannenhalli S, et al. Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. *J Virol*. 2005;79(11):6610–9. doi: 10.1128/JVI.79.11.6610-6619.2005.

88. Bushman F. Retroviral integration and human gene therapy. *J Clin Invest.* 2007;117(8):1–4. doi: 10.1172/JCI32949.based.
89. Kollewe. Production of recombinant proteins in insect cells. *Am J Biochem Biotechnol.* 2013;9(3):255–71. doi: 10.3844/ajbbsp.2013.255.271.
90. Qasim W, Vink CA, Thrasher AJ. Hybrid Lentiviral Vectors. *Mol Ther.* 2010;18(7):1263–7. doi: 10.1038/mt.2010.76.
91. Hacker DL & Balasubramanian S. Recombinant protein production from stable mammalian cell lines and pools. *Curr. Opin. Struct. Biol.* 2016;38: 129–136. doi: 10.1016/j.sbi.2016.06.005.
92. Oberbek A, Matasci M, Hacker DL, Wurm FM. Generation of stable, high-producing cho cell lines by lentiviral vector-mediated gene transfer in serum-free suspension culture. *Biotechnol Bioeng.* 2011;108(3):600–10. doi: 10.1002/bit.22968.
93. Brown AJ, James DC. Precision control of recombinant gene transcription for CHO cell synthetic biology. *Biotechnology Advances.* 2016;34:492–503. doi: 10.1016/j.biotechadv.2015.12.012.
94. Kuystermans D, Krampe ÆB, Al-rubeai HSÆM. Using cell engineering and omic tools for the improvement of cell culture processes. *Cytotechnology.* 2007:3–22. doi: 10.1007/s10616-007-9055-6.
95. Kim Y-G, Han YK, Kim JY, Lee EG, Lee HW, Lee GM. Effect of constitutively active ras overexpression on cell growth in recombinant chinese hamster ovary cells. *Biotechnol Prog.* 2011;27(2):577–80. doi: 10.1002/btpr.567.
96. Freshney RI. *Culture of Animal Cells. Culture of Animal Cells:A Manual of Basic Techniques and Specialized Application.* Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2010;346. 796 p.
97. Urlaub G, Mitchell PJ, Kas E, Chasin LA, Funanage VL, Myoda TT, et al. Effect of gamma rays at the dihydrofolate reductase locus: Deletions and inversions. *Somat Cell Mol Genet.* 1986;12(6):555–66. doi: 10.1007/BF01671941.
98. Demasi MA, de S. Molina E, Bowman-Colin C, Lojudice FH, Muras A, Sogayar MC. Enhanced Proteolytic Processing of Recombinant Human Coagulation Factor VIII B-Domain Variants by Recombinant Furins. *Mol Biotechnol.* 2016;58(6):404–14. doi: 10.1007/s12033-016-9939-9.
99. Colin C, Tobaruella FS, Correa RG, Sogayar MC, Demasi MA. Cloning and characterization of a novel alternatively spliced transcript of the human CHD7 putative helicase. *BMC Res Notes.* 2010;3(1):252. doi: 10.1186/1756-0500-3-252.
100. Tiscornia G, Singer O, Verma IM. Production and purification of lentiviral vectors. *Nat. Protoc.* 2006;1(1):241-5. doi: nprot.2006.37 [pii]r10.1038/nprot.2006.37.

101. Skulj M *et al.* Application of qPCR method for transgene copy number determination in recombinant CHO cells in biopharmaceutical industry. In: biopharmaceutical cell culture and biology molecular. Menges. 2009:1.
102. Maccani A, Ernst W, Grabherr R. Whole Genome Sequencing Improves Estimation Of Nuclear Dna Content Of Chinese Hamster Ovary Cells. *Cytom Part A*. 2013;83(10):893–5. doi: 10.1002/cyto.a.22331.
103. Kalyanaraman B, Zielonka J. Green fluorescent proteins induce oxidative stress in cells: A worrisome new wrinkle in the application of the GFP reporter system to biological systems?. *Redox Biol*. 2017;12:755–7. doi: 10.1016/j.redox.2017.03.019.
104. Gupta P & Lee KH. Genomics and proteomics in process development: opportunities and challenges. *Trends Biotechnol*. 2007;25: 324–330. doi: 10.1016/j.tibtech.2007.04.005.
105. Kim M, O’Callaghan PM, Droms KA, James DC. A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. *Biotechnol. Bioeng*. 2011;108: 2434–2446. doi: 10.1002/bit.23189.
106. Chatellard P, Pankiewicz R, Meier E *et al.* The IE2 promoter/enhancer region from mouse CMV provides high levels of therapeutic protein expression in mammalian cells. *Biotechnol. Bioeng*. 2007;96, 106–117. doi: 10.1002/bit.21172.
107. Girod P & Mermod N. Use of scaffold / matrix-attachment regions for protein production. *Gene Transf. Expr. Mamm. Cells*. 2003:359–379. doi: 10.1016/S0167-7306(03)38022-6.
108. Francis KE & Spiker S. Identification of *Arabidopsis thaliana* transformants without selection reveals a high occurrence of silenced T-DNA integration events. *Plant J*. 2005;41: 464–477. doi: 10.1111/j.1365-313X.2004.02312.x.
109. Mutskov V, Felsenfeld G. 2004. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO J*. 2004;23(1):138–149. doi: 10.1038/sj.emboj.7600013.
110. Sommeregger W, *et al.* Proteomic differences in recombinant CHO cells producing two similar antibody fragments. *Biotechnol. Bioeng*. 2016;113: 1902–1912. doi: 10.1002/bit.25957.
111. Kim JM, Kim JS, Park DH, Kang HS, Yoon J, Baek K, Yoon Y (2004) Improved recombinant gene expression in CHO cells using matrix attachment regions. *J Biotechnol*. 2004;107:95–105. doi:10.1016/j.jbiotec.2003.09.015.
112. Priola JJ, Calzadilla N, Baumann M, Borth N, Tate CG, Betenbaugh MJ. High-throughput screening and selection of mammalian cells for enhanced protein production. *Biotechnol J*. 2016;11(7):853–65. doi: 10.1002/biot.201500579.

113. Gaidukov L, Wroblewska L, Teague B, Nelson T, Zhang X, Liu Y, et al. A multi-landing pad DNA integration platform for mammalian cell engineering. *2018;46(8):4072–86. doi: 10.1093/nar/gky216.*
114. Jayaraman S, Haggie P, Wachter RM, Remington SJ, Verkman AS. Mechanism and cellular applications of a green fluorescent protein-based halide sensor. *J Biol Chem. 2000;275(9):6047-50. doi: 10.1074/jbc.275.9.6047.*
115. Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, Miyawaki A. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat Biotechnol. 2002;20(1):87–90. doi: 10.1038/nbt0102-87.*
116. Nagai K, Oubridge C, Kuglstatter A, Menichelli E, Isel C, Jovine L. Structure, function and evolution of the signal recognition particle. TL-22. *EMBO J. 2003;22(14):3479–85. doi: 10.1093/emboj/cdg337.*
117. Le Fourn V, Girod P-A, Buceta M, Regamey A, Mermoud N. CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion. *Metab Eng. 2014;21:91–102. Doi: 10.1016/J.YMBEN.2012.12.003.*
118. Sun HW, Lolis E. Salvaging recombinants from low-efficiency ligase reactions for more efficient subcloning. *Biotechniques. 1995;18(4):644–6, 648, 650. PMID:7598899.*
119. Leoni C, Volpicella M, De Leo F, Gallerani R, Ceci LR. Genome walking in eukaryotes. *FEBS J. 2011;278(21):3953–77. doi: 10.1111/j.1742-4658.2011.08307.x.*
120. Min G-S, Powell JR. Long-Distance Genome Walking Using the Long and Accurate Polymerase Chain Reaction. *Biotechniques. 1998;24(3):398–400. doi: 10.2144/98243bm13.*
121. Kuo C-C, Chiang AW, Shamie I, Samoudi M, Gutierrez JM, Lewis NE. The emerging role of systems biology for engineering protein production in CHO cells. *Curr Opin Biotechnol. 2018;51:64–9. doi: 10.1016/j.copbio.2017.11.015.*
122. Wiberg FC, Rasmussen SK, Frandsen TP, Rasmussen LK, Tengbjerg K, Coljee VW, et al. Production of target-specific recombinant human polyclonal antibodies in mammalian cells. *Biotechnol Bioeng. 2006; 94(2):396–405. doi: 10.1002/bit.20865.*
123. Hengen P. Vettorette, splinkerette and boomerang DNA amplification. *Trends Biochem Sci. 1995; 20(9):372–3. doi: 10.1016/S0968-0004(00)89079-9.*