

UNIVERSIDADE FEDERAL DO PARANÁ

ARTHUR SCHVEITZER FERREIRA

AVALIAÇÃO BIOFÍSICA E FUNCIONAL DE FATORES DE CRESCIMENTO DA  
FAMÍLIA EGF E FRAGMENTOS DE ANTICORPOS RECOMBINANTES

CURITIBA-PR

2022

ARTHUR SCHVEITZER FERREIRA

AVALIAÇÃO BIOFÍSICA E FUNCIONAL DE FATORES DE CRESCIMENTO DA  
FAMÍLIA EGF E FRAGMENTOS DE ANTICORPOS RECOMBINANTES

TESE DE DOUTORADO apresentada ao PROGRAMA DE  
PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E  
MOLECULAR, setor de CIÊNCIAS BIOLÓGICAS,  
Universidade Federal do Paraná, como parte dos requisitos  
necessários à obtenção do título de DOUTOR EM  
BIOLOGIA CELULAR E MOLECULAR

Orientador: Nilson Ivo Tonin Zanchin

CURITIBA-PR

2022

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)  
UNIVERSIDADE FEDERAL DO PARANÁ  
SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE CIÊNCIAS BIOLÓGICAS

Ferreira, Arthur Schweitzer.

Avaliação biofísica e funcional de fatores de crescimento da família EGF e fragmentos de anticorpos recombinantes. / Arthur Schweitzer Ferreira. – Curitiba, 2022.

1 recurso on-line : PDF.

Orientador: Nilson Ivo Tonin Zanchin.

Tese (Doutorado) – Universidade Federal do Paraná, Setor de Ciências Biológicas. Programa de Pós-Graduação em Biologia Celular e Molecular.

1. Fator de Crescimento Epidérmico. 2. Anticorpos. 3. Escherichia coli. I. Título. II. Zanchin, Nilson Ivo Tonin. III. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Biologia Celular e Molecular.

## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação BIOLOGIA CELULAR E MOLECULAR da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **ARTHUR SCHVEITZER FERREIRA** intitulada: **AVALIAÇÃO BIOFÍSICA E FUNCIONAL DE FATORES DE CRESCIMENTO DA FAMÍLIA EGF E FRAGMENTOS DE ANTICORPOS RECOMBINANTES**, sob orientação do Prof. Dr. NILSON IVO TONIN ZANCHIN, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa. A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 11 de Maio de 2022.

Assinatura Eletrônica

17/05/2022 17:20:26.0

NILSON IVO TONIN ZANCHIN  
Presidente da Banca Examinadora

Assinatura Eletrônica

11/07/2022 18:01:47.0

SHEILA MARIA BROCHADO WINNISCHOFER  
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

18/05/2022 14:38:20.0

MARCELO DE MACEDO BRÍGIDO  
Avaliador Externo (UNIVERSIDADE DE BRASÍLIA)

Assinatura Eletrônica

18/05/2022 10:15:22.0

ADRIANA FROHLICH MERCADANTE  
Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

19/05/2022 07:08:54.0

APARECIDA SADAÉ TANAKA  
Avaliador Externo (UNIVERSIDADE FEDERAL DE SÃO PAULO)



Dedico esse trabalho a todos aqueles que, assim como eu, se sacrificaram e se sacrificam em nome da ciência, da saúde e da humanidade em busca de tempos melhores.

## **AGRADECIMENTOS**

Aos meus pais, Adésia e Maurício, pelos exemplos infinitos de sabedoria e amor, os quais foram essenciais para o desenvolvimento da dita boa ciência.

Aos meus orientadores, Nilson Ivo Tonin Zanchin e Beatriz Gomes Guimarães, por me aceitarem nessa jornada de descobertas e de provações, as quais moldaram o cientista que eu pensava ser.

A Universidade Federal do Paraná e ao Instituto Carlos Chagas, pela estrutura e suporte necessários à realização desse trabalho e ao Programa de Pós-Graduação em Biologia Celular e Molecular da UFPR por essa etapa fundamental da minha formação acadêmica.

Também gostaria de deixar um agradecimento especial a cada um dos meus colegas do LBEP que viveram essa experiência comigo e dividiram o pesado fardo de fazer ciência nos tempos atuais.

Deixo meu agradecimento especial também aos meus amigos que independente da distância puderam ceder suas presenças e compartilhar momentos que guardarei com carinho na lembrança.

E por fim, um agradecimento ao meu companheiro e amor de vida Júnior, por me fortalecer quando eu fraquejei e não me permitiu desistir nessa empreitada.

“Eu, filho do carbono e do amoníaco,  
Monstro de escuridão e rutilância,  
Sofro, desde a epigênese da infância,  
A influência má dos signos do zodíaco.

Profundissimamente hipocondríaco,  
Este ambiente me causa repugnância...  
Sobe-me à boca uma ânsia análoga à ânsia  
Que se escapa da boca de um cardíaco.

Já o verme — este operário das ruínas —  
Que o sangue podre das carnificinas  
Come, e à vida em geral declara guerra,

Anda a espreitar meus olhos para roê-los,  
E há-de deixar-me apenas os cabelos,  
Na frialdade inorgânica da terra!”

(Augusto dos Anjos - Psicologia de um vencido)

## RESUMO

Fatores de crescimento da família EGF e NRG tem importantes papéis na sinalização celular para proliferação, diferenciação e sobrevivência. Todos os membros dessas famílias interagem com receptores ErbB através de seus domínios estruturais EGF. O fato de o domínio EGF dos diferentes fatores de crescimento EGF compartilharem a mesma conformação com três pontes dissulfeto na combinação C1-C3, C2-C4 e C5-C6 e se ligarem a mesma família de receptores apresenta algumas dificuldades tanto para a produção recombinante dos domínios EGF quanto para a sua detecção e identificação específica. Portanto, ferramentas bioquímicas robustas e específicas são necessárias para estudar os diferentes ligantes dessas famílias. Nesse sentido, existem os anticorpos e seus derivados que podem ser obtidos a partir de organismos vivos ou de bibliotecas sintéticas. Para isso, o objetivo deste trabalho foi desenvolver fragmentos de anticorpos do tipo scFv com sequências derivadas de anticorpos de origem animal sequenciados por espectrometria de massas e VHHs selecionados da literatura para o método de apresentação em bactéria. Os fragmentos derivados de anticorpos animais do tipo scFv foram expressos em diferentes construções, mas se mostraram instáveis e de pouco rendimento na avaliação por cromatografia por exclusão de tamanho, DLS supersensível e nanofluorimetria de varredura diferencial. Os anticorpos comerciais que deram origem aos scFvs foram avaliados pelas mesmas técnicas e demonstraram ter diferentes afinidades pelos domínios EGF em ensaios do tipo *dot blot* e qualidades estruturais distintas, tais como estabilidade e agregabilidade, como visto pelo DLS supersensível. O teste de apresentação em superfície bacteriana demonstrou que as construções de VHHs conseguem se ligar ao EGF e que quando o EGF é fusionado a mRFP há um impedimento da ligação. Ademais, a proteína de membrana inaQ apresentou níveis mais elevados de expressão de ambos VHHs propostos, sendo a mais adequada para o ensaio de apresentação em superfície bacteriana quando comparada à proteína Neae. Além disso, neste trabalho também foi desenvolvido um protocolo unificado para produção recombinante dos domínios EGF de sete fatores da família EGF e da NRG1 $\alpha$  em *Escherichia coli*. Estes domínios foram caracterizados por diversas técnicas biofísicas para avaliar sua conformação. Sua atividade foi demonstrada em ensaios funcionais utilizando cultura de células. Este protocolo representa uma solução técnica simples e eficiente para obtenção dos fatores da família EGF, facilitado o desenvolvimento de pesquisas sobre a função de proteínas que contém o domínio EGF.

Palavras-chave: domínio EGF, scFv, VHH, QbD, apresentação em superfície bacteriana

## ABSTRACT

Growth factors of EGF and NRG families have important roles in cell signaling for proliferation, differentiation, and survival. All members of these families interact with ErbB receptors through their EGF structural domains. The fact that the EGF domain of the different EGF growth factors share the same conformation with three disulfide bonds in the combination C1-C3, C2-C4 and C5-C6 and bind to the same type of receptors poses some difficulties both for recombinant production of the EGF domains and for their specific detection and identification. Therefore, robust, and specific biochemical tools are required to study the different ligands of these families. Antibodies and their fragments that can be obtained from living organisms or from synthetic libraries are used for this purpose. Considering this, the objective of this work was to develop scFv antibody fragments with sequences derived from mass spectrometry-sequenced animal antibodies and VHHs selected from the literature for bacterial surface display. scFv derived from animal antibodies were expressed using different constructs. However, expression resulted in either low yield or they turned out to be unstable according to evaluations performed by size exclusion chromatography, supersensitive DLS and differential scanning nanofluorimetry. Commercial antibodies that gave rise to the scFvs were evaluated by the same techniques and presented different affinities for EGF domains in dot blot assays and distinct structural qualities, such as stability and aggregability, as seen by supersensitive DLS. The bacterial surface display test revealed that the VHHs constructs can bind to EGF. However, when fused to mRFP, the EGF domain is no longer recognized by this VHH. The membrane protein inaQ showed higher expression levels for both proposed VHHs tested, presenting higher potential for bacterial surface display as compared to the protein NEAE. In this work, a unified protocol was developed for recombinant production of EGF domains of the EGF and NRG1 $\alpha$  families in *Escherichia coli*. All recombinant EGF domains were characterized using a set of biophysical techniques to verify their conformation and their biochemical activities demonstrated in cell growth and proliferation assays. This protocol provides a simple and efficient technical solution for recombinant production of EGF family domains, facilitating the development of research involving the EGF domain function.

Keywords: EGF domain, scFv, VHH, QbD, bacterial surface display

## LISTA DE ILUSTRAÇÕES

<b>FIGURA 1 - ESTRUTURA DO DOMÍNIO EGF</b> .....	3
<b>FIGURA 2. DOMÍNIOS E MODIFICAÇÕES DOS FATORES DA FAMÍLIA EGF</b> .....	7
<b>FIGURA 3. SEQUÊNCIA AMINOACIDICA DO DOMÍNIO EGF DAS FAMÍLIA EGF E NRG</b> .....	8
<b>FIGURA 4. ANTICORPOS E FRAGMENTOS DE ANTICORPOS</b> .....	10
<b>FIGURA 5. LOCALIZAÇÃO DOS RESÍDUOS DA ZONA DE VERNIER E CDRS</b> .....	14
<b>FIGURA 6. TESTES DE EXPRESÃO INICIAL DOS SCFVS</b> .....	51
<b>FIGURA 7. TESTES DE DIFERENTES TAMPÃO DE LISE</b> .....	52
<b>FIGURA 8. TESTE DE DIFERENTES D.O. NO MOMENTO DA INDUÇÃO DA EXPRESSÃO</b> .....	53
<b>FIGURA 9. TESTE DE DIFERENTES FONTES DE CARBONO NO MEIO DE CULTURA</b> .....	55
<b>FIGURA 10. TESTE DE DIFERENTES FONTES DE NITROGÊNIO OU SAL NO MEIO DE CULTURA</b> .....	56
<b>FIGURA 11. TESTE DE DIFERENTES MEIOS COMERCIAIS</b> .....	57
<b>FIGURA 12. TESTE DE DIFERENTES CONCENTRAÇÕES DE IPTG E TEMPERATURAS DE INDUÇÃO</b> .....	58
<b>FIGURA 13. TESTE DE DIFERENTES TEMPOS DE EXPRESSÃO</b> .....	58
<b>FIGURA 14. TESTE DE CO-EXPRESSÃO DOS SCFVS COM TRX-EGF</b> .....	59
<b>FIGURA 15. TESTE DE MUDANÇA DO TAMPÃO DE LISE E COMBINAÇÃO DE VARIÁVEIS PROMISSORAS DO SCFV OCG</b> .....	60
<b>FIGURA 16. TESTE DE MUDANÇA DO TAMPÃO DE LISE E COMBINAÇÃO DE VARIÁVEIS PROMISSORAS DO SCFV MMG</b> .....	61
<b>FIGURA 17. PURIFICAÇÃO DOS SCFVS FUSIONADOS A SGFP2 E ANÁLISE DE SGFP2 LIVRE</b> .....	62
<b>FIGURA 18. AGRUPAMENTO DE ESTRUTURAS DE SCFVS</b> .....	64
<b>FIGURA 19. SOBREPOSIÇÃO DOS MODELOS DOS SCFVS COM E SEM ALTERAÇÃO NOS RESÍDUOS DE VERNIER</b> .....	66
<b>FIGURA 20. ALINHAMENTO ESTRUTURAL DAS SEQUÊNCIAS DOS scFvs DO GRUPO MM</b> .....	67
<b>FIGURA 21. ALINHAMENTO ESTRUTURAL DAS SEQUÊNCIAS DOS scFvs DO GRUPO OC</b> .....	68
<b>FIGURA 22. EXPRESSÃO FINAL DOS SCFVS</b> .....	69
<b>FIGURA 23. CONFIRMAÇÃO DA EXPRESSÃO DOS SCFVS</b> .....	69
<b>FIGURA 24 CROMATOGRAFIAS DE EXCLUSÃO POR TAMANHO DEMONSTRANDO AGREGADOS SOLÚVEIS</b> .....	70
<b>FIGURA 25. PURIFICAÇÃO DOS SCFVS E IDENTIFICAÇÃO DE INSTABILIDADE</b> .	72
<b>FIGURA 26. PERFIL DE DISPERSIVIDADE DOS ANTICORPOS COMERCIAIS</b> .....	73
<b>FIGURA 27. NANOFUORIMETRIA DE VARREDURA DIFERENCIAL DOS ANTICORPOS COMERCIAIS</b> .....	73
<b>FIGURA 28. ANÁLISE DE SELETIVIDADE DOS ANTICORPOS ANTI-EGF SC E CS FRENTE AOS DIFERENTES DOMÍNIOS EGF</b> .....	74

<b>FIGURA 29. PERFIS DE DISPERSÃO DAS DIFERENTES FRAÇÕES CROMATOGRÁFICAS DO SCFV MM .....</b>	<b>76</b>
<b>FIGURA 30. COEFICIENTE DE DIFUSÃO EM RELAÇÃO A CONCENTRAÇÃO .....</b>	<b>77</b>
<b>FIGURA 31. NANOFLUORIMETRIA DE VARREDURA DIFERENCIAL DO scFv MM77</b>	
<b>FIGURA 32. TRATAMENTO TÉRMICO DOS SCFVS OC, 6CVKOC, 146OC e VERO78</b>	
<b>FIGURA 33. PURIFICAÇÃO DO EGF-MRFP .....</b>	<b>79</b>
<b>FIGURA 34. FLUORESCÊNCIA ANTES E APÓS O CONGELAMENTO DAS CÉLULAS. ....</b>	<b>80</b>
<b>FIGURA 35. CITOMETRIA DE FLUXO DAS CONSTRUÇÕES DE SUPERFÍCIE BACTERIANA INCUBADAS COM EGF-MRFP .....</b>	<b>80</b>
<b>FIGURA 36. TESTE DE INTERAÇÃO ENTRE NB1 E NB6 COM EGF .....</b>	<b>81</b>

## LISTA DE QUADROS

<b>QUADRO 1</b> CLASSIFICAÇÃO GERAL DO DOMÍNIO EGF.....	4
<b>QUADRO 2</b> SEQUÊNCIA DE OLIGONUCLEOTÍDEOS .....	20
<b>QUADRO 3</b> MISTURA DA PCR.....	25
<b>QUADRO 4</b> CICLOS DA PCR.....	25
<b>QUADRO 5</b> VARIÁVEIS DAS CONDIÇÕES DE EXPRESSÃO DOS TESTES .....	28
<b>QUADRO 6.</b> ESTRATÉGIA DE AQUISIÇÃO DE DADOS POR ESPECTROMETRIA DE MASSAS .....	34
<b>QUADRO 7.</b> CLASSIFICAÇÃO DE ESTRUTURAS CRISTALOGRÁFICAS E MODELOS EM GRUPOS .....	65



## LISTA DE ABREVIATURAS E SIGLAS

AREG – Anfiregulina

BTC - Betacelulina

CDR – Região determinante de complementariedade

CH – Domínio constante pesado

CL – Domínio constante leve

DNA – ácido desoxirribonucleico

D.O. – Densidade óptica

EGF – Fator de crescimento epidérmico

EPGN - Epigene

EPR – Epiregulina

Fab - Fragmento de ligação ao antígeno

Fc – Fração constante

Fv – Fração variável

HBEGF - Fator de crescimento do tipo heparina ligadora de EGF

HCAb – Anticorpo de apenas cadeias pesadas

MM – Sequência derivada do anticorpo comercial de origem de camundongo

NB – *Nanobody*

NRG1 $\alpha$  – Neuroregulina 1 alfa

OC – Sequência derivada do anticorpo comercial de origem de coelho

scFv - Fragmento de domínios variáveis de cadeia única

sdAf - Fragmento de anticorpo de domínio único

SGFP2 – *Superfolder green fluorescence protein*

TGF $\alpha$  – Fator de crescimento tumoral alfa

VH – Domínio variável pesado

VHH - Domínio variável de cadeia pesada de anticorpos de apenas cadeias pesadas

VL – Domínio variável leve

# SUMÁRIO

1. INTRODUÇÃO.....	1
2. REVISÃO DA LITERATURA.....	3
2.1. DOMÍNIO EGF.....	3
2.2. FAMÍLIAS EGF & NRG.....	4
2.3. ANTICORPOS.....	9
2.3.1. Anticorpos comerciais.....	11
2.3.2. Fragmentos derivados de anticorpos.....	13
2.3.2.1. Fragmentos de anticorpos hetero-oligoméricos.....	13
2.3.2.2. Fragmentos de anticorpos homo-oligoméricos.....	14
3. OBJETIVOS.....	17
3.1. OBJETIVOS ESPECÍFICOS.....	17
3.1.1. Objetivos específicos dos fatores de crescimento da família EGF e NRG1 $\alpha$ .....	17
3.1.2. Objetivos específicos de anticorpos.....	17
4. MATERIAIS & MÉTODOS.....	18
4.1. MATERIAIS & PROTOCOLOS.....	18
4.1.1. Linhagens celulares.....	18
4.1.1.1. Linhagens bacterianas.....	18
4.1.1.2. Linhagens de células de mamífero.....	18
4.1.2. Plasmídeos, genes sintéticos e oligonucleotídeos.....	18
4.1.2.1. pUC57.....	18
4.1.2.2. pET32a.....	19
4.1.2.3. pET28a.....	19
4.1.2.4. Genes sintéticos.....	19
4.1.2.5. Oligonucleotídeos.....	20
4.1.3. Reagentes.....	21
4.1.3.1. Soluções e tampões.....	21
4.1.3.2. Meios de cultura.....	23
4.1.3.3. Antibióticos.....	23
4.1.3.4. Géis.....	23
4.1.4. Protocolos.....	24
4.1.4.1. Preparo de células competentes de Escherichia coli.....	24

4.1.4.2. Transformação de cepas de Escherichia coli por choque térmico.....	24
4.1.4.3. Extração de DNA plasmideal por lise alcalina de Escherichia coli .....	25
4.1.4.4. 3.1.4.3 Reação em cadeia da polimerase (PCR).....	25
4.1.4.5. Reação de digestão de fragmentos de DNA .....	26
4.1.4.6. Reação de ligação de fragmentos de DNA.....	26
4.1.4.7. Expressão de proteínas recombinantes .....	26
4.1.4.8. Análise estrutural de scFvs com estrutura atômica resolvida.....	29
4.1.4.9. Desenho de construções híbridas.....	29
4.1.4.10. Expressão dos scFvs híbridos .....	30
4.1.4.11. Lise celular por sonicação e tratamento enzimático.....	30
4.1.4.12. Lise celular por alta pressão .....	30
4.1.4.13. Purificação de proteínas recombinantes .....	31
4.1.4.14. Western blot.....	32
4.1.4.15. Dot blot.....	33
4.1.4.16. Ensaio de proteólise limitada.....	33
4.1.4.17. Ensaio de dicroísmo circular .....	33
4.1.4.18. Caracterização proteica por espalhamento dinâmico de luz supersensível e nanofluorimetria de varredura diferencial .....	34
4.1.4.19. Mapeamento de pontes dissulfeto por espectrometria de massas.....	34
4.1.4.20. Ensaio de proliferação celular .....	35
4.1.4.21. Ensaio de regeneração de camada celular .....	36
4.1.4.22. Análises estatísticas .....	37
4.1.4.23. Cristalização do domínio EGF da NRG1 $\alpha$ .....	37
4.1.4.24. Desenho de VHHs anti-EGF .....	38
4.1.4.25. Teste de conceito da apresentação em superfície bacteriana.....	38
4.1.4.26. Citometria de fluxo.....	39
4.1.4.27. Análise de estruturas 3D de proteínas .....	40
4.2. PROCEDIMENTOS DOS FATORES DE CRESCIMENTO DA FAMÍLIA EGF E A NRG1 $\alpha$	40
4.2.1. Produção e caracterização biofísica e funcional dos fatores de crescimento da família EGF e a NRG1 $\alpha$ .....	40
4.3. PROCEDIMENTOS DOS ANTICORPOS .....	41
4.3.1. Produção dos scFvs e caracterização biofísica.....	41

4.3.2. Caracterização biofísica e de seletividade de anticorpos comerciais anti-EGF .....	42
4.4. PROCEDIMENTOS DO SISTEMA DE exposição EM SUPERFÍCIE BACTERIANA	42
4.4.1. Padronização da prova de conceito de sistema de expressão em superfície bacteriana .	42
5. RESULTADOS & DISCUSSÃO .....	43
5.1. FATORES DE CRESCIMENTO DA FAMÍLIA EGF E A NRG1 $\alpha$ .....	43
5.1.1. Artigo publicado na revista Scientific Reports.....	43
5.1.2. Discussão presente no artigo publicado na revista Scientific Reports .....	44
5.1.3. Manuscrito submetido à revista Molecular Biology Reports .....	47
5.1.4. Discussão presente no manuscrito submetido à revista Molecular Biology Reports. ....	48
5.1.5. Cristalização da NRG1 $\alpha$ .....	50
5.2. ANTICORPOS .....	50
5.2.1. Produção inicial dos scFvs anti-EGF .....	50
5.2.2. Otimização das condições de expressão dos scFvs fusionados a SGFP2 .....	51
5.2.3. Refinamento das condições de expressão dos scFvs fusionados a SGFP2 .....	59
5.2.4. Purificação e análise dos scFvs fusionados a SGFP2.....	62
5.2.5. Análise estrutural dos scFvs cujas estruturas atômicas estão descritas no PDB .....	63
5.2.6. Desenho das construções híbridas .....	65
5.2.7. Expressão final dos scFvs.....	68
5.2.8. Purificação dos scFvs .....	70
5.2.9. Caracterização biofísica dos anticorpos comerciais e sua reatividade .....	71
5.2.10. Caracterização biofísica dos scFvs .....	75
5.3. SISTEMA DE EXPRESSÃO EM SUPERFÍCIE BACTERIANA.....	78
5.3.1. Expressão e purificação do hEGF-mRFP .....	78
5.3.2. Testes de interação por citometria de fluxo.....	80
5.3.3. Testes de interação em placa .....	81
6. CONCLUSÃO.....	83
7. REFERÊNCIAS BIBLIOGRÁFICAS .....	84
8. APÊNDICE I – SEQUÊNCIAS DAS PROTEÍNAS EM NUCLEOTÍDEOS E AMINOÁCIDOS.....	91

# 1. INTRODUÇÃO

Fatores de crescimento tem diversas funções na manutenção da homeostasia , assim como no desenvolvimento de diferentes patologias (CHEN et al., 2016; SINGH; CARPENTER; COFFEY, 2016; ZENG; HARRIS, 2014).. O principal grupo de fatores de crescimento é a família de fatores de crescimento epidermal (EGF).

Todos os membros dessa família de proteínas possuem como domínio responsável pela ativação dos receptores celulares o domínio EGF, o qual possui uma alta conservação estrutural entre seus membros (WOUTERS et al., 2005a).

Apesar da família EGF ser amplamente estudada, novas associações com dados clínicos e laboratoriais são feitas e com isso, mais estudos sempre se fazem necessários para compreensão dos seus papéis como ferramentas terapêuticas (OKAMOTO et al., 2018), diagnósticas (BUSSER et al., 2011), prognósticas (PARK et al., 2014) e para o desenvolvimento de pesquisa básica (FERREIRA et al., 2022).

O desenvolvimento de ferramentas analíticas mais robustas precisa acompanhar as necessidades da pesquisa científica, sendo a principal classe de biomoléculas com esse intuito os anticorpos.

Os anticorpos podem ser produzidos inteiros ou em fragmentos, dependendo da sua finalidade e estratégia de produção (BECK et al., 2010). As dificuldades da produção do anticorpo inteiro é a necessidade de diversas modificações pós-traducionais e a necessidade de uma maquinaria celular complexa, além de um alto custo associado (TICKLE et al., 2009). Para seus fragmentos, estratégias mais simples são mais viáveis visando baratear e facilitar sua produção em maior escala (HELMA et al., 2015).

Os anticorpos e seus fragmentos podem possuir cadeias leve e pesada ou apenas cadeia pesada. Quando esses fragmentos possuem cadeias leve e pesada ligadas por um conector sintético, eles são denominados scFvs. Quando os fragmentos são apenas da cadeia pesada são denominados VHH.

Em geral, como ferramentas para pesquisa básica, os anticorpos e seus fragmentos possuem poucas informações disponíveis das moléculas sendo comercializadas, isso devido as baixas exigências das agências regulatórias e um alto custo na sua caracterização (VOSKUIL, 2014).

Sendo assim, esse projeto teve como foco a obtenção de fragmentos de anticorpos do tipo scFv derivados de anticorpos comerciais em diferentes construções para análise quanto a critérios de produção e utilização e a padronização de ensaios de apresentação em superfície

bacteriana de VHHs para desenvolvimento de novas moléculas. Exigiu-se alto grau de qualidade das proteínas obtidas visando o desenvolvimento de ferramentas de imunoenaios robustas.

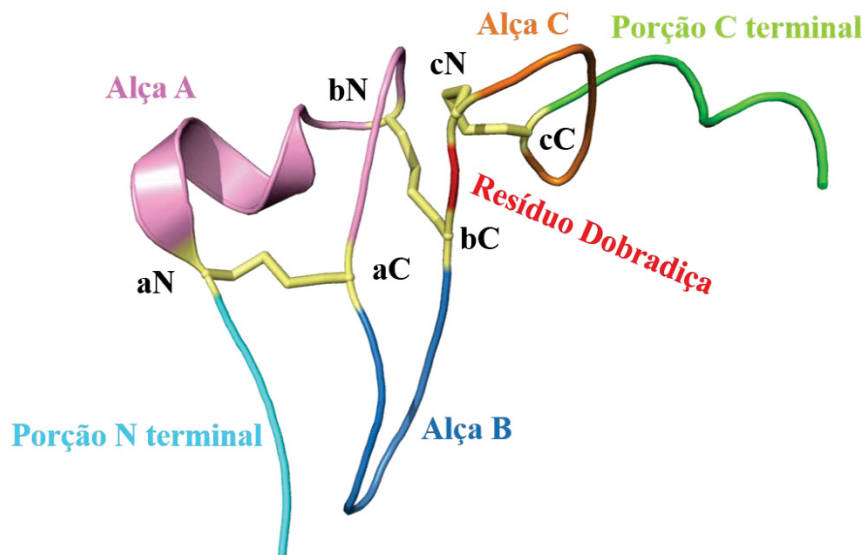
Para isso, foram estabelecidos protocolos de produção para os fatores da família EGF, exigindo-se um rigor frente à qualidade da conformação estrutural dos mesmos visando a geração de dados consistentes e reprodutíveis das ferramentas analíticas a serem desenvolvidas.

## 2. REVISÃO DA LITERATURA

### 2.1. DOMÍNIO EGF

O domínio EGF compreende entre 30 a 50 resíduos e é estruturalmente mantido por três pontes dissulfeto que formam três alças denominadas A, B e C, sendo a segunda e terceira separadas por um único resíduo, chamado de resíduo dobradiça, o que permite a mobilidade entre o segmento N-terminal, formado pelas alças A e B, e o segmento C-terminal, formado pela alça C. O motivo geral do domínio da superfamília EGF é  $CX_{4-14}CX_{3-8}CX_{4-14}CXC_{8-14}C$  (ZOELLEN et al., 2000). As cisteínas que formam as três pontes são denominadas aN e aC para a primeira, bN e bC para a segunda e cN e cC para a terceira (Figura 1) (WOUTERS et al., 2005b).

FIGURA 1 - ESTRUTURA DO DOMÍNIO EGF



Fonte: O próprio. Imagem gerada no programa PyMol v 4.5.0 usando o arquivo PDB: 1JL9 (hEGF). Em azul claro, a porção N terminal. Em amarelo, os resíduos de cisteína formando pontes dissulfeto. Em rosa, a alça A. Em azul escuro, a alça B. Em laranja, a alça C. Em verde, a porção C terminal. Em vermelho, o resíduo dobradiça.

Evolutivamente, esse domínio pode ser dividido em três subtipos, hEGF, cEGF1 e cEGF2, sendo derivados de um ancestral comum de quatro pontes dissulfeto, as quais ainda são conservadas nos domínios tipo-EGF da família das lamininas/integrinas. Os três subtipos são separados estruturalmente pelo tamanho, composição e modificações das suas sequências, como demonstrado na Tabela 1 (WOUTERS et al., 2005b).

**QUADRO 1 CLASSIFICAÇÃO GERAL DO DOMÍNIO EGF.**

hEGF	cEGF1	cEGF2
8-9 Resíduos entre cN E cC. cC se encontra na volta entre folhas $\beta$	<b>10+ Resíduos entre cN E cC cC se encontra na folha <math>\beta</math></b>	
4+ Resíduos entre bN E aC	<b>4+ Resíduos entre bN e aC</b>	<b>3 Resíduos entre bN e aC</b>
Hidroxilação em Asp fucosilação	<b>Hidroxilação em Asn fucosilação</b>	<b>Hidroxilação em Asn</b>

Fonte: O próprio.

Domínios do tipo EGF são encontrados isolados ou com múltiplas cópias em sequência, geralmente como subunidades em porções extracelulares de proteínas ligadas à membrana plasmática. Quando componentes de proteínas grandes, os domínios EGF são mais comumente localizados intercalando outros domínios, agindo como espaçadores estruturais, ou como moléculas ligadoras de cálcio, atuando em interações proteína-proteína, proteólise ou proteção à proteólise (ZOELEN et al., 2000). Os principais tipos de proteínas contendo o domínio EGF são fatores de crescimento, receptores presentes na membrana plasmática, proteínas de matriz extracelular, proteína de adesão, proteínas de transporte de colesterol e proteínas da cascata de coagulação. As funções do domínio EGF dependem dos seus subtipos, mas também do tipo celular e do contexto molecular (BJO et al., 2005; WOUTERS et al., 2005b). O domínio EGF é encontrado em diversos organismos, desde bactérias, fungos, tripanossomatídeos, eucariotos inferiores e superiores, além de proteínas virais (EMBL, 2020; ZOELEN et al., 2000).

Dentro da superfamília de proteínas EGF, localizam-se duas famílias de fatores de crescimento que compartilham de algumas similaridades, tais quais serem ambas do subtipo hEGF e serem as únicas famílias descritas em humanos a se ligarem nos receptores do tipo ErbB. Elas são a família EGF e a família das neuroregulinas (NRG) (ROSKOSKI, 2014a; WOUTERS et al., 2005b).

## 2.2. FAMÍLIAS EGF & NRG

As famílias EGF e NRG são responsáveis pelos ligantes da família de receptores do tipo ErbB. Os receptores do tipo ErbB, também conhecidos como HER, do inglês *human EGF receptors*, é constituída de quatro membros: ErbB1, também chamado de HER1 ou fator de crescimento epidermal (EGFR), ErbB2 ou HER2, ErbB3 ou HER3 e por fim ErbB4 ou HER4 (LEAHY, 2004).

Esses receptores possuem quatro domínios extracelulares, um domínio transmembrana e um domínio citoplasmático. Os domínios extracelulares I e III são homólogos e também chamados de L1 e L2, enquanto os domínios II e IV também são homólogos entre si,



podendo ser chamados de CR1 e CR2, do inglês *cystein rich domains*, uma vez que eles possuem mais de 50 cisteínas conservadas. A ligação ao receptor é mediada principalmente pelo domínio L2, com pequena contribuição do domínio L1, enquanto o domínio CR2 possui uma ação inibitória em conjunto com a CR1 (GARRETT et al., 2002; LEAHY, 2004).

A força de ligação entre os ligantes e os receptores e a preferência por formação de homo ou heterodímeros é uma combinação dual entre as propriedades intrínsecas de cada fator e de cada receptor. Por exemplo, o receptor HER2 é um receptor “órfão”, ou seja, sem ligante conhecido, no entanto, ele demonstra a maior capacidade de heterodimerização dentre todos os receptores por possuir uma alça intrinsecamente estendida que o torna constitutivamente disponível para dimerização, sendo assim o preferencial dentre todos os outros receptores da família para formar o par de dimerização (GARRETT et al., 2002; ZOELLEN et al., 2000).

Já o receptor HER3, que possui ligantes, não possui atividade tirosina-quinase na sua porção citoplasmática, sendo necessária a formação de heterodímeros para transdução de sinal. A sinalização pelo par ligante-receptor dimerizado leva à propagação de sinal intracelular para proliferação, diferenciação, sobrevivência, dentre outros. A atividade mitogênica desses fatores, no entanto, não está associada à afinidade entre ligante e receptor isoladamente, uma vez que interações fracas favorecem uma menor reciclagem dos receptores e dos ligantes, aumentando assim sua meia-vida na circulação. Um dos fatores principais que influencia a duração do sinal é a capacidade do ligante de formar um dímero simétrico ou assimétrico, sendo este primeiro um dímero com uma interação entre si forte e uma sinalização mais breve, enquanto o dímero assimétrico possui uma interação entre si mais fraca e uma sinalização mais longa. A força do sinal é responsável pelo efeito final sob a célula, uma vez que sinais fracos já foram demonstrados promoverem a diferenciação celular, enquanto sinais mais fortes tendem a promover a proliferação (ECCLES, 2011; FREED et al., 2017; ROSKOSKI, 2014a; SCHNEIDER; YARDEN, 2014; ZOELLEN et al., 2000).

O EGFR possui sete ligantes descritos, sendo eles o fator de crescimento epidermal (EGF), o fator de crescimento tumoral alfa (TGF- $\alpha$ ), anfiregulina (AREG), epiregulina (EPR), epigene (EPGN), betacelulina (BTC) e o fator de crescimento epidermal ligador de heparina (HB-EGF). Destes, a EPR, BTC e HB-EGF também conseguem se ligar ao Erb4 em níveis fisiológicos. Em níveis supra-fisiológicos, EGF e BTC conseguem levar a formação do heterodímero ErbB2/ErbB3. Como citado acima, suas afinidades com seus receptores são variadas (ALIMANDI et al., 1997; FREED et al., 2017; GARRETT et al., 2002; OGISO et al., 2002).

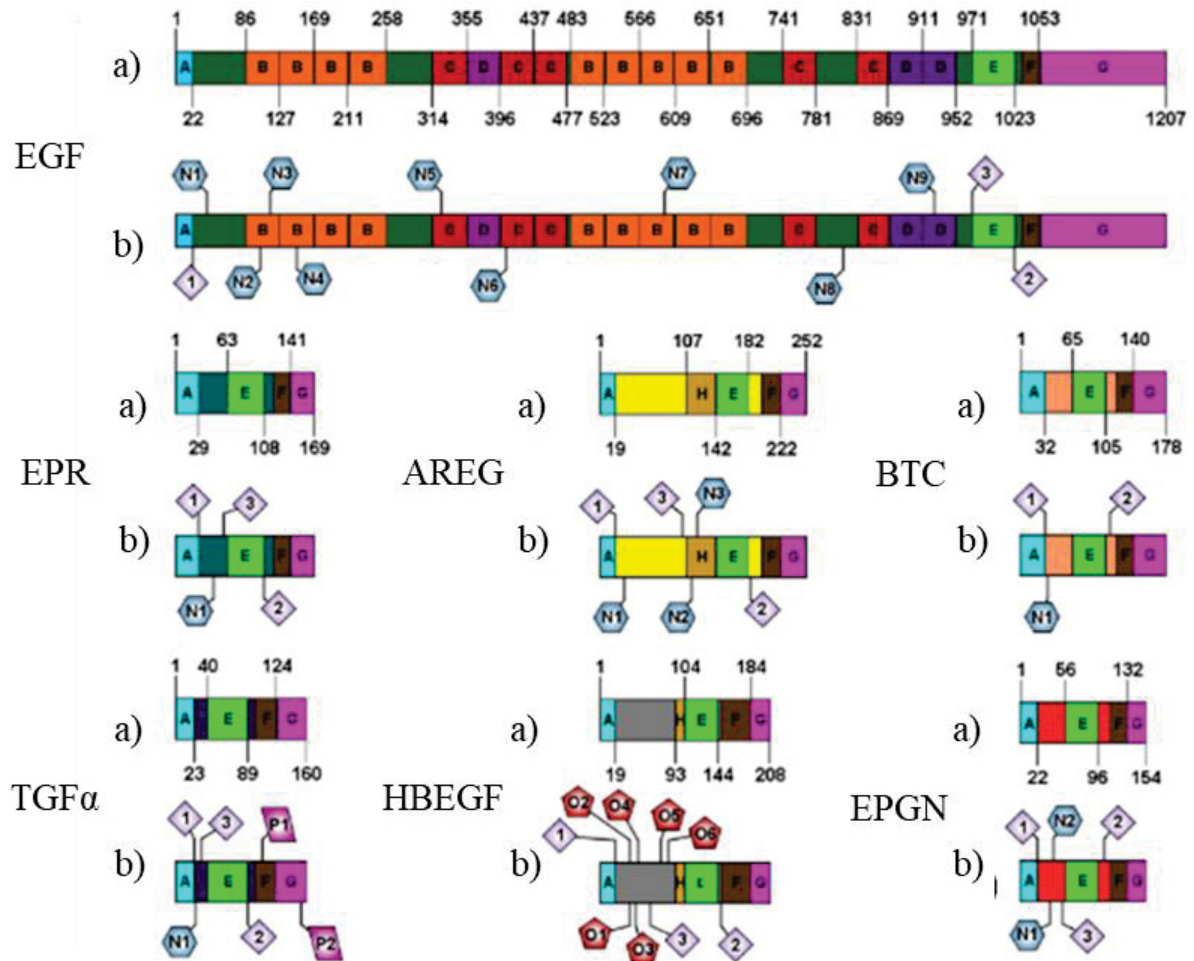
Todos os sete ligantes são sintetizados como proteínas precursoras transmembranares do tipo I, os quais possuem uma porção N-terminal, o domínio EGF, uma porção justaposta à membrana, um domínio transmembrana e uma porção C-terminal citoplasmática. Alguns fatores também possuem outros domínios como domínio ligante de heparina, receptores de LDL da classe B, domínios EGF-*like* e domínios EGF-*like* ligadores de cálcio. A proteína precursora do EGF possui oito repetições do domínio EGF-*like* além do domínio EGF propriamente dito, sendo denominado após a clivagem do peptídeo sinal (pré-pró-EGF) e da junção a membrana (pró-EGF) como EGF de alto peso molecular (HMW EGF, do inglês *high molecular weight* EGF). O HMW EGF, assim como o domínio EGF em si após seu último processamento, possuem a capacidade de ligação ao EGFR. Se a forma HMW EGF e EGF possuem diferentes funções ainda não está elucidado (Figura 2.). Todos os outros ligantes só possuem um domínio EGF (CHEN; JIN; MCINTYRE, 2017; SCHNEIDER; WOLF, 2009). O domínio EGF da família EGF e NRG é definido como CX<sub>7</sub>CX<sub>3-5</sub>CX<sub>10-12</sub>CXCX<sub>5</sub>GXRC (FIGURA 3) (ZENG; HARRIS, 2014).

O processamento desses fatores ocorre pela ação de metaloproteases ligadoras de zinco da família das metaloproteases da matriz extracelular (MMP) e da família das desintegrinas e metaloproteases (ADAM), as quais agem na porção extracelular, liberando um ectodomínio solúvel. (HARRIS; CHUNG; COFFEY, 2003).

As NRGs são outra família de ligantes que também possuem o domínio EGF e se ligam ao Erb3 e ao Erb4 (CARVALHO et al., 2016). Essa família possui pelo menos 15 isoformas de splicing alternativos de 4 genes, sendo as principais: NRG1 tipos I à IV, NRG2, NRG3 e NRG4. Todas as neuroregulinas podem ser divididas em subtipos alfa, beta ou gama, devido a splicings alternativos da porção C-terminal da sequência do seu domínio EGF, alterando assim sua afinidade com seus receptores (COTE et al., 2005; GUMÀ et al., 2010; SCHAEFER; FITZPATRICK; SLIWKOWSKI, 1997). A membra mais estudada é a NRG1 $\alpha$ , a qual será a única membra dessa família a ser abordada nesse trabalho.

Os ligantes do EGFR e as NRGs podem atuar em vias de sinalização parácrina, justácrina, autócrina, endócrina e ExTRAcrine, do inglês *exosomal targeted receptor activation*, além de funcionar como reguladores gênicos através de suas porções citoplasmáticas que também podem sofrer ação de proteólises, atuando como uma sinalização bimodal extra e intracelular (COTE et al., 2005; JAY, 2016; SCHNEIDER; WOLF, 2009; SINGH; CARPENTER; COFFEY, 2016).

FIGURA 2. DOMÍNIOS E MODIFICAÇÕES DOS FATORES DA FAMÍLIA EGF

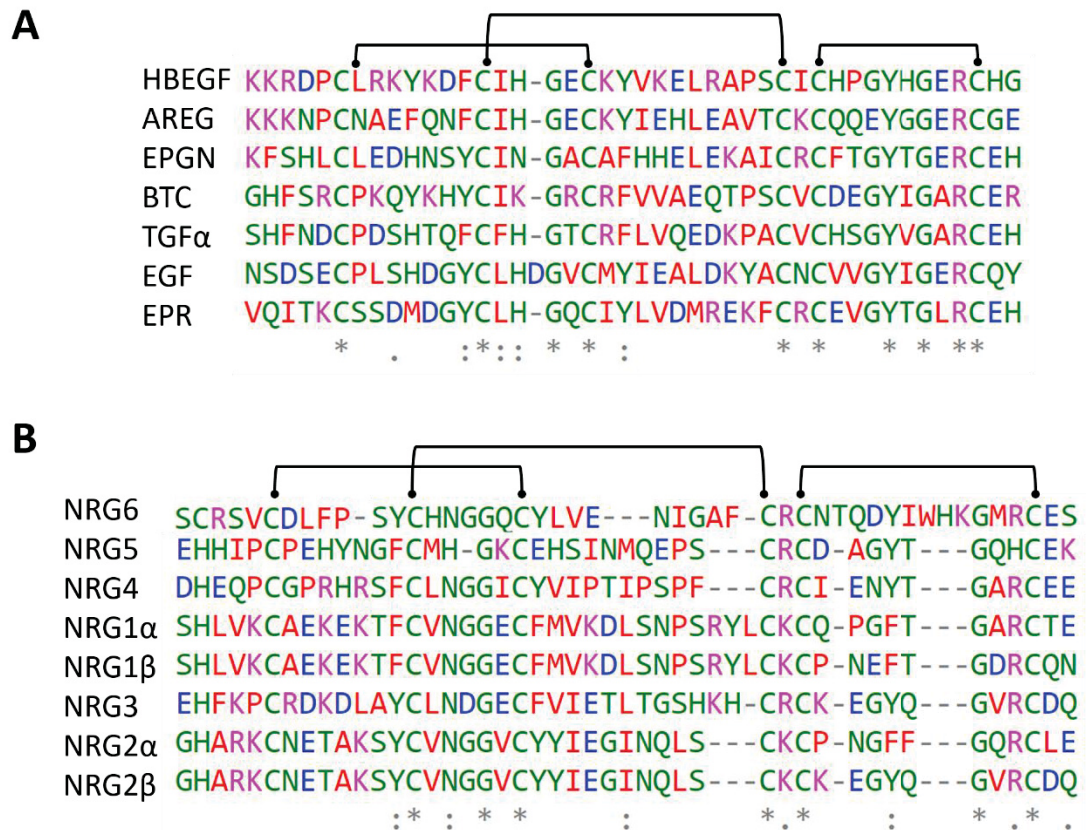


FONTE: O próprio. Legenda: Esquema dos fatores da família EGF com seus domínios, processamentos e modificações numerados. a) Estruturas e seus domínios. b) Pontos de processamento e modificações pós-traducionais. Losangos brancos: Pontos de processamento numerados por ordem de processamento. Hexágonos azuis: N-glicosilações numeradas por posição sequencial. Pentágonos vermelhos: O-glicosilações numeradas por posição sequencial. Trapézios rosas: Palmitoilação numeradas por posição sequencial. A) peptídeo-sinal B) Receptor LDL da classe B C) Domínio EGF-like D) Domínio EGF-like ligador de cálcio E) Domínio EGF F) Domínio transmembrana G) Domínio citoplasmático H) Domínio ligante de heparina. Regiões coloridas não sinalizadas não possuem domínios conhecidos.

Ambas as famílias atuam em diversas vias de sinalização envolvidas no desenvolvimento embrionário, pré-implantação do embrião, desenvolvimento placentário, organogênese, especialmente do pulmão, coração, sistema nervoso central, ossos, rins e pele, desenvolvimento pós-natal, homeostase da pele e do sistema gastrointestinal, cicatrização, desenvolvimento dos órgãos reprodutores, produção de leite materno, regeneração hepática, resposta imune, dentre outras (CHEN et al., 2016). Conseqüentemente, elas também estão relacionadas com numerosas patologias. Dentre elas pode-se mencionar neoplasias, tais como adenocarcinoma de pulmão, glioblastoma, câncer de próstata, câncer de ovário, problemas de desenvolvimento, doença de Alzheimer, doença de Parkinson, esclerose múltipla, hipertensão,

fibrose cística, hiperglicemia crônica, dentre outras (CHEN et al., 2016; SINGH; CARPENTER; COFFEY, 2016; ZENG; HARRIS, 2014).

**FIGURA 3.** SEQUÊNCIA AMINOACÍDICA DO DOMÍNIO EGF DAS FAMÍLIA EGF E NRG



FONTE: O próprio. A. Alinhamento estrutural das sequências da família EGF. B. Alinhamento estrutural das sequências da família NRG. Em arcos pretos, a localização das pontes dissulfeto. Resíduos em vermelho: apolares. Resíduos em verde: polares. Resíduos azuis: carregados negativamente. Resíduos em rosa: carregados positivamente. Asteriscos: resíduos conservados em todas as sequências. Dois pontos: Resíduos de alta similaridade conservados. Um ponto: Resíduos de baixa similaridade conservados.

Sendo assim, o desenvolvimento de ferramentas analíticas para o estudo dessas proteínas frente a diferentes contextos se faz sempre necessário. Tanto o FDA quanto a ANVISA definem como boas ferramentas analíticas aquelas que atendem os parâmetros de precisão, exatidão, linearidade, limite de detecção e quantificação, especificidade, reprodutibilidade, estabilidade e recuperação adequados (ANVISA, 2003; ECCLES, 2011; FDA, 2015; OKINES; CUNNINGHAM; CHAU, 2011; ROSKOSKI, 2014b).

Dentre as ferramentas que se enquadram dentro desses parâmetros estão os anticorpos e seus derivados.

### 2.3. ANTICORPOS

Anticorpos são moléculas produzidas por células do sistema imune e podem ser divididos em hetero-oligoméricos, também chamados de imunoglobulinas ou gamaglobulinas, que são aqueles de estrutura formada por duas cadeias pesadas e duas cadeias leves conectadas entre si, ou homo-oligoméricos, possuindo apenas a cadeia pesada (ABBAS; LICHTMAN; PILLAI, 2011). A partir de ambos os tipos é possível obter fragmentos derivados funcionais utilizando os domínios de reconhecimento do antígeno (FIGURA 4).

Ambos os tipos têm sua estrutura dividida entre domínios constantes e domínios variáveis, sendo estes últimos responsáveis pela ligação ao alvo. Dentro da região variável, existe uma região hipervariável denominada região determinante da complementariedade (CDR), onde estão localizados comumente a maior parte da força de ligação com o alvo. Essas regiões podem reconhecer dois tipos de epítomos. O primeiro deles é o epítomo linear, o qual é uma região onde os resíduos que serão reconhecidos estão linearmente justapostos, ou seja, a estrutura primária é responsável pela ligação antígeno-anticorpo. Já o segundo tipo é o epítomo conformacional, o qual é formado por uma região tridimensional, sendo assim o enovelamento da proteína é responsável pela proximidade de resíduos sequencialmente distantes, permitindo assim o reconhecimento do anticorpo (ABBAS; LICHTMAN; PILLAI, 2011).

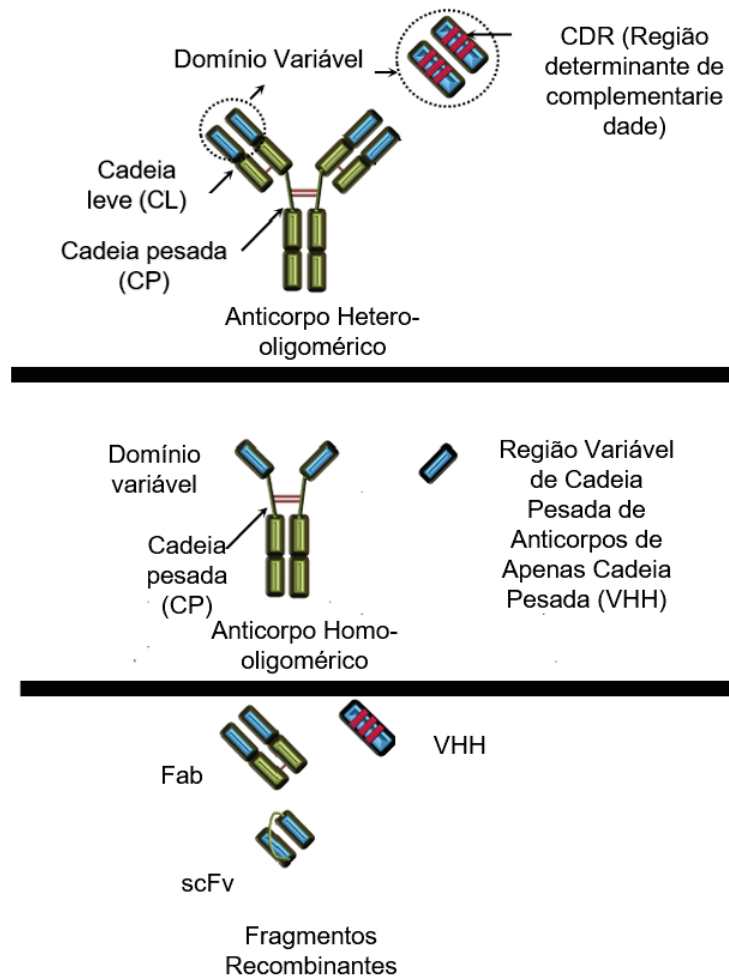
A força de ligação de um anticorpo pelo seu alvo é calculada pela constante de dissociação, a qual é influenciada pela quantidade de contatos e conseqüentemente das forças envolvidas entre o anticorpo e seu alvo. Anticorpos dependendo da sua classe podem variar sua valência, ou seja, a quantidade de regiões determinantes antigênicas. Anticorpos inteiros sempre possuem pelo menos duas regiões determinantes antigênicas, mas esse valor pode ser tão grande quanto dez. O somatório da força de ligação de todos os determinantes antigênicos é denominada avidéz do anticorpo (ABBAS; LICHTMAN; PILLAI, 2011).

Quando as CDRs estão na presença de estruturas semelhantes é possível ocorrer o fenômeno da reatividade cruzada, quando um anticorpo gerado contra uma proteína, reconhece outra. Esse fenômeno é intimamente ligado a quantidade de similaridade entre duas seqüências, seja ela uma similaridade espacial no caso de epítomos estruturados ou similaridade sequencial no caso de epítomos lineares (FRANK, 2002; SELA-CULANG; KUNIK; OFRAN, 2013).

Anticorpos de maneira geral podem ser obtidos pela forma clássica, a partir de um linfócito ou hibridoma, ou produzidos por engenharia biotecnológica (HUDSON; SOURIAU, 2003; KÖHLER; MILSTEIN, 1975).



**FIGURA 4. ANTICORPOS E FRAGMENTOS DE ANTICORPOS**



Fonte: Imagem modificada de ((ROVIS; LEGNAME, 2014). Parte superior demonstrando porções mais importantes de anticorpos hetero-oligoméricos. Parte central demonstrando porções mais importantes de anticorpos homo-oligoméricos. Parte inferior demonstrando diferentes tipos de fragmentos de anticorpos. Fab: fragmento de ligação a antígeno. ScFv: fragmento de cadeia única.

Anticorpos obtidos da forma clássica tendem a atingir uma afinidade limite pelo seu alvo em ensaios *in vivo* devido aos limites de recombinação de sequências do linfócito B, muitas vezes devido a mecanismos regulatórios de autoimunidade ou tempo de seleção. Sendo assim, é necessário o melhoramento *in vitro* por evolução dirigida (BODER; MIDELFORT; WITTRUP, 2000) ou modificações estruturais para aumentar a sua afinidade, permitindo assim uma maior afinidade por seus ligantes (LIPPOW; WITTRUP; TIDOR, 2007).

Os anticorpos podem ser modificados utilizando-se engenharia de anticorpos para se obter derivados, uma vez que estes têm sua expressão facilitada em sistemas mais simples e baratos, tais como bactérias e leveduras. Os fragmentos derivados possuem vantagem frente ao anticorpo inteiro por permitir uma produção mais rápida e barata, no entanto, a falta da porção Fc leva a uma maior chance de instabilidade, agregação e degradação dos mesmos (NELSON, 2010).

### 2.3.1. Anticorpos comerciais

Dentre os grandes desafios dos anticorpos disponíveis no mercado e o desenvolvimento de novos anticorpos estão a falta de especificidade, o ruído, a qualidade do sinal, o controle de qualidade, a validação e a reprodutibilidade dos produtos disponíveis (VOSKUIL, 2014; WELLER, 2016).

Enquanto dados dos anticorpos terapêuticos projetam um lucro de 300 bilhões de dólares em 2025, os anticorpos comerciais não terapêuticos para uso em pesquisa estão na casa dos milhões (GRILO; MANTALARIS, 2019). Esse último grupo é o principal responsável por prover ferramentas para pesquisa e desenvolvimento nas mais diversas áreas da ciência, sendo os critérios de qualidade para esse grupo bastante baixos, uma vez que o lucro gerado comparado ao dinheiro investido é muito menor. Um anticorpo policlonal se encontra na casa dos mil dólares para mg de anticorpo vendido, enquanto anticorpos monoclonais se encontram em torno de dez mil dólares e anticorpos recombinantes por volta de cinquenta mil dólares, sendo assim, o corte de custo no desenvolvimento é a caracterização das moléculas, situação nomeada “O Dilema dos Anticorpos”, o qual descreve “*Antibody haste, research waste*” como uma realidade no desenvolvimento atual (ASCOLI; BIRABAHARAN, 2015).

Os problemas de falta de especificidade de um determinado anticorpo ao longo do seu desenvolvimento muitas vezes não são considerados e também muitas vezes não podem ser solucionados, uma vez que similaridade de epítomos são uma característica inerente da natureza dos alvos. Nesses casos, cabe ao desenho experimental propor combinações de possibilidades para se inferir a origem da ligação do anticorpo, uma vez que epítomos similares ainda podem ter diferenças de afinidade que podem ser explorados para diminuir sinais inespecíficos (VOSKUIL, 2014). No entanto, a falta de cuidado dos fabricantes também deve ser destacada, uma vez que a maior parte dos anticorpos vendidos não possuem dados de reatividade cruzada que podem ser considerados válidos. Uma das causas é o preço para o levantamento desses dados, em consequência, o maior problema se deve a falta da caracterização do epítomo de ligação, o que é considerado uma caracterização parcial e de baixo critério, impossibilitando os testes de reatividade cruzada válidos, tornando experimentos do tipo sanduíche ou western blot facilmente inválidos (WELLER, 2016). Nos piores casos, os anticorpos vendidos não podem ser usados para o que eles são descritos (HERRERA et al., 2013). Outro problema ligado a isso é a não disponibilização da origem do antígeno usado para a geração do anticorpo (WELLER, 2016).

Apesar da afinidade ser uma das características mais importantes dos anticorpos, tendo-se desenvolvido muitas técnicas capazes da determinação dessa característica, é um dado bastante raro de aparecer na ficha técnica dos anticorpos comercializados, o que limita a capacidade de interpretação dos dados gerados pelos imunoenaios e gera mais experimentos para a padronização, uma vez que as diluições do anticorpo estão relacionadas com sua força de ligação ao alvo, assim como os tipos de ligações devem ser consideradas na adição de sais e detergentes nos tampões (VOSKUIL, 2014; WELLER, 2016).

A aquisição e rastreamento dos anticorpos também tem se tornado cada dia mais difíceis, uma vez que apenas 44% dos anticorpos usados nos artigos podem ser rastreados, independente de fator de impacto. Isso é corroborado pelo fato que mesmo dentro do meio acadêmico, os critérios de qualidade para técnicas utilizando anticorpos são quase nulos, assim como, artigos sobre caracterização e validação de anticorpos são raros, sendo muito mais comuns aqueles que descrevem moléculas novas com pouca caracterização (WELLER, 2016). Outra dificuldade é a variação dentre os lotes, uma vez que a maior parte das companhias não produz todos os anticorpos do seu catálogo e adquirem eles de terceiros, mantendo os dados de qualidade apenas fornecidos por terceiros de um lote fixo, dificultando a reprodutibilidade dos dados. Também existem vários anticorpos no mercado com diferentes nomes e dados de qualidade, que na realidade, provêm de um mesmo anticorpo em diferentes lotes. Outro problema na aquisição de anticorpos é a maneira que eles são armazenados e preservados, desde sua fabricação até a estocagem no laboratório, uma vez que anticorpos estocados em equipamentos refrigeradores de uso frequente tendem a sofrer sucessivos congelamentos/descongelamentos (VOSKUIL, 2014).

O problema de rastreabilidade não é necessariamente culpa do desenvolvimento da molécula, mas também das fusões entre companhias que levam a junção dos catálogos, mas de maneira pouco clara e dos muitos revendedores que não discriminam as informações corretamente (SCHONBRUNN, 2014; WELLER, 2016). Infelizmente, não se é claro os motivos que levam a um anticorpo a ser descontinuado pelas companhias, uma vez que são produtos de alto custo para pesquisa e desenvolvimento, no entanto, mostra a vantagem do desenvolvimento dos anticorpos recombinantes, uma vez que dados da sequência preservam a identidade do anticorpo e permitem sua rastreabilidade (WELLER, 2016).



## 2.3.2. Fragmentos derivados de anticorpos

### 2.3.2.1. Fragmentos de anticorpos hetero-oligoméricos

Os derivados de anticorpos hetero-oligoméricos mais comuns são fragmentos de ligação ao antígeno ou Fabs (do inglês *fragment antigen-binding*) e os fragmentos de domínios variáveis de cadeia única ou scFvs (do inglês *single-chain variable domain antibody fragment*), os quais são formados sem os domínios constantes e possuem as regiões variáveis das cadeias leves e pesadas unidas entre si por conectores sintéticos (BECK et al., 2010).

A dificuldade de obtenção de quantidades trabalháveis de fragmentos de anticorpos hetero-oligoméricos produzidos de maneira recombinante é um desafio devido à complexidade desse grupo de proteínas. Para ultrapassar esse desafio, a maturação de biomoléculas pode ser feita para melhorar a afinidade e/ou solubilidade, assim como a construção de moléculas híbridas com o enxerto de CDRs em arcabouços diferentes do de origem e a padronização rigorosa das condições de expressão por metodologias que adotem qualidade baseada no projeto (QbD, do inglês *quality by design*) (KUMAR et al., 2019a; MAKABE et al., 2008; RAZAI et al., 2005).

Dentre as estratégias mais usadas para a maturação molecular estão a evolução dirigida (RAZAI et al., 2005), na qual se baseia na alteração aleatória da sequência aminoacídica, mutações sítio dirigidas ou modificações baseadas em análises estruturais (DUCANTEL; MULLER, 2012), enquanto as construções híbridas em geral são feitas com o uso de arcabouços estáveis para enxerto de CDRs, os quais podem possuir ou não alterações em resíduos do arcabouço denominados resíduos da Zona de Vernier (MAKABE et al., 2008; SAERENS et al., 2005).

Os resíduos da Zona de Vernier são encontrados em regiões próximas as CDRs (FIGURA 5) e em regiões importantes para interação entre as cadeias dos anticorpos hetero-oligoméricos. Apesar de não serem responsáveis pela interação com o antígeno, trabalhos demonstram sua importância na mobilidade entre as cadeias e das próprias CDRs, influenciando assim a força de ligação ao alvo (MAKABE et al., 2008).

Estratégias de padronização de condições de expressão que se baseiam em desenho de experimentos geralmente são utilizadas na produção industrial de biomoléculas tais como proteínas de interesse comercial. Nesse contexto, a QbD visa, utilizando de uma abordagem sistemática, entender o processo de produção para a melhoria da qualidade do produto final, uma vez que as análises detalhadas das condições de produção podem otimizar e tornar o processo mais robusto. Dentre as estratégias utilizadas a mais comum é a OFAT (OFAT, do

inglês *one factor at a time*), a qual é utilizada para a triagem inicial das variáveis importantes no processo de produção de biomoléculas. A partir das variáveis identificadas, combinações buscando sinergismo podem ser usadas para a melhoria do processo como um todo (KUMAR et al., 2019b).

FIGURA 5. LOCALIZAÇÃO DOS RESÍDUOS DA ZONA DE VERNIER E CDRS



Fonte: Imagem modificada de: (MAKABE et al., 2008). Sequências dos anticorpos anti-EGFR m528VH e sua versão humanizada h528VH. Sequência superior: anticorpo murino. Anticorpo humano. VH: Região variável pesada. VL: Região variável leve. Em azul claro, resíduos da zona de vernier. Em vermelho, CDRs.

Independentemente do método de obtenção e produção dos anticorpos ou derivados, a robustez da ferramenta está relacionada com a sequência proteica e com o mecanismo de reconhecimento adotado pelas CDRs.

Dois anticorpos comerciais contra o EGF, um de origem murínica e outro de origem leporínica, foram sequenciados e suas sequências obtidas em genes sintéticos para expressão em leveduras, no entanto, essa estratégia não apresentou resultados satisfatórios (FERREIRA, 2018). Sendo assim, esse trabalho desenvolveu novas construções de fragmentos de anticorpos para sua avaliação.

### 2.3.2.2. Fragmentos de anticorpos homo-oligoméricos

Os fragmentos derivados dos anticorpos homo-oligoméricos, denominados sdAfs (do inglês **s**ingle **d**omain **a**ntibody **f**ragment), VHH (do inglês *variable heavy chain from heavy chain only antibodies*) ou nanobodies® (primeiro nome comercial popularizado) possuem vantagens na produção biotecnológica quando comparados com outros tipos de fragmentos de

anticorpos modificados, tal como o scFv. Para se produzir um scFv necessita-se de uma construção em que os fragmentos das regiões variáveis sejam conectados um no outro, intercalados por um conector sintético, o qual não se pode prever o comprimento ou sequência ideal. Sendo assim, a dimerização do scFv pode ser prejudicada ou mesmo a sua interação com o antígeno. De forma diferente, os sdAf por serem de domínio único, não apresentam estas limitações (AHMAD et al., 2012; MUYLDERMANS, 2013).

Os anticorpos e fragmentos podem ser obtidos diretamente dos linfócitos B, comercialmente ou através de bibliotecas virgens, imunes e sintéticas. Com a exceção das bibliotecas semi-sintéticas e sintéticas, todas as outras formas estão condicionadas às limitações dos linfócitos B. Anticorpos e derivados provenientes de bibliotecas sintéticas, no entanto, não possuem as limitações intrínsecas do linfócito B, uma vez que a afinidade e seletividade estarão limitadas principalmente pelo tamanho do repertório da biblioteca e o método de triagem da mesma (RAJPAL et al., 2005).

As bibliotecas sintéticas utilizam-se de sequências diversas para o arcabouço e/ou para as CDRs da porção variável. Devido a sua natureza sintética, sequências não possíveis de serem adquiridas pelas limitações do sistema imune podem ser selecionadas positivamente. O processo de geração de bibliotecas sintéticas de VHHs é simplificado quando comparado a bibliotecas sintéticas de fragmentos hetero-oligoméricos, uma vez que moléculas monoméricas facilitam o trabalho de obtenção de proteínas solúveis e mais estáveis (MUYLDERMANS, 2001; SAERENS et al., 2005; YAN et al., 2014).

Dentre os métodos de triagem de bibliotecas sintéticas, existem aqueles que se baseiam em apresentação na superfície de bactérias e leveduras dos fragmentos de anticorpos, os quais são fusionados a moléculas de superfície, possuindo a vantagem de fazer a triagem no ambiente externo da célula, o qual é geralmente oxidante, além de permitir usos de diversas ferramentas de marcação e etiquetas que facilitam a purificação dessas moléculas da superfície celular (TANHA et al., 2002; WEAVER-FELDHAUS et al., 2004).

A estratégia de apresentação na superfície de bactéria é recente, sendo as proteínas de interesse fusionadas a porções de proteínas da membrana externa, as quais são responsáveis por carrear as proteínas para a superfície. Em geral a estrutura para essa estratégia utiliza da sequência sinal de exportação para o periplasma, seguido da porção N-terminal da proteína de membrana externa, a qual vai ser inserida por mecanismos próprios da *Escherichia coli*. A maioria dos trabalhos insere entre a proteína de membrana e a proteína de interesse um sítio de proteólise (SALEMA; FERNÁNDEZ, 2017; XU et al., 2017). A vantagem desse sistema para VHHs é a simplicidade e velocidade do sistema quando comparado com o de apresentação em

leveduras, uma vez que eles não precisam de mecanismos complexos de tradução e modificações pós-traducionais.

### **3. OBJETIVOS**

Produção e caracterização dos fatores de crescimento das famílias EGF e NRG1 $\alpha$  e a produção de anticorpos derivados de construções comerciais e de novas construções provenientes de uma biblioteca sintética e engenharia de proteínas.

#### **3.1. OBJETIVOS ESPECÍFICOS**

##### **3.1.1. Objetivos específicos dos fatores de crescimento da família EGF e NRG1 $\alpha$**

1. Produzir os domínios EGF da família EGF e NRG1 $\alpha$  de forma recombinante.
2. Caracterização biofísica e funcional dos domínios EGF.

##### **3.1.2. Objetivos específicos de anticorpos**

1. Produzir e purificar scFvs derivados de sequências de anticorpos comerciais anti-EGF.
2. Realizar caracterização biofísica dos scFvs.
3. Análise estrutural de arcabouços e desenho de construções híbridas.
4. Produzir scFvs híbridos e realizar caracterização biofísica.
5. Caracterização biofísica e funcional dos anticorpos comerciais parentais dos scFvs.
6. Produzir VHHs anti-hEGF em sistema de apresentação em superfície bacteriana.

## 4. MATERIAIS & MÉTODOS

### 4.1. MATERIAIS & PROTOCOLOS

#### 4.1.1. Linhagens celulares

##### 4.1.1.1. Linhagens bacterianas

❖ *Escherichia coli* SHuffle

Genótipo: *F' lac, pro, lacIq / Δ(ara-leu)7697 araD139 fhuA2 lacZ::T7 gene1 Δ(phoA)PvuII phoR ahpC\* galE (or U) galK λatt::pNEB3-r1-cDsbC (SpecR, lacIq) ΔtrxB rpsL150(StrR) Δgor Δ(malF)3*

❖ *Escherichia coli* BL21 Star (DE3)

Genótipo: *F- ompT gal dcm lon hsdSB(rB-mB-) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS)*

❖ *Escherichia coli* BL21 Star (DE3) ΔslyD pRARE2

Genótipo: *F- ompT gal dcm lon hsdSB(rB-mB-) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS) ΔslyD pRARE2(CamR)*

❖ *Escherichia coli* DH5α™.

Genótipo: *F- Φ80lacZΔM15 Δ (lacZYA-argF) U169 recA1 endA1 hsdR17 (rK-, mK+) phoA supE44 λ- thi-1 gyrA96 relA1.*

##### 4.1.1.2. Linhagens de células de mamífero

❖ HeLa

Origem: Adenocarcinoma cervical humano retirado de Henrietta Lacks em 1951. Adquirida do Banco de Células do Instituto Carlos Chagas.

❖ NHDF

Origem: Fibroblastos dérmicos neonatais humanos. Adquirida de Lonza CC-2509.

#### 4.1.2. Plasmídeos, genes sintéticos e oligonucleotídeos

##### 4.1.2.1. pUC57

O plasmídeo pUC57 é um vetor de subclonagem de *Escherichia coli* com tamanho de 2710 pares de base (Anexo A).

O pUC57 contém sítio de múltipla clonagem, inserido na sequência codificadora de *lacZα* que permite seleção de clones recombinantes por diferenciação entre colônias azuis e brancas na presença de X-gal e IPTG (completar XXX), e resistência a ampicilina pela expressão do gene *bla*(Ap<sup>R</sup>), o qual expressa uma beta-lactamase.

#### 4.1.2.2. pET32a

O plasmídeo pET32a é um vetor de expressão de *Escherichia coli* com tamanho de 5900 pares de base (Anexo B).

O pET32a contém sítio de múltipla clonagem, o promotor da RNA polimerase do fago T7 e operador *lac* e resistência a ampicilina pela expressão gene *bla*(Ap<sup>R</sup>), o qual expressa beta-lactamase. O códon da metionina do sítio de reconhecimento pela enzima *NcoI* está em pauta de leitura com a proteína de tioredoxina (Trx), a qual é utilizada como proteína de fusão para auxiliar na produção e enovelamento de proteínas que estão fusionadas ao seu C terminal. Além dela, também se encontram em pauta uma sequência de hexa-histidinas, um sítio de reconhecimento da protease trombina, uma etiqueta S e um sítio de reconhecimento da enteroquinase.

#### 4.1.2.3. pET28a

O plasmídeo pET28a é um vetor de expressão de *Escherichia coli* com tamanho de 5369 pares de base (Anexo C).

O pET28a contém sítio de múltipla clonagem, o promotor da RNA polimerase do fago T7 e operador *lac* e resistência a canamicina pela expressão do gene *kanR*, o qual expressa uma aminoglicosídeo-fosfotransferase. O códon da metionina do sítio de reconhecimento pela enzima *NcoI* dita a pauta de leitura das sequências inseridas no sítio de múltipla clonagem nesse trabalho. A sequência de hexa-histidina na região C-terminal está em pauta com o sítio de reconhecimento pela enzima *XhoI*, assim como o códon de terminação.

#### 4.1.2.4. Genes sintéticos

As sequências codificadoras dos genes sintéticos 4NIK-MM, 146MM, 148MM, VERMM, 6CVKOC, 146OC, VEROc, NB1 e NB6 foram adquiridos da empresa GenOne já subclonados em vetor pET28a. O gene sintético FR-CCDB foi adquirido da empresa GenOne no vetor pUC57. Os genes sintéticos iNAQ e NEAE foram adquiridos da empresa FastBio já

subclonados em vetor pET28a. Os genes sintéticos da AREG, BTC, EGF, EPGN, HB-EGF, NRG1 $\alpha$  e TGF $\alpha$  foram adquiridos da empresa FastBio já subclonados em vetor pET32a. O gene sintético da EPR foi adquirido da empresa GenCust já subclonado em vetor pET32a. O gene sintético da SGFP2 e OPN foram adquiridos da empresa GenScript e são provenientes de outros trabalhos do grupo. A sequência da mRFP foi adquirida do repositório Addgene no vetor pDsRed-Monomer-C1. Todas as sequências utilizadas nesse trabalho podem ser encontradas no Apêndice I.

#### 4.1.2.5. Oligonucleotídeos

As sequências moldes para a construção dos scFvs MM e OC foram as dos plasmídeos construídos em outro trabalho (FERREIRA, 2018). De maneira resumida, as sequências das cadeias pesadas MMH e OCH presentes nos plasmídeos de levedura pPIC9K foram moldes para os primers scFvMMH e scFvOCH, os quais contém em seus primers anti-senso parte do conector sintético. Para as sequências das cadeias leves MML e OCL presentes nos plasmídeos de levedura pPICZ $\alpha$  foram moldes para os primers scFvMML e scFvOCL, os quais contém em seus primers senso parte do conector sintético. A SGFP2 foi amplificada do plasmídeo 137-SGP2 presente no banco de plasmídeos do laboratório. Esses produtos de PCR foram posteriormente subclonados em pET28a, pET28a com SGFP2 no C-terminal e no pET32a, sendo posteriormente chamadas as construções como MM/OC, MMG/OCG e 32MM/32OC, respectivamente. A lista de oligonucleotídeos usadas se encontra no Quadro 2.

**QUADRO 2 SEQUÊNCIA DE OLIGONUCLEOTÍDEOS**

<b>ScFvMMH Senso</b>	<b>GACTCCATGGAAATCCAATTGCAACAATCAGGTCCAGAATTGGTGAAACC</b>
<b>scFvMMH Anti-Senso</b>	<b>GACTGAATTCACCGCCACCACCACCGCTACCACCGCCAGCAGAAACAGTAA CCAAAGTACCTTGTCGCCAGTAAGC</b>
<b>scFvMML Senso</b>	<b>GACTGAATTCGGCGGTTCTGGTGGCGGTGGCTCTGATGTTATGATGACTCA AACTCCATTGTCTCTTCCAGTTTCTCTTGG</b>
<b>scFvMML Anti-Senso</b>	<b>GACTCTCGAGCTTAATTTCCAACCTTAGTACCACCACCAAAAGTCCATGGAA AATGAGAAC</b>
<b>scFvOCH Senso</b>	<b>GACTCCATGGAAATCCAATTGCAACAATCAGGTCCAGAATTGGTGAAACC</b>
<b>ScFvOCH Anti-Senso</b>	<b>GACTGAATTCACCGCCACCACCACCGCTACCACCGCCAGAAGAAACAGTAA CCAAAGTACCTGGACCCCACTTAC</b>
<b>scFvOCL Senso</b>	<b>GACTGAATTCGGCGGTTCTGGTGGCGGTGGCTCTCAACTTATGACTAAGAC CCCATCCCCAGTCTCCG</b>
<b>ScFvOCL Anti-Senso</b>	<b>GACTCTCGAGCTTAACAACAACCTTCAGTACCACCACCGAAGGCG</b>
<b>SGFP2 Senso</b>	<b>TATAGGATCCGGATCCTCTGGTCTGGTATGGTTTCCAAA</b>
<b>SGFP2 Anti-Senso</b>	<b>TATACTCGAGCTCGAGTTATCAGGAATTCGCTTTGTACAGTTC</b>
<b>mRFP Senso</b>	<b>TATAAAGCTTGGCGGTGGTAGCGACAACACCGAGGACGTCATCAAG</b>
<b>mRFP Anti-Senso</b>	<b>TATACTCGAGCTACTGGGAGCCGGAGTGG</b>

Fonte: O próprio.



### 4.1.3. Reagentes

#### 4.1.3.1. Soluções e tampões

Solução de preparo celular do processo 1 (P1):

Preparar solução para concentração final de 25 mM de Tris.HCl pH 8 e 50 mM de EDTA.

Solução de lise alcalina do processo 2 (P2):

Preparar solução para concentração final de 200 mM de NaOH e 1% de SDS.

Solução de neutralização do processo 3 (P3):

Misturar 8,6 mL de ácido acético glacial 3 M com 45 mL de acetato de potássio 5 M e 21,4 mL de água ultrapura. Manter em geladeira ou em 4 °C.

Solução PBS (do inglês Phosphate-Buffered Saline) 10X

Misturar 1,4 M de NaCl com 74 mM g de  $\text{Na}_2\text{HPO}_4$  e 296 mM de  $\text{NaH}_2\text{PO}_4$  e qsp 1 L de água ultrapura.

Solução PBS-T (do inglês Phosphate-Buffered Saline + Tween)

Misturar 899 mL de água ultrapura com 100 mL PBS 10X e 1 mL Tween 20 (100 %).

Tampão para Western blot (WB) 10X

Misturar 248 mM de tris base com 2 M de glicina. Completar volume para 1 L com água ultrapura.

Tampão de Transferência

Misturar 100 mL de tampão para Western blot 10X com 200 mL metanol e completar q.s.p 1 L com água ultrapura.

Solução de bloqueio

Misturar 5 mL PBS 10X com 2,5 g leite em pó desnatado MOLICO® (NESTLÉ).

#### Azul de Coomassie

Dissolver 300 mg de azul de Coomassie R-250 em 135 mL de metanol. Adicionar 30 mL de ácido acético e 135 mL de água ultrapura. Filtrar em papel Whatman 1 mm ou 3 mm. Guardar em frasco escuro em temperatura ambiente.

#### Solução descorante

Preparar mistura com 30 % de metanol, 10 % de ácido acético glacial e 60 % de água filtrada.

#### Tampão de eletroforese para géis de poliacrilamida 5X

Misturar 12,5 mM de tris base, 1 M de glicina e 50 mL de SDS 10 %. Completar volume para 1 L com água ultrapura.

#### Tampão TBE 10X (Tris/Ácido bórico glacial/EDTA)

Dissolver 893 mM de tris base com 55 mL de ácido acético glacial e 40 mL de EDTA. Completar volume com água filtrada para 1 L.

#### Solução de acrilamida

Pesar 99 g de acrilamida, 2,7 g de bis-acrilamida e acertar o volume para 300 mL com água ultrapura.

#### Tampão de amostra de DNA 6X:

Pesar 3 g Ficoll 400. Aquecer 17 mL de água à aproximadamente 45 °C. Adicionar o Ficoll aos poucos e esperar dissolver com agitação lenta.

Pesar 20 mg de xilenocianol e 20 mg de azul de bromofenol. Dissolver em 500 µL de água.

Adicionar à solução de Ficoll depois que este já tenha dissolvido. Acertar o volume para 20 mL se necessário.

Tampão de amostra para proteínas 4X:

Para o preparo de 20 mL de tampão de amostra, misturar 3,2 mL de Tris.HCl 1 M pH 6,6 com 8 mL de SDS 10 %, 2 mL de  $\beta$ -mercaptoetanol 14,7 M, 4,8 mL de glicerol e 4 mg de azul de bromofenol.

#### 4.1.3.2. Meios de cultura

Meio de cultura de lisogenia (LB, do inglês *lysogeny broth*)

Para preparação de 1 L, dissolver 10 g de triptona, 10 g de cloreto de sódio e 5 g de extrato de levedura e completar volume com água ultrapura. Se meio sólido, adicionar 15 g de ágar. Autoclavar a 120 °C por 30 min.

#### 4.1.3.3. Antibióticos

A concentração de canamicina utilizada foi de 0,025 mg a cada 1 mL de placa ou inóculo, enquanto de ampicilina e clorafenicol foram de 0,1 mg a cada 1 mL de placa ou inóculo.

#### 4.1.3.4. Géis

Gel de agarose

Preparar gel de agarose dissolvendo-se agarose em tampão TBE 1X para a concentração desejada. Misturar amostra com tampão de amostra de DNA 6X e aplicar nos poços do gel.

Gel de corrida 13 %

Preparar gel de corrida 13 % utilizando 1,9 mL de solução de acrilamida 33 % misturado com 0,8 mL de Tris.HCl pH 8,8 1 M, 50  $\mu$ L de SDS 10 %, 35  $\mu$ L de persulfato de amônio 10 %, 3,5  $\mu$ L de TEMED e completar q.s.p 10 mL com água ultrapura. Para outras concentrações, ajustar quantidade de solução de acrilamida e água.

Gel de corrida 15%

Preparar gel de corrida 15% utilizando 2,2 mL de solução de acrilamida 33% misturado com 0,8 mL de Tris HCl pH 8,8 1M, 50  $\mu$ L de SDS 10 %, 35  $\mu$ L de persulfato de amônio 10 %, 3,5  $\mu$ L de TEMED e completar q.s.p 10 mL com água ultrapura.

### Gel de empilhamento

Preparar gel de empilhamento 6,5 % utilizando 0,35 mL de solução de acrilamida 33 % misturado com 0,3 mL de Tris.HCl pH 6,8 5 M, 25  $\mu$ L de SDS 10 %, 17,5  $\mu$ L de persulfato de amônio 10 %, 3,5  $\mu$ L de TEMED e completar q.s.p 2 mL com água ultrapura. Para outras concentrações, ajustar quantidade de solução de acrilamida e água.

#### 4.1.4. Protocolos

##### 4.1.4.1. Preparo de células competentes de *Escherichia coli*

Sobre uma placa de LB contendo o antibiótico apropriado, a partir de um estoque em glicerol da cepa a ser utilizada estriou-se a placa que foi incubada 18 horas a 37 °C. Uma colônia foi transferida para 5 mL de meio LB contendo, se necessário, o antibiótico adequado, sendo incubada a 37 °C com agitação de 200 rpm.

Um mL do inóculo foi transferido para 100 mL de meio LB contendo, se necessário, o antibiótico adequado. A cultura foi incubada a 37 °C com agitação de 200 rpm entre 2 a 3 horas até atingir a densidade ótica a 600 nm de 0,6. A cultura foi transferida para um tubo de centrífuga e incubada em gelo por 30 min em repouso.

Em seguida, a cultura foi centrifugada a 4000 g por 5 min a 4 °C. O sobrenadante foi descartado com cuidado e as células ressuspendidas em 50 mL de solução estéril de CaCl<sub>2</sub> 100 mM + Hepes 10 mM pH 7. As células foram incubadas em gelo por 30 min em repouso e centrifugadas novamente.

O sobrenadante foi descartado novamente e as células ressuspendidas em 2 mL de solução estéril de CaCl<sub>2</sub> 100 mM Hepes 10 mM pH 7 + glicerol 10%. As células foram então aliqüotadas em tubos cônicos de 1,5 mL em aliqüotas de 50  $\mu$ L. As aliqüotas foram então armazenadas em -80 °C.

##### 4.1.4.2. Transformação de cepas de *Escherichia coli* por choque térmico

Cada aliqüota de células competentes congeladas foi colocada em gelo por 5 min. A cada aliqüota foi adicionado pelo menos 50 ng de plasmídeo e deixada incubando por 30 min em gelo. A aliqüota então foi colocada em termobloco a 42 °C por exatos 2 min. Após o aquecimento, a aliqüota foi colocada em gelo por 2 min. Então, adicionou-se 1 mL de meio LB e incubou-se sob agitação por 1 h a 37 °C com agitação de 200 rpm. Transformações de DNA plasmideal resultante de ligações foram centrifugadas por 10 min a 3000 g e ressuspendidas em

100  $\mu$ L de LB e todo o volume foi plaqueado. Transformações convencionais não foram centrifugadas e apenas 100  $\mu$ L foram plaqueados. As placas foram incubadas em estufa 37 °C por 18h.

#### 4.1.4.3. Extração de DNA plasmideal por lise alcalina de *Escherichia coli*

Culturas de 2 mL foram centrifugadas a 4000 g por 10 min. Os sobrenadantes foram descartados. As células foram ressuspendidas em 250  $\mu$ L de P1. Adicionou-se 100  $\mu$ g/ml de RNase. Incubou-se por 10 min e então adicionou-se 200  $\mu$ L de P2 e homogeneizou-se suavemente por inversão. Incubou-se por 5 min. Por fim, adicionou-se 250  $\mu$ L de P3 e misturou-se suavemente até formação de grumos opacos. Centrifugou-se por 30 min a 20000 g a 4 °C. Retirou-se o sobrenadante com cuidado e adicionou-se 0,6 V de isopropanol. Centrifugou-se novamente e descartou-se o sobrenadante. Os DNAs precipitados foram então lavados com 1 mL de etanol 70%. Centrifugou-se por 10 min e descartou-se o sobrenadante, deixando-se os DNAs precipitados secar. Por fim, ressuspendeu-se os DNAs em 50  $\mu$ L de água.

#### 4.1.4.4. 3.1.4.3 Reação em cadeia da polimerase (PCR)

O protocolo para amplificação foi o mesmo para todas as PCRs de subclonagem utilizadas nesse projeto, como descrito nos quadros 3 e 4:

**QUADRO 3 MISTURA DA PCR**

Reagente:	Quantidade ( $\mu$ L):
Água	10,5
Q5 Tampão 5X	10
dNTP 0,2 mM	1
Oligonucleotídeo senso (10pmol)	1
Oligonucleotídeo anti-senso (10pmol)	1
Molde (0,5-1,0 ng)	1
Q5 <i>High-Fidelity</i> DNA polimerase	0,5

Fonte: O próprio.

**QUADRO 4 CICLOS DA PCR**

Ciclo:	Temperatura:	Tempo:
Desnaturação Inicial:	98 °C	30 segundos
Desnaturação (x35):	98 °C	10 segundos
Anelamento (x35):	58 °C	30 segundos
Extensão (x35):	72 °C	30 segundos
Extensão final:	72 °C	2 minutos
Resfriamento:	16 °C	-

Fonte: O próprio.

Após o término, a reação de amplificação foi verificada aplicando-se 1  $\mu$ L da reação misturado em tampão de amostra de DNA em gel de agarose 2%, o qual foi submetido a uma

voltagem de 80 V por 1 hora. Após esse tempo, o gel foi submerso em solução de brometo de etídeo para sucessiva revelação por luz UV 616 nm.

#### 4.1.4.5. Reação de digestão de fragmentos de DNA

Misturar 1 unidade de enzima de restrição (Anza) para cada 1 µg de DNA plasmídeo, utilizando o tampão recomendado pelo fabricante na concentração final 1X e água ultrapura para completar o volume. As reações foram incubadas à 37 °C por 16 horas.

#### 4.1.4.6. Reação de ligação de fragmentos de DNA

As reações de ligação foram feitas utilizando uma razão molar de inserto-vetor de 3:1. Misturou-se o inserto e o vetor, adicionando-se 0,5 µL de enzima T4 DNA ligase (Anza), tampão T4 DNA ligase (Anza) na concentração final de 1X e água ultrapura para o volume final desejado. As reações foram incubadas à 4 °C por 16 horas.

#### 4.1.4.7. Expressão de proteínas recombinantes

##### ❖ Fatores de Crescimento da Família EGF & NRG1 $\alpha$

Escolheu-se uma colônia isolada da placa para a preparação de pré-inóculo de 2 mL de meio LB com 100 µg/mL de ampicilina, o qual foi mantido a 37 °C por 16-18 horas sob agitação de 200 rpm. Em seguida, transferiu-se 5 mL para frasco de Erlenmeyer de 250 mL contendo 50 mL de meio LB com 100 µg/mL de ampicilina. Manteve-se a solução a 37 °C sob agitação de 200 rpm até que a densidade óptica (D.O.) a 600 nm fosse de 0,6. Então, diminuiu-se a temperatura para 16 °C e, após 30 min, adicionou-se 0,5 mM de IPTG. Manteve-se a solução por 24 h nessas condições e, então, centrifugou-se a 6000 g por 15 minutos a 4 °C. O sedimento celular foi mantido congelado até o momento da lise e purificação.

##### ❖ Expressão Inicial dos scFv MM, OC, MMG, OCG, 32OC e 32MM

Escolheu-se uma colônia isolada de cada transformação dos plasmídeos MM, MMG, 32MM, OC, OCG e 32OC em cepa *Escherichia coli* BL21 Star (DE3)  $\Delta$ slyD pRARE2 da placa para a preparação de pré-inóculo de 10 mL de meio LB com 50 µg/mL de canamicina e 37 µg/mL de clorafenicol, que foi mantido a 37 °C por 16-18 horas sob agitação de 200 rpm.

Em seguida, 1% de pré-inóculo foi adicionado ao inóculo de 50 mL em frasco de Erlenmeyer de 250 mL. A cultura foi mantida à 37 °C sob agitação de 200 rpm até a D.O. a 600 nm de 0,5-0,7. Após alcançar a D.O. desejada, a cultura foi aclimatada à 16 °C por meia hora antes da adição de IPTG na concentração de 0,5 mM. O tempo de expressão foi de 20 h.

Após o tempo de expressão, centrifugou-se a cultura a 7000 g por 10 minutos a 4 °C, lavou-se com tampão contendo 50 mM Tris HCl pH 8, 300mM NaCl e 5 mM EDTA, centrifugou-se novamente, descartou-se o sobrenadante e ressuspendeu-se em tampão de lavagem. Então, congelou-se a solução de células até o momento do processamento.

#### ❖ Otimização da Expressão dos Fragmentos de Anticorpos fusionados a SGFP2

Para a otimização da expressão dos fragmentos de anticorpos foram utilizadas as construções fusionadas a SGFP2 devido a facilidade de visualização do resultado pela fluorescência ou não das culturas. Para isso adotou-se a estratégia de QbD OFAT, a qual se baseia na alteração de uma condição por vez. Essa técnica além de ajudar a sistematizar testes de expressão também demonstra quais variáveis realmente são importantes para a produção de uma determinada proteína em uma determinada cepa bacteriana.

As condições de cada teste foram iguais às do padrão, com exceção das variáveis destacadas na Tabela 3, tendo cada teste sido realizado de maneira simultânea entre os scFvs OCG e MMG na cepa BL21(DE3)  $\Delta$ SlyD pRARE2. Para controle foi utilizada a construção OPN-SGFP2, devido ao seu natural alto rendimento e facilidade de expressão adquirida em outros trabalhos do grupo. A única condição que não consta na Tabela 3 foi a expressão com colônias co-transformadas simultaneamente com cada um dos scFvs e o pET32-EGF, mas que se utilizou da condição padrão para seu teste.

Os nomes dos meios de cultivo diferentes se referem a substituição de fonte de carbono (Condições 2 a 4), a quantidade da fonte de nitrogênio (Condição 5), substituindo a fonte de nitrogênio (Condições 5 a 9), o tipo de sal (Condição 10) do meio de lisogenia (LB, do inglês *lysogeny broth*) nas concentrações descritas na Tabela 3. Meios prontos também foram testados (Condições 11 a 13), assim como diferentes D.O. de indução (Condições 14 e 15), diferentes concentrações finais de IPTG (Condições 16 e 17), diferentes temperaturas de expressão (Condições 18 e 19) e diferentes tempos de expressão (Condições 20 e 21). Todos os espaços em branco na Tabela 3 são as condições padrão.

Também foi otimizado o tampão de lise. Foram testados diferentes aditivos para verificar uma melhora da proteína de interesse na fração solúvel. Os aditivos foram testados

para as concentrações finais de: SDS 0,5%, SDS 0,1%, Glicerol 10%, NP-40 0,1% e sarcosil 0,2%. O tampão de lise celular original é composto por 50 mM Tris HCl pH 8 300 mM NaCl 20 mM imidazol, o qual tem a mesma composição do tampão de afinidade A. O intuito final é achar uma condição de expressão melhor e a partir dela tentar retirar os aditivos, uma vez que eles têm grandes chances de interferirem nos testes de interação pretendidos.

**QUADRO 5 VARIÁVEIS DAS CONDIÇÕES DE EXPRESSÃO DOS TESTES**

Condição	Meio de cultivo	D.O. de indução a 600 nM	Concentração final de IPTG	Temperatura de expressão	Tempo de expressão
Padrão	LB. Fonte de Carbono (FC): Peptona 1% (m/v)/ Fonte de Nitrogênio (FN): Extrato de Levedura 1% (m/v)/ Sal: NaCl 250 mM	0,6	0,5 mM	16 °C	20 h
1	LB-Glicose (1%)				
2	LB-Lactose (1%)				
3	LB-Glicerol (1%)				
4	LB-Tween 80 (7,5%)				
5	LB-Extrato de levedura (2%)				
6	LB-Ureia (3%)				
7	LB-Sulfato de Amônio (1%)				
8	LB-Peptona (1%)				
9	LB-Infusão de Fígado (1%)				
10	Lb-CaCl 10 mM				
11	TB ( <i>Terrific broth</i> ) FC: Triptona 1,2% (m/v) e Glicerol 0,5% (m/v). FN: Extrato de Levedura 2,4% (m/v). Sal: Tampão Fosfato (KH <sub>2</sub> PO <sub>4</sub> /K <sub>2</sub> HPO <sub>4</sub> ) 89 mM				
12	SB ( <i>Super broth</i> ) FC: Triptona 3% (m/v). FN: Extrato de Levedura 2% (m/v). Sal: MOPS 1% (m/v).				
13	SOB ( <i>Super optimal broth</i> ) FC: Triptona 2% (m/v). FN: Extrato de levedura 0,5% (m/v). Sal: NaCl 10 mM mM KCl 2,5 MgSO <sub>4</sub> 20 mM				
14		0,06			
15		1,2			
16			0,1 mM		
17			1 mM		
18				4°C	
19				37°C	
20					3h
21					28h

Fonte: O próprio.

❖ Expressão Final dos Fragmentos de Anticorpos MM, OC, MMG, OCG, 32OC e 32MM

Uma colônia isolada de cada transformação dos plasmídeos MM, MMG, 32MM, OC, OCG e 32OC em cepa *Escherichia coli* Shuffle foi escolhida da placa para a preparação de pré-inóculo de 10 mL de meio LB com 50 µg/mL de canamicina, o qual foi mantido à 37 °C por 16-18 horas sob agitação de 200 rpm.

Em seguida, 1% de pré-inóculo foi adicionado ao inóculo de 50 mL em frasco de Erlenmeyer de 250 mL. A cultura foi mantida à 37 °C sob agitação de 200 rpm até a D.O. a 600



nm de 0,5-0,7. Após a D.O. desejada, a cultura foi aclimatada à 16 °C por meia hora antes da adição de IPTG na concentração de 0,5 mM. O tempo de expressão foi de 20 h.

Após o tempo de expressão, a cultura foi centrifugada à 7000 g por 10 minutos à 4 °C, lavada com tampão contendo 50 mM Tris HCl pH 8, 300mM NaCl, centrifugada novamente, descartado o sobrenadante e ressuspendido em tampão de lavagem. A solução de células foi então congelada até o momento do processamento e extração das proteínas.

#### 4.1.4.8. Análise estrutural de scFvs com estrutura atômica resolvida

Com o intuito de desenvolver novas construções que tivessem seus parâmetros de solubilidade e rendimento melhorados, foram buscadas todas as estruturas no *Protein Data Bank* contendo scFvs cristalizados a partir da fração solúvel, sem passar por corpos de inclusão ou contendo etiquetas de aumento de solubilidade.

Para isso, utilizando-se dos filtros para busca em texto com a palavra “scFv” e filtro para organismo de expressão com a palavra “coli”. A partir dos resultados foram filtrados manualmente aqueles que de fato se encontravam dentro das condições necessárias para este estudo.

As estruturas então foram analisadas utilizando-se do software PyMol e dos modelos gerados utilizando-se da plataforma online Kotai Antibody Builder para definir as estruturas mais compatíveis com as sequências originais, verificando-se alinhamento sequencial estrutural, sobreposição estrutural, tamanho das CDRs, comparação entre os tamanhos de CDRs teóricos e práticos, tamanho dos conectores sintéticos entre cadeias, organismo de origem e resíduos da zona de Vernier.

#### 4.1.4.9. Desenho de construções híbridas

Levando em conta a análise anterior, escolheu-se sequências compatíveis para serem utilizadas de arcabouços para a inserção das CDRs das cadeias leve e pesadas no formado scFv.

Para o desenho das construções híbridas foram escolhidas as sequências com maior compatibilidade de sequência do arcabouço e tamanho e constituição das CDRL1 e CDRH3. Além destas, foram incluídas construções levando em conta sequências de outros scFvs os quais o laboratório já possuía experiência em expressão, mas que não se adequavam nesses critérios para intuito de comparação. Também foram encomendadas duas construções que além da

substituição das sequências das CDRs no arcabouço escolhido, os resíduos da zona de Vernier foram modificados.

No total, foram encomendadas para a empresa GenOne em pET28a sete construções, quatro derivadas do scFv MM e três derivadas do scFv OC.

Para as construções MM foram escolhidas uma com sequência compatível da análise anterior (4NIKMM), duas com sequências de scFvs que o laboratório já produziu (146MM e 148MM) e uma com a sequência da análise anterior com mudanças também nos resíduos da zona de Vernier (VERMM).

Para as construções OC foram escolhidas uma com a sequência compatível da análise anterior (6CVKOC), uma com sequência de scFv que o laboratório já produziu (146OC) e uma com a sequência da análise anterior com mudanças também nos resíduos da zona de Vernier (VEROC).

#### 4.1.4.10. Expressão dos scFvs híbridos

Foram testadas duas cepas de expressão para a expressão dos scFvs híbridos, sendo uma a cepa BL21  $\Delta$ SlyD pRARE2 e a Shuffle. Para o teste de expressão, foram realizados inóculos de 50 mL com 1% de pré-inóculo com 50  $\mu$ L/mL de canamicina. Incubou-se a 37 °C com agitação de 200 rpm até a D.O. à 600 nm atingir 0,6, quando os inóculos foram aclimatados por 30 min à 16 °C. Após esse período, as culturas foram induzidas com 0,5 mM de IPTG e incubadas por 24 h. As expressões foram comparadas por western blot e eletroforese em gel de poliacrilamida com dodecil sulfato de sódio (SDS-PAGE) corado por azul de Coomassie.

#### 4.1.4.11. Lise celular por sonicação e tratamento enzimático

As culturas congeladas foram descongeladas em gelo e então ressuspensas em 1 mL de tampão de afinidade A contendo lisozima para concentração final de 0,05 mg/mL.

#### 4.1.4.12. Lise celular por alta pressão

O sedimento celular foi ressuspendido em 10 mL de tampão de afinidade A. As células foram lisadas por alta pressão utilizando-se de um microfluidificador (Microfluidics™, EUA) a 18000 psi. As amostras então foram centrifugadas a 20000 g por 30 minutos à 4 °C

para separação das frações solúveis e insolúveis. As frações solúveis foram mantidas em gelo para purificação por cromatografia de afinidade ao níquel.

#### 4.1.4.13. Purificação de proteínas recombinantes

- ❖ Cromatografia de afinidade ao níquel dos fatores de crescimento da família EGF e NRG1 $\alpha$

As frações solúveis dos extratos celulares foram transferidas para um superloop de 50 ml e montado em um Sistema de cromatografia ÄKTA Pure M25 (Cytiva, Marlborough, EUA). As proteínas recombinantes foram inicialmente purificadas por cromatografia de afinidade usando uma coluna HisTrap de 5 mL (Cytiva 17524802) pré-equilibrada em tampão A. A taxa de fluxo para todas as etapas foi de 5 mL/min. Após o carregamento da amostra, a coluna foi lavada com 3 volumes (CV) de tampão A, seguido por um gradiente de 0 a 10% de tampão B (50 mM Tris.HCl pH 8,0; 300 mM NaCl; 500 mM de imidazol) em 2 CV e, a proteína eluída usando um gradiente de 10 a 100% de tampão B em 4 CV. Frações de 1 mL foram coletados e as frações de pico analisadas por SDS-PAGE.

- ❖ Digestão das proteínas de fusão para separação da Trx das proteínas de interesse

Para a etapa de retirada da Trx, as frações foram reunidas e digeridas com trombina. Para isso, um volume de 5 mL de amostras de pico da cromatografia de afinidade foi diluído para um volume final de 15 mL usando tampão de diluição (Tris.HCl 25 mM pH 8,0; NaCl 150 mM). As digestões foram realizadas usando 1 U de trombina (Cytiva 27084601) para 10 mg de proteína de fusão recombinante a 25°C por 24 h. Uma vez que a digestão foi confirmada por SDS-PAGE, os produtos da digestão foram carregados em um superloop de 50 mL e a Trx em fusão com hexahistidina ou as proteínas não digeridas foram capturadas em uma coluna HisTrap de 5 mL (Cytiva 17524802) pré-equilibrada em tampão A usando um sistema ÄKTA Pure M25 (Cytiva, Marlborough, MA, EUA). Após a injeção da amostra, a coluna foi lavada usando 3 CV de Tampão A. A fração não ligada contendo a proteína de interesse foi coletado e concentrado usando concentradores Amicon Ultra-0,5 mL 3K até um volume final de 500  $\mu$ L.

- ❖ Cromatografia de afinidade ao níquel dos fragmentos de anticorpos

As frações solúveis foram transferidas para um superloop de 50 ml e montado em um Sistema de cromatografia ÄKTA Pure M25 (Cytiva, Marlborough, EUA). As proteínas recombinantes foram inicialmente purificadas por cromatografia de afinidade usando uma coluna HisTrap de 1 mL (Cytiva 17524802) pré-equilibrada em tampão A. A taxa de fluxo para todas as etapas foi de 5 mL/min. Após o carregamento da amostra, a coluna foi lavada com 3 volumes (CV) de tampão A, seguido por um gradiente de 0 a 10% de tampão B (50 mM Tris.HCl pH 8,0; 300 mM NaCl; 500 mM de imidazol) em 2 CV e, a proteína eluída usando um gradiente de 10 a 100% de tampão B em 4 CV. Frações de 1 mL foram coletados e as frações de pico analisadas por SDS-PAGE. Em seguida, as frações foram concentrada utilizando-se de concentradores Amicon Ultra-0,5 mL 3K até um volume final de 500 µL.

#### ❖ Cromatografia de Exclusão por Tamanho

O concentrado de 500 µL as amostras foi purificado por cromatografia de exclusão por tamanho. Para isso, a coluna Superdex 75 10/300 GL (Cytiva 29148721) foi equilibrada com tampão de cromatografia de exclusão por tamanho (25 mM Tris.HCl pH 8,0; 150 mM NaCl) usando um Sistema ÄKTA Pure M25 (Cytiva, Marlborough, MA, EUA) com fluxo de 0,8 mL/min. Frações de 0,5 mL foram coletadas e as frações do pico cromatográfico foram analisadas por SDS-PAGE.

#### 4.1.4.14. Western blot

Géis de poliacrilamida após corrida eletroforética contendo amostras de interesse foram empilhados com membrana de nitrocelulose ativada por metanol com três folhas de papel de filtro em cima e três embaixo, com a membrana voltada ao polo positivo do sanduíche e o gel voltado ao polo negativo. Todas as camadas foram previamente embebidas em tampão de transferência. Então a transferência ocorreu por 1 h a 20 V em sistema de transferência úmido.

A membrana foi mergulhada em solução de bloqueio por 1 h a 4 °C. A membrana foi lavada três vezes com solução PBS-T 1X e então incubada com anticorpo primário diluído em PBS-T 1X conforme o fabricante. A membrana foi novamente lavada três vezes com solução PBS-T 1X e então incubada com anticorpo secundário marcado com fluoróforo diluído em PBS-T 1X conforme o fabricante. Por fim, a membrana foi lavada com PBS 1X e então lida em LiCor Odyssey Infrared Imaging System (LI-COR Biosciences, Nebraska, EUA). Quando necessário, as bandas foram quantificadas pelo próprio software do aparelho.

#### 4.1.4.15. Dot blot

Em membrana de celulose foram aplicados 2  $\mu$ L de proteína em triplicata e esperou-se 30 min. A membrana foi mergulhada em solução de bloqueio por 1 h a 4 °C. A membrana foi lavada três vezes com solução PBS-T 1X e então incubada com anticorpo primário diluído em PBS-T 1X conforme o fabricante (nome do fabricante, local sede do fabricante).

A membrana foi novamente lavada três vezes com solução PBS-T 1X e então incubada com anticorpo secundário marcado com fluoróforo diluído em PBS-T 1X conforme o fabricante. Por fim, a membrana foi lavada com PBS 1X e então lida em LiCor Odyssey Infrared Imaging System (LI-COR Biosciences, Nebraska, EUA). Quando necessário, as bandas foram quantificadas pelo próprio software do aparelho.

#### 4.1.4.16. Ensaio de proteólise limitada

A proteólise limitada foi realizada com os fatores de crescimento recombinantes em condições redutoras e não redutoras. Para redução das ligações dissulfeto, os fatores de crescimento foram tratados com 10 mM DTT (Merck 3483123). As amostras foram incubadas com 0,1% (p/p) de tripsina (Sigma-Aldrich 59427C) por 30 e 60 min a 37°C. Os controles sem tripsina foram incubados por 60 min. Os produtos de digestão e reações de controle foram analisados por SDS-PAGE.

#### 4.1.4.17. Ensaio de dicroísmo circular

Para análises de dicroísmo circular, o tampão de amostra dos fatores de crescimento recombinantes foi alterado por ultrafiltração para 4,2 mM Tris-HCl pH 8,0 e 25 mM NaCl usando concentradores Amicon Ultra-0,5 mL 3 K. Os espectros de CD foram adquiridos usando fatores de crescimento recombinantes a 0,2 mg/mL em cubeta de 1 mm em um espectrômetro de dicroísmo circular Jasco J-815 (JASCO, Tóquio, Japão). A velocidade de varredura foi de 50 nm/min.

Cada espectro foi acumulado pelo menos cinco vezes. Os espectros foram adquiridos de 195 a 260 nm a 20°C e após elevar a temperatura de 20 a 70°C, 75° ou 85°C, dependendo do limite de voltagem. Um novo conjunto de espectros foi adquirido após resfriamento a 20°C. Os dados brutos foram convertidos usando a fórmula de elipticidade molar residual média e o autozero foi definido usando o valor em 250 nm.

#### 4.1.4.18. Caracterização proteica por espalhamento dinâmico de luz supersensível e nanofluorimetria de varredura diferencial

Os fatores de crescimento recombinantes purificados foram analisados em um dispositivo nanoDSF Prometheus Panta (NanoTemper Technologies, Munique, Alemanha) para análise de espalhamento dinâmico de luz (DLS) supersensível e análises de efeito térmico sob condições redutoras e não redutoras usando capilares padrão. As medições de DLS supersensível foram realizadas a 20 °C usando a configuração de alta sensibilidade (10 aquisições de 5 s cada) e dois capilares independentes. Para análise do efeito térmico, as amostras sem agentes redutores foram submetidas a uma rampa térmica linear de 20 a 95°C com medição paralela de DLS.

#### 4.1.4.19. Mapeamento de pontes dissulfeto por espectrometria de massas

Os fatores de crescimento recombinantes foram digeridos sem redução de cisteínas utilizando diferentes proteases selecionadas de acordo com os sítios de clivagem esperados presentes em sua sequência primária e os peptídeos foram analisados por métodos LC-MS/MS com as combinações descritas no quadro 6.

**QUADRO 6. ESTRATÉGIA DE AQUISIÇÃO DE DADOS POR ESPECTROMETRIA DE MASSAS**

<b>Fator:</b>	<b>Proteases:</b>	<b>LC-MS/MS</b>
<b>AREG</b>	<b>Lys-C Glu-C</b>	<b>nano LC EThcD, ESI LC HCD</b>
<b>BTC</b>	<b>Lys-C Glu-C</b>	<b>nano LC EThcD, ESI LC HCD, nano LC HCD</b>
<b>EGF</b>	<b>Lys-C Glu-C</b>	<b>nano LC EThcD, ESI LC HCD, nano LC HCD</b>
<b>EPGN</b>	<b>Lys-C Tripsina Pepsina</b>	<b>nano LC EThcD, ESI LC HCD</b>
<b>EPR</b>	<b>Tripsina Pepsina</b>	<b>nano LC EThcD, ESI LC HCD</b>
<b>HB-EGF</b>	<b>Lys-C Glu-C Pepsina</b>	<b>nano LC EThcD, ESI LC HCD, nano LC HCD</b>
<b>NRG1<math>\alpha</math></b>	<b>Lys-C Glu-C</b>	<b>nano LC EThcD, ESI LC HCD, nano LC HCD</b>
<b>TGF<math>\alpha</math></b>	<b>Lys-C Pepsina</b>	<b>nano LC EThcD, ESI LC HCD</b>

Fonte: O próprio.

Para a preparação dos peptídeos, 5  $\mu$ g de fatores de crescimento recombinantes não reduzidos foram incubados em solução de 6 M de ureia + 5 mM N-etilmaleimida (NEM) por 15 min à temperatura ambiente, digeridos por 4 h a 37°C com 0,1  $\mu$ g Lys-C (exceto EPR), diluído com tampão compatível com ambas proteases e incubado após esse período com a outra protease. As razões de protease/proteína recombinante variaram de 1:20 a 1:50. A segunda e terceira reações proteolíticas foram incubadas a 37°C durante a noite e inativadas por acidificação ou aquecimento. Os produtos de proteólise foram dessalinizados usando ponteiras C18 e os peptídeos eluídos das ponteiras concentrados por evaporação *speed vac*. Cem nanogramas de peptídeos digeridos foram submetidos a LC-MS/MS utilizando um

cromatógrafo Ultimate 3000 RSLCnano acoplado a um espectrômetro Orbitrap Fusion Lumos (Thermo Fischer Scientific, Waltham, MA, EUA).

As análises de cromatografia líquida foram realizadas utilizando um fluxo de 100  $\mu\text{L}/\text{min}$  (ESI LC HCD) ou 250  $\text{nL}/\text{min}$  (nano LC HCD) e gradiente de 5 a 40% de MeCN, 0,1% de ácido fórmico por 90 min. A aquisição de dados foi realizada no modo DDA com espectros MS1 e MS2 (fragmentação pós HCD) adquiridos com o orbitrap configurado para resolução de 120k com controle automático de ganho de  $4 \times 10^5$  e tempo máximo de injeção de 200 ms ou 240 ms, respectivamente. Nas análises de nano LC EThcD, gradientes de 5 a 40% de MeCN, 0,1% de ácido fórmico foram realizados por 30 min.

Os espectros MS1 foram adquiridos no modo DDA com o orbitrap configurado para resolução de 120k com padrão de controle automático de ganho e tempo máximo de injeção de 50 ms. A fragmentação foi realizada usando EThcD MS2 com o orbitrap ajustado para resolução de 30k usando padrão de controle automático de ganho e tempo máximo de injeção de 54 ms.

Os dados brutos foram analisados usando o programa computacional BioPharma Finder 3.0 e a identificação dos locais de ligação dissulfeto deveria atender aos seguintes critérios: até 3 ligações dissulfeto por íon; Identificações MS2, erro de massa do precursor  $\leq 5$  ppm e pontuação de confiança  $> 0$ .

#### 4.1.4.20. Ensaio de proliferação celular

Células HeLa foram cultivadas em DMEM (ThermoFisher 41965062) suplementado com 10% de BFS até a confluência usando uma incubadora Forma Series II Water Jacketed CO2 (Thermo Fischer Scientific, Waltham, MA, EUA). As células foram cultivadas a 37°C, com umidade e 5% de CO2. Posteriormente,  $10^5$  células por poço foram transferidas para uma placa de 24 poços e incubadas por mais 24 h, quando o meio foi substituído por DMEM sem BFS, e as culturas foram tratadas com 50 ng/mL de fatores de crescimento EGF recombinantes em triplicados em paralelo com controles não tratados também em triplicata. Após 48 h, os poços foram lavados com PBS e tratados com 200  $\mu\text{L}$  de PBS contendo tripsina a 1% (p/v) (Merck T8003). Após 5 min de ação da tripsina a 37°C, as células liberadas foram coletadas por centrifugação a 500 g por 5 min e suspensas em 280  $\mu\text{L}$  de PBS e 20  $\mu\text{L}$  de CountBright™ Absolute Counting Beads (ThermoFisher C36950). O número de células foi determinado em um BD FACS Canto II (BD Life Sciences, New Jersey, EUA) e os resultados analisados utilizando o programa computacional FlowJo™ v10.8 (BD Life Sciences,



Franklin Lakes, NJ, EUA). Pelo menos 1.000 contagem foram adquiridas e um máximo de 3.000 contagens foi definido como limite superior. A quantificação foi realizada conforme descrito pelo manual CountBright™ Absolute Counting Beads (Invitrogen MAN0018850).

Para ensaios de dose-resposta,  $5 \times 10^4$  células HeLa por poço de uma cultura confluenta foram cultivadas em placas de 96 poços por 24 h usando uma incubadora Forma Series II Water Jacketed CO<sub>2</sub> (Thermo Fischer Scientific, Waltham, MA, EUA). As células foram cultivadas a 37°C, com umidade e 5% de CO<sub>2</sub>. Posteriormente, foram tratados com cinco concentrações crescentes de fatores de crescimento recombinantes em triplicado: 0,005 ng/mL, 0,05 ng/mL, 0,5 ng/mL, 5 ng/mL e 50 ng/mL. Após 48 h de incubação, os poços foram lavados com PBS, tratados com 50 µL de MTT (Invitrogen M6494) a 5 mg/mL e 50 µL de meio sem soro e incubados a 37°C por 3 h. Após este tempo, foram adicionados 150 µL da solução solvente (4 mM HCl, 0,1% NP40 em isopropanol) e a incubação continuou por mais uma hora. A absorbância foi lida a 570 nm usando um leitor híbrido Synergy H1 (BioTek Instruments, Winooski, VT, EUA). As curvas de dose-resposta foram ajustadas usando ajuste não linear dos dados.

#### 4.1.4.21. Ensaio de regeneração de camada celular

Células NDFH foram cultivadas em DMEM (ThermoFisher 41965062) suplementado com 10% de BFS até a confluência usando uma incubadora Forma Series II Water Jacketed CO<sub>2</sub> (Thermo Fischer Scientific, Waltham, MA, EUA). As células foram cultivadas a 37°C, com umidade e 5% de CO<sub>2</sub>. Posteriormente, o meio foi alterado para DMEM sem BFS e a incubação continuou por mais 24 h. Neste momento, foi feito um arranhão, 50 ng de cada fator recombinante foram adicionados às culturas em triplicata e imagens das células não coradas foram adquiridas. Após 48 h de incubação, as células foram coradas com cristal violeta (Synth C.I. 42555) de acordo com as instruções do fornecedor. Resumidamente, as células foram fixadas usando metanol 100% gelado por 10 min. Em seguida, 0,1 mL de uma solução de cristal violeta a 0,5% em metanol a 25% foi adicionado a cada poço, incubando-se por 10 min e lavando com PBS em seguida. As imagens das células foram adquiridas nos tempos 0 e 48 h após o tratamento com os fatores de crescimento em um microscópio invertido Nikon Eclipse TE300 (Nikon, Tóquio, Japão) usando uma objetiva de 10X e o software QCapture Pro 6.0 (Teledyne QImaging, Surrey, Canadá). A taxa de regeneração da camada celular foi calculada usando o programa computacional ImageJ. As imagens dos arranhões foram alinhadas com a base da imagem e um retângulo de 900 × 600 pixels foi cortado ao redor



do arranhão. A área no tempo 0 foi calculada usando a ferramenta de seleção de polígonos e a área em 48 h foi calculada usando a área entre as bordas da área de risco convertida em imagens de 8 bits e usando a opção *sharpened*. A ferramenta *find edge* foi utilizada e a imagem foi convertida para binária. As partículas foram então analisadas usando tamanhos entre 10 e 100.000 pixels com contornos nus, incluindo furos para calcular as áreas.

#### 4.1.4.22. Análises estatísticas

Os gráficos e as análises estatísticas foram realizados no software GraphPad Prism 5 (Graph Pad, San Diego, EUA). Os dados dos ensaios em triplicata para cada fator de crescimento foram comparados com os dados de controle usando análise de variância One-Way e pós-teste de Dunnett, onde  $p < 0,05$  foi considerado estatisticamente significativo.

#### 4.1.4.23. Cristalização do domínio EGF da NRG1 $\alpha$

Cristalização de proteínas é uma etapa inicial para elucidação da estrutura tridimensional através de cristalografia de raios-x. A cristalização consiste na supersaturação das moléculas proteicas em solução, levando a uma nucleação e crescimento de um cristal proteico. Essa supersaturação é atingida através do uso de propriedades físico-químicas e bioquímicas, como pH, temperatura e aditivos, tais como sais, detergentes, cofatores ou parceiros de interação (MCPHERSON; CUDNEY, 2014).

Diferentes técnicas podem ser empregadas para a cristalização, sendo a mais popular a difusão por vapor. Nessa técnica, uma gota contendo a proteína misturada com a solução precipitante é colocada numa câmara selada contendo solução precipitante pura, tornando possível a difusão do vapor de água da gota para o meio até o equilíbrio de osmolaridade do sistema. A desidratação da gota leva a concentração da proteína e do agente precipitante da gota permitindo assim a transição de fase para a nucleação do cristal. A gota em questão pode estar apoiada em um pedestal ou pendurada na tampa (MCPHERSON; CUDNEY, 2014). Para o experimento de cristalização foi escolhido o domínio EGF da hNRG1 $\alpha$ , o qual pelas análises estruturais demonstrou-se bastante estável.

Foram utilizadas três diferentes concentrações: 6,6 mg/mL, 10 mg/mL e 15 mg/mL. Para cada concentração foram utilizados dois kits comerciais contendo 96 soluções precipitantes cada, sendo eles: Index HT (Hampton Research) e JCSG Plus (Molecular Dimension). Em ambos os casos 200 nL de proteína foram misturados com 200 nL de solução precipitante para a formação da gota. A mistura de cada solução precipitante com a proteína e

a aplicação da gota na placa de cristalização foi realizado pelo mosquito <sup>®</sup> crystal (SPT Labtech). Os poços foram fotografados utilizando-se do sistema automatizado ROCK IMAGER <sup>®</sup> (FORMULATRIX).

#### 4.1.4.24. Desenho de VHHs anti-EGF

Devido à dificuldade de obtenção de scFvs anti-hEGF derivados de anticorpos comerciais de forma solúvel, estável e funcional, novas opções de construções foram avaliadas. Na literatura existe até o momento apenas um trabalho descrevendo sequências de CDRs anti-EGF, no entanto, a sequência dos arcabouços do VHH não foi divulgada (GUARDIOLA et al., 2018). Para isso então buscou-se na literatura e optou por utilizar o arcabouço universal proposto (SAERENS et al., 2005), havendo uma análise de compatibilidade *in silico* utilizando-se de modelos gerados pela plataforma SWISSMODEL, usando o template PDB:4TVS (BIENERT et al., 2017). Foram escolhidos o VHH NB1 e o VHH NB6 (Apêndice I) devido a maior quantidade de informações disponíveis no trabalho em questão, além de ambos se mostrarem funcionais em inibir a ligação do EGF ao hEGFR.

#### 4.1.4.25. Teste de conceito da apresentação em superfície bacteriana

A apresentação em superfície bacteriana é uma técnica relativamente nova que consiste na expressão de proteínas de interesse fusionadas a proteínas de membrana externa de bactéria. O transporte do citoplasma, para o espaço entre membranas e enfim ao exterior celular é feito através das sequências sinalizadoras da proteína de membrana. Nesse trabalho foram testadas duas diferentes proteínas de membrana externa de *Escherichia coli* com diferentes propriedades estruturais.

A primeira é a porção N terminal da proteína intimina (eae), nesse trabalho denominada Neae. Ela é uma proteína de membrana externa responsável pela ligação e lesão em tecidos pela *E. coli*. A região N terminal é formada por um domínio LysM, seguido de uma região de barril  $\beta$  inserido na membrana externa e por fim um domínio do tipo imunoglobulina D0. O complexo SEC é responsável pelo transporte do citoplasma para o espaço periplasmático, onde o complexo BAM ancora o barril  $\beta$  na membrana e transloca o domínio D0 e quaisquer proteínas fusionadas a ele para o lado externo da membrana (SALEMA et al., 2013).

A segunda é a porção N terminal da proteína de nucleação gélida (inaQ), nesse trabalho denominada inaQ. Ela é uma proteína de membrana externa responsável pela nucleação de cristais de gelo em temperaturas acima de 0 °C, permitindo assim a bactérias Gram

negativas danificar tecidos vegetais. Ela, ao contrário da intimina, é uma proteína intrinsecamente desestruturada, sendo sua região de sequências não repetitivas denominada de porção N terminal, seguida de uma porção central de sequências repetitivas e uma porção C-terminal de sequências não repetitivas. Seu mecanismo de translocação é independente de peptídeos sinais clássicos, mas seu mecanismo de ancoramento a membrana externa ainda não foi elucidado (XU et al., 2017).

Para a prova de conceito foram clonadas as sequências do NB1 e NB6 no lugar da marca tóxica ccdB. Como controle experimental também foi clonada a sequência da proteína OPN com a proteína de membrana inaQ. As construções foram comparadas frente a sua + pelo grau de fluorescência da SGFP2 e foram testadas frente a ligação ao EGF e ao EGF-mRFP por imunoenaios de interação em placa e em citometria de fluxo. O imunoenasão em placa foi realizado de duas maneiras. A primeira, adsorvendo o EGF no fundo da placa de 96 poços de fundo plano e borda escura através da mistura da proteína com 50 mM de tampão de carbonato de sódio pH 9,5 na concentração final de proteína de 1 mg/mL e aplicando-se 100 µL. A incubação foi feita por 16 horas à 4 °C. Após esse período, a placa foi lavada três vezes com PBS-T e incubada com as células numa D.O. a 600 nm de 1.0, aplicando-se 100 µL. A incubação foi feita por 1 hora e a placa foi então lavada três vezes com PBS e lida no sistema Invitrogen™ iBright™ CL1500 (Waltham, Massachusetts, EUA).

O segundo método foi adicionando-se 100 µL de células numa D.O. a 600 nm de 1.0, centrifugando por 4000 g 5 min. A placa foi lavada três vezes com PBS-T com centrifugações entre cada lavagem de 4000 g por 5 min e adicionando-se 100 µL da proteína EGF-mRFP numa concentração de 1 mg/mL. Incubou-se por 1 hora à temperatura ambiente. Lavou-se três vezes com PBS, com centrifugações entre cada lavagem de 4000g por 5 min e a placa então foi lida no sistema Invitrogen™ iBright™ CL1500 Waltham, Massachusetts, EUA).

#### 4.1.4.26. Citometria de fluxo

Para os ensaios de citometria de fluxo, culturas frescas foram centrifugadas à 4000g por 5 min e ressuspensas para D.O. a 600 nm final de 1,0 em 1 mL. Foram centrifugadas então novamente as culturas e lavadas duas vezes com 1 mL de PBS. Os volumes foram então ajustados para 400 µL. Metade desse volume foi incubado com 100 nM de EGF-mRFP e o volume completado para 400 µL. Incubou-se por 1 h a 4 °C no escuro. Depois, lavou-se mais duas vezes com 1 mL de PBS para retirada de proteínas não ligadas. O volume foi então ajustado para 1 mL e 100000 eventos foram adquiridos no citômetro. Controles de compensação

foram feitos utilizando as células BL21  $\Delta$ slyD como controle negativo de fluorescência, BL21 (DE3)  $\Delta$ slyD pRARE2 com OPN-inaQ expressa na membrana como controle positivo de fluorescência da SGFP2 e BL21  $\Delta$ slyD com Trx-EGF-mRFP expressa no citoplasma como controle positivo de fluorescência da mRFP.

#### 4.1.4.27. Análise de estruturas 3D de proteínas

As estruturas dos fatores de crescimento da família EGF e dos complexos de fatores de crescimento da família EGF e os receptores de EGF foram visualizadas usando PyMOL Molecular Graphics System versão 4.5.0, Schrödinger, LLC.

## 4.2. PROCEDIMENTOS DOS FATORES DE CRESCIMENTO DA FAMÍLIA EGF E A NRG1A

### 4.2.1. Produção e caracterização biofísica e funcional dos fatores de crescimento da família EGF e a NRG1 $\alpha$

A metodologia escolhida para a construção dos fatores de crescimento e de sua expressão foi baseada na literatura (KADO et al., 2016). Todos passaram por três etapas de purificação: afinidade, remoção da tioredoxina e exclusão por tamanho. Para comprovação do correto enovelamento foram realizados ensaios biofísicos e funcionais.

Para os ensaios biofísicos, primeiramente foi feito o ensaio de proteólise limitada, visando fazer uma análise inicial da presença ou ausência de pontes dissulfeto. O ensaio se baseia em analisar a ação de uma protease frente uma proteína em seu estado reduzido e não-reduzido, esperando-se menor ação proteolítica em uma proteína em seu estado não-reduzido.

Em seguida, foram realizados ensaios de dicroísmo circular, que visa avaliar o sinal das diferentes estruturas secundárias e desordenadas. Os domínios EGF não possuem sinais fortes no dicroísmo circular devido à falta de estruturas secundárias, no entanto, foram adquiridos espectros antes, durante e depois o aquecimento das proteínas a temperaturas de 70, 80 ou 85 °C, dependendo do limite da voltagem. A comparação entre os espectros é capaz de dar informações sobre a rigidez dos domínios e resistência térmica.

Também foram realizadas análises de nanoDSF para comparar a estabilidade das proteínas entre estados reduzidos e não-reduzidos, propriedades termodinâmicas, tamanho de partículas e suas fluorescências intrínsecas. O esperado era que as proteínas não-reduzidas fossem mais resistentes a temperatura, menor tamanho particular e mais estáveis.

Por fim, as pontes dissulfeto foram analisadas por espectrometria de massas, analisando-as proteínas digeridas por proteases específicas a quantidade e característica das sequências primárias, evitando-se processos redutores para a obtenção dos espectros. A etapa de preparação, aquisição e análise dos dados foi feita em parceria com a Plataforma de Espectrometria de Massas do Instituto Carlos Chagas. A composição dos peptídeos é capaz de informar precisamente a presença de pontes dissulfeto e entre que resíduos.

Para os ensaios funcionais, foram realizados ensaios de proliferação, devido à natureza dos domínios EGF em ativar vias de sinalização com essa finalidade, sendo analisados por citometria de fluxo a quantificação das células e também por ensaios de conversão de sais de formazan, verificando o efeito em diferentes concentrações das proteínas. Também foram realizados ensaios de regeneração de camada celular, uma vez que essas proteínas também tem a capacidade de ativar a migração celular junto da proliferação.

### 4.3. PROCEDIMENTOS DOS ANTICORPOS

#### 4.3.1. Produção dos scFvs e caracterização biofísica

Devido as experiências passadas dentro do grupo, sequências de Fabs foram transferidas nesse trabalho para o formado scFv, visando uma maior simplicidade para otimizar a obtenção de proteínas funcionais em bactérias. Nesse mesmo sentido, diferentes construções foram produzidas de início: scFvs sem proteínas de fusão, scFvs fusionados com a proteína verde fluorescente SGFP2 no C-terminal e por fim scFvs fusionados com a tioredoxina no seu N-terminal. Todas as construções foram confirmadas por sequenciamento.

Foram utilizadas estratégias de QbD para tentar otimizar a expressão das construções em fusão com SGFP2 devido a facilidade de verificar o efeito da expressão das proteínas pela fluorescência e devido ao desejo inicial de utilizá-las para ensaios de interação. Essa estratégia foi alterada depois dos resultados preliminares. Foram realizados estudos estruturais para o desenvolvimento de construções híbridas, se baseando na transferência das CDRs originais para outros arcabouços, para isso foram utilizados como base estruturas disponíveis no bando de dados do *Protein Data Bank* (PDB) para avaliar outras possibilidades de obtenção de proteínas funcionais.

Duas etapas de purificação foram utilizadas para todas as proteínas obtidas de maneira solúvel e em quantidades razoáveis de trabalho. A primeira, cromatografia de afinidade e a segunda cromatografia de exclusão por tamanho.

Então, utilizou-se de análises de nanoDSF em diferentes condições para avaliar a estabilidade, agregabilidade e a qualidade das proteínas recombinantes. Ensaios de *dot blot* também foram usados para avaliar de maneira inicial a capacidade de ligação ao alvo dos scFvs.

Também foram feitas tentativas de cristalização do domínio EGF da NRG1 $\alpha$  devido aos resultados obtidos de estabilidade e rendimento.

#### **4.3.2. Caracterização biofísica e de seletividade de anticorpos comerciais anti-EGF**

Também foram utilizados os anticorpos comerciais que foram sequenciados em trabalho prévio (FERREIRA, 2018) e que deram origem as sequências dos scFvs iniciais com intuito de comparação e demonstrar as propriedades originais dos anticorpos inteiros e assim avaliar suas propriedades termodinâmicas, estabilidade e qualidade utilizando-se também de nanoDSF.

Suas seletividades frente aos diferentes membros da família de fatores de crescimento EGF e a NRG1 $\alpha$  também foi avaliada utilizando-se de *western blot* e *dot blot*.

### **4.4. PROCEDIMENTOS DO SISTEMA DE EXPOSIÇÃO EM SUPERFÍCIE BACTERIANA**

#### **4.4.1. Padronização da prova de conceito de sistema de expressão em superfície bacteriana**

Por fim, utilizaram-se partes das proteínas de membrana externa de *Escherichia coli* para fusionar fragmentos de apenas cadeia pesada retirados da literatura (GUARDIOLA et al., 2018) em pauta com a SGFP2 em seu C-terminal para padronizar a utilização de um sistema de expressão em superfície bacteriana para a descoberta de novos fragmentos de anticorpos. Para isso, utilizaram-se ensaios de citometria de fluxo e de imunofluorescência direta em placa.

Para isso, foi desenvolvida uma construção de Trx-EGF fusionada a mRFP no seu C-terminal, a fim de utilizar de dupla fluorescência para comprovar a ligação ao alvo.

## **5. RESULTADOS & DISCUSSÃO**

### **5.1. FATORES DE CRESCIMENTO DA FAMÍLIA EGF E A NRG1A**

Os resultados e discussão da obtenção dos fatores de crescimento se encontram no artigo aceito e no submetido a seguir, com exceção da tentativa de cristalização da NRG1 $\alpha$  que se encontra no item subsequente.

#### **5.1.1. Artigo publicado na revista Scientific Reports**



OPEN

# A toolkit for recombinant production of seven human EGF family growth factors in active conformation

Arthur Schweitzer Ferreira<sup>1,2</sup>, Amanda Lopacinski<sup>1,2</sup>, Michel Batista<sup>3</sup>, Priscila Mazzocchi Hiraiwa<sup>1</sup>, Beatriz Gomes Guimarães<sup>1</sup> & Nilson Ivo Tonin Zanchin<sup>1,2</sup>✉

Epidermal growth factors (EGF) play a wide range of roles in embryogenesis, skin development, immune response homeostasis. They are involved in several pathologies as well, including several cancer types, psoriasis, chronic pain and chronic kidney disease. All members share the structural EGF domain, which is responsible for receptor interaction, thereby initiating transduction of signals. EGF growth factors have intense use in fundamental research and high potential for biotechnological applications. However, due to their structural organization with three disulfide bonds, recombinant production of these factors in prokaryotic systems is not straightforward. A significant fraction usually forms inclusion bodies. For the fraction remaining soluble, misfolding and incomplete disulfide bond formation may affect the amount of active factor in solution, which can compromise experimental conclusions and biotechnological applications. In this work, we describe a reliable procedure to produce seven human growth factors of the EGF family in *Escherichia coli*. Biophysical and stability analyses using limited proteolysis, light scattering, circular dichroism and nanoDSF show that the recombinant factors present folded and stable conformation. Cell proliferation and scratch healing assays confirmed that the recombinant factors are highly active at concentrations as low as 5 ng/ml.

Several diseases are associated to disruption of growth factor homeostasis. This is so for the family of epidermal growth factors (EGF) whose altered expression has been linked to several pathogeneses, such as lung adenocarcinoma and several other cancer types, chronic pain, chronic kidney disease and psoriasis<sup>1–5</sup>. Signaling by EGF growth factors is mediated by the ErbB family of G protein-coupled receptors, which comprises four members (ErbB1 to ErbB4). ErbBs have three conserved domains: an ectodomain responsible for ligand binding, a transmembrane domain, and a cytoplasmic tyrosine kinase domain responsible for signal transduction. Binding of the EGF growth factors to the receptors is mediated by the EGF domain, leading to homo or heteroligomerization of the receptors and activating cross-phosphorylation of their intracellular C-terminal domain. Receptor activation may trigger signal transduction to several pathways involved in differentiation, proliferation, migration or survival, with a wide range of targets including PLCY, STAT5, SHIP1, Abl, PTP-2c, STAT1, Syk, Vav2, RasA1, SH3BGR1, Grb2, Src, Crk, Nck, Shc and Cbl<sup>6</sup>. ErbB receptors are found in many cell types and tissues, such as blood cells, hepatocytes, cardiac cells, as well as in epithelial, endothelial, nervous, bone and adipose tissues<sup>7</sup>. EGF growth factors are likewise present in many cell types and tissues, such as epithelial cells, fibroblasts, blood cells, platelets, endothelial cells, colon, glands, mineralized tissues, among others<sup>8</sup>.

Structurally, all EGF growth factors share the EGF domain, which comprises six cysteine residues, with a conserved nucleus of constant number of residues, CX7CX4-5CX10-13CX8C. The six conserved cysteines form three disulfide bonds with the C1-C3, C2-C4, and C5-C6 combination. This arrangement results in formation of the A, B, and C loops, which divide the structure in the N-terminal portion, composed of loops A and B and the C-terminal portion, comprising loop C. Loops B and C are separated by a single residue, the so-called hinge residue, which allows flexibility between the two segments<sup>9</sup>. Both the N- and C-regions have distinct roles and are important for receptor selectivity due to differences in affinity of each region for the many different receptors. It is, thus, important that the entire domain is well-structured for efficient binding<sup>10,11</sup>.

<sup>1</sup>Laboratory of Structural Biology and Protein Engineering, Carlos Chagas Institute, FIOCRUZ Paraná, Curitiba, PR, Brazil. <sup>2</sup>Cellular and Molecular Biology Graduate Program, Federal University of Paraná, Curitiba, PR, Brazil. <sup>3</sup>Mass Spectrometry Facility RPT02H, Carlos Chagas Institute, FIOCRUZ Paraná, Curitiba, PR, Brazil. ✉email: nilson.zanchin@fiocruz.br



EGF growth factors can bind to different ErbB receptors and induce formation of different receptor dimers. However, they are usually grouped according to their biological activity and receptor specificity<sup>9,12,13</sup>. In this work, we focused on the seven members of the EGF family of growth factors that bind exclusively to ErbB1, also known as HER1 (human EGFR related) and as EGFR (epidermal growth factor receptor)<sup>8,9,12</sup>. In addition to epidermal growth factor (EGF) itself, this group includes amphiregulin (AREG), betacellulin (BTC), epiregulin (EPR), epigen (EPGN), Heparin-binding epidermal growth factor (HBEGF) and the transforming growth factor alpha (TGF $\alpha$ ). In human cells, these seven EGF growth factors are synthesized as type I transmembrane precursor proteins, composed of an N-terminal extension, the EGF domain, a membrane juxtaposed portion, a transmembrane domain, and a cytoplasmic C-terminal tail. The EGF precursor protein, however, is an exception, presenting nine repeats of the EGF domain<sup>14</sup> although, after processing only the one closest to the membrane can actively bind to the receptors. AREG and HBEGF have also a Heparin-Binding domain adjacent to the EGF domain<sup>8</sup>.

ErbB1 ligand EGF growth factors have wide utilization from fundamental research to state-of-the-art therapies<sup>15–17</sup>. However, recombinant production of EGF domain growth factors has been a challenging task, demanding either eukaryotic systems, such as yeast<sup>17,18</sup>, mammalian cell lineages<sup>19–21</sup>, or fusion with ancillary proteins<sup>22–26</sup> or even in vitro renaturing procedures<sup>27,28</sup>. Although the pro-forms of EGF family members have post-translational modifications even before membrane processing, none of the post-translational modifications is located at the EGF domain, allowing for their production in *Escherichia coli*, a system for which highly tested vectors for regulated expression are available, being also less time-consuming and more cost-effective<sup>29</sup>. Despite these advantages, recombinant production of complex proteins in *E. coli* may lead to incorrect folding, inclusion body formation, disulfide bond mismatch, among other limitations.

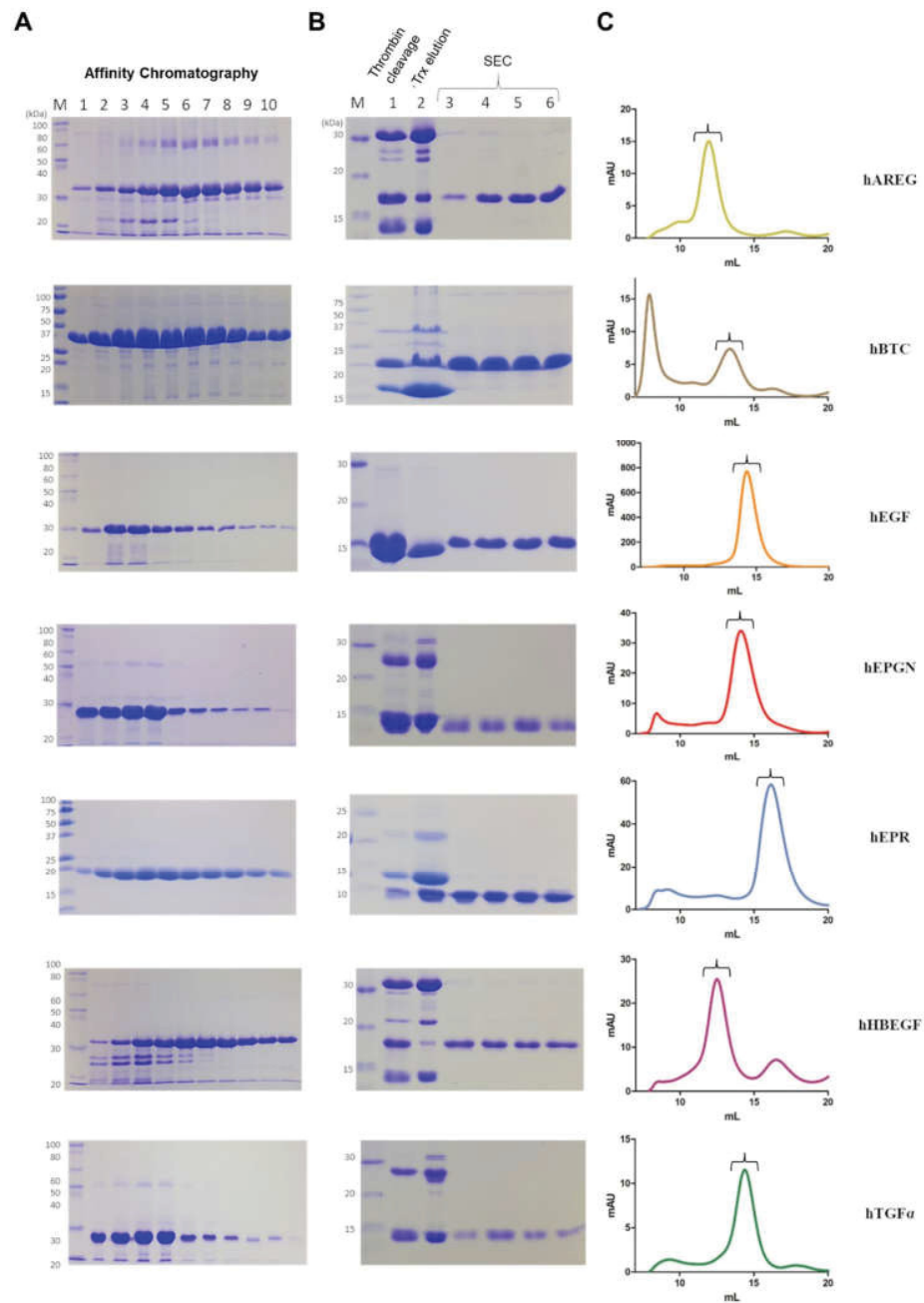
Disulfide bond formation is a critical step to obtain functional EGF growth factors. Therefore, many strategies have been attempted to overcome this problem, such as utilization of the *E. coli* Shuffle<sup>30</sup> and Origami 2 (DE3)<sup>31</sup> strains to try to control redox environment, secretion to the periplasm<sup>32</sup>, and fusion with chaperone-proteins, such as oleosin<sup>33</sup>, glutathione S-transferase<sup>34</sup>, SUMO<sup>35</sup>, B1 domain of streptococcal protein G<sup>34</sup>, protein disulfide isomerase (PDI)<sup>35</sup>, N-utilization substance protein A<sup>35</sup>, maltose-binding protein<sup>35</sup>, b'a' domain of PDI (PDIb'a')<sup>35</sup>, ELK16<sup>36</sup> and thioredoxin (Trx)<sup>37</sup>. In general, these studies have demonstrated the requirement of a fusion partner for proper folding of the EGF growth factors, which can be efficiently separated by proteolytic cleavage. Among the various fusion proteins used in previous studies, Trx and SUMO were shown to be efficient fusion partners for *E. coli* expression of hEPR<sup>22,37</sup> and hHBEGF<sup>25</sup>, respectively.

In addition to the technical limitations related to choosing the appropriate fusion partner for *E. coli* expression, there is also a need for consistent information on the procedures to purify the EGF domain growth factors, as well as for a description of the biophysical methods ensuring the correct structural conformation and, functional demonstration of their activity. Therefore, in this work, we developed a simple and easily reproducible method for recombinant expression, purification, biophysical characterization, and demonstration of activity for the seven members of the EGF growth factor family that are known to bind ErbB1. The same experimental set up and reagents can be used for production of all seven EGF factors. We demonstrate by using biophysical analyses that the recombinant factors present folded and stable conformation. Furthermore, we demonstrate their usefulness in cell proliferation and scratch healing assays.

## Results

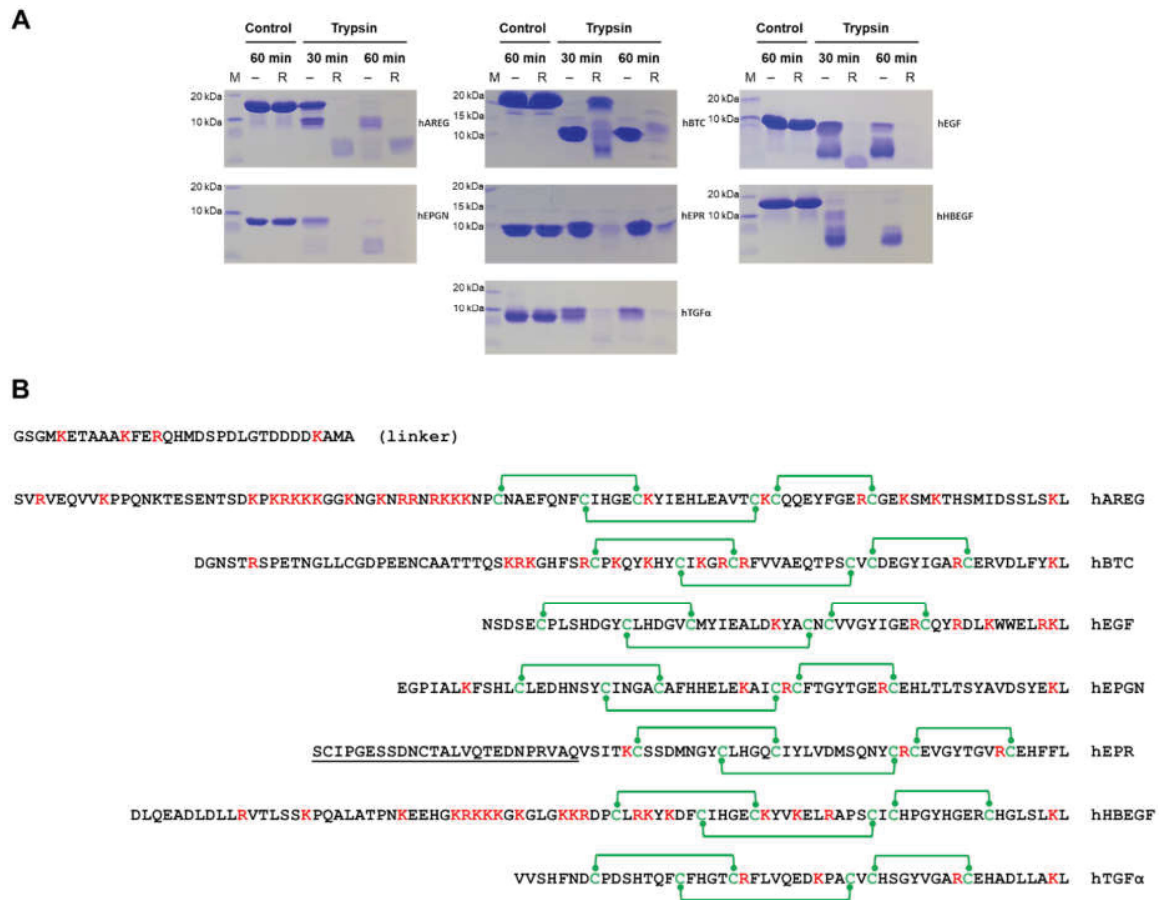
**Expression and purification of the EGF growth factors.** Expression of the EGF domain growth factors hAREG, hBTC, hEGE, hEPGN, hEPR, hHBEGF and hTGF $\alpha$  in fusion with Trx in the *E. coli* strain Shuffle performed overnight at 16 °C was highly efficient. An initial immobilized nickel affinity chromatography was used to purify the fusion proteins (Fig. 1A). Following separation of the histidine-tagged Trx from the EGF growth factors by thrombin digestion, the histidine-tagged Trx was captured on an immobilized nickel column. The growth factors recovered in the flow through were further fractionated by size-exclusion chromatography (Fig. 1B,C). hEGE, hEPGN, hEPR and hTGF $\alpha$  eluted in a single peak with the expected volume for a monomeric conformation while hAREG and hHBEGF eluted with lower volumes indicating dimerization as previously described<sup>38</sup>. hBTC, on the other hand, showed partial aggregation with approximately 50% of the sample eluting in the void volume. Only the hBTC fractions eluting in the expected volume for monomeric conformation were used in the assays. The secondary peak in the hHBEGF profile corresponds to contaminants. All factors showed a high yield and purity > 95% after the last step of chromatography (Fig. 1B). 500-mL cultures yielded approximately 20 mg of Trx-growth factor fusion proteins after the first affinity purification chromatography. The final yield after digestion of 10 mg of the fusion proteins with thrombin and size exclusion chromatography was: 1–3 mg for hAREG, 3–5 mg for hBTC, 6–8 mg for hEGE, 3–5 mg for hEPGN, 3–5 mg for hEPR, 1–3 mg for hHBEGF and 1–3 mg for hTGF $\alpha$ .

**Stability and conformational analysis of the recombinant EGF growth factors.** The conformation of the seven recombinant EGF growth factors was initially assessed by limited proteolysis assays. Considering that the EGF domain comprises the constant sequence CX7CX4-5CX10-13CXCX8C with disulfide bonds linking the cysteine residues in the combinations C1-C3, C2-C4, C5-C6, the limited proteolysis assays were performed with recombinant factors previously treated with DTT in parallel with control samples that were not treated with reducing agent (Fig. 2A). All reduced samples showed high sensitivity to trypsin, being totally degraded after 30 min of incubation time, except for BTC that showed some resistance, although it also was completely digested after 60 min. On the other hand, in the absence of reducing agents, the growth factors showed more resistance to trypsin proteolysis (Fig. 2A). hEGE, hEPR, and hTGF $\alpha$  were the most resistant, followed by hAREG and hEPGN with hBTC and hHBEGF being more sensitive. Considering only the non-reduced samples, higher resistance to trypsin seems to correlate with the lower number of trypsin cleavage sites found in the



**Figure 1.** Purification of the EGF growth factors. **(A)** Images of Coomassie stained SDS–polyacrylamide gels showing the peak fractions of the initial affinity chromatography (1 to 10). **(B)** Images of Coomassie stained SDS–polyacrylamide gels showing the result of thrombin protease digestion (lane 1), elution fraction from the purification of Trx-his6 after thrombin cleavage (lane 2) and peak fractions after the size exclusion chromatography (SEC) (lanes 3–6). Sizes (kDa) of the molecular mass markers (M) are indicated on the left. Only the relevant parts of the gels are shown. Original uncropped images of the gels are presented in the Supplementary Information file, Figure S1. **(C)** Elution profiles of the size exclusion chromatography analyses on a Superdex 75 10/300 GL column. Brackets indicate the region of the peak analyzed by SDS-PAGE in **(B)**. Each row represents the purification of the EGF growth factor indicated on the right.

EGF domain (Fig. 2B). Except for hEPR, all non-reduced samples showed intermediate bands (Fig. 2A) that we interpreted as the EGF domain nucleus maintained by the disulfide bonds. The intermediate band of hBTC shows particularly high resistance to degradation. In the case of hBTC, this intermediate band remains stable up to 60 min of trypsin treatment. It is unclear why proteolysis of hEPR does not produce such an intermediate band. In conclusion, the higher resistance to proteolysis shown by the growth factors in the absence of treatment

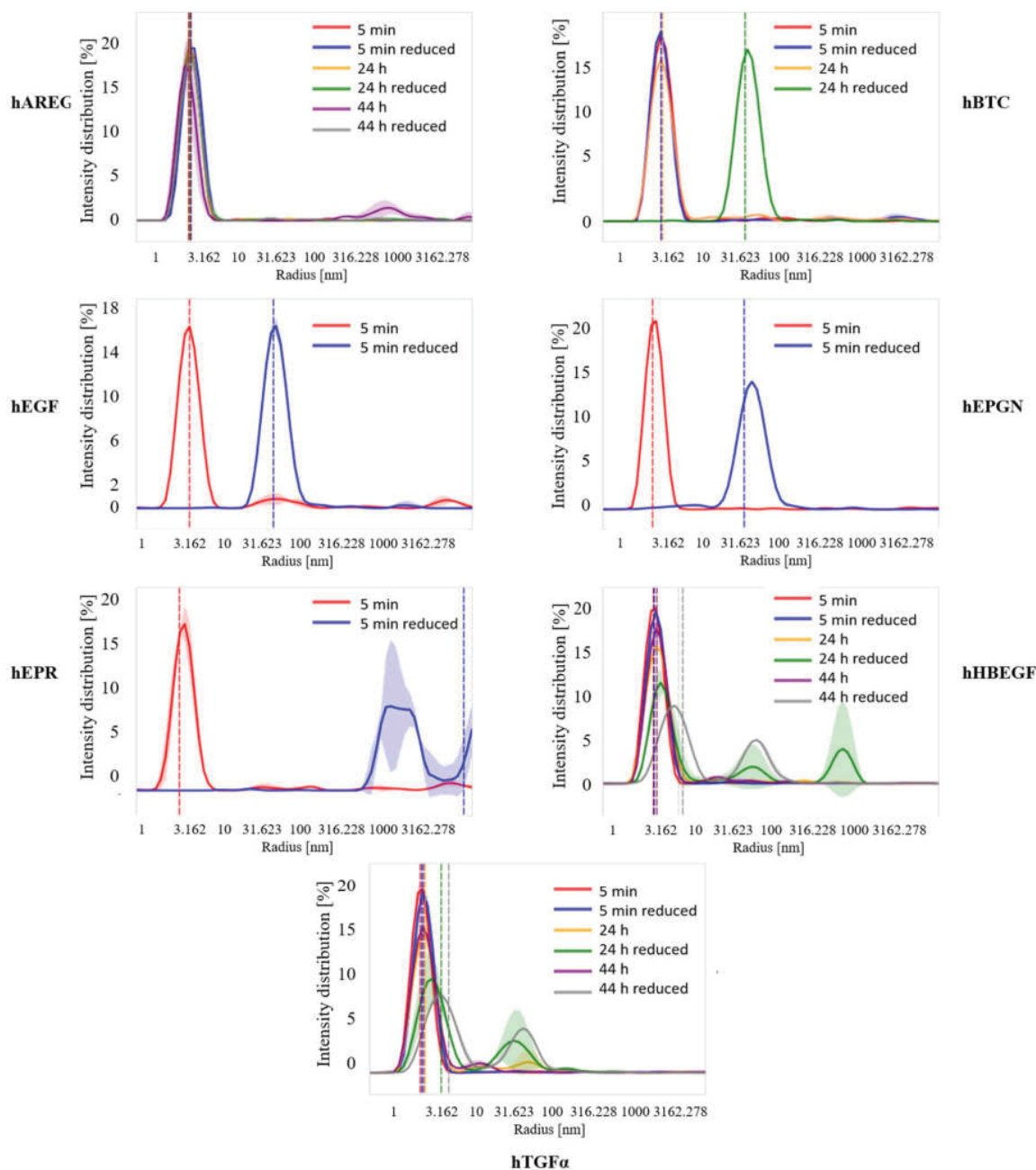


**Figure 2.** Stability analysis by limited proteolysis using trypsin. The proteolytic assays were performed with samples both untreated and treated with the reducing agent DTT and incubated for 30 and 60 min. Control non-reduced and reduced samples without trypsin treatment were incubated under the same conditions. (A) The panels show Coomassie stained SDS-polyacrylamide gels used for the analysis of the proteolysis products and control reactions of the EGF growth factors as identified on the right side of each panel. M, molecular mass markers. -, non-reducing condition. R, reducing condition. Only the relevant parts of the gels are shown. Original uncropped images of the gels are presented in the Supplementary Information file, Figure S2. (B) Sequences of the EGF domains with the basic residues target of trypsin marked in red and the cysteine residues marked in green. Green lines indicate the disulfide bonds. The sequence identified by (linker) represents the part of the linker encoded by the vector which remains attached to the growth factors after thrombin digestion. The underlined EPR region indicates the segment of the hEPR propeptide required for stable expression of this EGF growth factor.

with reducing agents gives an indication that the disulfide bonds are formed and possibly also display the right combination.

The effect of the reducing agent on the conformation of the growth factors was also investigated by dynamic light scattering (DLS, Fig. 3). Under non-reducing conditions all growth factors showed a single peak with radius size centered at ~ 3 nm. According to the DLS measurements, hEGF, hEPGN and hEPR respond quickly to treatment with the reducing agent. Immediately after adding DTT, the radius size of hEGF and hEPGN shifts from ~ 3 to 30–35 nm while the hEPR radius size increases to ~ 1000 nm, indicating aggregation after reduction of the disulfide bonds (Fig. 3). In the case of hBTC, an effect on radius size, similar to the one seen for hEGF and hEPGN, was detected only after 24 h of incubation with the reducing agent (Fig. 3). Surprisingly, the hAREG radius size was not affected by the reducing agent even after 44 h of incubation while hHBEGF and hTGFα shown partial increase in the radius size after 24 and 44 h of incubation with the reducing agents (Fig. 3). In general, the effect of the reducing agent on the radius size indicates that before reducing the disulfide bonds, the growth factors are properly folded.

Thermal stability of the growth factors under non-reducing conditions was initially accessed by nano-differential scanning fluorimetry (nanoDSF) combined with parallel radius size analysis by DLS. Since EGF has two tryptophan residues and other EGFs have tyrosine and phenylalanine residues, the fluorescence signal was followed at 330 nm. The fluorescence signals did not show a clear transition between unfolded and folded states with temperature increase (Supplementary Figure S3). For hEGF, hEPGN, hEPR and hTGFα, the radius size did not change significantly indicating that there is no or little aggregation despite temperature increase (Fig. 4).

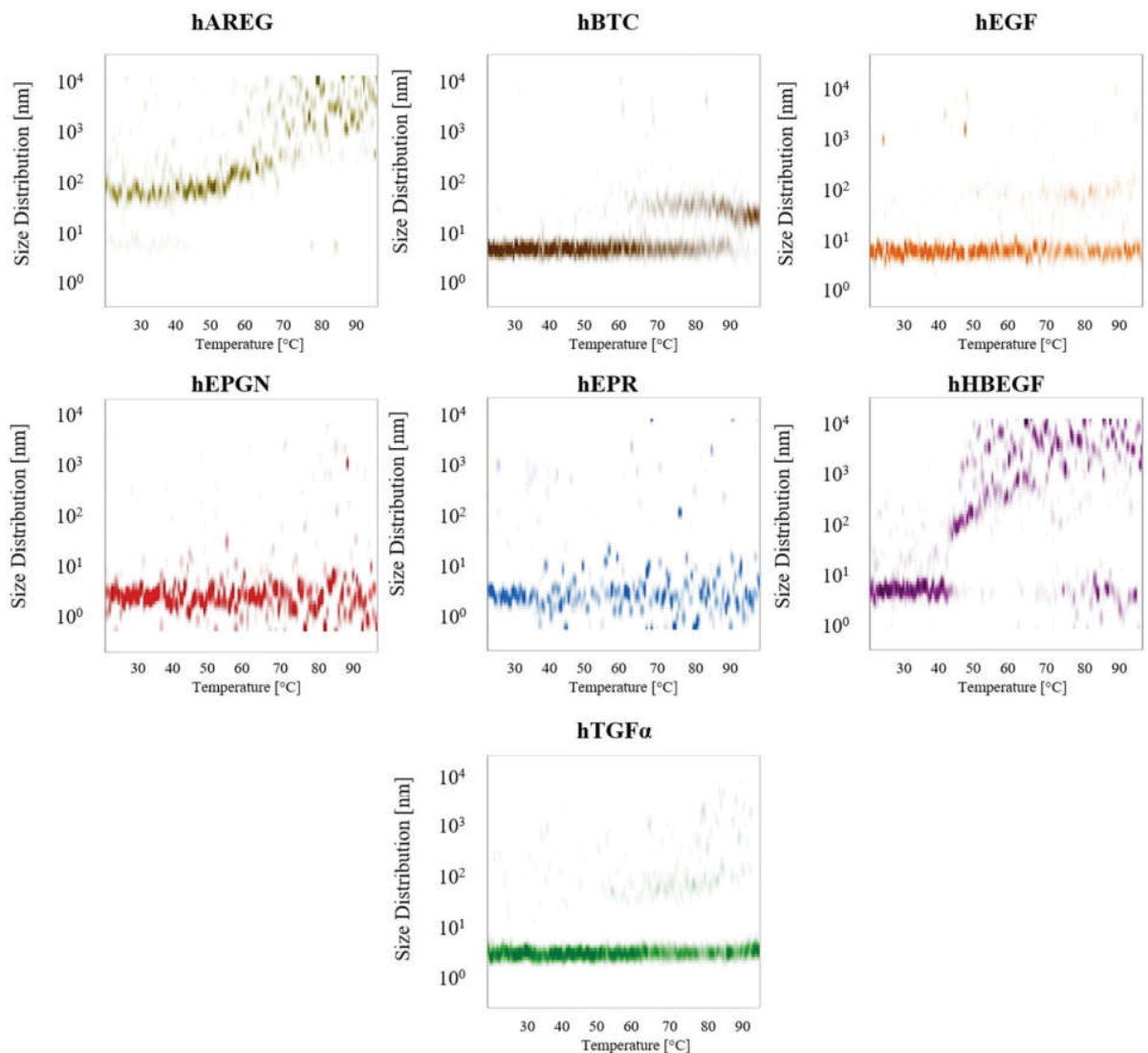


**Figure 3.** Radius size and aggregation analysis by DLS. Profiles of radius size distribution of the purified EGF growth factors under reducing and non-reducing conditions as determined by dynamic light scattering analyses. The dashed vertical lines represent the center of the cumulative radius of each sample. All assays were performed at 20 °C. Times and treatment are indicated in the figure. The graphics show the intensity distribution (%) as a function of particle radius (nm).

For hAREG and hBTC there is an increase in the particles size at ~ 55 to 60 °C while for hHBEGF starts showing aggregation at ~ 45 °C (Fig. 4). Complete data can be found in the Supplementary Information file.

The circular dichroism (CD) spectrum for each EGF growth factor under non-reducing conditions was acquired at 20 °C, followed by temperature increase (70–85 °C) and after returning the temperature down to 20 °C (Fig. 5). Similar to the analysis by nanoDSF, temperature increase monitored by CD did not reveal transition states for the EGF growth factors. Actually, the CD spectra of hBTC, hEPGN and hTGF $\alpha$  did not really change with temperature. hAREG, hEGF and hHBEGF showed a reduction of CD signal at 85 °C as compared to the signal at 20 °C. After reducing the temperature, the negative peak of the hEGF and hEPR spectra shifted slightly to a lower wavelength, indicating that they do not return completely to the same conformation observed at 20 °C. The fact that the temperature increase did not produce transition between folded and unfolded states is





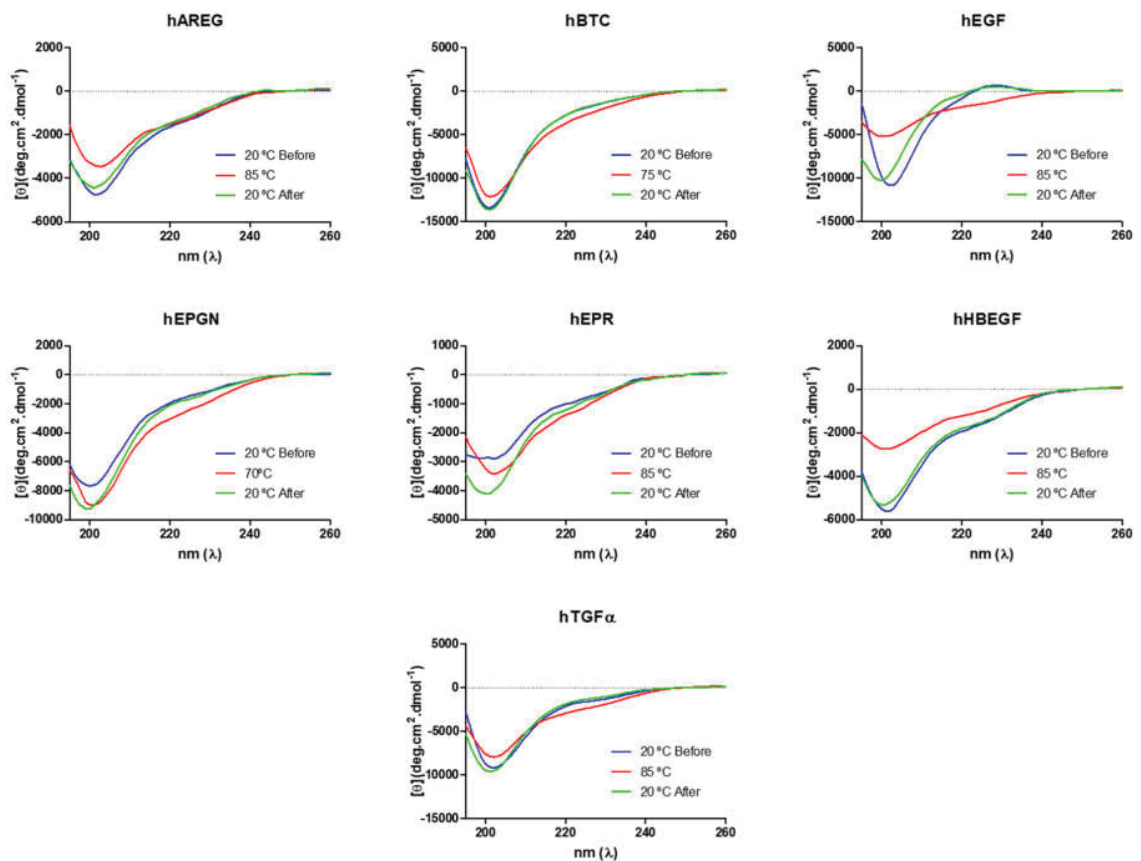
**Figure 4.** Particle size distributions (nm) for the seven growth factors as a relation of temperature increase from 20 to 95 °C as determined by DLS using non-reduced proteins.

consistent with the rigid structure of the EGF domains maintained by the three disulfide bonds. The resistance to thermal denaturation also indicates that the disulfide bonds should be formed in the proper combination.

**Functional analysis of the recombinant EGF growth factors.** The activity of the recombinant EGF growth factors was initially evaluated by cell proliferation assays. Their capacity to stimulate cell proliferation was tested *in vitro* using HeLa cells with a fixed concentration of 50 ng/mL (Fig. 6A). All factors led to strong induction of HeLa cells proliferation with at least a threefold increase relative to the control untreated cells. hEPR showed the highest induction capacity, followed by hBTC, hAREG, hEGF, hTGFα, hEPGN and, hHBEGF with the lowest stimulation.

A dose response curve was performed using serial dilutions of the recombinant growth factors, starting at 50 ng/mL, and the proliferation of HeLa cells determined indirectly by the MTT method (Fig. 6B,C). The dose–response curve showed that all EGF growth factors induce cell proliferation at the concentrations of 50 and 5 ng/mL. hEPGN, hEPR and hTGFα showed statistically difference from control also at 0.5 ng/mL whereas hEPGN and hTGFα induced cell proliferation at concentration as low as 0.05 ng/mL. These results show that the recombinant EGF growth factors are highly potent to induce cell proliferation.

The activity of the EGF growth factors was further evaluated by scratch wound assays using the fibroblast NDFH cell line. This assay indicates the capacity of proliferation and migration of fibroblasts. Artificial wounds were created by scratching confluent layers of NDFH cells, following by photographing the wound gap of the live cultures and addition of 50 ng of each recombinant EGF growth factor. The cultures were incubated after 48 h, stained with crystal violet and photographed again. The percentage of scratch closure was determined by the ratio between the area of the gap at the beginning of the experiment and the area of the gap after 48 h (Fig. 7).



**Figure 5.** Circular dichroism spectra of the seven EGF growth factors. The blue lines correspond to the spectra acquired at 20 °C, the red lines to the spectra acquired after heating to the indicated temperatures and the green lines correspond to the spectra acquired after the temperature was reduced to 20 °C. The far-UV circular dichroism spectra (195–260 nm) were recorded using a J-815 spectropolarimeter with five accumulations and using a 1 mm path length cell.

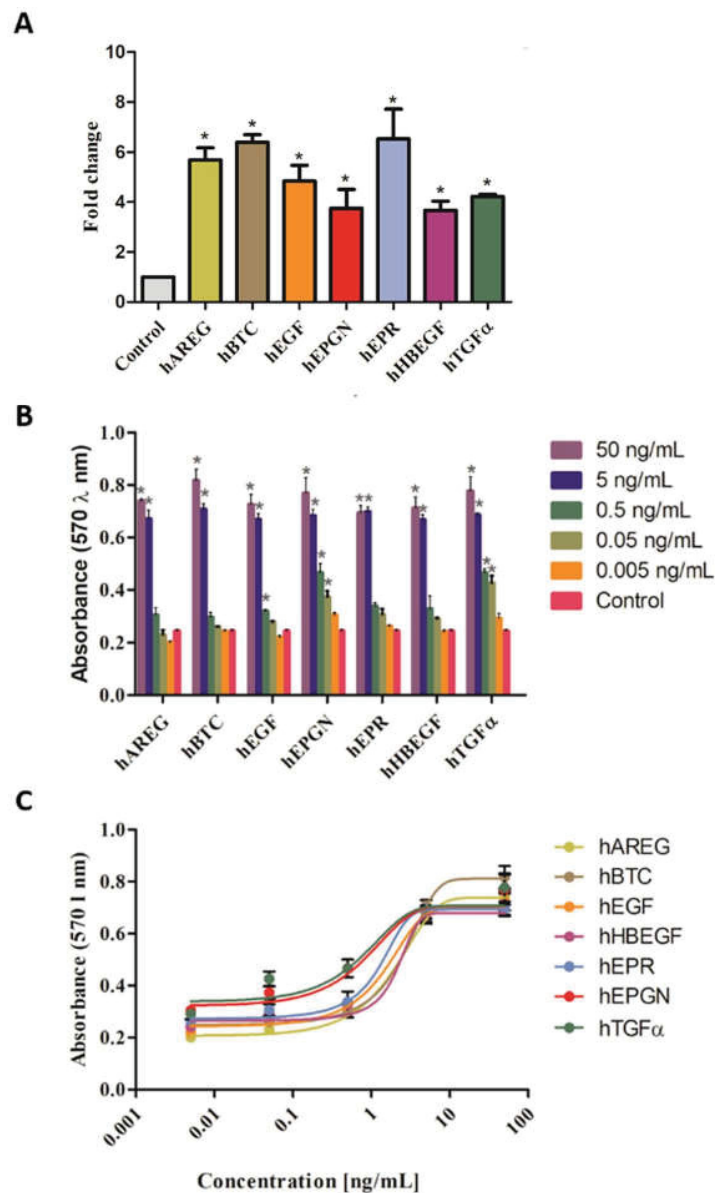
The rate of artificial wound closure represents the combined induction of proliferation and migration. All EGF domains strongly induced wound closure when compared with the control. hEGF showed the highest rate of closure, followed by hEPR, hEPGN, hAREG, hBTC, hTGF $\alpha$  and, hHBEGF showing the lowest closure percentage. In summary, the activity analyses show that the recombinant EGF growth factors are highly active, being able to induce cell proliferation and migration.

## Discussion

The members of the EGF family of growth factors share the same structural domain, EGF, and are able to interact with cellular receptors and activate pathways leading to proliferation of different cells and tissues. The presence of three disulfide bonds in the EGF domain, with alternate combination for the first two pairs, renders a protein structure that is difficult to reconstitute in a functional conformation in prokaryotic organisms. In this work, we have developed genetic constructs and described detailed procedures to obtain seven human growth factors of the EGF family in a simple and easy fashion, using a single protocol and the same experimental set up and reagents. The genetic constructs containing the growth factors fused to Trx in pET32a proved to be very efficient for soluble expression in the *E. coli* Shuffle strain. The subsequent steps of affinity purification, thrombin digestion and recovery of the cleaved growth factors have worked very efficiently for all EGF growth factors tested. Our data show that utilization of Trx as fusion partner is very practical not requiring time-consuming methods for refolding in vitro of *E. coli* expressed insoluble proteins neither the use of eukaryotic cells as expression systems.

Concerning the N-terminal extension of the linker encoded by the vector pET32a that remains attached to the growth factor after digestion with thrombin, we have examined the binding mode of hEGF, hEPGN, hEPR and hTGF $\alpha$  to the receptor using the structures of their respective complexes with the extracellular domain of the EGF receptor deposited at PDB<sup>39–41</sup>. In all cases, the N- and C-terminal ends of the ligands are located outside the binding cavity, far from the binding interface, in positions which N- and C-terminal extensions are unlikely to interfere with binding (Fig. 8).

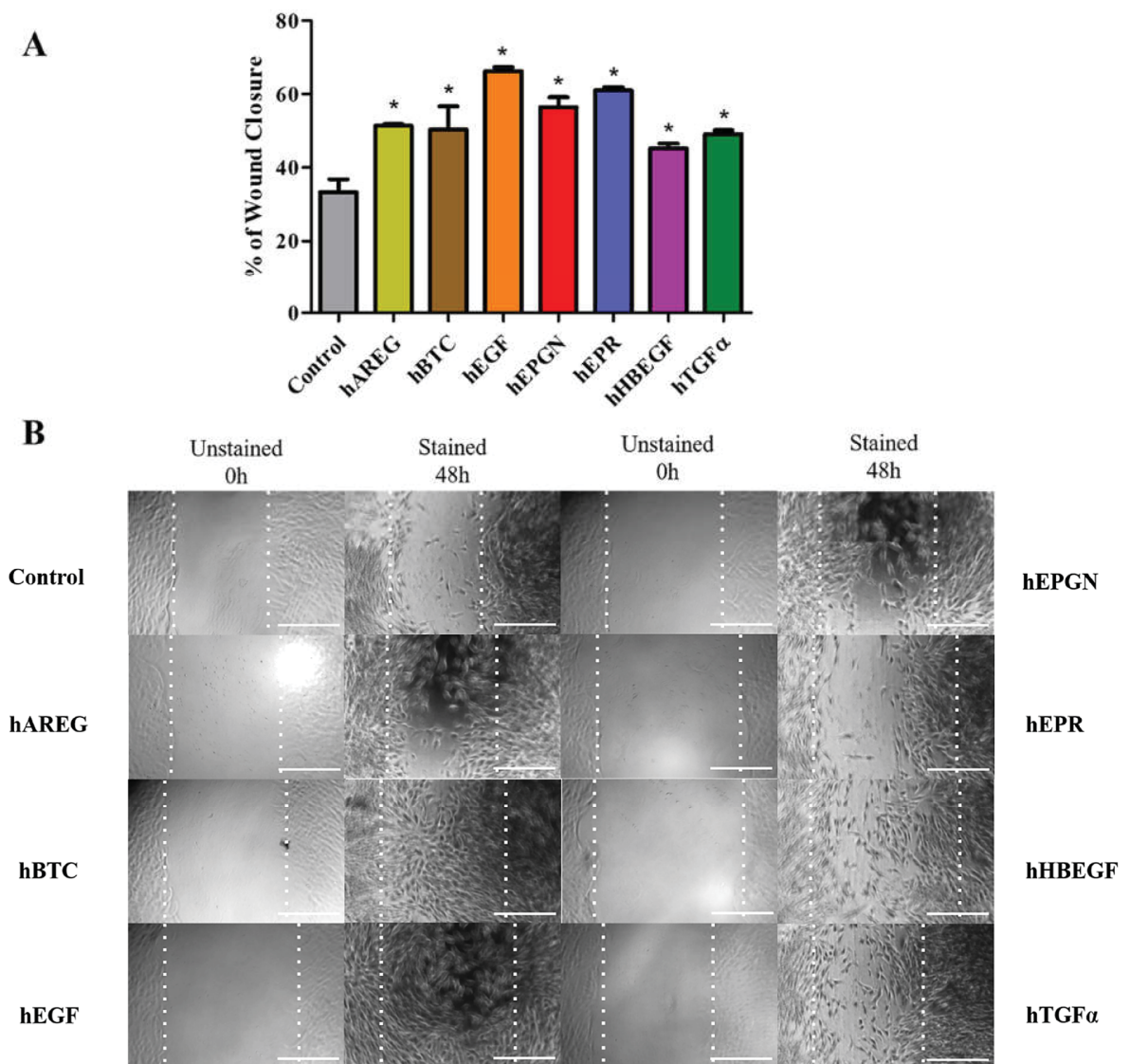
A major challenge for *E. coli* expression of small disulfide-bond-rich proteins is the correct pairing of the cysteine residues for disulfide bond formation so that the protein can assume a native folding. Previous studies have shown that Trx and SUMO were successfully used as fusion partners for *E. coli* expression of hEPR<sup>22,26</sup> and



**Figure 6.** Activity of the seven EGF growth factors on cell proliferation. **(A)** Induction of HeLa cell proliferation by the seven EGF growth factors. Triplicates of HeLa cell cultures were treated with 50 ng/mL of each growth factor and the fold change of the proliferation rate relative to untreated cells (control) was determined after 48 h. The cells were quantified by flow cytometry and normalized using the CountBright™ Absolute Counting Beads method. \* $p < 0.05$ . **(B)** HeLa cell proliferation induced by the seven EGF growth factors as determined by the MTT assay. Triplicate cultures of HeLa cells were treated with the indicated amounts of the respective EGF growth factors for 48 h and quantified using the MTT assay. \* $p < 0.05$ . **(C)** Cell cultures absorbance as a function of the growth factor concentration. The mean values were derived from the assay described in (B) and curves were fitted with nonlinear regression.

hHBEGF<sup>25</sup>, respectively. However, it is important to point out that for expression of hEPR, an extension towards the N-terminus of pro-EPR needs to be included in the fusion protein for stability. The protein amenable to crystallization includes 25 residues of the propeptide<sup>26</sup>. Expression of the EGF domain of hEPR without N-terminal extension resulted in unstable protein (data not shown). Therefore, in order to improve the success rate aiming both soluble expression and stable protein after cleavage of Trx, the N- and C-terminal limits for cloning the EGF growth factors was based on the 3D structures deposited at PDB (Table 1).

Conformational and stability analyses were performed by indirect and direct methods to evaluate the folding state of the recombinant growth factors. Limited proteolysis and dynamic light scattering, which were performed in parallel with non-reduced and reduced samples, revealed striking differences between the two states. In the non-reduced state, all EGF growth factors showed much higher resistance to proteolysis, as well single peaks for

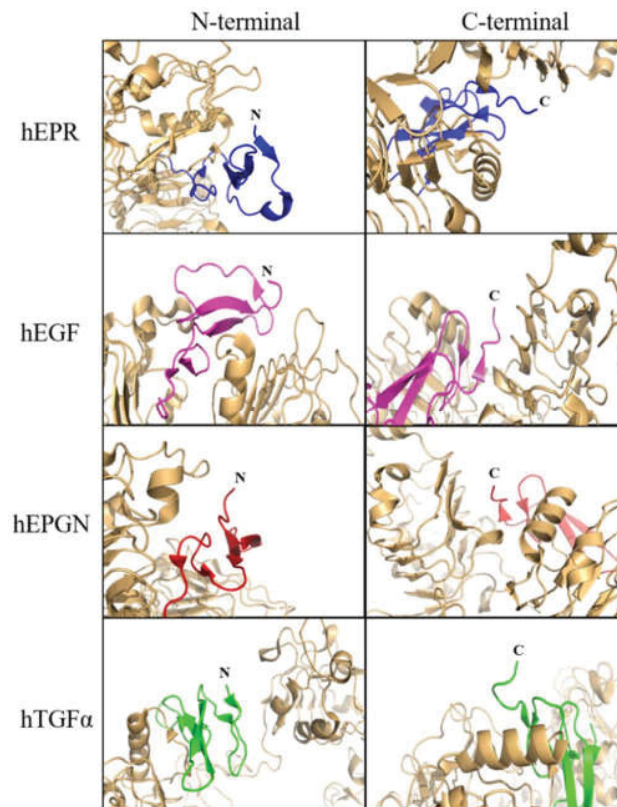


**Figure 7.** Scratch wound closure induced by the seven EGF growth factors. **(A)** Confluent cultures of NDFH cells were scratched with a sterile pipette tip and treated with 50 ng/mL of each EGF growth factor for 48 h. For quantification of the gap closure rates, the area of the gap at 48 h was compared to the area of the initial gap. Gap areas were determined with the ImageJ software using microscopic images as described in the Methods section. \* $p < 0.05$ . **(B)** Representative microscopy images at 100 X magnification acquired at time 0 (unstained living cells after the scratch) and after 48 h of fixed cells stained with violet crystal. The bar size is 100  $\mu\text{m}$ .

radius size distribution at 20 °C, which are indications that the disulfide bonds are correctly formed. Interestingly, hAREG showed a divergent behavior under reducing conditions. DLS analyses indicated that hAREG does not aggregate after disulfide bond reduction even after long incubation times while all others showed aggregation under reducing conditions, although hBEGF and hTGF $\alpha$  required longer times. Assays with increasing temperatures monitored both by nanoDSF and CD, did not detect clear transition states, which is consistent with the stable nature of the EGF domain sustained by the three disulfide bonds. The CD spectra of hEGF and hTGF $\alpha$  were similar to the CD spectra described in the literature<sup>42</sup> for these factors and are consistent with spectra of proteins with none or low alpha helical content. hEGF, hEPGN, hEPR and hTGF $\alpha$  also showed strong resistance to aggregation under high temperatures while hBTC, hAREG and hBEGF showed intermediate resistance to aggregation under heating. The conformational and stability analyses were complemented by mass spectrometry, which revealed the three correct disulfide bonds for AREG and correct bonds for the C1-C2 cysteine pairs of BTC, and C5-C6 for EPGN and EPR (Supplementary Figures S4–S7). The mass spectrometry analysis seems to be limited mostly by the low efficiency of the proteolytic digestion of the EGF domain, which does not generate di-peptides carrying disulfide bonds that are suitable for MS/MS.

The biological activity of the recombinant EGF growth factors was demonstrated by cell proliferation and migration assays. All of them induced cell proliferation at concentrations as low as 5 ng/mL. The cell proliferation





**Figure 8.** Analysis of the position of the N- and C-terminal ends of EGF growth factors bound to EGF receptors according to the crystal structures of the complexes available at PDB. The EGF receptors are shown in light orange. EGFR and EPR (blue) from PDB: 5WB7. EGFR and EGF (pink) from PDB: 1NQL.EPGN (red) from PDB: 5WB8. EGFR and TGF $\alpha$  (green) from PDB: 1MOX. The N- and C-terminals of the ligands are indicated. The images shown that the N- and C-terminals of the ligands are exposed to the solvent and not involved in the interaction with the receptor.

Growth factor	Uniprot code	Amino acid range	PDB code
hAREG	P15514	101–198 (188–198 part of the propeptide)	2RNL
hBTC	P35070	32–111	1IP0
hEGF	P01133-1	971–1023	1JL9
hEPGN	Q6UW88-1	49–109	5WB8
hEPR	O14944	38–108 (38–62 part of the propeptide)	5E8D
hHB-EGF	Q99075	63–148	1XDT
hTGF $\alpha$	P01135-1	39–89	2TGF

**Table 1.** Sequences of the growth factors used in this study.

effect obtained in our work for hEGF, hAREG, hBTC, hHBEGF and hTGF $\alpha$  is similar to the results published in previous studies for HeLa cells<sup>43,44</sup>, and in the case of hEGF also for other cell lines<sup>45</sup>. Considering the concentrations between 5 and 50 ng mL required for induction of proliferation, our findings for hEGF are similar to the ones described for MIO-M1 cell line<sup>46</sup> and, for hEPGN and hTGF $\alpha$  described for myeloma cells stably expressing ErbB-1<sup>45</sup>. In addition to proliferation, the recombinant EGF growth factors can induce cell migration, as shown by the artificial scratch wound healing assays using NDFH cells. This finding is consistent with previous reports showing that hAREG, hHBEGF and hEGF are capable of inducing closure of scratches on other types of cell lines<sup>46–49</sup>. In general, the activity assays show that the recombinant EGF growth factors are highly active. This, together with the detailed information provided about the genetic constructs, purification procedures and biophysical analyses for quality control of folding state of the recombinant growth factors represent handy tools for quick and reliable in-house production of the seven members of the EGF growth factor family.

## Methods

**Plasmid constructs.** The EGF domain of hAREG, hBTC, hEGF, hEPGN, hHBEGF and hTGF $\alpha$  were identified using Uniprot annotations and PDB structures, as shown in Table 1. Synthetic genes for these six factors were obtained from *GenScript* (Piscataway, NJ, USA) optimized for expression in *Escherichia coli* and cloned into the *NcoI* and *XhoI* restriction sites of plasmid pET32a<sup>50</sup>. The hEPR synthetic gene was designed as previously described by Kado and co-workers<sup>26</sup> and acquired from GenCust (Boynes, France).

**Protein expression.** All pET32a-derived plasmids were transformed into the *E. coli Shuffle* strain<sup>30</sup> using heat shock<sup>51</sup>. Transformed cells were grown on LB-agar<sup>52</sup> plates containing ampicillin (100  $\mu$ g/mL) and incubated overnight at 37 °C. A single colony from each clone was used to inoculate 10 mL of LB containing ampicillin (100  $\mu$ g/mL) and incubated overnight at 37 °C. 5 mL of this culture were used to inoculate 500 mL of LB containing ampicillin (100  $\mu$ g/mL). When the O.D. at 600<sub>nm</sub> reached 0.6, the incubation temperature was reduced to 16 °C and expression induced by adding 0.5 mM IPTG. The incubation was continued for further 24 h and the cells harvested by centrifugation (6000  $\times$  g for 10 min) and stored at -80 °C.

**Protein purification.** For purification, the cell pellet of 500-mL cultures was suspended in 10 mL of Buffer A (50 mM Tris.HCl pH 8.0; 300 mM NaCl; 20 mM imidazole) and the cells lysed using an M-110L microfluidizer (Microfluidics, Newton, MA, USA). After ten cycles of 18,000 psi, the extracts were clarified by centrifugation at 20,000  $\times$  g for 30 min. The supernatant was transferred to a 50 mL super-loop and assembled onto an ÄKTA Pure M25 chromatography system (Cytiva, Marlborough, USA). The recombinant proteins were initially purified by affinity chromatography using a 5 mL HisTrap column (Cytiva 17524802) pre-equilibrated in buffer A. The flow rate for all steps was 5 mL/min. After loading the sample, the column was washed with 3 column volumes (CV) of buffer A, followed by a gradient from 0 to 10% of Buffer B (50 mM Tris.HCl pH 8.0; 300 mM NaCl; 500 mM imidazole) in 2 CV and, the protein eluted using a 10 to 100% Buffer B gradient in 4 CV. Fractions of 1 mL were collected and the peak fractions analyzed by SDS-PAGE.

For digestion with thrombin, a 5 mL volume of peak samples was diluted to a final volume of 15 mL using dilution buffer (25 mM Tris.HCl pH 8.0; 150 mM NaCl). The digestions were performed using 1 U of thrombin (Cytiva 27084601) for 10 mg of recombinant fusion protein (Trx-EGF factor) at 25 °C for 24 h. Once the digestion was confirmed by SDS-PAGE, the products of digestion were loaded onto a 50 mL super-loop and the his-tagged Trx was captured on a 5 mL HisTrap column (Cytiva 17524802) pre-equilibrated in buffer A using an ÄKTA Pure M25 system (Cytiva, Marlborough, MA, USA). After sample injection, the column was washed using 3 CV of Buffer A. The flowthrough containing the EGF growth factor was collected and concentrated using Amicon Ultra-0.5 mL 3 K concentrators up to a final volume of 500  $\mu$ L. Subsequently, the concentrated 500- $\mu$ L samples were fractionated by size exclusion chromatography. For this purpose, a Superdex 75 10/300 GL (Cytiva 29148721) was equilibrated with size exclusion buffer (25 mM Tris.HCl pH 8.0; 150 mM NaCl) using an ÄKTA Pure M25 system (Cytiva, Marlborough, MA, USA) at flow rate of 0.8 mL/min. Fractions of 0.5 mL were collected and samples from the chromatographic peak fractions were analyzed by SDS-PAGE as described below.

**SDS-PAGE analysis of proteins.** Cells extracts, chromatography fractions, purified proteins and proteolysis reactions were analyzed by electrophoresis using the Laemmli buffer system<sup>53</sup>. Poly-acrylamide concentrations in the stacking and separating gels were 6.5% and 15%, respectively. 0.75 mm-thick gels were assembled on Mini-PROTEAN Tetra Cells (Bio-Rad Laboratories, Hercules, USA). Protein samples were mixed with Laemmli sample buffer, heated to 95 °C for 5 min and applied onto the gel wells. Electrophoresis was performed at 100 V. Gels were stained using Coomassie Brilliant Blue R-250 and destained using a methanol/acetic acid/water (30/10/60) solution<sup>54</sup>.

**Limited proteolysis assay.** Limited proteolysis was performed with the recombinant growth factors under reducing and non-reducing conditions. For reduction of the disulfide bonds, the growth factors were treated with 10 mM DTT (Merck 3483123). The samples were incubated with 0.1% (w/w) trypsin (Sigma-Aldrich 59427C) for 30 and 60 min at 37 °C. Controls without trypsin were incubated for 60 min. The products of digestions and control reactions were analyzed by SDS-PAGE as described above.

**Dynamic Light Scattering and effect of temperature monitored by nano-differential scanning fluorimetry (nanoDSF).** The purified recombinant growth factors were analyzed on a nanoDSF Prometheus Panta device (NanoTemper Technologies, Munich, Germany) for DLS and thermal effect analyses under reducing and non-reducing conditions using standard capillaries. DLS measurements were performed at 20 °C using the high sensitivity setup (10 acquisitions of 5 s each) and two independent capillaries. For analysis of the thermal effect, the samples without reducing agents were subjected to a linear thermal ramp from 20 to 95 °C with parallel measurement of DLS.

**Circular dichroism.** For circular dichroism analyses, the sample buffer was changed by ultrafiltration to 4.2 mM Tris-HCl pH 8.0 and 25 mM NaCl using Amicon Ultra-0.5 mL 3 K concentrators. CD spectra were acquired using recombinant growth factors at 0.2 mg/mL in 1 mm path cuvette on a Jasco J-815 circular dichroism spectrometer (JASCO, Tokyo, Japan). The scanning speed was 50 nm/min. Each spectrum was accumulated at least five times. Spectra were acquired from 195 to 260 nm at 20 °C and after raising the temperature from 20 to 70 °C, 75 °C or 85 °C, depending on the voltage limit. A new set of spectra was acquired after cooling to 20 °C.

The raw data were converted using the mean residue molar ellipticity formula and autozero was set by using the value at 250 nm.

**Cell proliferation assay.** HeLa cells (Carlos Chagas Institute cell bank) were cultivated in DMEM (ThermoFisher 41965062) supplemented with 10% BFS until confluence using a Forma Series II Water Jacketed CO<sub>2</sub> incubator (Thermo Fischer Scientific, Waltham, MA, USA). Cells were cultivated at 37 °C, with humidity and 5% CO<sub>2</sub>. Subsequently, 10<sup>5</sup> cells per well were transferred to a 24 well plate and incubated for additional 24 h, when the medium was replaced by DMEM without BFS, and the cultures were treated with 50 ng/mL of recombinant EGF growth factors in triplicates in parallel with untreated controls also in triplicates. After 48 h, the wells were washed with PBS and treated with 200 µL of PBS containing trypsin 1% (w/v) (Merck T8003). After 5 min of trypsin action at 37 °C, the released cells were collected by centrifugation at 500 × g for 5 min and suspended in 280 µL of PBS and 20 µL of *CountBright™ Absolute Counting Beads* (ThermoFisher C36950). The cell number was determined on a BD FACS Canto II (BD Life Sciences, New Jersey, USA) and the results analyzed by FlowJo™ v10.8 (BD Life Sciences, Franklin Lakes, NJ, USA). At least 1000 beads were acquired and a maximum of 3000 beads was set as upper limit. Quantification was performed as described by the *CountBright™ Absolute Counting Beads* manual (Invitrogen MAN0018850).

For dose–response assays, 5 × 10<sup>4</sup> HeLa cells (Carlos Chagas Institute cell bank) per well from a confluent culture were cultivated in 96-well plates for 24 h using a Forma Series II Water Jacketed CO<sub>2</sub> incubator (Thermo Fischer Scientific, Waltham, MA, USA). The cells were cultivated at 37 °C, with humidity and 5% CO<sub>2</sub>. Subsequently, they were treated with five increasing concentrations of recombinant growth factors in triplicates: 0.005 ng/mL, 0.05 ng/mL, 0.5 ng/mL, 5 ng/mL and 50 ng/mL. After 48 h of incubation, the wells were washed with PBS, treated with 50 µL of MTT (Invitrogen M6494) at 5 mg/mL and 50 µL of serum free medium and incubated at 37 °C for 3 h. After this time, 150 µL of the solvent solution (4 mM HCl, 0.1% NP40 in isopropanol) was added and the incubation continued for another hour. Absorbance was read at 570 nm using a Synergy H1 Hybrid Reader (BioTek Instruments, Winooski, VT, USA). Dose–response curves were adjusted using nonlinear fit of the data.

**Scratch closure assay.** NDFH cells (Lonza CC-2509) were cultivated in DMEM (ThermoFisher 41965062) supplemented with 10% BFS until confluence using a Forma Series II Water Jacketed CO<sub>2</sub> incubator (Thermo Fischer Scientific, Waltham, MA, USA). Cells were cultivated at 37 °C, with humidity and 5% CO<sub>2</sub>. Subsequently, the medium was changed to DMEM without BFS, and the incubation continued for another 24 h. At this time, a scratch was made, 50 ng of each recombinant factor was added to cultures in triplicates and images of the unstained cells were acquired. After 48 h of incubation, the cells were stained using crystal violet (Synth C.I. 42555) according to the supplier's instructions. Briefly, the cells were fixed using ice-cold 100% methanol for 10 min. Afterwards, 0.1 mL of a 0.5% crystal violet solution in 25% methanol was added to each well, incubated for 10 min and washed with PBS. Images of the cells were acquired at time 0 and 48 h after the treatment with the growth factors on a Nikon Eclipse TE300 Inverted Microscope (Nikon, Tokyo, Japan) using a 10X objective and the QCapture Pro 6.0 software (Teledyne QImaging, Surrey, Canada). The rate of artificial wound healing was calculated using ImageJ<sup>55</sup>. The scratches were aligned with the base of the picture and a 900 × 600-pixel rectangle was cropped around the scratch. The area at time 0 was calculated using the polygon selection tool and the area at 48 h was calculated using the area between the edges of the scratch area converted to 8-bit images and sharpened. The find edges tool was used and the image was converted to binary. The particles were then analyzed using sizes between 10 and 100,000 with bare outlines, including holes to calculate the areas.

**Statistical analysis.** Graphs and statistical analyses were performed using the GraphPad Prism 5 software (Graph Pad, San Diego, USA). The data from assays in triplicate for each growth factor were compared to the control data using One-Way analysis of variance and Dunnett posttest, where p < 0.05 was considered statistically significant.

**Analysis of protein 3D structures.** The structures of the EGF growth factors and of the complexes of EGF growth factors and EGF receptors were visualized using PyMOL Molecular Graphics System version 4.5.0, Schrödinger, LLC.

### Data availability

Original datasets and reagents generated during this study are available from the corresponding author on reasonable request.

Received: 7 December 2021; Accepted: 14 March 2022

Published online: 23 March 2022

### References

1. Yang, C. C. & Chang, K. W. Eicosanoids and HB-EGF/EGFR in cancer. *Cancer Metastasis Rev.* **37**, 385–395 (2018).
2. Zhao, J. *et al.* Betacellulin enhances ovarian cancer cell migration by up-regulating Connexin43 via MEK-ERK signaling. *Cell. Signal.* **65**, 109439 (2020).
3. Carvalho, S. *et al.* An antibody to amphiregulin, an abundant growth factor in patients' fluids, inhibits ovarian tumors. *Oncogene* **35**, 438–447 (2016).
4. Verma, V. *et al.* The dichotomous role of epiregulin in pain. *Pain* **161**, 1052–1064 (2020).
5. Liu, Y. Epidermal growth factor on the healing of human subacute tympanic membrane perforation. *Am. J. Otolaryngol.* **41**, 102399 (2020).
6. Roskoski, R. The ErbB/HER family of protein-tyrosine kinases and cancer. *Pharmacol. Res.* **79**, 34–74 (2014).

7. Ho, J., Moyes, D. L., Tavassoli, M. & Naglik, J. R. The role of ErbB receptors in infection. *Trends Microbiol.* **25**, 942–952 (2017).
8. Singh, B., Carpenter, G. & Coffey, R. J. EGF receptor ligands: Recent advances. *Fl000Research* **5**, 2270 (2016).
9. Schneider, M. R. & Wolf, E. The epidermal growth factor receptor ligands at a glance. *J. Cell. Physiol.* **218**, 460–466 (2009).
10. Barbacci, E. G. *et al.* The structural basis for the specificity of epidermal growth factor and heregulin binding. *J. Biol. Chem.* **270**, 9585–9589 (1995).
11. Beerli, R. R. & Hynes, N. E. Epidermal growth factor-related peptides activate distinct subsets of ErbB receptors and differ in their biological activities. *J. Biol. Chem.* **271**, 6071–6076 (1996).
12. Dreux, A. C., Lamb, D. J., Modjtahedi, H. & Ferns, G. A. A. The epidermal growth factor receptors and their family of ligands: Their putative role in atherogenesis. *Atherosclerosis* **186**, 38–53 (2006).
13. Yarden, Y. & Slivkowsky, M. X. Untangling the ErbB signalling network. *Nat. Rev. Mol. Cell Biol.* **2**, 127–137 (2001).
14. Harris, R. C., Chung, E. & Coffey, R. J. EGF receptor ligands. *Exp. Cell Res.* **284**, 2–13 (2003).
15. Zhao, B. *et al.* A treatment combination of igf and egf promotes hair growth in the angora rabbit. *Genes (Basel)*. **12**, 1–9 (2021).
16. Silva, C. O. *et al.* EGF functionalized polymer-coated gold nanoparticles promote EGF photostability and EGFR internalization for photothermal therapy. *PLoS ONE* **11**, 1–29 (2016).
17. Sundqvist, A. *et al.* TGF $\beta$  and EGF signaling orchestrates the AP-1- and p63 transcriptional regulation of breast cancer invasiveness. *Oncogene* **39**, 4436–4449 (2020).
18. Eissazadeh, S. *et al.* Production of recombinant human epidermal growth factor in *Pichia pastoris*. *Braz. J. Microbiol.* **48**, 286–293 (2017).
19. Davis, K. M. *et al.* Production of glycosylated heparin-binding EGF-like growth factor in HeLa cells using vaccinia virus. *Protein Exp. Purif.* **8**, 57–67 (1996).
20. Shin, S., Li, N., Kobayashi, N., Yoon, J. W. & Jun, H. S. Remission of diabetes by  $\beta$ -cell regeneration in diabetic mice treated with a recombinant adenovirus expressing betacellulin. *Mol. Ther.* **16**, 854–861 (2008).
21. Khan, M. A. *et al.* Expression line approach to recombinant human epidermal growth factor into the yeast, *Pichia pastoris* from Huh-7 cell line. *Mol. Biol. Rep.* **41**, 1445–1451 (2014).
22. Sato, K. *et al.* Solution structure of epiregulin and the effect of its C-terminal domain for receptor binding affinity. *FEBS Lett.* **553**, 232–238 (2003).
23. Ferrer Soler, L., Cedano, J., Querol, E. & De Llorens, R. Cloning, expression and purification of human epidermal growth factor using different expression systems. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **788**, 113–123 (2003).
24. Khalili, M., Soleyman, M. R., Baazm, M. & Beyer, C. High-level expression and purification of soluble bioactive recombinant human heparin-binding epidermal growth factor in *Escherichia coli*. *Cell Biol. Int.* **39**, 858–864 (2015).
25. Lu, W., Cao, P., Lei, H. & Zhang, S. High-level expression and purification of heparin-binding epidermal growth factor (HB-EGF) with SUMO fusion. *Mol. Biotechnol.* **44**, 198–203 (2010).
26. Kado, Y. *et al.* Epiregulin recognition mechanisms by anti-epiregulin antibody 9E5: Structural, functional, and molecular dynamics simulation analyses. *J. Biol. Chem.* **291**, 2319–2330 (2016).
27. Thompson, S. A., Harris, A., Hoang, D., Ferrer, M. & Johnson, G. R. COOH-terminal extended recombinant amphiregulin with bioactivity comparable with naturally derived growth factor. *J. Biol. Chem.* **271**, 17927–17931 (1996).
28. Miura, K. *et al.* Solution structure of betacellulin, a new member of EGF-family ligands. *Biochem. Biophys. Res. Commun.* **294**, 1040–1046 (2002).
29. Ma, Y. *et al.* High efficient expression, purification, and functional characterization of native human epidermal growth factor in *Escherichia coli*. *Biomed Res. Int.* **2016** (2016).
30. Lobstein, J. *et al.* SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Fact.* **11**, 1 (2012).
31. Kaplan, O. *et al.* Enhanced mitogenic activity of recombinant human vascular endothelial growth factor VEGF121 expressed in *E. coli* Origami B (DE3) with molecular chaperones. *PLoS ONE* **11**, 1–22 (2016).
32. Abdull Razis, A. F. *et al.* Expression of recombinant human epidermal growth factor in *Escherichia coli* and characterization of its biological activity. *Appl. Biochem. Biotechnol.* **144**, 249–261 (2008).
33. Qiang, W. *et al.* Molecular pharming of the recombinant protein hegf-hegf concatenated with oleosin using transgenic arabidopsis. *Genes (Basel)*. **11**, 1–13 (2020).
34. Zheng, X., Wu, X., Fu, X., Dai, D. & Wang, F. Expression and purification of human epidermal growth factor (hEGF) fused with GB1. *Biotechnol. Biotechnol. Equip.* **30**, 813–818 (2016).
35. Nguyen, M. T. *et al.* Prokaryotic soluble overexpression and purification of human VEGF165 by fusion to a maltose binding protein tag. *PLoS ONE* **11**, 1–15 (2016).
36. Zhao, Q., Xu, W., Xing, L. & Lin, Z. Recombinant production of medium- to large-sized peptides in *Escherichia coli* using a cleavable self-aggregating tag. *Microb. Cell Fact.* **15**, 1–9 (2016).
37. Kado, Y. *et al.* Epiregulin recognition mechanisms by 9E5(Fab) antibody. *J. Biol. Chem.* **5**, 1–30 (2015).
38. Macdonald-Obermann, J. L. & Pike, L. J. Different epidermal growth factor (EGF) receptor ligands show distinct kinetics and biased or partial agonism for homodimer and heterodimer formation. *J. Biol. Chem.* **289**, 26178–26188 (2014).
39. Garrett, T. P. J. *et al.* Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor  $\alpha$ . *Cell* **110**, 763–773 (2002).
40. Ferguson, K. M. *et al.* EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Mol. Cell* **11**, 507–517 (2003).
41. Freed, D. M. *et al.* EGFR ligands differentially stabilize receptor dimers to specify signaling kinetics. *Cell* **171**, 683–695.e18 (2017).
42. Prestrelski, S. J. *et al.* Solution structure and dynamics of epidermal growth factor and transforming growth factor  $\alpha$ . *J. Biol. Chem.* **267**, 319–322 (1992).
43. Hang, Q. *et al.* Integrin  $\alpha 5$  suppresses the phosphorylation of epidermal growth factor receptor and its cellular signaling of cell proliferation via N-glycosylation. *J. Biol. Chem.* **290**, 29345–29360 (2015).
44. Henriksen, L., Grandal, M. V., Knudsen, S. L. J., van Deurs, B. & Grøvdal, L. M. Internalization mechanisms of the epidermal growth factor receptor after activation with different ligands. *PLoS One* **8**, (2013).
45. Kochupurakkal, B. S. *et al.* Epigen, the last ligand of ErbB receptors, reveals intricate relationships between affinity and mitogenicity. *J. Biol. Chem.* **280**, 8503–8512 (2005).
46. Hu, Z. X. *et al.* PI3K-mediated glioprotective effect of epidermal growth factor under oxidative stress conditions. *Int. J. Ophthalmol.* **7**, 413–420 (2014).
47. You, D. H. & Nam, M. J. Effects of human epidermal growth factor gene-transfected mesenchymal stem cells on fibroblast migration and proliferation. *Cell Prolif.* **46**, 408–415 (2013).
48. Xu, K. P., Ding, Y., Ling, J., Dong, Z. & Yu, F. S. X. Wound-induced HB-EGF ectodomain shedding and EGFR activation in corneal epithelial cells. *Investig. Ophthalmol. Vis. Sci.* **45**, 813–820 (2004).
49. Wang, L. *et al.* AREG mediates the epithelial–mesenchymal transition in pancreatic cancer cells via the EGFR/ERK/NF- $\kappa$ B signaling pathway. *Oncol. Rep.* **43**, 1558–1568 (2020).
50. LaVallie, E. R. *et al.* A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Nat. Biotechnol.* **11**, 187–193 (1993).



51. Chang, A. Y., Chau, V. W., Landas, J. A. & Yvonne. Preparation of calcium competent *Escherichia coli* and heat-shock transformation. *J. Exp. Microbiol. Immunol.* **1**, 22–25 (2017).
52. Bertani, G. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* **62**, 293–300 (1951).
53. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
54. Sambrook, J. & Russel, D. W. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harboc Laboratory Press **3**, (2000).
55. Abràmoff, M. D., Magalhães, P. J. & Ram, S. J. Image processing with imageJ. *Biophotonics Int.* **11**, 36–41 (2004).

### Acknowledgements

This work was partially funded by the ICC-CNPq-PROEP Grant Number 442320/2019-1. ASF and AL were supported by CNPq (140923/2018-7) and CAPES graduate fellowships, respectively. NITZ and BGG are supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) research career fellowships (304167/2019-3 and 304788/2018-0). The authors acknowledge the FIOCRUZ program of technical platforms through the facilities for Protein Purification and Characterization (RPT-15A), Mass Spectrometry (RPT-02H) and Cytometry (RPT-08L).

### Author contributions

N.I.T.Z., B.G.G. and A.S.F. conceived the work and designed the experiments. A.S.F. and A.L. performed the protein expression and purification experiments. A.S.F. and B.G.G. performed the biophysical analyses. A.S.F. and P.M.H. performed the cell activation assays. M.B. performed the mass spectrometry analysis. A.S.F. and N.I.T.Z. wrote the main manuscript and A.S.F. prepared all the figures. All authors reviewed the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-09060-9>.

**Correspondence** and requests for materials should be addressed to N.I.T.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Supplementary Information to the paper

## **A Toolkit for Recombinant Production of all Seven Human EGF Family Growth Factors in Active Conformation**

Arthur Schweitzer Ferreira <sup>1,2</sup>; Amanda Lopacinski <sup>1,2</sup>, Michel Batista <sup>3</sup>, Priscila Mazzocchi Hiraiwa, Beatriz Gomes Guimarães <sup>1</sup>, Nilson Ivo Tonin Zanchin <sup>1,2, #</sup>

<sup>1</sup>Laboratory of Structural Biology and Protein Engineering, Carlos Chagas Institute, FIOCRUZ Paraná – Curitiba-PR, Brazil

<sup>2</sup>Cellular and Molecular Biology Graduate Program, Federal University of Paraná – Curitiba-PR, Brazil.

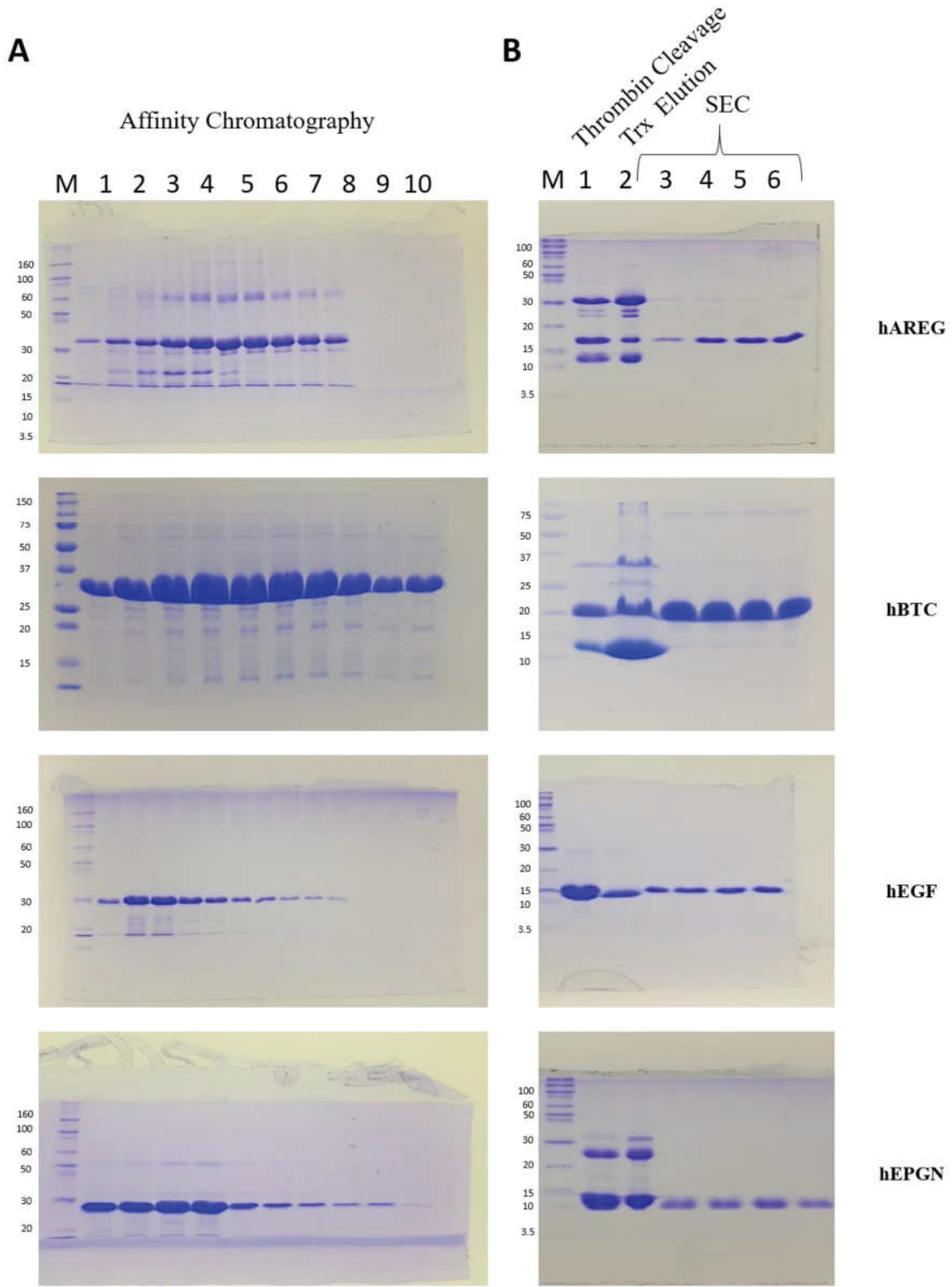
<sup>3</sup>Mass Spectrometry Facility RPT02H, Carlos Chagas Institute, FIOCRUZ Paraná – Curitiba-PR, Brazil.

# Corresponding author. Email address: [nilson.zanchin@fiocruz.br](mailto:nilson.zanchin@fiocruz.br)

### **Supplementary information content:**

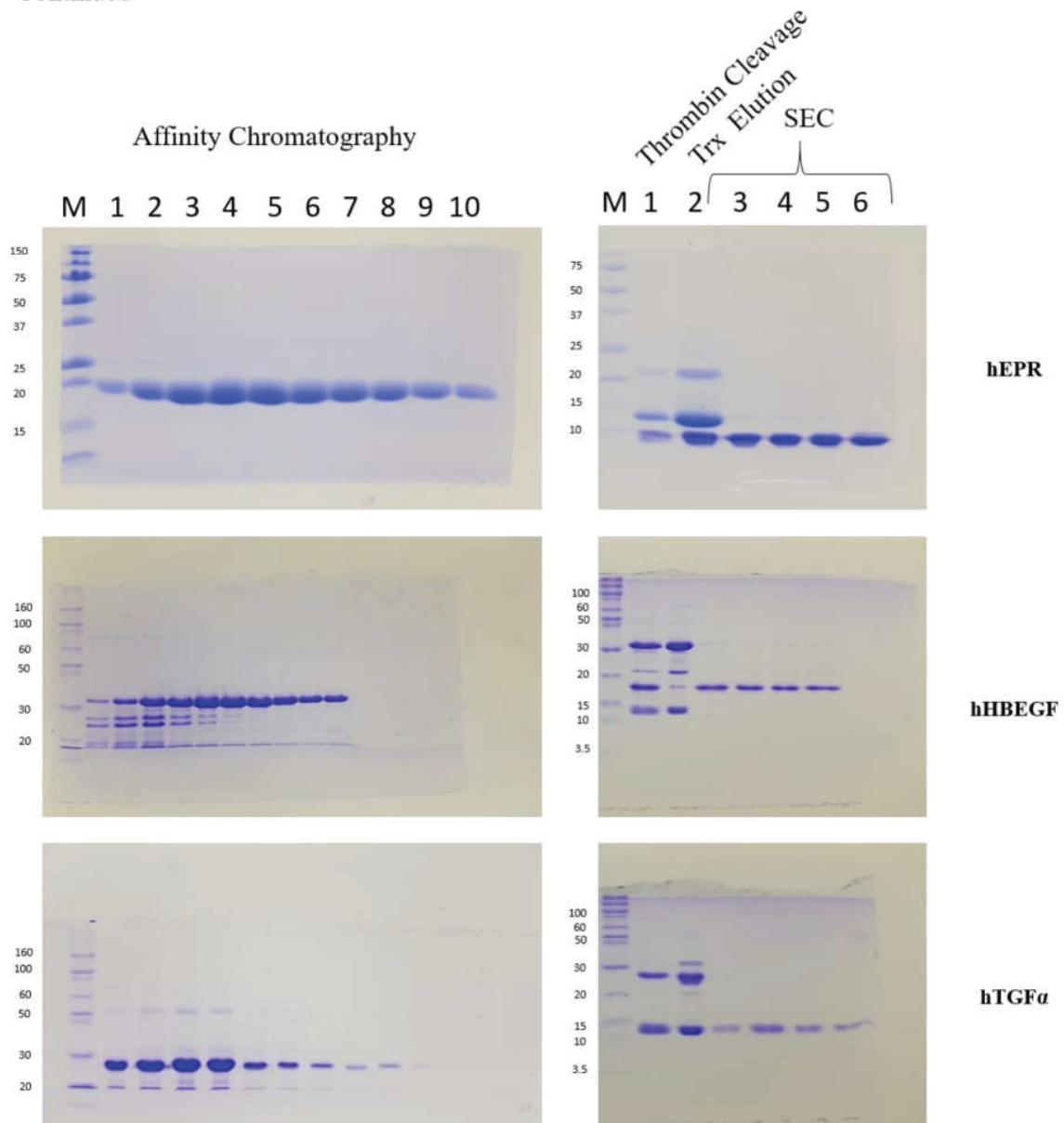
1. Complete images of the polyacrylamide gel electrophoresis analyses performed during the purification of the EGF growth factors.
2. Complete images of the polyacrylamide gel electrophoresis analyses of the limited proteolysis assays.
3. NanoDSF plots showing the effect of temperature increase on the seven EGF factors.
4. Mass spectrometry analysis of disulfide bonds

1. Complete images of the polyacrylamide gel electrophoresis analyses performed during the purification of the EGF growth factors.



Continued on the next page.

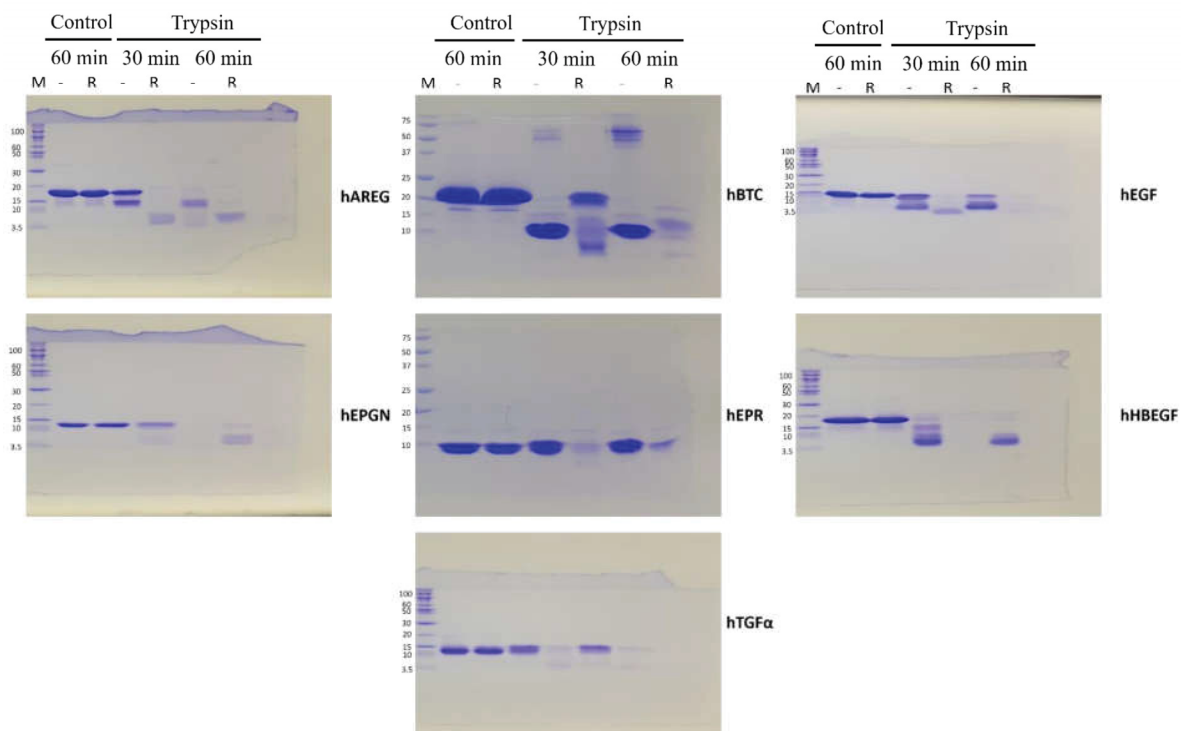
Continued



**Figure S1.** Complete images of the analyses performed during the purification of the EGF growth factors **(A)** Original images of the full-length Coomassie stained SDS-polyacrylamide gels showing the peak fractions of the initial affinity chromatography. **(B)** Original Images of the full-length Coomassie stained SDS-polyacrylamide gels showing the result of thrombin protease digestion (thrombin), the separation of Trx-His6 by capture on an IMAC column (Trx-His6) and the size exclusion chromatography. 1, products of the thrombin protease digestion; 2, elution fraction from the purification of Trx-his6 after thrombin cleavage; 3-6, peak fractions after the size exclusion chromatography (SEC) of the growth factors. M, molecular mass markers. Masses in kDa are indicated on the left.

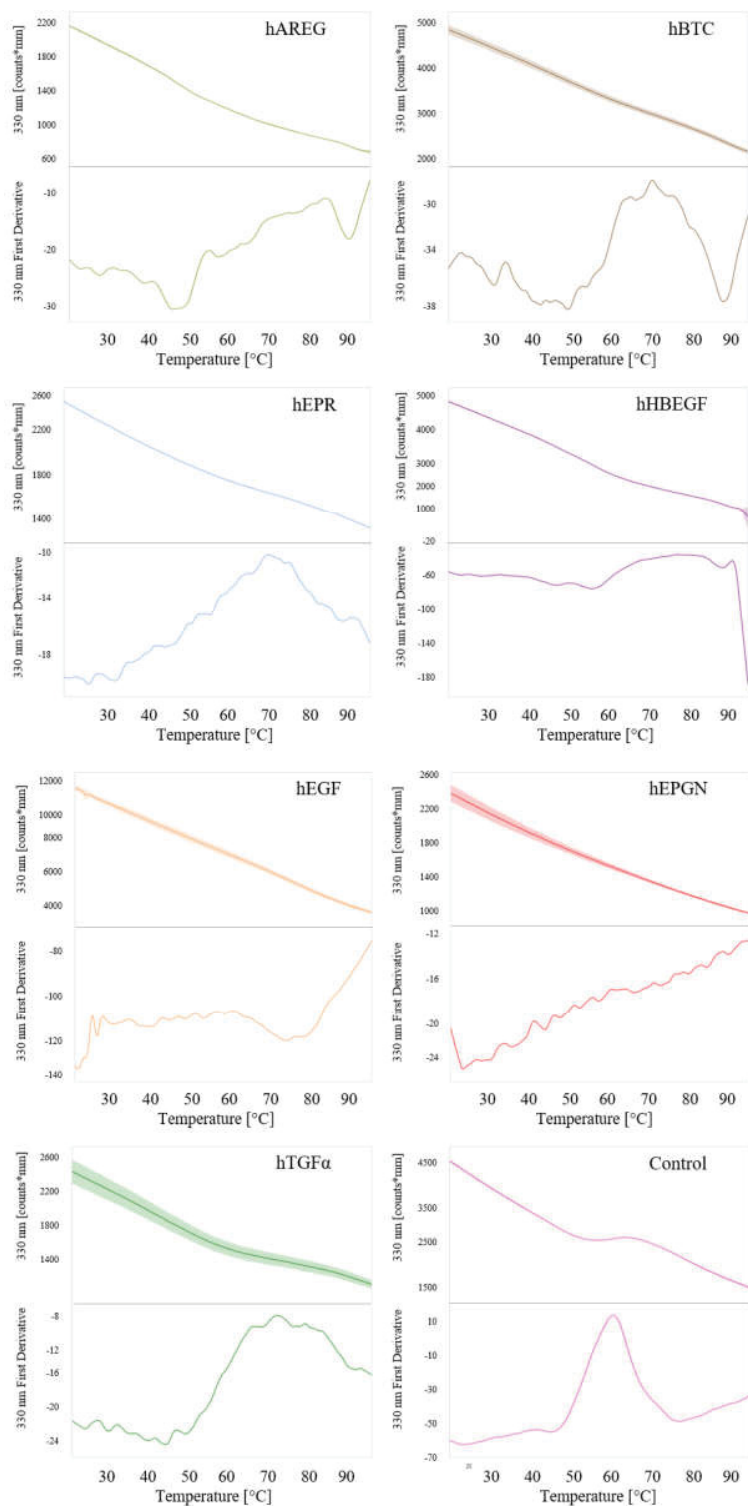


**2. Complete images of the polyacrylamide gel electrophoresis analyses of the limited proteolysis assays.**



**Figure S2.** Complete images of the Coomassie stained SDS-polyacrylamide gels used for the analysis of the proteolysis products and control reactions of the EGF growth factors as identified on the right side of each panel. Stability analysis by limited proteolysis using trypsin. The proteolytic assays were performed with samples both untreated (-) and treated with the reducing agent DTT and incubated for 30 and 60 min. Control non-reduced and reduced samples without trypsin treatment were incubated under the same conditions. M, molecular mass markers. Masses in kDa are indicated on the left. -, non-reducing condition. R, reducing condition.

### 3. NanoDSF plots showing the effect of temperature increase on the seven EGF factors



**Figure S3.** Effect of temperature increase on the seven EGF factors monitored by nanoDSF. Top and bottom panels show, respectively, the fluorescence at 330 nm as a function of temperature and its first derivative. In contrast to the control sample (bottom right), a transition temperature could not be determined for the growth factors.

#### 4. Mass spectrometry analysis of disulfide bonds

The seven recombinant growth factors were digested without reduction of cysteines using different proteases selected according to the expected cleavage sites present in their primary sequence and the peptides were analyzed by LC-MS/MS methods with the combinations described below:

Factor	Proteases	LC-MS/MS
AREG	Lys-C plus Glu-C	nano LC EThcD, ESI LC HCD
BTC	Lys-C plus Glu-C	nano LC EThcD, ESI LC HCD, nano LC HCD
EGF	Lys-C plus Glu-C	nano LC EThcD, ESI LC HCD, nano LC HCD
EPGN	Lys-C plus trypsin, plus pepsin	nano LC EThcD
EPR	Pepsin plus trypsin	nano LC EThcD
HBEGF	Lys-C plus pepsin; Lys-C plus Glu-C	nano LC EThcD, ESI LC HCD, nano LC HCD
TGF $\alpha$	Lys-C plus pepsin	nano LC EThcD

For peptide preparations, 5  $\mu$ g of unreduced recombinant EGF domain factors were incubated in 6 M urea, 5 mM NEM for 15 min at RT, digested for 4 h at 37 °C with 0.1  $\mu$ g Lys-C (except EPR), diluted with protease compatible buffer and incubated with another protease. The protease/EGF domain factor ratios ranged from 1:20 to 1:50. The second or third proteolytic reactions were incubated at 37 °C overnight and inactivated by acidification or by heating. Proteolysis products were desalted using C18 stage tips and the eluates from the stage tips concentrated by speed vac. Hundred nanograms of digested peptides were submitted to LC-MS/MS using an Ultimate 3000 RSLCnano coupled to an Orbitrap Fusion Lumos (Thermo Fischer Scientific, Waltham, MA, USA). The liquid chromatography analyses were performed using a flow of 100  $\mu$ L/min (ESI LC HCD) or 250 nL/min (nano LC HCD) and gradient of 5 to 40% MeCN, 0.1% formic acid for 90 min. Data acquisition was performed in the DDA mode with MS1 and MS2 (post HCD fragmentation) spectra acquired with the orbitrap set to 120k resolution with automatic gain control of  $4 \times 10^5$  and maximum injection time of 200 ms or 240 ms, respectively. In the nano LC EThcD analyses, gradients from 5 to 40% MeCN, 0.1% formic acid were carried out for 30 min. MS1 spectra were acquired in the DDA mode with the orbitrap set to 120k resolution with automatic gain control standard and maximum injection time of 50 ms. Fragmentation was performed using EThcD MS2 with the orbitrap set to 30k resolution using automatic gain control standard, and maximum injection time of 54 ms.

Raw data were analyzed using BioPharma Finder 3.0 and the disulfide bond sites identification was required to fit the follow criteria: up to 3 disulfide bonds per ion; MS2 identifications, precursor mass error  $\leq$  5 ppm and, confidence score  $>$  0. With these criteria, we have directly identified di-

peptides linked by a single disulfide bond for AREG, BTC, EPGN, and EPR (Figures S4 to S7). Unfortunately, it was not possible to map disulfide bonds for EGF, HBEGF and TGF $\alpha$ . EGF and HBEGF analyses mapped bonds (C38-C52 for EGF, and C78-C91 for HBEGF) that did not reach the requirements of the criteria described above. According to the LC-MS/MS chromatogram, EGF seems to be more resistant to digestion than the other factors. HBEGF and TGF $\alpha$  were digested with pepsin combined with Lys-C or Trypsin. However, since pepsin can use a relatively high number of amino acid residues as cleavage sites and at different rates, the digestion efficiency did not correlate with peptide identification, hampering mapping of disulfide bonds. The precise mapping of all predicted disulfide bonds present in the seven growth factors here evaluated is an overwhelming challenge. A possible way to get through this, may require the digestion of milligram amounts of each factor, followed by multiple cycles of digestion and purification of the peptides to obtain di-peptides carrying a single disulfide bond, that can be unambiguously identified by LC-MS/MS.

## AREG

Amino acid sequence:

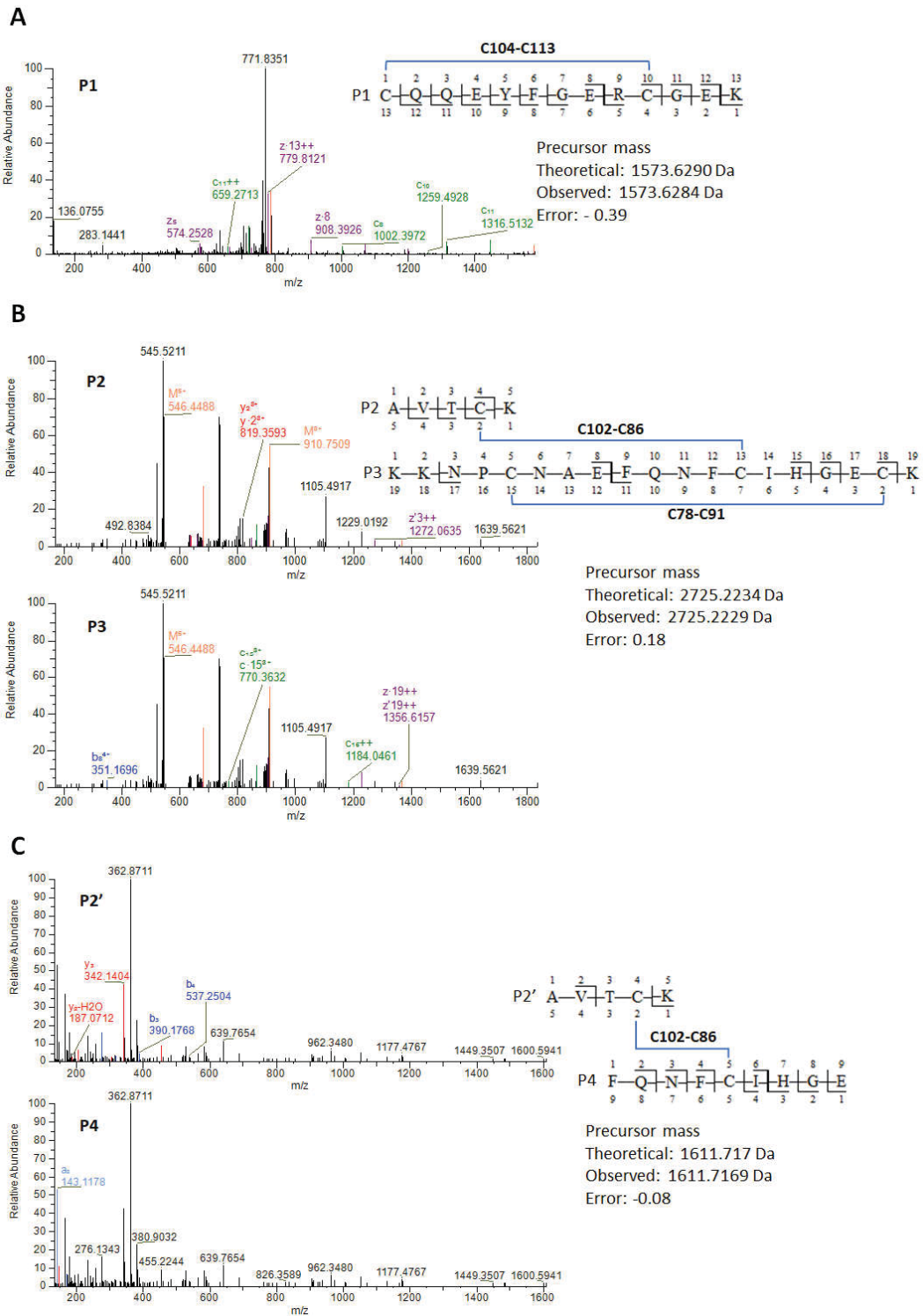
GSGMKETA~~AAK~~FERQHMDSPDLGTDDDDKAMASVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNGKNRRNRK  
 KKNPCNAE~~FQNF~~CIHGECKYIEHLEAVTCKCQOEYFGERCGEKSMKTHSMIDSSLSKL

Predicted disulfide bonds: C78-C91, C86-C102, C104-C113

Sites identified:

Site	Modification	Delta (ppm)	Confidence Score	ID Type	MS Area	method	Identification
1:C104,C113	1ss	-0.39	100	MS2	6.79E+07	nano LC EThcD	C104-C113 MS2 confirmatory spectrum
1:C102/1:C78,C86,C91	2ss	0.18	91.9	MS2	1.54E+07	nano LC EThcD	C78-C91 by inference
1:C102/1:C86	1ss	-0.08	100	MS2	2.67E+04	ESI LC HCD	C86-C102 MS2 confirmatory spectrum

C104-C113 and C86-C102 were identified directly by MS2 confirmatory spectra. C78-C91 was identified by inference since it is part of a di-peptide containing a directly identified bond.



**Figure S4.** Assignment of the MS/MS spectra of the disulfide bond-containing peptides of the recombinant AREG EGF domain. The three disulfide bonds C104-C113 (A), C102-C86/C78-C91 (B) and C102-C86 (C) were identified by MS/MS. Peptides P2 and P3 were identified in the same spectrum (B), however they were annotated separately. The same is valid for P2' and P4 (C).

## BTC

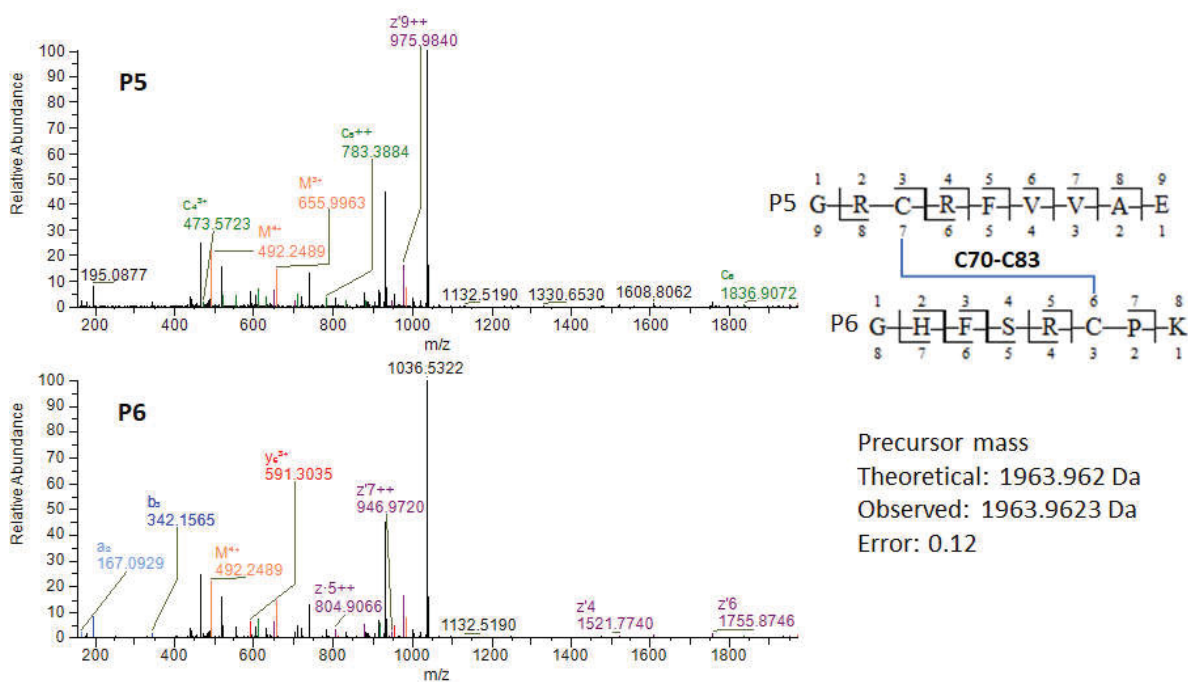
Amino acid sequence:

GSGMKETAAAKFERQHMDSPDLGTDDDDKAMADGNSTRSPETNGLLCGDPEENCAATTTQSKRKGHF<sup>S</sup>RCPKQ  
**YKHYCIKGRCRFVVAEQTPSCVCEGYIGARCERVDLFYKL**

Predicted disulfide bonds: C70-C83, C78-C94, C96-C105

Site identified:

Site	Modification	Delta (ppm)	Confidence Score	ID Type	MS Area	method	Identification
1:C70/1:C83	1ss	0.12	98.6	MS2	4.25E+08	nano LC ETHcD	C70-C83 MS2 confirmatory spectrum



**Figure S5.** Assignment of the MS/MS spectrum of the disulfide bond-containing peptide of recombinant BTC. The disulfide bond C70-C83 was identified by MS/MS. Peptides P5 and P6 were identified in the same spectrum, however they were annotated separately.

## EPGN

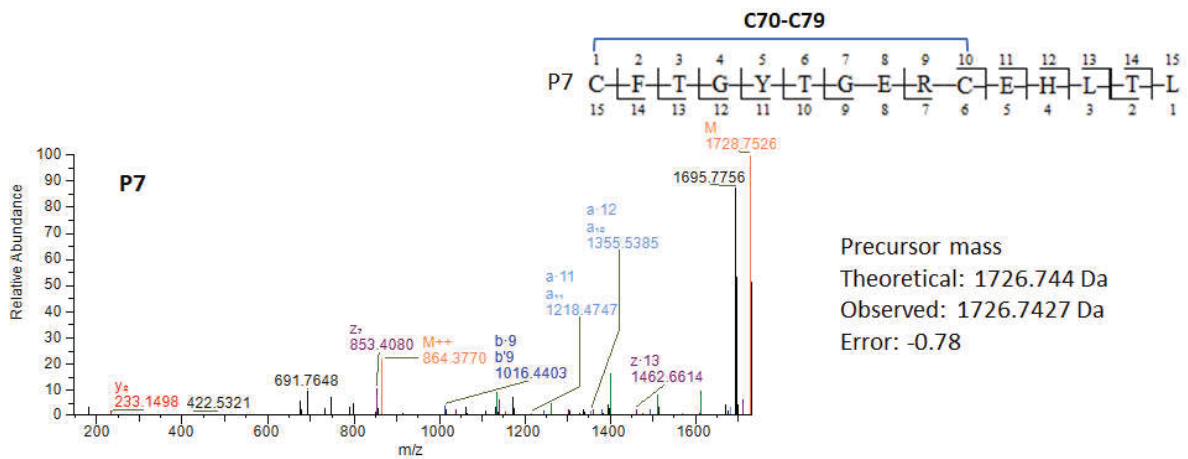
Sequence:

GSGMKETAAAKFERQHMDSPDLGTDKDDDDKAMAEGPIALKFSHLCLEDHNSYCI NGACAFHHELEKAI CRCF TG  
YTGERCEHLTLTSYAVDSYE

Predicted disulfide bonds: C44-C57, C52-C68, C70-C79.

Site identified:

Site	Modification	Delta (ppm)	Confidence Score	ID Type	MS Area	Method	Identification
1:C70,C79	1ss	-0.78	100	MS2	9.94E+07	nano LC ETHcD	C70-C79 MS2 confirmatory spectrum



**Figure S6.** Assignment of the MS/MS spectrum of the disulfide bond-containing peptide of recombinant EPGN. The disulfide bond C70-C79 was identified by MS/MS.

## EPR

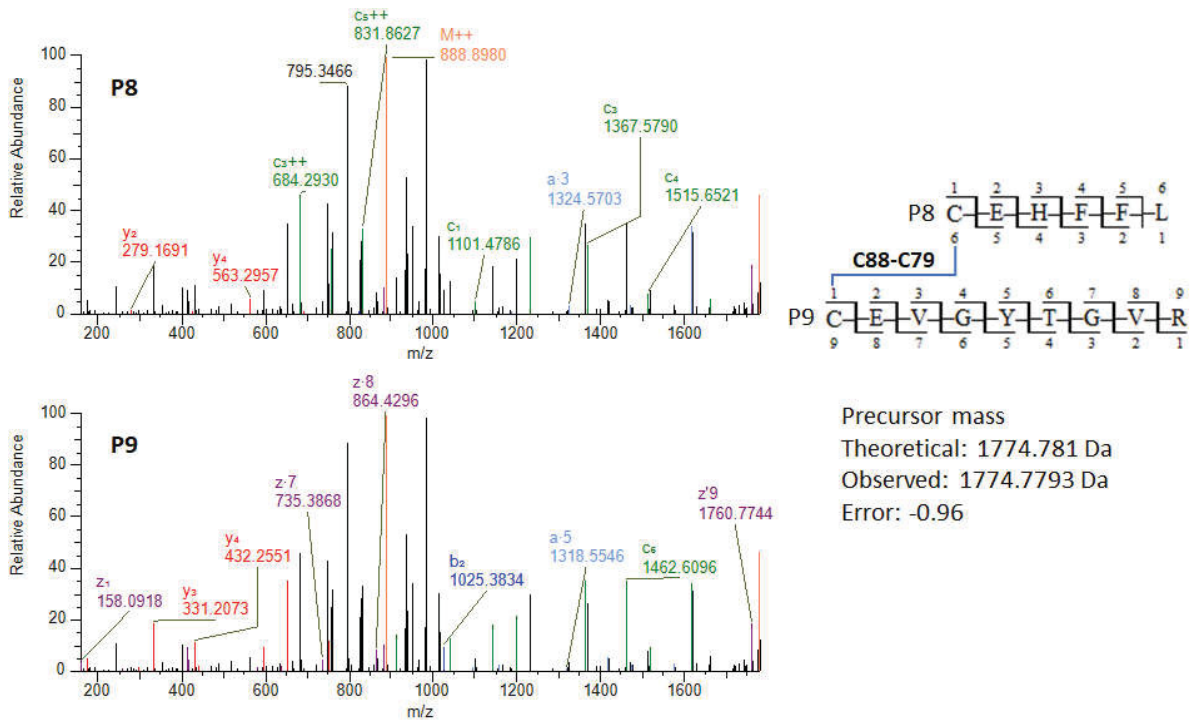
Amino acid sequence:

GSGMKETAAAKFERQHMDSPDLSCIPGESSDNCTALVQTEDNPRVAQVSI TKCSSDMNGYCLHGQC IYLVDMSSQNYRCREVG YTGVRCEHFFL

Predicted disulfide bonds: C53-C66, C61-C77, C79-C88

Site identified:

Site	Modification	Delta (ppm)	Confidence Score	ID Type	MS Area	Method	Identification
1:C88/1:C79	1ss	-0.96	71	MS2	1.55E+06	nano LC EThcD	C79-C88 MS2 confirmatory spectrum



**Figure S7.** Assignment of the MS/MS spectrum of the disulfide bond-containing peptide of recombinant EPR. The disulfide bond C79-C88 was identified by MS/MS. Peptides P8 and P9 were identified in the same spectrum, however they were annotated separately.



### 5.1.2. Discussão presente no artigo publicado na revista *Scientific Reports*

Os membros da família de fatores de crescimento EGF compartilham o mesmo domínio estrutural EGF e são capazes de interagir com receptores celulares e ativar vias que levam à proliferação de diferentes células e tecidos. A presença de três ligações dissulfeto no domínio EGF, com combinação alternada para os dois primeiros pares, torna uma estrutura proteica difícil de reconstituir em uma conformação funcional em organismos procarióticos.

Neste trabalho desenvolvemos construções genéticas e descrevemos procedimentos detalhados para obter sete fatores de crescimento humano da família EGF de forma simples e fácil, usando um único protocolo e a mesma configuração experimental e reagentes. As construções genéticas contendo os fatores de crescimento fundidos a Trx em pET32a mostraram-se muito eficientes para expressão solúvel na cepa Shuffle de *E. coli*. As etapas subsequentes de purificação por afinidade, digestão com trombina e recuperação dos fatores de crescimento clivados funcionaram de forma muito eficiente para todos os fatores de crescimento EGF testados.

Nossos dados mostram que a utilização de Trx como parceiro de fusão é muito prática, não exigindo métodos demorados para redobrar *in vitro* de proteínas insolúveis expressas por *E. coli* nem o uso de células eucarióticas como sistemas de expressão. Em relação à extensão N-terminal do ligante codificado pelo vetor pET32a que permanece ligado ao fator de crescimento após digestão com trombina, examinamos o modo de ligação de hEGF, hEPGN, hEPR e hTGF $\alpha$  ao receptor usando as estruturas de seus respectivos complexos com o domínio extracelular do receptor de EGF depositado em PDB.

Em todos os casos, as extremidades N- e C-terminais dos ligandos estão localizadas fora da cavidade de ligação, longe da interface de ligação, em posições nas quais as extensões N- e C-terminais provavelmente não interferem na ligação. Um grande desafio para a expressão de *E. coli* de pequenas proteínas ricas em ligações dissulfeto é o emparelhamento correto dos resíduos de cisteína para a formação de ligações dissulfeto para que a proteína possa assumir um dobramento nativo. Estudos anteriores mostraram que Trx e SUMO foram usados com sucesso como parceiros de fusão para a expressão de *E. coli* de hEPR (KADO et al., 2015; SATO et al., 2003) e hHBEGF (LU et al., 2010), respectivamente.

No entanto, é importante salientar que para a expressão de hEPR, uma extensão em direção ao terminal N de pró-EPR precisa ser incluída na proteína de fusão para estabilidade. A proteína passível de cristalização inclui 25 resíduos do propeptídeo. A expressão do domínio EGF de hEPR sem extensão N-terminal resultou em proteína instável (dados não mostrados).

Portanto, a fim de melhorar a taxa de sucesso visando tanto a expressão solúvel quanto a proteína estável após a clivagem de Trx, os limites N- e C-terminais para clonagem dos fatores de crescimento EGF foram baseados nas estruturas 3D depositadas no PDB.

Análises conformacionais e de estabilidade foram realizadas por métodos indiretos e diretos para avaliar o estado de dobramento dos fatores de crescimento recombinantes. Proteólise limitada e espalhamento dinâmico de luz, que foram realizados em paralelo com amostras não reduzidas e reduzidas, revelaram diferenças marcantes entre os dois estados. No estado não reduzido todos os fatores de crescimento de EGF mostraram resistência muito maior à proteólise, bem como picos únicos para distribuição de tamanho de raio a 20 °C, que são indicações de que as ligações dissulfeto estão formadas corretamente. Curiosamente, hAREG mostrou um comportamento divergente sob condições redutoras.

As análises de DLS indicaram que hAREG não se agrega após a redução da ligação dissulfeto, mesmo após longos tempos de incubação, enquanto todos os outros mostraram agregação sob condições redutoras, embora HBEGF e TGF $\alpha$  precisassem de tempos mais longos. Ensaios com temperaturas crescentes monitoradas tanto por nanoDSF quanto por CD não detectaram estados de transição claros, o que é consistente com a natureza estável do domínio EGF sustentado pelas três ligações dissulfeto. Os espectros de CD de hEGF e hTGF $\alpha$  foram semelhantes aos espectros de CD descritos na literatura para esses fatores e são consistentes com espectros de proteínas com nenhum ou baixo conteúdo alfa-helicoidal. hEGF, hEPGN, hEPR e hTGF $\alpha$  também mostraram forte resistência à agregação sob altas temperaturas enquanto hBTC, hAREG e hHBEGF mostraram resistência intermediária à agregação sob aquecimento.

As análises conformacional e de estabilidade foram complementadas por espectrometria de massa, que revelou as três ligações dissulfeto corretas para AREG e ligações corretas para os pares de cisteína C1-C2 de BTC e C5-C6 para EPGN e EPR. A análise por espectrometria de massa parece ser limitada principalmente pela baixa eficiência da digestão proteolítica do domínio EGF, que não gera dipeptídeos portadores de pontes dissulfeto adequadas para MS/MS. A atividade biológica dos fatores de crescimento do EGF recombinante foi demonstrada por ensaios de proliferação e migração celular. Todos eles induziram a proliferação celular em concentrações tão baixas quanto 5 ng/mL.

O efeito de proliferação celular obtido em nosso trabalho para hEGF, hAREG, hBTC, hHBEGF e hTGF $\alpha$  é semelhante aos resultados publicados em estudos anteriores para células HeLa (HANG et al., 2015; HENRIKSEN et al., 2013), e no caso de hEGF também para outras linhagens celulares (HENRIKSEN et al., 2013). Considerando as concentrações entre 5 e 50 ng

mL necessárias para a indução da proliferação, nossos achados para hEGF são semelhantes aos descritos para linha celular MIO-M146 e, para hEPGN e hTGF $\alpha$  descritos para células de mieloma expressando ErbB-145 de forma estável. Além da proliferação, os fatores de crescimento de EGF recombinantes podem induzir a migração celular, como demonstrado pelos ensaios artificiais de cicatrização de feridas usando células NDFH. Esse achado é consistente com relatos anteriores mostrando que hAREG, hHBEGF e hEGF são capazes de induzir o fechamento de arranhões em outros tipos de linhagens celulares.

Em geral, os ensaios de atividade mostram que os fatores de crescimento de EGF recombinantes são altamente ativos. Isso, juntamente com as informações detalhadas fornecidas sobre as construções genéticas, procedimentos de purificação e análises biofísicas para controle de qualidade do estado de dobramento dos fatores de crescimento recombinantes representam ferramentas úteis para a produção interna rápida e confiável dos sete membros da família de fatores de crescimento EGF.

### **5.1.3. Manuscrito submetido à revista *Molecular Biology Reports***

**Functional EGF domain of the human neuregulin 1 $\alpha$  (hNRG1 $\alpha$ ) produced in *Escherichia coli* with accurate disulfide bonds**

Arthur Schweitzer Ferreira<sup>1,3\*</sup>, Amanda Lopacinski<sup>1,3\*</sup>, Michel Batista<sup>2</sup>, Priscila Mazzocchi Hiraiwa<sup>1</sup>  
Beatriz Gomes Guimarães<sup>1</sup> and Nilson I. T. Zanchin<sup>1,3,#</sup>

<sup>1</sup>Laboratory of Structural Biology and Protein Engineering, Carlos Chagas Institute, FIOCRUZ Paraná – Curitiba-PR, Brazil

<sup>2</sup>Mass Spectrometry Facility, Carlos Chagas Institute, FIOCRUZ Paraná – Curitiba-PR, Brazil

<sup>3</sup>Cellular and Molecular Biology Graduate Program, Federal University of Paraná – Curitiba-PR, Brazil.

NITZ - ORCID: 0000-0002-1153-0694

BGG – ORCID: 0000-0002-8546-2460

\*These authors contributed equally to this work.

# Corresponding author. Email address: [nilson.zanchin@fiocruz.br](mailto:nilson.zanchin@fiocruz.br)

## Abstract

**Background:** Neuregulins comprise a large family of growth factors containing an epidermal growth factor (EGF) domain. NRG1 acts in signaling pathways involved in proliferation, apoptosis, migration, differentiation, and adhesion of many normal cell types and in human diseases. The EGF domain of NRG1 mediates signaling by interaction with members of the ErbB family of receptors. Easy access to functional hNRG1 $\alpha$  EGF domain can be a valuable tool to investigate its function in different cell types.

**Materials and methods:** The EGF domain of hNRG1 $\alpha$  was produced in *Escherichia coli* in fusion with TrxA and purified after cleavage of TrxA. Conformation and stability analyses were performed by using biophysical methods and the disulfide bonds mapped by mass spectrometry. The activity of the hNRG1 $\alpha$  EGF domain was demonstrated in cell proliferation and migration assays.

**Results:** Approximately 3.3 mg of hNRG1 $\alpha$  EGF domain were obtained starting from a 0.5 L of *E. coli* culture. Heat denaturation assays monitored by circular dichroism and dynamic light scattering revealed that it is a highly stable protein. Correct formation of the three disulfide bonds was demonstrated by mass spectrometry with high accuracy. The recombinant EGF domain of hNRG1 $\alpha$  purified in this work is highly active, inducing cell proliferation at concentration as low as 0.05 ng/mL. It induces also cell migration as demonstrated by a gap closure assay.

**Conclusions:** The EGF domain of hNRG1 $\alpha$  was produced in *E. coli* with the correct disulfide bonds and presented high stimulation of HeLa cell proliferation and NDFH cell migration.

**Keywords:** human Neuregulin-1, recombinant EGF domain, disulfide bond assignment, hNRG1 $\alpha$  EGF domain activity ·

## Introduction

Neuregulins (NRG) comprise a large family of growth factors encoded by six genes (NRG1-NRG6) with the number of isoforms being amplified by a large number of alternative splicing events. The isoforms differ in amino acid sequence, overall structure, and levels of expression in different tissues [1]. They contain an epidermal growth factor (EGF) domain that mediates cell signaling by interaction with members of the ErbB family of cellular receptors [2–4]. The EGF domain of NRG1 stimulates ErbB tyrosine kinase receptors of signaling pathways involved in cell proliferation, apoptosis, migration, differentiation, and adhesion [2–5]. NRG1 has been implicated in development of breast, heart, muscle, and brain tissues and in human diseases including breast cancer [5]. Furthermore, NRG1–ErbB4 signaling is involved in important processes required for brain development and plasticity and neuropsychiatric disorders [2–4].

Six different types of NRG1 (Types I–VI) have been identified, containing distinct N-terminal regions. They are synthesized as transmembrane precursor proteins, which subsequently undergo proteolytic processing on the C-terminal region of the EGF, releasing the EGF signaling domain with the type-specific N-terminal region, except for type III, that remains bound to the membrane due to a transmembrane domain in its N-terminal region [2, 3, 6]. The EGF domain of neuregulins was shown to be sufficient to activate ErbB receptors in a specific manner and induce cellular responses [7, 8]. This favors the functional studies which require the utilization of recombinant neuregulin EGF domains, since some features of this domain, such as small size and absence of glycosylation sites, facilitates its recombinant production in prokaryotes such as *Escherichia coli* [9, 10]. In this scenario, the advantages of using *E. coli* as the host system are well known and include fast growth, cheapest growth media components, simplest laboratory setup, easy and fast transformation with exogenous DNA [11]. Nonetheless, this expression system has its specific disadvantages, particularly, the formation of inclusion bodies, which in turn, makes the process of protein purification more complex and time-consuming by requiring *in vitro* refolding of the recombinant proteins, which is not always possible. Several studies have reported overexpression of the EGF domain alone for various members of the EGF family of growth factors [12–14], however, only relatively low yields were obtained.

Despite small size, the EGF domain displays a complex structure maintained by three disulfide bonds formed by the C<sub>1</sub>X<sub>7</sub>C<sub>2</sub>X<sub>4</sub>-5C<sub>3</sub>X<sub>10</sub>-13C<sub>4</sub>X<sub>5</sub>X<sub>8</sub>C<sub>6</sub> conserved sequence, and with its correct folding being dependent on the accurate formation of disulfide bonds linking the cysteine residues in the combinations C1-C3, C2-C4 and C5-C6 [15]. In fact, accurate formation of the disulfide bonds may be one of the major limitations to produce soluble and functional EGF domains in prokaryotes such as

*E. coli*. However, it is possible to overcome this limitation by using fusion proteins, which can function as folding and solubility chaperons or modified *E. coli* strains designed to favor disulfide bond formation.

In this work, we describe the expression of the human NRG1 $\alpha$  EGF domain fused to TrxA in *E. coli* and a two-step procedure to separate the fusion protein and purify the hNRG1 $\alpha$  EGF domain. Yields in the range of ~6.5 mg per liter of culture were obtained after purification. We demonstrate by mass spectrometry that the disulfide bonds are formed between the correct cysteine pairs. In addition, we show that the recombinant hNRG1 $\alpha$  EGF domain presents biological activity in cell proliferation and migration assays. Our study describes a handy method to obtain biologically active EGF domains.

## **Materials and Methods**

### **Protein Expression and Purification**

We used the NMR structure of the hNRG1 $\alpha$  EGF domain deposited at the Protein Data Bank with the code 1HRE [16] to define the amino acid sequence to be expressed. A synthetic gene encoding this 67-residue segment was acquired from GenScript (Piscataway, NJ, USA) cloned between the *Nco*I and *Xho*I restriction sites of plasmid pET32a (Novagen/Merck, Darmstadt, Germany). In this construction, the hNRG1 $\alpha$  EGF domain is expressed in fusion with thioredoxin and a hexa-histidine tag. Thrombin and enterokinase cleavage sites allow separation of the hNRG1 $\alpha$  EGF domain from TrxA-His6. The resulting plasmid, pET32a-hNRG1 $\alpha$ , was transformed into the *E. coli* expression strain SHuffle (NEB C3026J) [17] using a heat shock protocol as previously described [18]. The transformed cells were plated on Lysogeny Broth (LB)-agar plates containing 100  $\mu$ g/mL of ampicillin and incubated overnight at 37 °C. A single colony was used to inoculate 10 mL of LB containing 100  $\mu$ g/mL of ampicillin. After incubation of this starting culture overnight at 37 °C, 5 mL were used to inoculate a 500 mL LB culture containing 100  $\mu$ g/mL of ampicillin. This culture was maintained at 37 °C until the optical density at 600<sub>nm</sub> reached ~0.6. Subsequently, the culture was transferred to 16 °C, induced with 0.5 mM IPTG and incubated further at 16 °C for 24 h with rapid agitation (300 rpm).

Following induction, the bacterial cells were harvested by centrifugation at 6,000 x g for 10 min at 4 °C. The cells were suspended in 10 mL of buffer A (50 mM TrisHCl pH 8.0; 300 mM NaCl; 20 mM imidazole), lysed using a microfluidizer (Microfluidics, Newton, MA, USA) and the extract was



cleared by centrifugation at 20,000 x g for 30 min. The Trx-hNRG1 $\alpha$  EGF domain fusion protein was purified by affinity chromatography using a HisTrap 5 mL column (Cytiva 17524802) on an ÄKTA Pure M25 chromatography system (Cytiva, Marlborough, USA). After loading the sample, the column was washed using 3 column volumes (CV) of buffer A, followed by a 0 to 10% gradient of buffer B (50 mM TrisHCl pH 8.0; 300 mM NaCl; 500 mM imidazole) in two CV. The target protein was eluted using a 10 to 100 % gradient of buffer B in four CV. The peak fractions were combined after analysis by SDS-PAGE. 5 mL of the sample were diluted to a final volume of 15 mL using 50 mM Tris-HCl pH 8.0 and 300 mM NaCl. Protein concentration was determined based on absorbance at 205 nm and the extinction coefficient using a NanoDrop ONE<sup>c</sup> Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The thioredoxin-tag was removed by digestion with one unit of thrombin (Cytiva 27084601) for each 10 mg of recombinant TrxA-hNRG1 $\alpha$  EGF domain fusion protein at 25 °C for 20 h. After thrombin digestion, TrxA containing the hexa-histidine tag was removed by affinity chromatography using the same protocol described above. The flowthrough sample, containing the hNRG1 $\alpha$  EGF domain devoid of the fusion sequences, was concentrated using an Amicon Ultra-0.5 mL 3K (Merk Millipore, Cork Ireland) to a final volume of 500  $\mu$ L and purified further by size-exclusion chromatography using a Superdex 75 10/300 GL column (Cytiva 29148721). After each chromatographic step, the samples were analyzed by SDS-PAGE. Poly-acrylamide concentrations in the stacking and separating gels were 6.5% and 15%, respectively. Electrophoresis was performed with 0.75 mm-thick gels assembled on Mini-PROTEAN Tetra Cells (Bio-Rad Laboratories, Hercules, USA) at 100 V. Gels were stained using Brilliant Blue R-250 and destained using a methanol/acetic acid/water (30%/10%/60%, v/v) solution.

### **Limited Proteolysis Assay**

Limited proteolysis assays were performed under reducing and non-reducing conditions. For reduction of the disulfide bonds, the recombinant hNRG1 $\alpha$  EGF domain was treated with 10 mM of dithiothreitol (DTT, Merck 3483123). The limited proteolysis assays were performed with a 100:1 (w/w) ratio of recombinant hNRG1 $\alpha$  EGF domain and trypsin (Sigma 59427C) and incubated for 30 and 60 min at 37°C. Control reactions without trypsin were run in parallel. The products of the reactions were analyzed by SDS-PAGE.

### **Determination of Disulfide Bonds by Mass Spectrometry**

For sample preparation, 5 µg of unreduced recombinant hNRG1α EGF domain were incubated in 6 M urea, 5 mM NEM for 15 min at RT, digested for 4 h at 37 °C with 0.1 µg Lys-C, diluted 1:8 with PBS, incubated with 0.17 µg of Glu-C at 37 °C overnight and inactivated with 0.5% TFA. Proteolysis products were desalted using C18 stage tips and the eluates from the stage tips concentrated by speed vac. Hundred nanograms of digested peptides were submitted to LC-MS/MS using an Ultimate 3000 RSLCnano system coupled to an Orbitrap Fusion Lumos (Thermo Fischer Scientific, Waltham, MA, USA). For ESI LC HCD, the liquid chromatography analysis was performed using a flow of 100 µL/min and gradient from 5 to 40% MeCN, in 0.1% formic acid for 90 min. The acquisition was done in DDA mode with MS1 and MS2 (post HCD fragmentation) spectra acquired in the orbitrap set to 120k resolution with automatic gain control of  $4 \times 10^5$  and maximum injection time of 200 ms or 240 ms, respectively. In nano LC EThcD analysis, the flow was 250 nL/min, gradient from 5 to 40% MeCN, 0.1% formic acid for 30 min. The acquisition was done in DDA mode with MS1 spectra acquired in the orbitrap set to 120k resolution with automatic gain control standard and maximum injection time of 50 ms. Fragmentation was done using EThcD, MS2 in the orbitrap set to 30k resolution with automatic gain control standard, and maximum injection time of 54 ms. Raw data were analyzed in BioPharma Finder 3.0 and the disulfide bond sites identified fit the follow criteria: up to 3 disulfide bonds per peptide; MS2 identification, and precursor mass error  $\leq 5$  ppm.

### **Dynamic Light Scattering and Thermal Unfolding Analyses**

Dynamic Light Scattering (DLS) and thermal unfolding were performed on a Prometheus Panta device (Nanotemper Technologies, Munich, Germany) using standard capillaries. DLS measurements were performed at 20 °C using the high sensitivity setup (10 acquisitions of 10 seconds each) in duplicates. Non-reduced samples of the hRG1α EGF domain were also submitted to a linear ramp of 1 °C/min from 20 °C to 95 °C, and back to 20 °C, with parallel measurement of DLS.

### **Circular Dichroism**

For circular dichroism analyses, the sample buffer was changed by ultrafiltration to 4.2 mM Tris-HCl pH 8.0 and 50 mM NaCl using Amicon Ultra-0.5 mL 3K concentrators (Merk Millipore, Cork Ireland). CD spectra were acquired using the hRG1α EGF domain at 0.2 mg/mL using a 1 mm path cuvette on a Jasco J-815 circular dichroism spectrometer (JASCO, Tokyo, Japan) with a scanning speed of 50 nm/min at 20 °C. Each spectrum was accumulated at least five times in the range from 190 nm to 260 nm. For thermal denaturation, the temperature was raised from 20 °C to 80 °C. A new set of

spectra was acquired after cooling to 20 °C. The raw data were converted using the mean residue molar ellipticity formula and autozero was set by using the value at 250 nm.

### **Cell Proliferation Assay**

HeLa cells (Carlos Chagas Institute cell bank) were cultivated in DMEM (Thermo Fisher Scientific 41965062) supplemented with 10% bovine fetal serum (BFS) at 37 °C with 80% humidity and 5% CO<sub>2</sub> using a Forma Series II Water Jacketed CO<sub>2</sub> incubator (Thermo Fischer Scientific, Waltham, MA, USA). Subsequently, cells from a confluent culture were transferred to a 24-well plate at a concentration of 10<sup>5</sup> cells per well and incubated for 24 h. The medium was changed for DMEM supplemented with 3.3 % BFS and treated with 50 ng/mL of recombinant hNRG1α EGF domain in triplicates in parallel with untreated controls also in triplicates. After 48h, the wells were washed with PBS (137 mM NaCl, 2.7 mM KCl, 8 mM Na<sub>2</sub>HPO<sub>4</sub>, and 2 mM KH<sub>2</sub>PO<sub>4</sub>) and treated with 200 µL of PBS containing trypsin 1% (w/v) (Merck T8003). After 5 min of incubation at 37 °C, the released cells were collected, centrifuged at 500 x g for 5 min and suspended in 280 µL of PBS and 20 µL of *CountBright™ Absolute Counting Beads* (Thermo Fisher C36950). The cell number was determined on a BD FACS Canto II (BD Life Sciences, New Jersey, USA) and the results analyzed using the software FlowJo™ v10.8 (BD Life Sciences, Franklin Lakes, NJ, USA). At least 1000 beads were acquired and a maximum of 3000 beads was set as upper limit. Quantification was performed following the protocol described at the *CountBright™ Absolute Counting Beads* manual (Invitrogen MAN0018850). The one-way ANOVA with Dunnett post-test was performed by comparing the means between the groups.

A dose response curve of the effect of the recombinant hNRG1α EGF domain on HeLa cell proliferation was determined by the 3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide (MTT) viability cell assay. For this purpose, HeLa cells were cultivated in 96-well plates in manner similar manner to the proliferation assay. The assay was performed in triplicate utilizing 5x10<sup>4</sup> cells per well and increasing concentrations of the recombinant hNRG1α EGF domain: 0.005 ng/mL, 0.05 ng/mL, 0.5 ng/mL, 5 ng/mL and 50 ng/mL. After 48h of incubation, the wells were washed with PBS, treated with 50 µL of MTT (Invitrogen M6494) at 5 mg/mL and 50 µL of serum free medium and incubated at 37 °C for 3 h. Subsequently, 150 µL of a solvent solution (4 mM HCl, 0.1% NP40 in isopropanol) was added to each well and incubated for another hour. Absorbance was read at 570 nm using a Synergy H1 Hybrid Reader (BioTek Instruments, Winooski, VT, USA). The dose-response curve was adjusted using a nonlinear fit of the data.

### **Scratch closure assay**

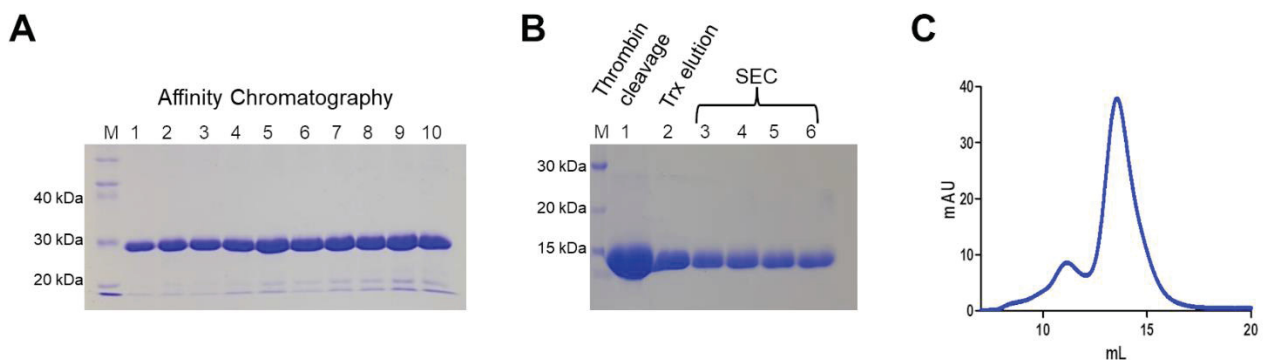
NDFH cells (Lonza CC-2509) were maintained in DMEM (Thermo Fisher 41965062) supplemented with 10 % BFS at 37 °C with 80% humidity and 5% CO<sub>2</sub> using a Forma Series II Water Jacketed CO<sub>2</sub> incubator (Thermo Fischer Scientific, Waltham, MA, USA). For scratch closure assays, the cells were allowed to grow to ~100% confluence when the medium was changed to DMEM without BFS and, the incubation continued for further 24 h. At this time, a scratch was made at the center of the well, 50 ng of recombinant hNRG1 $\alpha$  EGF domain were added to cultures and images of the unstained cells were acquired. Each treatment was performed in triplicate. After 48 h, the cells were stained using the crystal violet protocol described by Feoktistova *et al* [19]. Briefly, the cells were initially fixed using ice-cold 100% methanol for 10 min, stained with a 0.5% (w/v) crystal violet solution in 20% (v/v) methanol for 10 min and washed with PBS. Images of the stained cells were acquired on a Nikon Eclipse TE300 Inverted Microscope (Nikon, Tokyo, Japan) using a 10X objective and the QCapture Pro 6.0 software (Teledyne QImaging, Surrey, Canada). The rate of wound healing was calculated using ImageJ [20]. The scratches were aligned with the base of the image and a 900x600-pixel rectangle was cropped around the scratch. The area at time 0 was calculated using the polygon selection tool and the area at 48 h was calculated using the area between the edges of the scratch converted to an 8-bit image, which is sharpened afterwards. The find edges tool was applied and the image was converted to binary format. The particles were then analyzed using areas between 30-100000 inch<sup>2</sup> with bare outlines, including holes and the values obtained using the summarize tool.

## **Results**

### **High yields of hNRG1 $\alpha$ EGF domain obtained from *E. coli* expression**

The EGF domain of hNRG1 $\alpha$  (amino acid sequence according to Protein Data Bank code 1HRE) [16] was expressed in fusion with thioredoxin A (TrxA), which contributes for correct protein folding and solubility [21], using the pET32a plasmid and the *E. coli* expression strain SHuffle (NEB C3026J) [17]. The segment between TrxA and the hNRG $\alpha$  EGF domain contains a hexa-histidine tag to facilitate purification by affinity chromatography on immobilized metal columns and, thrombin and enterokinase cleavage sites allowing for subsequent cleavage and separation of the fusion protein.

High levels of purified fusion protein were obtained after affinity chromatography, which migrates as a single band of approximately 30 kDa in SDS-PAGE (Figure 1A). TrxA was efficiently cleaved by thrombin digestion. However, TrxA-Hisx6 and the hNRG1 $\alpha$  EGF domain co-migrate with apparent molecular masses of ~15 kDa in 15% SDS-PAGE (Figure 1B, lane 1). Nevertheless, TrxA-Hisx6 was efficiently separated from the hNRG $\alpha$  EGF domain by a second round of purification on an immobilized nickel column (Figure 1B), which retained TrxA-Hisx6. The hNRG1 $\alpha$  EGF domain found in the flow through of the second affinity chromatography was further purified by size-exclusion chromatography (Figure 1B and 1C). A major peak eluted with the expected volume for the monomeric conformation of the hNRG1 $\alpha$  EGF domain (Figure 1C). TrxA has a distinct elution profile in the Superdex 75 10/300 column, which is clearly distinguishable from the hNRG1 $\alpha$  EGF domain elution profile (data not shown). Although it is a two-step purification protocol, this process is significantly less cumbersome than *in vitro* refolding of denatured proteins, especially considering the need for accurate disulfide bond formation. We obtained ~3.3 mg of recombinant hNRG1 $\alpha$  EGF domain starting from 10 mg of protein before thrombin digestion and from 500 mL of *E. coli* cell culture.



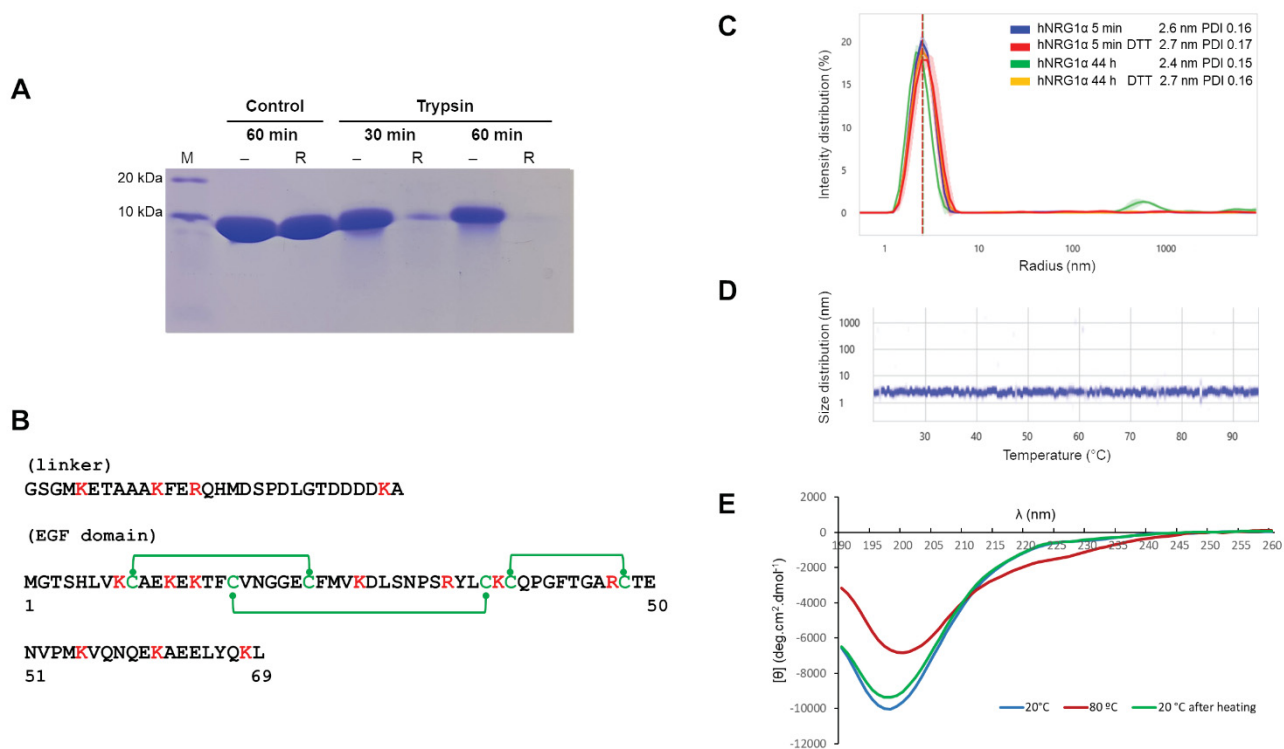
**Figure 1.** Analysis of the hNRG1 $\alpha$  EGF domain purification steps. **(A)** SDS-PAGE analysis of the fusion protein TrxA-hNRG1 $\alpha$  EGF domain purified by affinity chromatography on an immobilized nickel column. **(B)** SDS-PAGE analysis of the thrombin digestion product (lane 1), TrxA-Hisx6 eluted after a second round of nickel affinity chromatography (lane 2), and hNRG1 $\alpha$  EGF domain fractions eluted from size exclusion chromatography (lanes 3-6). M, molecular mass. **(C)** Elution profile of the hNRG1 $\alpha$  EGF domain fractionated on a Superdex 75 10/300 GL size-exclusion column. The samples loaded in the lanes 3-6 shown in (B) correspond to the major peak of the chromatogram.

### **Stability and conformational analysis of the recombinant hNRG1 $\alpha$ EGF domain**

Conformation of the recombinant hNRG1 $\alpha$  EGF domain was initially assessed by limited proteolysis assays. This type of assay can provide indirect information on protein conformation since unstructured or unfolded regions are more accessible to the protease, being digested at faster rates than compact, well-structured regions. Considering that the EGF domain comprises the conserved sequence C<sub>1</sub>X7C<sub>2</sub>X4-5C<sub>3</sub>X10-13C<sub>4</sub>XC<sub>5</sub>X8C<sub>6</sub> [15] and its correct folding is dependent on the correct formation of disulfide bonds linking the cysteine residues in the combinations C1-C3, C2-C4, C5-C6, limited proteolysis assays were performed with recombinant factors previously treated with DTT in parallel with control samples, which were not treated with reducing agent (Figure 2A). Treatment of the hNRG1 $\alpha$  EGF domain with reducing agents, such as DTT, is expected to disrupt the disulfide bonds, which may lead to complete unfolding of the domain. Consistently, the reduced hNRG1 $\alpha$  EGF domain showed high sensitivity to trypsin, being totally degraded after 30 min of incubation, while the non-reduced control was resistant to trypsin for up to 60 min of incubation (Figure 2A). This result provides indication that the cysteines are forming disulfide bonds and that hNRG1 $\alpha$  EGF domain presents a compact conformation.

The effect of the reducing agent on the conformation of the hNRG1 $\alpha$  EGF domain was also evaluated by dynamic light scattering (DLS). Surprisingly, in all conditions tested, including incubation with a reducing agent for 44 hours, the hNRG1 $\alpha$  EGF domain showed a single peak with radius size centered at  $\sim$ 2.6 nm, except for one minor peak after 44 hours indicating partial aggregation (Figure 2C). These results show that the hNRG1 $\alpha$  EGF domain is not prone to aggregation even after disruption of the disulfide bonds. Increasing the temperature from 20 to 95 °C also did not result in an increase in the size distribution of hNRG1 $\alpha$  EGF domain (Figure 2D).

To obtain additional information about the conformational stability, we measured circular dichroism (CD) spectra of the hNRG1 $\alpha$  EGF domain in the far-UV region at 20 °C, after sample heating to 80 °C and, after returning the temperature down to 20 °C (Figure 2E). Thermal denaturation is indicated by a reduction of the CD signal in the 200 nm wavelength range at 80 °C. However, the CD spectrum acquired after returning the temperature to 20 °C overlaps with the spectrum previously obtained at 20 °C (Figure 2E), indicating that the hNRG1 $\alpha$  EGF domain refolds properly after thermal denaturation.



**Figure 2.** Stability and conformational analysis of the hNRG1 $\alpha$  EGF domain. **(A)** Limited proteolysis assays performed with samples untreated (-) and treated (+) with the reducing agent DTT and incubated for 30 and 60 min with trypsin. Control refers to the indicated conditions in the absence of trypsin. M, molecular mass markers. Only the relevant part of the gel is shown. **(B)** Sequence of the hNRG1 $\alpha$  EGF domain with the basic residues targeted by trypsin marked in red and the cysteine residues marked in green. Green lines indicate the disulfide bonds. The sequence identified by (linker) represents the part of the linker encoded by the vector which remains attached to the growth factors after thrombin digestion. **(C)** Profiles of radius of the hNRG1 $\alpha$  EGF domain in presence or absence of DTT as determined by dynamic light scattering analyses. The graphic shows the intensity distribution (%) as a function of particle radius (nm). All assays were performed at 20 °C. Times and treatment are indicated in the figure as well as the corresponding cumulant radius and polydispersity index (PDI). **(D)** Size distribution (nm) of the hNRG1 $\alpha$  EGF domain as determined by DLS as a function of the temperature. **(E)** Circular dichroism analysis of the hNRG1 $\alpha$  EGF domain. The blue line corresponds to the spectrum acquired at 20 °C, the red line to the spectrum acquired 80 °C and the green line corresponds to the spectrum acquired at 20°C after sample heating up to 80 °C. The far-UV circular dichroism spectra (195-260 nm) were recorded using a J-815 spectropolarimeter with five accumulations and using a 1 mm path length cell.

### Recombinant hNRG1 $\alpha$ EGF domain presents accurate disulfide bonds

To confirm the correct pairing of cysteines forming disulfide bonds, the purified hNRG1 $\alpha$  EGF domain was submitted to mass spectrometry analysis. The conserved C1-C3, C2-C4, C5-C6 cysteine pairs are numbered according to the sequence shown in Figure 2B, corresponding to C39-C53, C47-C67 and C69-C78. The proteases Lys-C and Glu-C were selected for digestion based on the amino acid

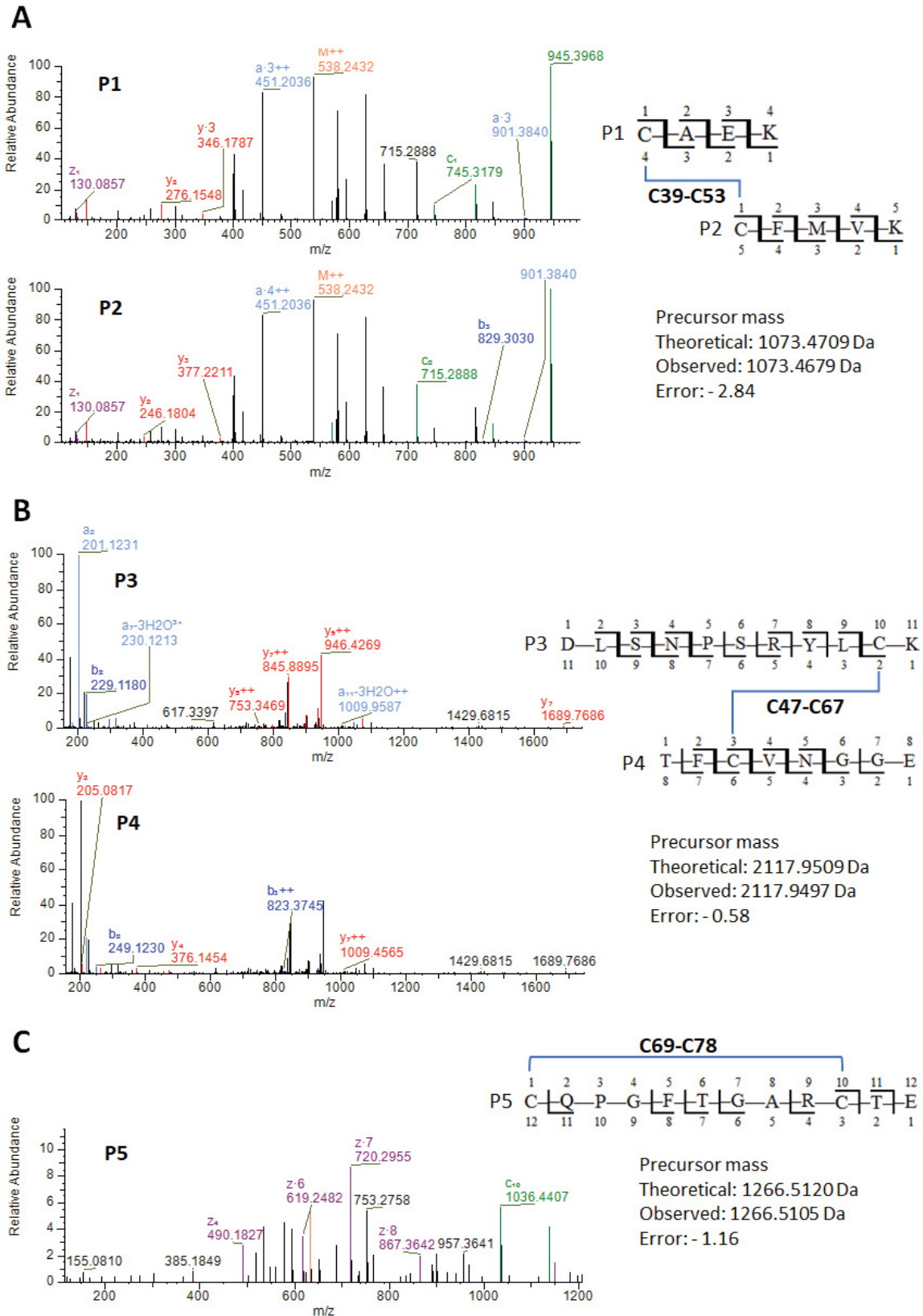


sequence of the hNRG1 $\alpha$  EGF domain. The hinge residue between C67 and C69, as well as the sequence between C47 and C53 are the major points of constraint regarding site-specific proteolytic digestion of the hNRG1 $\alpha$  EGF domain since, under non-reducing conditions, ineffective digestion leads to complex mass and peptides with more than two cysteines in the top score peptides. Nevertheless, all three disulfide bonds peptides were detected in large peptide abundance (MS1 area > 10<sup>6</sup>) (Table 1), with C69-C78 being the most abundant peptide containing disulfide bond among all, as shown in Figure 3.

Table 1: List of peptides identified with one disulfide bond each.

Site	Peptide Sequence	M/Z	Charge State	MS Area	Method
C39/1:C53	CAEK/CFMVK	358.83	3	2.66E+07	nano LC EThcD
C47/1:C67	TFCVNGGE/DLSNPSRYLCK	707.325	3	4.81E+06	ESI LC HCD
C69,C78	CQPGFTGARCTE	634.263	2	2.44E+09	nano LC EThcD



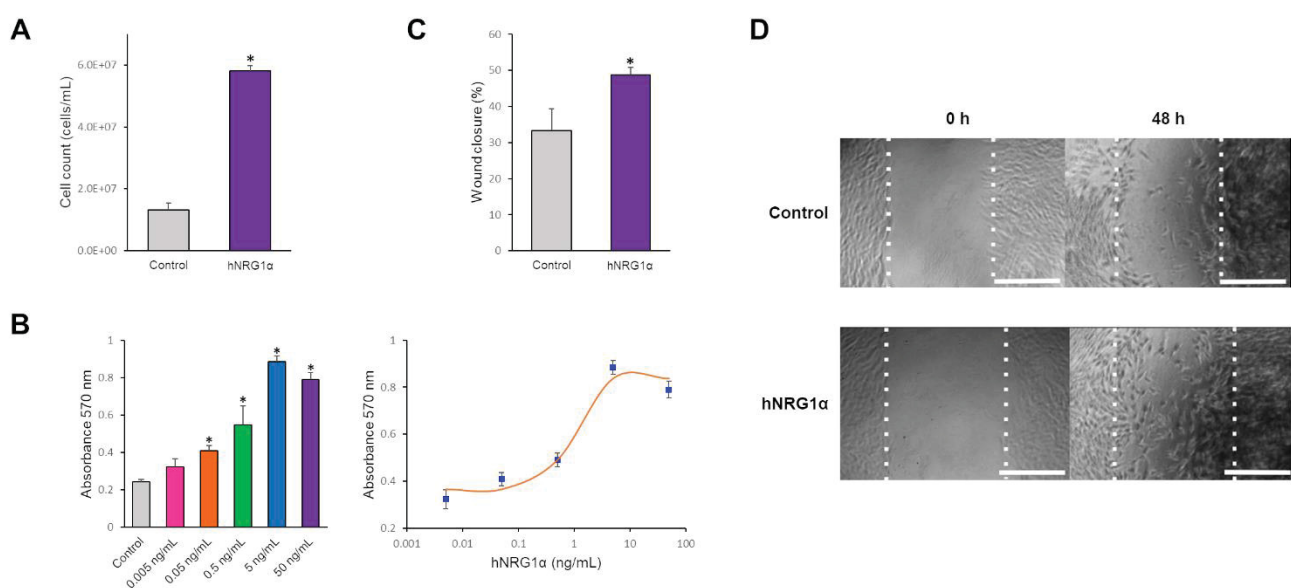


**Figure 3.** Assignment of the MS/MS spectra of the disulfide bond-containing peptides of the recombinant hNRG1 $\alpha$  EGF domain. The three disulfide bonds C39-C53 (A), C47-67 (B) and C68-C78 (C) were identified by MS/MS. Peptides P1 and P2 were identified in the same spectra (A), however they were annotated separately. The same is valid for P3 and P4 (B). MS2 obtained with nano LC ETHcD (A and C) or ESI LC HCD (B).

## The hNRG1 $\alpha$ EGF domain is biologically active and promotes cell proliferation and migration

To determine if the recombinant hNRG1 $\alpha$  EGF domain possesses intrinsic biological activity, its effect on cell proliferation was initially examined by a flow cytometry assay. The same number of HeLa cells was inoculated in both hNRG1 $\alpha$  EGF domain-treated and untreated wells, with the treated wells receiving 50 ng/mL of hNRG1 $\alpha$  EGF domain. The effect on proliferation was evaluated after 48 h, showing over a three-fold increase in cell number in the cultures treated with the recombinant hNRG1 $\alpha$  EGF domain (Figure 4A). Then, we performed a MTT based assay, which measures the reduction of a tetrazolium component (MTT) into an insoluble formazan product by the mitochondria, indirectly reflecting the number of viable cells. HeLa cells were cultivated with increasing concentrations of recombinant hNRG1 $\alpha$  EGF domain to determine its capability to induce cell proliferation. The growth rate of cultures treated with concentrations as low as 0.05-0.5 ng/mL is significantly higher than the growth rate of control cells (Figure 4B).

The activity of the recombinant hNRG1 $\alpha$  EGF domain was also evaluated regarding cell migration by determining the gap closure rate of a scratch generated in NDFH cell monolayers (Figure 4C-D). The number of cells in the scratch zone was determined in the moment of gap generation and after 48 h. The gap closure rate was significantly higher in the NDFH cultures treated with the recombinant hNRG1 $\alpha$  EGF domain as compared to the control cultures (Figure 4D). These experiments show that the recombinant hNRG1 $\alpha$  can induce both cell proliferation and migration.



**Figure 4.** Analysis of the biological activity of the recombinant hNRG1 $\alpha$  EGF domain on cell proliferation and migration. **(A)** Comparison of the proliferation rate of HeLa cells treated with saturating concentration (50 ng/mL) of recombinant hNRG1 $\alpha$  EGF domain with HeLa cell control cultures. Asterisk indicates  $p < 0.05$ . **(B)** HeLa cell proliferation induced by the hNRG1 $\alpha$  EGF domain as determined by the MTT assay. Triplicate cultures of HeLa cells were treated with the indicated amounts of the hNRG1 $\alpha$  EGF domain, and absorbance at 570 nm was measured after 48 h. The right panel shows the data fitting using a nonlinear regression. Asterisk indicates  $p < 0.05$ . **(C)** Analysis of the EGF-domain of hNRG1 $\alpha$  activity on cell migration based on a scratch closure assay. The area of the wound in the moment of scratch generation was compared with the area after 48 h of incubation with 50 ng/mL of hNRG1 $\alpha$  EGF domain. Control samples correspond to the same experimental conditions in absence of hNRG1 $\alpha$  EGF domain. The assays were performed in triplicate. **(D)** Representative microscopy images of the scratch closure assay. The images at 100 X magnification were acquired at time 0 from native cells and after 48h from fixed cells stained with violet crystal. The vertical dotted lines indicate the border of the gap at time 0. Scale bar: 100  $\mu\text{m}$ .

## Discussion

Neuregulins comprise a family of growth factors which, together with their ErbB receptors are involved in numerous neurodevelopmental processes and nervous system activities (for reviews see [2–4, 22]). NRG1 is expressed in neuronal and non-neuronal cell types and is responsible for cell maintenance, proliferation and migration [2, 3]. NRG2 promotes neuronal survival [23] and NRG3 has been associated with schizophrenia [24]. Six neuregulin genes have been identified in human cells, each given rise to multiple isoforms originating from alternative splicing. Most neuregulins are synthesized as transmembrane proteins, which subsequently undergo proteolytic processing releasing the EGF signaling domain with different N-terminal domains [2, 3, 6]. The EGF domain of neuregulins was shown to be sufficient to activate ErbB receptors in a specific manner [8] and induce cellular responses *in vitro* [7].

The EGF domain comprises six cysteine residues separated by a constant number of residues (CX7CX4-5CX10-13CXCX8C) with the six conserved cysteines forming three disulfide bonds in the C1-C3, C2-C4, and C5-C6 combination [15]. Previous studies have described the expression in *E. coli* of the EGF domains fused only to affinity purification tags for some members of the EGF family [12–14]. However, the downside of such an approach was the limited solubility of the final product, and therefore, the necessity to purify inclusion bodies and to perform refolding *in vitro* to obtain a biologically active protein. A major reason to avoid larger N- and C-terminal tags is because they can cause a steric hindrance during the ligand-receptor interaction. To overcome both solubility and large-size tag problems, it is possible to express the protein of interest in fusion with chaperone

proteins and including cleavage sites for specific proteases in the linker connecting the two proteins for separation of the protein of interest after purification. In this work, we have used the pET32a expression vector with the EGF domain of hNRG1 $\alpha$  fused to the C-terminus of thioredoxin-A (TrxA). Upstream of the hNRG1 $\alpha$  EGF domain, the genetic construction includes cleavage sites for enterokinase and thrombin proteases and a hexa-histidine tag. As expected, TrxA has a strong contribution for high levels of soluble expression (Figure 1), which indicates also proper folding of the recombinant fusion protein. To avoid subsequent interference of the TrxA moiety, the fusion protein was separated by digestion with thrombin and TrxA, which keeps the hexa-histidine tag, was removed by affinity using an immobilized nickel column. This strategy turned out to be very efficient for production of soluble hNRG1 $\alpha$  EGF domain.

Conformational stability and correct pairing of cysteines forming disulfide bonds were evaluated using several different approaches. Narhi *et al* [25] demonstrated that when EGF is reduced, it goes through a considerable conformational change. Thus, when submitted to a reduced milieu these conformational changes can possibly lead to an easier access to the proteolytic sites by trypsin, which was confirmed by the comparison of the degradation pattern presented by the hNRG1 $\alpha$  EGF domain in presence and absence of a reducing agent. A higher resistance to proteolytic digestion in absence of reducing agent is an indication of a compact conformation. Dynamic light scattering analyses in presence and absence of reducing agent revealed only small changes in the hydrodynamic radius, which increased from 2.4 nm to 2.7 nm. The hydrodynamic behavior of the hNRG1 $\alpha$  EGF domain did not change even after long incubation times in presence of reducing agent neither after submission to a thermal ramp from 20°C to 95°C. Conformation changes on the hNRG1 $\alpha$  EGF domain were detected only by circular dichroism at high temperature (Figure 2). These results indicate a highly stable conformation and solubility for the recombinant hNRG1 $\alpha$  EGF domain.

Mapping disulfide bonds is a challenging task due to the reactive nature of thiol groups. In the case of the EGF domain, the compact nature of its folded structure adds up to the difficulties to obtain high quality peptide preparations for mass spectrometry. Nevertheless, peptides containing correct cysteine pairing were detected in high abundance for all three disulfide bonds, further supporting the evidence that the recombinant hNRG1 $\alpha$  EGF domain produced in this work is properly folded.

After obtaining evidence that the hNRG1 $\alpha$  EGF domain presents proper folding, we assessed its biological activity by determining its ability to induce cellular proliferation and migration cell culture assays. Both the proliferation rate determined by direct cell count and cell viability determined by MTT assays revealed a steady increase in cell growth. The data obtained in this work

agree with previous observations where members of the EGF family recombinantly produced were shown to positively induce proliferation processes in several different cell lineages in the same concentration range [26–28]. In a similar fashion, scratch gap closure assays were performed to measure the migratory capacities of the cells treated with saturating amounts of the hNRG1 $\alpha$  EGF domain. Gap closure assays are useful to determine cell migration although in the case of fibroblast lineages such as NDFH (normal human dermal fibroblasts), proliferation and migration are usually linked. These assays showed that the hNRG1 $\alpha$  EGF domain significantly induces faster gap closure than untreated control cells. Together, these assays show that the recombinant hNRG1 $\alpha$  EGF domain presents the expected biological activity on cell proliferation and migration.

In addition to its role in the development and activity of the nervous system, NRG1 has been shown to play a central function in other processes including in nervous system injury recovery and repair [29](Alizadeh et al, 2017; Kataria et al 2019), cardiac remodeling [30] and on the proliferation, migration, and angiogenesis of human periodontal ligament stem cells [10](Li et al., 2022). The pathways underlining the processes modulated by NRG1/ErbB signaling are currently being investigated using reconstituted *in vitro* systems. In this context, the recombinant hNRG1 $\alpha$  EGF domain can be a handy tool for studies designed to fill the molecular gaps between the step of ligand-receptor interaction up to the final biological outcome.

In summary, the strategy adopted in this work, combining high expression mediated by an expression chaperone with proteolytic separation of the chaperone consists in an efficient procedure to obtain the recombinant hNRG1 $\alpha$  EGF domain, displaying correct disulfide bonds and functional properties similar to those described for its biologically functional counterparts. This recombinant growth factor can be readily produced to be used in functional experiments intended to expand the knowledge of the NRG1-ErbB signaling pathway.

### **Supplementary Information**

The online version contains supplementary material available.

### **Acknowledgements**

The authors acknowledge the FIOCRUZ program of technical platforms for the support provided by the facilities for Protein Purification and Characterization (RPT15A), Mass Spectrometry (RPT02H) and Cytometry (RPT08L).

**Author contributions:**

All authors contributed to the study conception and design. AL and ASF performed protein expression and purification experiments. ASF and BGG performed the biophysical characterizations and data analysis. MB performed mass spectrometry data acquisition and analysis. ASF and PMH performed the activity assays. ASF, BGG, MB, PMH and NITZ analyzed the results. AL and NITZ prepared the first draft of the manuscript. All authors reviewed the manuscript and approved its final version.

**Funding**

This work was partially funded by the ICC-CNPq-PROEP grant number 442320/2019-1. ASF and AL were supported by CNPq (140923/2018-7) and CAPES graduate fellowships, respectively. NITZ and BGG are supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) research career fellowships (304167/2019-3 and 304788/2018-0).

**Data availability:** The datasets and reagents generated during this study are available from the corresponding author on reasonable request.

**Statements & Declarations****Competing Interests**

The authors have no relevant financial or non-financial interests to disclose.

**Ethical approval/ Consent to participate/Consent to publish**

Not applicable

**References**

1. Meyer D, Yamaal T, Garratt A, et al (1997) Isoform-specific expression and function of neuregulin. *Development* 124:3575–3586. <https://doi.org/10.1242/dev.124.18.3575>
2. Mei L, Xiong WC (2008) Neuregulin 1 in neural development, synaptic plasticity and schizophrenia. *Nat Rev Neurosci* 9:437–452. <https://doi.org/10.1038/nrn2392>
3. Mei L, Nave KA (2014) Neuregulin-ERBB signaling in the nervous system and



- neuropsychiatric diseases. *Neuron* 83:27–49. <https://doi.org/10.1016/j.neuron.2014.06.007>
4. Cespedes JC, Liu M, Harbuzariu A, Annette N, et al (2018) Neuregulin in Health and Disease. *Int J Brain Disord Treat* 4:. <https://doi.org/10.23937/2469-5866/1410024>
  5. Falls DL (2003) Neuregulins: Functions, forms, and signaling strategies. *Exp Cell Res* 284:14–30. [https://doi.org/10.1016/S0014-4827\(02\)00102-7](https://doi.org/10.1016/S0014-4827(02)00102-7)
  6. Kao WT, Wang Y, Kleinman JE, et al (2010) Common genetic variation in neuregulin 3 (NRG3) influences risk for schizophrenia and impacts NRG3 expression in human brain. *Proc Natl Acad Sci U S A* 107:15619–15624. <https://doi.org/10.1073/pnas.1005410107>
  7. Pinkas-Kramarski R, Shelly M, Glathe S, et al (1996) Neu differentiation factor/neuregulin isoforms activate distinct receptor combinations. *J Biol Chem* 271:19029–19032. <https://doi.org/10.1074/jbc.271.32.19029>
  8. Carraway KL, Weber JL, Unger MJ, Ledesma J, Yu N, Gassmann M, Lai C (1997) Neuregulin-2, a new ligand of ErbB3/ErbB4-receptor tyrosine kinases. *Nature* 387:512–516. [https://doi.org/10.1007/3-540-29623-9\\_3120](https://doi.org/10.1007/3-540-29623-9_3120)
  9. Zhang D, Sliwkowski MX, Mark M, et al (1997) Neuregulin-3 (NRG3): A novel neural tissue-enriched protein that binds and activates ErbB4. *Proc Natl Acad Sci U S A* 94:9562–9567. <https://doi.org/10.1073/pnas.94.18.9562>
  10. Jones JT, Akita RW, Sliwkowski MX (1999) Binding specificities and affinities of egf domains for ErbB receptors. *FEBS Lett* 447:227–231. [https://doi.org/10.1016/S0014-5793\(99\)00283-5](https://doi.org/10.1016/S0014-5793(99)00283-5)
  11. Rosano GL, Ceccarelli EA (2014) Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Front Microbiol* 5:1–17. <https://doi.org/10.3389/fmicb.2014.00172>
  12. Mautino B, Costa LD, Gambarotta G, et al (2004) Bioactive recombinant neuregulin-1, -2, and -3 expressed in *Escherichia coli*. *Protein Expr Purif* 35:25–31. <https://doi.org/10.1016/j.pep.2003.12.012>
  13. Abdull Razis AF, Ismail EN, Hambali Z, et al (2008) Expression of recombinant human epidermal growth factor in *Escherichia coli* and characterization of its biological activity. *Appl Biochem Biotechnol* 144:249–261. <https://doi.org/10.1007/s12010-007-8019-9>
  14. Song IJ, Ikram M, Subhan F, et al (2015) Molecular characterization and expression analysis of mouse epidermal growth factor-like domain 8. *Int J Mol Med* 36:541–550. <https://doi.org/10.3892/ijmm.2015.2252>
  15. Schneider MR, Wolf E (2009) The epidermal growth factor receptor ligands at a glance. *J Cell Physiol* 218:460–466. <https://doi.org/10.1002/jcp.21635>
  16. Nagata K, Kohda D, Hatanaka H, et al (1994) Solution structure of the epidermal growth factor-like domain of heregulin- $\alpha$  a ligand for p180(erbB-4). *EMBO J* 13:3517–3523. <https://doi.org/10.1002/j.1460-2075.1994.tb06658.x>
  17. Lobstein J, Emrich CA, Jeans C, et al (2012) SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb Cell Fact* 11:1–16. <https://doi.org/10.1186/s12934-016-0512-9>
  18. Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166:557–580. [https://doi.org/10.1016/S0022-2836\(83\)80284-8](https://doi.org/10.1016/S0022-2836(83)80284-8)

19. Feoktistova M, Geserick P, Leverkus M (2016) Crystal violet assay for determining viability of cultured cells. *Cold Spring Harb Protoc* 2016:343–346. <https://doi.org/10.1101/pdb.prot087379>
20. Abràmoff MD, Magalhães PJ, Ram SJ (2005) Image processing with ImageJ Part II. *Biophotonics Int* 11:36–43
21. LaVallie ER, DiBlasio EA, Kovacic S, et al (1993) A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Bio/Technology* 11:187–193. <https://doi.org/10.1038/nbt0293-187>
22. Buonanno A (2010) 基因的改变 NIH Public Access. *Brain Res Bull* 83:122–131. <https://doi.org/10.1016/j.brainresbull.2010.07.012>.The
23. Kim JA, Jayabalan AK, Kothandan VK, et al (2016) Identification of Neuregulin-2 as a novel stress granule component. *BMB Rep* 49:449–454. <https://doi.org/10.5483/BMBRep.2016.49.8.090>
24. Wang YC, Chen JY, Chen ML, et al (2008) Neuregulin 3 Genetic Variations and Susceptibility to Schizophrenia in a Chinese Population. *Biol Psychiatry* 64:1093–1096. <https://doi.org/10.1016/j.biopsych.2008.07.012>
25. NARHI LO, ARAKAWA T, MCGINLEY MD, et al (1992) Circular dichroism of reduced and oxidized recombinant human epidermal growth factor. *Int J Pept Protein Res* 39:182–187. <https://doi.org/10.1111/j.1399-3011.1992.tb00786.x>
26. Mvula B, Moore TJ, Abrahamse H (2010) Effect of low-level laser irradiation and epidermal growth factor on adult human adipose-derived stem cells. *Lasers Med Sci* 25:33–39. <https://doi.org/10.1007/s10103-008-0636-1>
27. Yoneda N, Sato Y, Kitao A, et al (2011) Epidermal growth factor induces cytokeratin 19 expression accompanied by increased growth abilities in human hepatocellular carcinoma. *Lab Invest* 91:262–272. <https://doi.org/10.1038/labinvest.2010.161>
28. Inatomi O, Andoh A, Yagi Y, et al (2006) Regulation of amphiregulin and epiregulin expression in human colonic subepithelial myofibroblasts. *Int J Mol Med* 18:497–503. <https://doi.org/10.3892/ijmm.18.3.497>
29. Kataria H, Karimi-Abdolrezaee S (2017) Neuregulin-1: A novel regulator of glial response in spinal cord injury. *Neural Regen Res* 12:1616–1617. <https://doi.org/10.4103/1673-5374.217331>
30. Dugaucquier L, Feyen E, Mateiu L, et al (2020) The role of endothelial autocrine NRG1/ERBB4 signaling in cardiac remodeling. *Am J Physiol - Hear Circ Physiol* 319:H443–H455. <https://doi.org/10.1152/ajpheart.00176.2020>



**Supplementary Table S1.** Abundance of the disulfide bond peptides.

Site	MS Area
<b>C69,C78</b>	2.71E+09
C39/C67	4.86E+08
C47,C53	3.22E+08
C39/~C39,~C47,~C53	1.28E+08
C53/C47	1.08E+08
C53/C53	6.91E+07
C39/~C69,~C78/C39	5.38E+07
<b>C39/C53</b>	3.42E+07
~C69,~C78/~C47,~C53	1.69E+07
C67/C67	6.28E+06
C39/~C47,~C53,~C67	6.17E+06
<b>C47/C67</b>	5.22E+06
C39/~C69,~C78/~C67,~C69,~C78	4.19E+06
C47/~C67,~C69,~C78	2.29E+06
~C67,~C69,~C78	1.21E+06
~C39,~C47/~C53,~C67,~C69,~C78	8.67E+05
C53/~C47,~C53/~C67,~C69,~C78	4.87E+05
C53/~C69,~C78/~C69,~C78/~C39,~C47	3.41E+05
~C69,~C78/~C47,~C53/~C69,~C78	2.70E+05
C53/C67	2.49E+05
C47/~C39,~C47,~C53	1.12E+05
Total	3.96E+09

Table shows all the peptides identified by MS2 containing disulfide bonds. The MS area is the sum of areas from nano LC EThcD and ESI LC HCD analysis. ~ represents possible localization of disulfide bonds.

#### 5.1.4. Discussão presente no manuscrito submetido à revista *Molecular Biology Reports*.

As neurogulinas compreendem uma família de fatores de crescimento que, juntamente com seus receptores ErbB, estão envolvidos em vários processos de desenvolvimento neurológico e atividades do sistema nervoso. NGR1 é expressa em tipos de células neuronais e não neuronais e é responsável pela manutenção, proliferação e migração celular (MEI; NAVE, 2014; MEI; XIONG, 2008).

A NRG2 promove a sobrevivência neuronal (KIM et al., 2016) e a NRG3 tem sido associado à esquizofrenia (WANG et al., 2008). Seis genes de neuregulina foram identificados em células humanas, cada um dando origem a múltiplas isoformas originárias de *splicing* alternativo. A maioria das neurogulinas são sintetizadas como proteínas transmembranares, que posteriormente, sofrem processamento proteolítico liberando o domínio de sinalização EGF com diferentes domínios N-terminais (KAO et al., 2010; MEI; NAVE, 2014; MEI; XIONG, 2008).

O domínio EGF das neurogulinas mostrou-se suficiente para ativar os receptores ErbB de maneira específica (CARRAWAY et al., 1997) e induzir respostas celulares *in vitro* (PINKAS-KRAMARSKI et al., 1996). O domínio EGF compreende seis resíduos de cisteína separados por um número constante de resíduos (CX7CX4-5CX10-13CXCX8C) com as seis cisteínas conservadas formando três ligações dissulfeto na combinação C1-C3, C2-C4 e C5-C6 (FRIEDEL et al., 2015).

Estudos anteriores descreveram a expressão em *E. coli* dos domínios EGF fundidos a marcadores de purificação de afinidade para alguns membros da família EGF (ABDULL RAZIS et al., 2008; MAUTINO et al., 2004; SONG et al., 2015). No entanto, a desvantagem de tal abordagem foi a solubilidade limitada do produto final e, portanto, a necessidade de purificar os corpos de inclusão e realizar o redobramento *in vitro* para obter uma proteína biologicamente ativa.

Uma das principais razões para evitar etiquetas N- e C-terminais maiores é porque eles podem causar um impedimento estérico durante a interação ligante-receptor. Para superar os problemas de solubilidade e de etiquetas de tamanho grande, é possível expressar a proteína de interesse em fusão com proteínas chaperonas, incluindo sítios de clivagem para proteases específicas no ligante que conecta as duas proteínas para separação da proteína de interesse após purificação.

Neste trabalho, utilizamos o vetor de expressão pET32a com o domínio EGF de hNRG1 $\alpha$  fundido ao C-terminal da tiorredoxina-A (TrxA). A montante do domínio EGF hNRG1 $\alpha$ , a construção genética inclui sítios de clivagem para proteases de enteroquinase e trombina e um marcador de hexa-histidina. Como esperado, TrxA tem uma forte contribuição para altos níveis de expressão solúvel, o que indica também o enovelamento adequado da proteína de fusão recombinante. Para evitar a interferência subsequente da porção TrxA, a proteína de fusão foi separada por digestão com trombina e TrxA, que mantém o marcador hexa-histidina, foi removida por afinidade usando uma coluna de níquel imobilizada.

Esta estratégia mostrou-se muito eficiente para a produção do domínio solúvel hNRG1 $\alpha$  EGF. A estabilidade conformacional e o emparelhamento correto de cisteínas formando pontes dissulfeto foram avaliados usando várias abordagens diferentes. Foi demonstrado que quando o EGF é reduzido, ele sofre uma considerável mudança conformacional (NARHI et al., 1992). Assim, quando submetidas a um meio reduzido, essas alterações conformacionais possivelmente podem levar a um acesso mais fácil aos sítios proteolíticos pela tripsina, o que foi confirmado pela comparação do padrão de degradação apresentado pelo domínio hNRG1 $\alpha$  EGF na presença e ausência de um agente redutor.

Uma maior resistência à digestão proteolítica na ausência de agente redutor é uma indicação de uma conformação compacta. Análises de espalhamento dinâmico de luz na presença e ausência de agente redutor revelaram apenas pequenas mudanças no raio hidrodinâmico, que aumentou de 2,4 nm para 2,7 nm. O comportamento hidrodinâmico do domínio EGF hNRG1 $\alpha$  não se alterou mesmo após longos tempos de incubação na presença de agente redutor nem após submissão a uma rampa térmica de 20°C a 95°C. Alterações de conformação no domínio EGF hNRG1 $\alpha$  foram detectadas apenas por dicróismo circular em alta temperatura. Estes resultados indicam uma conformação e solubilidade altamente estáveis para o domínio recombinante hNRG1 $\alpha$  EGF.

Mapear ligações dissulfeto é uma tarefa desafiadora devido à natureza reativa dos grupos tiol. No caso do domínio EGF, a natureza compacta de sua estrutura dobrada soma-se às dificuldades de obter preparações peptídicas de alta qualidade para espectrometria de massa. No entanto, peptídeos contendo o emparelhamento correto de cisteína foram detectados em alta abundância para todas as três ligações dissulfeto, apoiando ainda mais a evidência de que o domínio EGF hNRG1 $\alpha$  recombinante produzido neste trabalho é devidamente dobrado.

Após obter evidências de que o domínio EGF hNRG1 $\alpha$  apresenta dobramento adequado, avaliamos sua atividade biológica determinando sua capacidade de induzir proliferação celular e ensaios de cultura de células de migração. Tanto a taxa de proliferação determinada por contagem direta de células quanto a viabilidade celular determinada por ensaios de MTT revelaram um aumento constante no crescimento celular.

Os dados obtidos neste trabalho concordam com observações anteriores onde os membros da família EGF produzidos de forma recombinante mostraram induzir positivamente processos de proliferação em várias células diferentes.

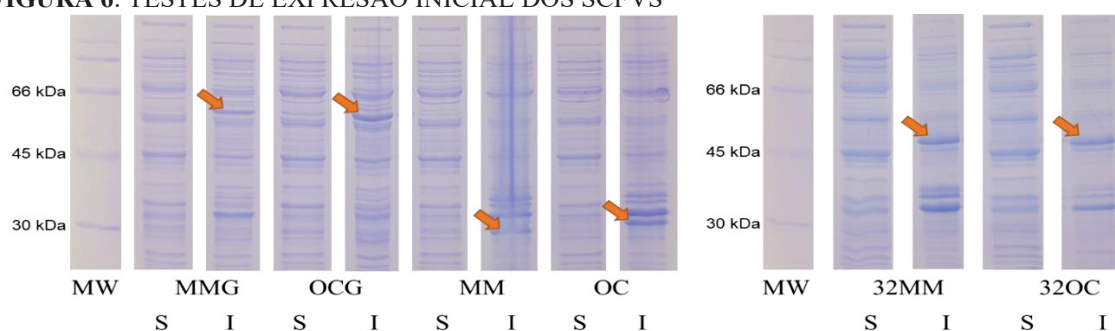
### **5.1.5. Cristalização da NRG1 $\alpha$**

Os ensaios de cristalização da NRG1 $\alpha$  resultaram em precipitados amorfos, que é um estado resultante de um estado supersaturado desorganizado, gota clara, a qual é ausência de supersaturação, e separação de fases. As placas foram observadas no dia 8 e no dia 110 após o preparo, não havendo mudanças nos poços. Sendo assim, decidiu-se por interromper as tentativas de cristalização até ter-se um parceiro de interação, uma vez que quase todas as estruturas resolvidas com o domínio EGF foram co-cristalizadas com parceiros de interação.

## **5.2. ANTICORPOS**

### **5.2.1. Produção inicial dos scFvs anti-EGF**

Os testes de expressão iniciais podem ser observados na FIGURA 6. Os scFvs recombinantes foram observados majoritariamente nas frações insolúveis com baixos níveis de expressão solúvel nas condições iniciais testadas. Para determinar as melhores condições de expressão, optou-se por seguir apenas com as construções OCG e MMG fusionadas a SGFP2, pelo desejo inicial de utilizar essas construções em ensaios de interação.

**FIGURA 6. TESTES DE EXPRESSÃO INICIAL DOS SCFVS**

Fonte: O próprio. SDS-PAGE corado com azul de Coomassie. MW: Marcador de massa molecular. S: Solúvel. I: Insolúvel. Setas laranja: bandas esperadas para as proteínas de interesse.

### 5.2.2. Otimização das condições de expressão dos scFvs fusionados a SGFP2

Para otimizar as condições iniciais, foi utilizada a estratégia de QbD OFAT, seguida da combinação das melhores condições a fim de se alcançar um maior rendimento de expressão solúvel. Como controle de expressão foi utilizada a proteína OPN fusionada a SGFP2, aqui chamada de OPN-SGFP2.

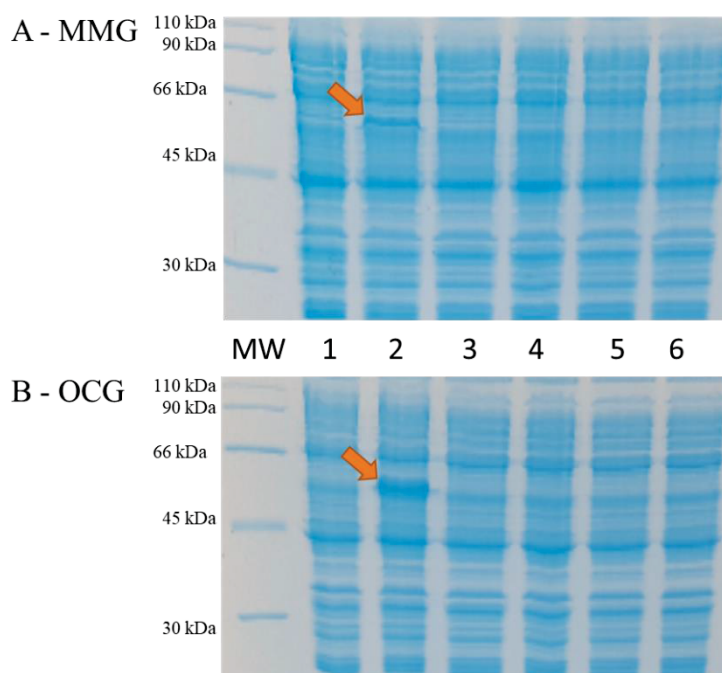
Inicialmente, foram testadas diferentes composições de tampões de lise para determinar qual teria a maior rendimento das proteínas de interesse na fração solúvel. Dentre as condições testadas foram um tampão salino sem aditivos (50 mM Tris pH 8 300 mM NaCl) e o mesmo tampão com aditivos, tais como 0,2% sarcosil, 0,5% SDS, 0,6M ureia, 10% glicerol e 0,1% NP-40. Sarcosil é um surfactante aniônico utilizado para impedir que proteínas solúveis do citoplasma interajam com proteínas de membrana no momento de lise, tornando-as insolúveis em co-agregados (FRANKEL; SOHN; LEINWAND, 1991). Assim como o sarcosil, SDS também é um surfactante aniônico que tem propriedades anfifílicas e interage com proteínas a formar complexos carregados negativamente, tendo capacidade tanto intramolecular, quanto intermolecular, de desfazer interações hidrofóbicas, iônicas e de hidrogênio, dependendo da sua concentração e da proteína (HOU; HE; WANG, 2020; SCHMID et al., 2017). Ureia é um agente caotrópico, capaz de desfazer interações hidrofóbicas e mais eficientemente, pontes de hidrogênio (HOU; HE; WANG, 2020). Já o glicerol é um co-solvente, capaz de inibir agregação proteica no processo de enovelamento ou re-enovelamento de proteínas, além de ajudar a estabilizar proteínas já enoveladas (VAGENENDE; YAP; TROUT, 2009). Por fim, NP-40 é um surfactante não iônico não-desnaturante, comumente utilizado em tampões de lise por seu efeito solubilizante (JI, 2010). A melhor condição testada foi do tampão de lise contendo 0,5% de SDS para MMG e OCG, como visto na FIGURA 7.

Não houve maior ou menor extração de nenhuma das proteínas nativas no extrato solúvel, evidenciando uma influência direta na obtenção dos scFvs frente as outras condições.

A seguir, diversas condições de expressão foram testadas em comparação com a condição padrão (Meio LB. D.O. no momento da indução: 0,6. Quantidade de IPTG: 0,5 mM. Temperatura de expressão: 16 °C. Tempo de expressão: 20 h), utilizando-se do tampão de lise com 0,5% de SDS para a extração das proteínas solúveis.

As variáveis mudadas foram a composição do meio de cultura, seja na fonte de carbono, fonte de nitrogênio, quantidade e tipos de sais, D.O. no momento da indução, quantidade do agente indutor IPTG, temperatura de expressão, tempo de expressão e tentativa de co-expressão com o ligante.

**FIGURA 7. TESTES DE DIFERENTES TAMPÃO DE LISE**



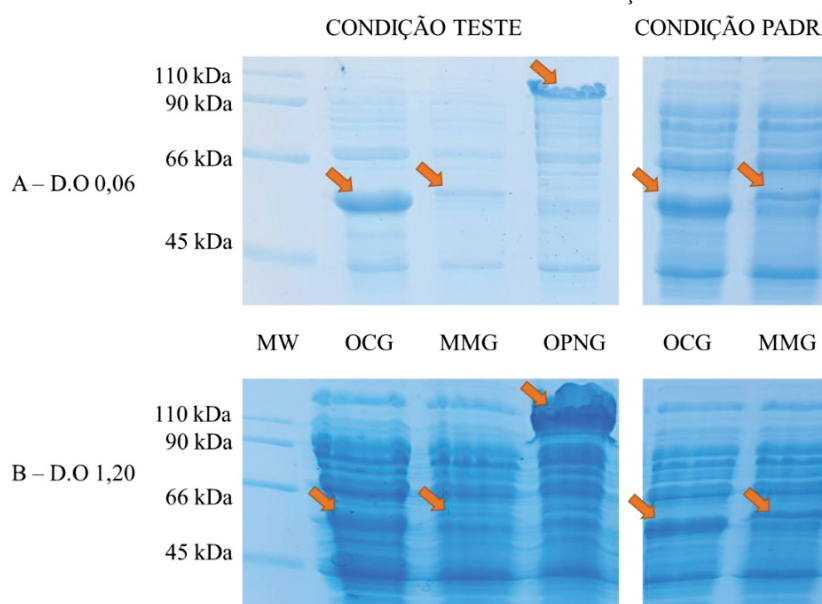
Fonte: O próprio. SDS-PAGE corado com azul de Coomassie das frações solúveis. MW: Marcador de massa molecular. 1: Tampão contendo 0,2% de sarcosil. 2: Tampão contendo 0,5% de SDS. 3: Tampão contendo 0,6 M de ureia. 4: Tampão contendo 10% glicerol. 5: Tampão contendo 0,1% de NP-40. 6: Tampão sem aditivos.

A composição do meio de cultura tem grande impacto na expressão de proteínas recombinantes, seja alterando a quantidade ou o tipo de fonte de carbono, nitrogênio e sais (CUI et al., 2006; MARINI et al., 2014; ROSANO; CECCARELLI, 2014; TSENG; LENG, 2012; VAZ et al., 2011), essa mesma lógica se aplica na substituição completa dos ingredientes na mudança de um meio de cultura por outro (ZARSCHLER et al., 2013). Outros fatores de igual relevância na produção de proteínas recombinantes são a D.O. no momento de indução, a qual influencia a massa final de células e consequentemente sua capacidade de esgotar o meio de

cultura de nutrientes para sua sobrevivência e síntese proteica assim como acúmulo de proteínas recombinantes no seu citoplasma (ROSANO; CECCARELLI, 2014), a quantidade de agente indutor, o tempo e temperatura de expressão que estão ligados a cinética e equilíbrio entre expressão, enovelamento e metabolismo celular, assim como a toxicidade e desgaste energético por parte do organismo (FRAGOSO-JIMÉNEZ et al., 2019; KRAMER et al., 1996; MARINI et al., 2014; ROSANO; CECCARELLI, 2014; RYAN et al., 1996).

Nos testes de diferentes D.O. no momento da indução da expressão (FIGURA 8), para a construção OCG a D.O. de 0,06 mostrou um mesmo nível de expressão que a condição padrão, no entanto, com menos bandas de contaminantes, sendo assim, uma condição promissora. Já para a construção MMG, nenhuma das duas D.O. mostrou melhorias quando comparadas com os controles. O controle OPN-SGFP2 foi omitido dos outros géis por não agregar novas informações, no entanto, foi utilizado como controle de produção dos reagentes, extração e corrida eletroforética em todos os experimentos.

**FIGURA 8.** TESTE DE DIFERENTES D.O. NO MOMENTO DA INDUÇÃO DA EXPRESSÃO



Fonte: O próprio. SDS-PAGE corado com azul de Coomassie. Do lado esquerdo, géis das condições testadas cuja indução foi feita a uma D.O. de 0,06 (A) e a uma D.O. de 1,2 (B). Do lado direito, géis da condição padrão (Meio LB. D.O. no momento da indução: 0,6. Quantidade de IPTG: 0,5 mM. Temperatura de expressão: 16 °C. Tempo de expressão: 20 h) para comparação. MW: Marcador de massa molecular. Setas laranjas indicam as proteínas de interesse.

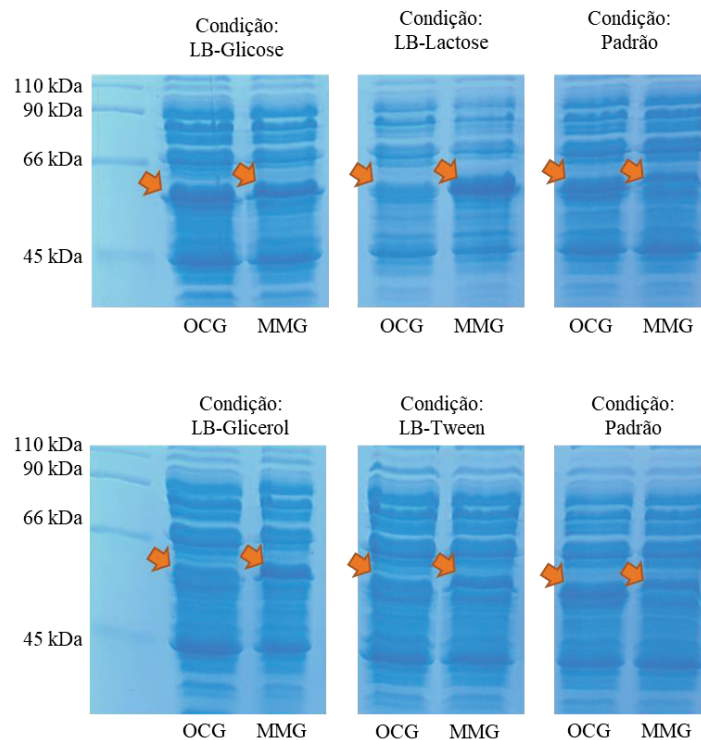
Nos testes de diferentes fontes de carbono (FIGURA 9) verificamos que o scFv OCG teve uma ligeira melhora no meio LB-Glicose, uma piora no meio LB-Lactose e nenhuma diferença relevante nos meios LB-Tween e LB-Glicerol. Meio ricos em glicose podem ter um impacto negativo em *Escherichia coli*, devido a auxiliar na taxa de proliferação, levando a um desbalanço metabólico devido ao acúmulo de acetil-coenzima A, um ácido orgânico tóxico



(FRAGOSO-JIMÉNEZ et al., 2019). No entanto, quando o pH do meio de cultura consegue neutralizar a acidez excessiva gerada pelo subproduto do ciclo do ácido tricarboxílico, isso pode levar a uma melhora energética na produção de proteínas recombinantes, o que pode ser o caso, uma vez que o pH 8 do meio utilizado é considerado o mais adequado para esse intuito (WANG et al., 2014).

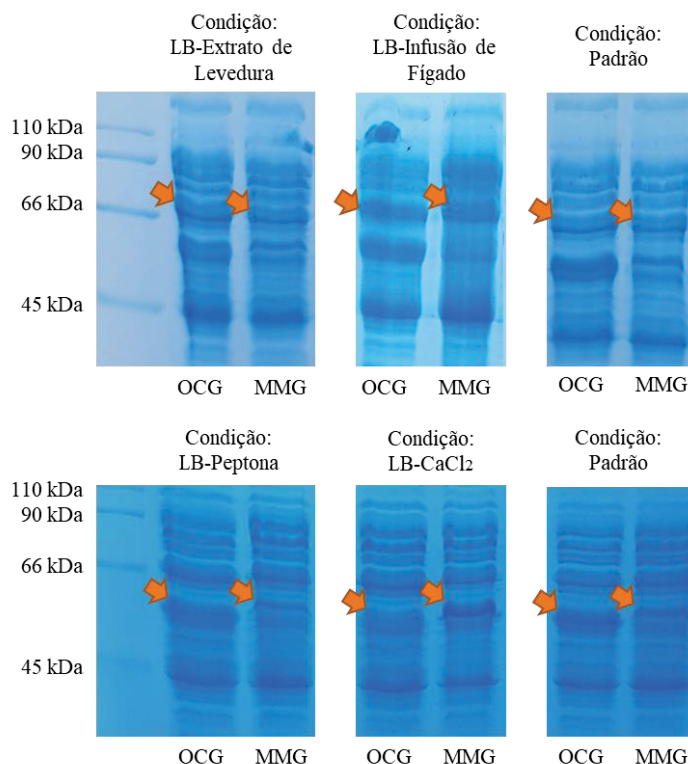
A piora da expressão no meio LB-Lactose pode ser devido ao fato da lactose, além de ser uma fonte de carbono, também atuar se ligando ao operador lac, permitindo assim a expressão de proteínas nativas, assim como a T7 RNA polimerase, aumentando assim a expressão gênica nos plasmídeos e podendo levar a uma sobrecarga de expressão (DVORAK et al., 2015; JAMES et al., 2021). Já o scFv MMG, teve aumento de expressão no meio LB-Glicose, LB-Lactose, LB-Glicerol e LB-Tween, evidenciando assim uma preferência por outras fontes de carbono ao invés da peptona ou mesmo um efeito negativo da peptona na produção deste scFv. A peptona, além de seu papel como fonte de carbono, também tem a função de regulação osmótica, uma vez que é uma fonte rica em prolina, a qual pode ser acumulada intracelularmente para compensar uma alta osmolaridade externa. No entanto, esse efeito também leva a célula a continuar sua proliferação mesmo em meios com alta osmolaridade (AMEZAGA; BOOTH, 1999). Meios com glicerol como fonte de carbono já foram descritos por promover viabilidade de culturas de *Escherichia coli* quando comparados com meios contendo glicose como fonte de carbono, assim como maior produtividade por ativar diferentes vias metabólicas, não ter um efeito catabólito e aumentar o fluxo metabólico em direção a produção de fragmentos de anticorpos recombinantes (KOPP et al., 2018).



**FIGURA 9. TESTE DE DIFERENTES FONTES DE CARBONO NO MEIO DE CULTURA**

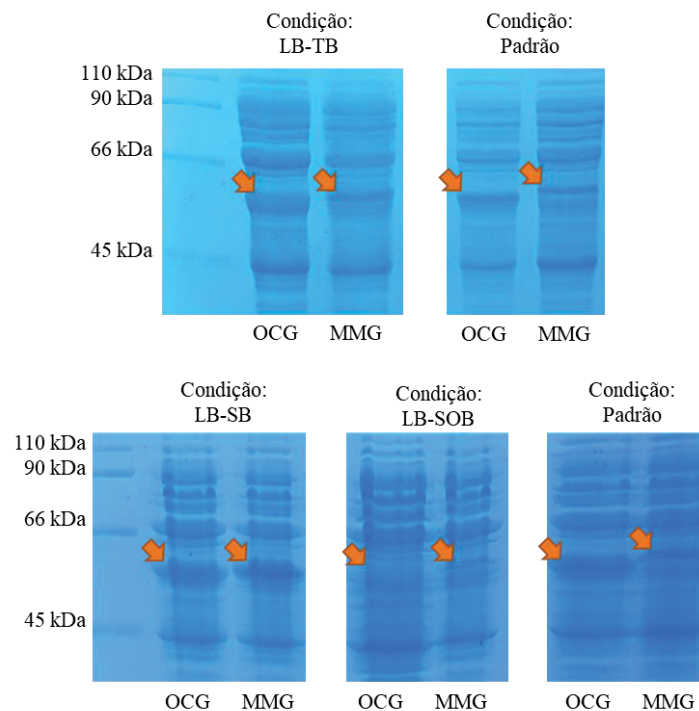
Fonte: O próprio. SDS-PAGE corado com azul de Coomassie. MW: Marcador de massa molecular. Setas laranjas indicam as proteínas de interesse. Condição padrão (Meio LB. D.O. no momento da indução: 0,6. Quantidade de IPTG: 0,5 mM. Temperatura de expressão: 16 °C. Tempo de expressão: 20 h).

Para as diferentes fontes de nitrogênio e uso de um sal diferente do padrão (FIGURA 10), percebemos apenas uma melhora na expressão do scFv MMG no meio LB-CaCl<sub>2</sub>. A presença de cloreto de cálcio no meio de cultura tem efeito sobre a expressão de porinas na superfície da membrana externa da *Escherichia coli*, assim como em ATPases que estão ligadas ao efluxo de componentes intracelulares, podendo assim estar atuando em colaborar com a saída de metabólitos negativos para a produção proteica (HAO et al., 2016; MARTINS et al., 2011). Outra possibilidade é a entrada de DNA do meio de cultura de células mortas para o periplasma, utilizando-se do auxílio do cloreto de cálcio em desestabilizar a membrana externa e/ou pelas porinas e o material nucleico ser usado como fonte alternativa de carbono e nitrogênio, como já demonstrado (CHANG et al., 2017; HUANG et al., 2022).

**FIGURA 10.** TESTE DE DIFERENTES FONTES DE NITROGÊNIO OU SAL NO MEIO DE CULTURA

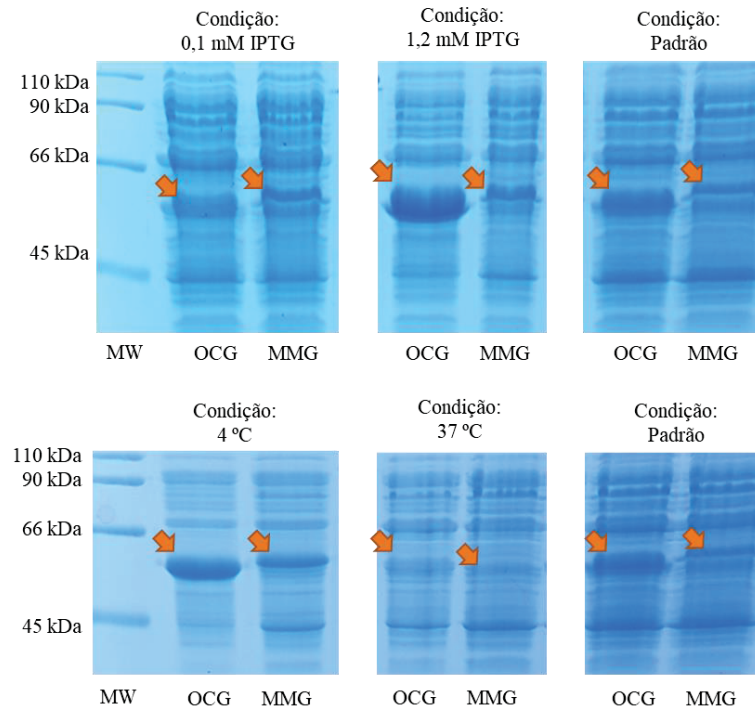
Fonte: O próprio. SDS-PAGE corado com azul de Coomassie. MW: Marcador de massa molecular. Setas laranjas indicam as proteínas de interesse. Condição padrão (Meio LB. D.O. no momento da indução: 0,6. Quantidade de IPTG: 0,5 mM. Temperatura de expressão: 16 °C. Tempo de expressão: 20 h).

Os outros meios comerciais, no entanto, não demonstraram nenhum aumento nos níveis de expressão dos scFvs (FIGURA 11). Sendo assim, a preferência do scFv MMG por diferenças no meio de cultura não são extremas, demonstrando mais uma vez a vantagem da estratégia OFAT, uma vez que a mudança total da composição do meio de cultura não alterou a expressão proteica, assim não se sabe se todos os componentes dos novos meios são neutros para a expressão ou se existe uma soma de fatores positivos e negativos nesses meios. Os meios LB-Ureia e LB-Sulfato de Amônio foram letais para o crescimento bacteriano, o que contradiz a literatura que propõe como fontes de nitrogênio substitutas para *Escherichia coli*, no entanto, não foram encontrados trabalhos com a cepa SHuffle especificamente (LIPNIČANOVÁ et al., 2019; TAABODI et al., 2020). Meios contendo infusão de fígado e peptona como fontes de nitrogênio já foram propostos para expressão de proteínas recombinantes em condições fermentativas, no entanto, não houve efeito positivo nas condições propostas (LONG et al., 2018; SINGH et al., 2017; ZHANG et al., 2022).

**FIGURA 11. TESTE DE DIFERENTES MEIOS COMERCIAIS**

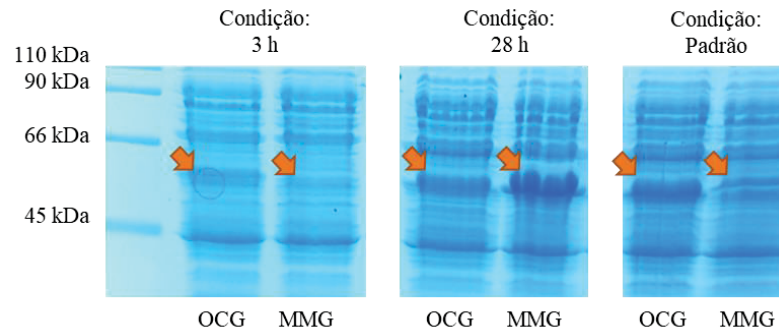
Fonte: O próprio. SDS-PAGE corado com azul de Coomassie. MW: Marcador de massa molecular. Setas laranjas indicam as proteínas de interesse. Condição padrão (Meio LB. D.O. no momento da indução: 0,6. Quantidade de IPTG: 0,5 mM. Temperatura de expressão: 16 °C. Tempo de expressão: 20 h).

Para as diferentes quantidades de IPTG e temperatura (FIGURA 12), vemos um aumento da produção do scFv OCG com maiores quantidades de IPTG e uma melhoria de ambos os scFvs na expressão à 4 °C, uma vez que apesar das bandas de expressão não estarem mais intensas, as bandas de contaminantes estão menos intensas, sendo um efeito positivo ao se considerar o processo de produção de proteínas recombinantes puras. Ambas as variáveis são as mais comumente alteradas no processo de estabelecimento de protocolos de expressão de proteínas recombinantes, não sendo uma surpresa que ambas tiveram grande impacto, visto que a temperatura afeta o metabolismo bacteriano como um todo e a intensidade da indução também afeta as vias de produção e processamento proteicos (AKBARI et al., 2015; FARSHDARI et al., 2020).

**FIGURA 12. TESTE DE DIFERENTES CONCENTRAÇÕES DE IPTG E TEMPERATURAS DE INDUÇÃO**

Fonte: O próprio. SDS-PAGE corado com azul de Coomassie. MW: Marcador de massa molecular. Setas laranjas indicam as proteínas de interesse. Condição padrão (Meio LB. D.O. no momento da indução: 0,6. Quantidade de IPTG: 0,5 mM. Temperatura de expressão: 16 °C. Tempo de expressão: 20 h).

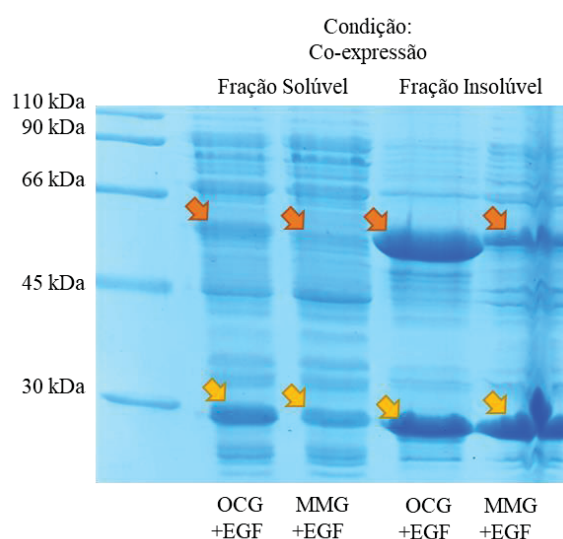
Por fim, a mudança do tempo de expressão (FIGURA 13) se mostrou negativa em ambos os casos com a diminuição do tempo de expressão, demonstrando assim uma necessidade por um tempo mínimo para a sua produção e no caso do scFv MMG, um ganho de expressão com um aumento prolongado do tempo de expressão. Outros casos utilizando-se de metodologias semelhantes para estudo de variáveis também encontraram vantagem em maiores tempos de expressão (AKBARI et al., 2015).

**FIGURA 13. TESTE DE DIFERENTES TEMPOS DE EXPRESSÃO**

Fonte: O próprio. SDS-PAGE corado com azul de Coomassie. MW: Marcador de massa molecular. Setas laranjas indicam as proteínas de interesse. Condição padrão (Meio LB. D.O. no momento da indução: 0,6. Quantidade de IPTG: 0,5 mM. Temperatura de expressão: 16 °C. Tempo de expressão: 20 h).

Para o teste de co-expressão (FIGURA 14), foi utilizado para comparação a fração insolúvel de cada scFv, uma vez que não possuíamos controles para co-expressão estabelecidos. Em ambos os casos, o efeito da expressão da proteína Trx-EGF foi negativo para a produção dos scFvs, evidenciado pelas bandas intensas na fração insolúvel de ambos os testes e principalmente pela grande produção da Trx-EGF, evidenciando uma possível preferência para sua expressão e/ou processamento desta proteína frente ao scFv. Apesar de possível a co-expressão de fragmentos de anticorpos com seus ligantes em *Escherichia coli* (BOSSI et al., 2010), ela não é um método simples e direto para obtenção de fragmentos de anticorpos de maneira solúvel devido à alta complexidade.

**FIGURA 14.** TESTE DE CO-EXPRESSÃO DOS SCFVS COM TRX-EGF



Fonte: O próprio. SDS-PAGE corado com azul de Coomassie. MW: Marcador de massa molecular. Setas laranjas indicam os scFvs. Setas amarelas indicam o Trx-EGF.

O próximo passo foi um refinamento das condições, tentando se combinar diferentes variáveis positivas a fim de explorar um possível efeito sinérgico. Nessa etapa já se iniciou a tentativa de diminuição da quantidade de SDS no tampão de lise, a fim de garantir um menor impacto na utilização dos scFvs no futuro.

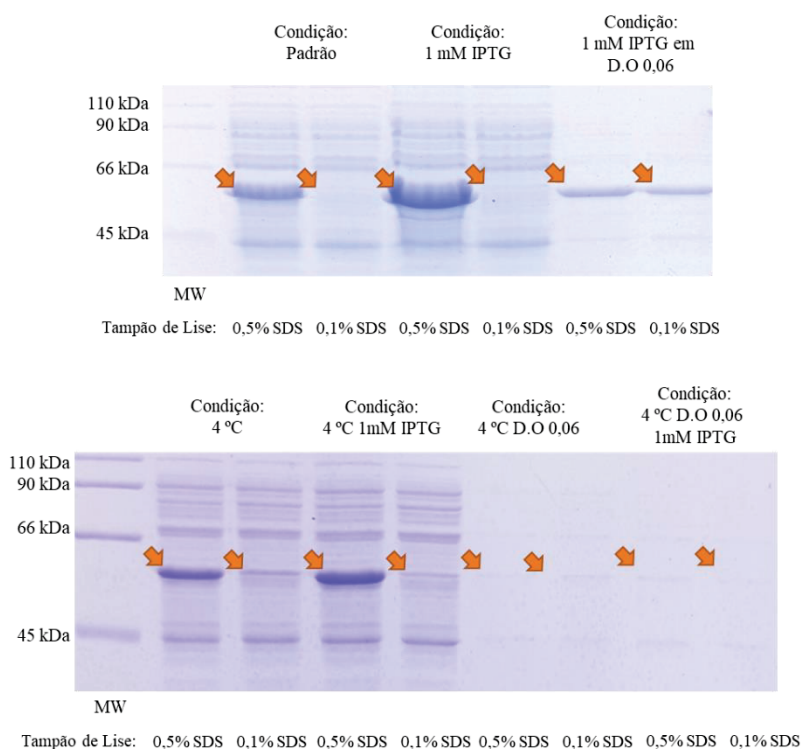
### 5.2.3. Refinamento das condições de expressão dos scFvs fusionados a SGFP2

Para a etapa de refinamento, foram escolhidas para o scFv OCG as variáveis “1 mM IPTG”, “D.O. 0,06” e “4°C” e para o scFv MMG as variáveis “LB-Lactose”, “28 h” e “4°C” e apesar de não ter sido uma das condições iniciais, verificada a melhoria em uma expressão mais longa, também foi introduzida a variável “48 h”.

Também foi testado em paralelo a preparação de frações solúveis com um tampão de lise com 0,1% de SDS em vez de 0,5% na tentativa da retirada do detergente. Foi utilizada a condição padrão para demonstrar o efeito dessa mudança e as variáveis promissoras individuais ou em combinação (FIGURA 15).

Para a combinação das variáveis do scFv OCG “1 mM IPTG” e “D.O. 0,06”, verificamos que a mudança no tampão de lise não teve efeito sobre o produto final (Figura 14), no entanto, devido à baixa expressão proteica dessa combinação não se pode afirmar de fato que o aditivo de fato foi diminuído devido a um provável excesso de SDS mesmo na concentração de 0,1% em tão pouca proteína na fração solúvel. No entanto, vemos claramente que a mudança do aditivo nas condições padrão e na condição promissora de 1 mM de IPTG o desaparecimento da proteína de interesse na fração solúvel.

**FIGURA 15. TESTE DE MUDANÇA DO TAMPÃO DE LISE E COMBINAÇÃO DE VARIÁVEIS PROMISSORAS DO SCFV OCG**



Fonte: O próprio. SDS-PAGE corado com azul de Coomassie. MW: Marcador de massa molecular. Setas laranjas indicam as proteínas de interesse. Condição padrão (Meio LB. D.O. no momento da indução: 0,6. Quantidade de IPTG: 0,5 mM. Temperatura de expressão: 16 °C. Tempo de expressão: 20 h).

Vemos, no entanto, que na condição “4 °C” e na combinação “4°C” e “1 mM IPTG”, a presença da proteína na fração solúvel com 0,1% de SDS, apesar da combinação das variáveis ter diminuído a quantidade proteica solúvel total quando comparada com a variável isolada de “1 mM IPTG”. A combinação dupla de “D.O. 0,06” e “4 °C” e a combinação tripla “D.O. 0,06”,

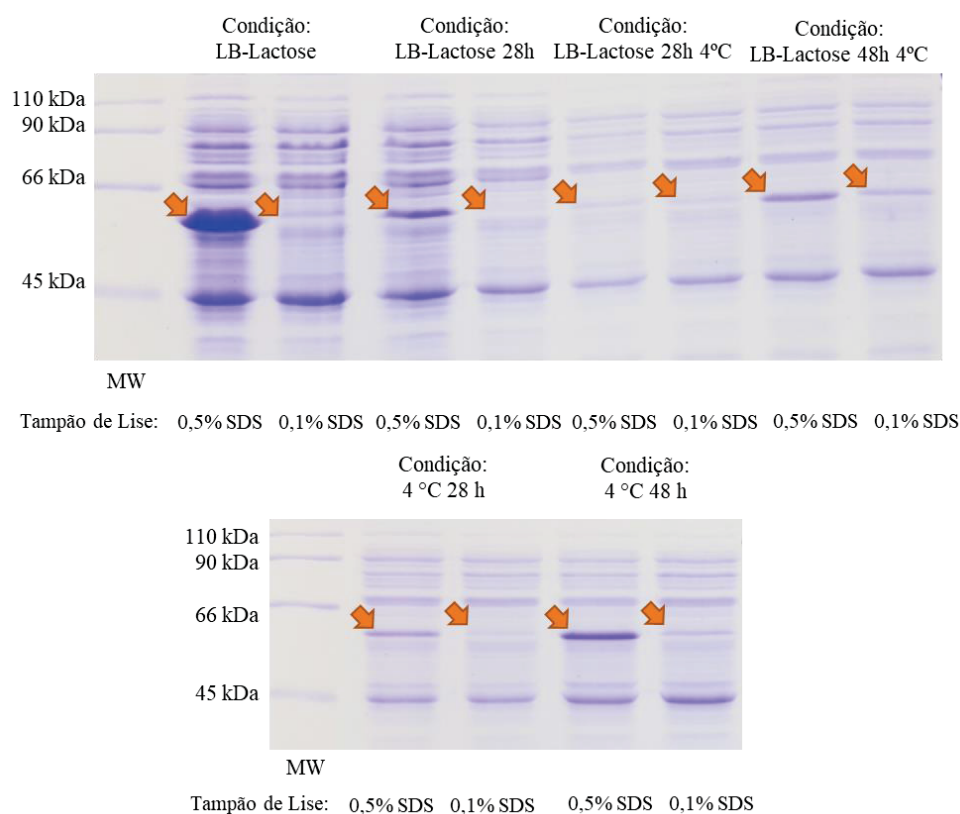


“1 mM IPTG” e “4 °C” se mostraram negativas, muito provavelmente devido a um baixo número de células bacterianas e uma temperatura muito reduzida que impede sua proliferação a um ponto de uma produção recombinante satisfatória.

Sendo assim, a melhor condição para o scFv OCG para o tampão 0,5% de SDS continuou sendo com 1 mM IPTG e no tampão com 0,1% de SDS a expressão a 4 °C.

Para o scFv MMG, o efeito da redução do SDS no tampão de lise também teve um efeito negativo, no entanto, a maior parte das condições continuou tendo uma quantidade de proteína recombinante na fração solúvel (FIGURA 16).

**FIGURA 16.** TESTE DE MUDANÇA DO TAMPÃO DE LISE E COMBINAÇÃO DE VARIÁVEIS PROMISSORAS DO SCFV MMG



Fonte: O próprio. SDS-PAGE corado com azul de Coomassie. MW: Marcador de massa molecular. Setas laranjas indicam as proteínas de interesse.

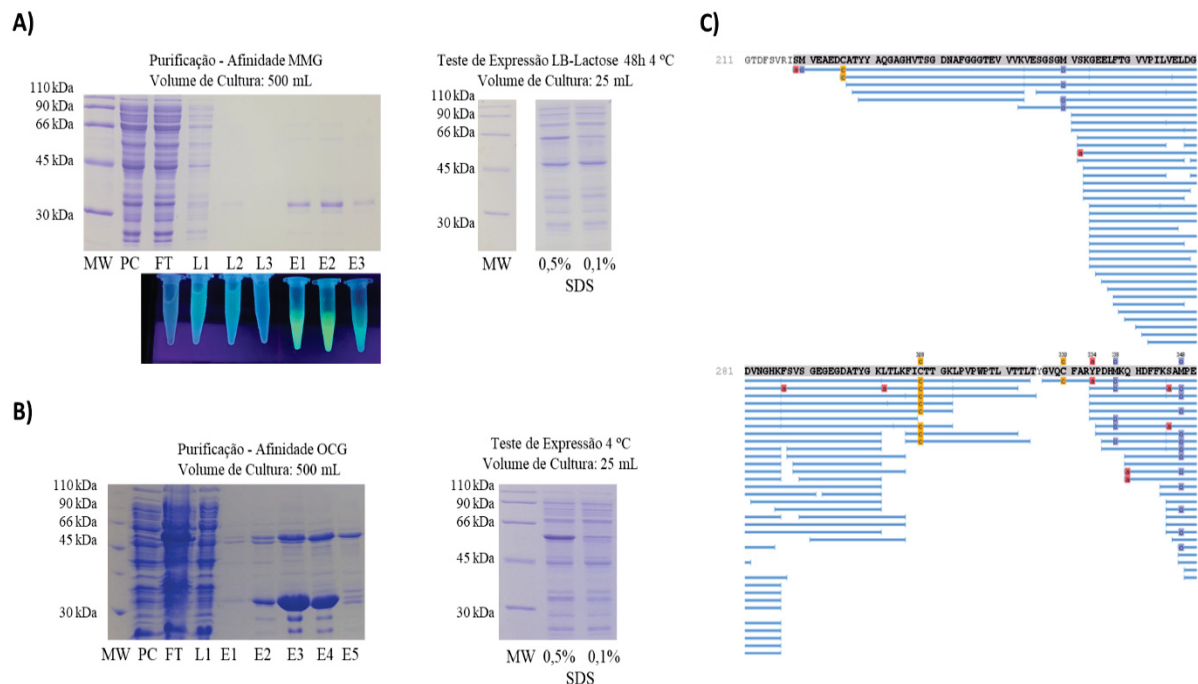
A combinação do meio LB-Lactose com o tempo de expressão 28 h teve um efeito negativo na produção da proteína recombinante, assim como a expressão em meio LB-Lactose por 28 h a 4 °C. No entanto, após 48 h percebeu-se um aumento do nível de expressão e um menor nível de contaminantes, sendo assim uma condição equiparável com o meio LB-Lactose isolado das outras variáveis. Esse mesmo efeito foi evidenciado utilizando-se do meio LB com 28 h e 48 h de expressão a 4 °C. Isso demonstra a complexidade da cinética da produção proteica no contexto do metabolismo bacteriano.

Sendo assim, a melhor condição para o scFv MMG para o tampão 0,5% de SDS foi o LB-Lactose e no tampão com 0,1% de SDS a combinação LB-Lactose, 48 h 4 °C.

#### 5.2.4. Purificação e análise dos scFvs fusionados a SGFP2

Para a próxima etapa, optou-se por começa a padronização da purificação pelo scFv OCG e MMG na condição de 0,1% de SDS e em um volume maior de cultura a fim de ter um rendimento mínimo após a etapa de purificação por afinidade. Inesperadamente, notou-se uma grande quantidade de proteína perto dos 30 kDa com afinidade ao níquel, condizendo com o tamanho que seria esperado de uma SGFP2 livre (FIGURA 17). Essa presença não pode ser notada em nenhum gel anterior ou mesmo no extrato solúvel antes da cromatografia.

**FIGURA 17. PURIFICAÇÃO DOS SCFVS FUSIONADOS A SGFP2 E ANÁLISE DE SGFP2 LIVRE**



Fonte: O próprio. A) SDS-PAGEs corados com azul de Coomassie e frações de eluição de cromatografia expostos a luz.UV. MW: Marcador de massa molecular. PC: Pré-coluna. FT: Fração não ligada a coluna. L1 a L3 = Lavagens. E1 a E3: Eluições. B) SDS-PAGEs corados com azul de Coomassie. MW: Marcador de massa molecular. PC: Pré-coluna. FT: Fração não ligada a coluna. L1 = Lavagem. E1 a E5: Eluições. C) Cobertura de peptídeos do sequenciamento da banda de ~30 kDa da purificação do scFv OCG.

Considerando que a sequência de hexahistidina fica no C-terminal da SGFP2 fusionada ao scFv e a fluorescência alta das frações eluídas da tentativa de purificação do scFv MMG, decidimos analisar por espectrometria de massas as bandas de E3 e E4 cortadas do gel da purificação do scFv OCG para confirmar a sequência da SGFP2 e, se possível, determinar sua extremidade N-terminal. Notamos que a maior parte dos peptídeos identificados na análise



por espectrometria de massas são enriquecidos próximos da metionina da SGFP2, indicando uma grande probabilidade de existir um início de tradução no interior do mRNA que codifica a proteína de fusão recombinante, seja pela própria metionina ou códons alternativos que já foram demonstrados ser utilizados tanto por anticorpos em *Escherichia coli*, quanto por outras proteínas (BAEZ et al., 2019; LEITH et al., 2019; TEN-CATEN et al., 2018). Trabalhos paralelos do nosso grupo com outros scFvs apresentaram o mesmo problema. Sendo assim, decidimos mudar a estratégia de produção dos fragmentos de anticorpos para as construções sem SGFP2 e sem proteínas de fusão.

Para contornar os problemas de expressão, buscamos cepas alternativas de *Escherichia coli* que favorecessem a formação de pontes dissulfeto, etapa crítica para o bom enovelamento dessas moléculas, uma vez que a estratégia de otimização de condições de expressão utilizada para as construções fusionadas a SGFP2 demandaram bastante tempo.

Em paralelo, outras estratégias envolvendo biologia estrutural combinada com engenharia de proteínas foram sendo desenvolvidas caso a mudança de cepa não solucionasse o problema de produção das proteínas de interesse na fração solúvel.

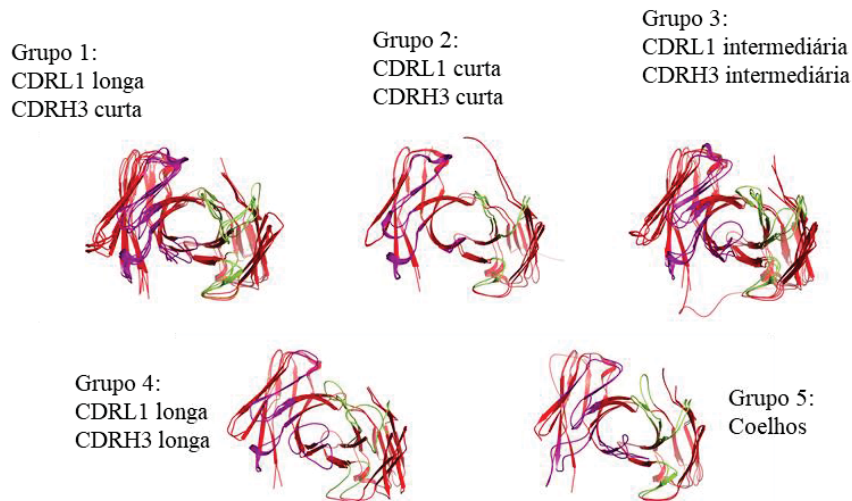
### **5.2.5. Análise estrutural dos scFvs cujas estruturas atômicas estão descritas no PDB**

Obtiveram-se inicialmente 86 resultados da busca utilizando os marcadores booleanos “scFv” e “coli” no PDB, apenas 13 dos quais se enquadravam nas especificações da busca (Quadro 7). Isso se deve em grande parte por muitas das estruturas não terem sido publicadas em artigos científicos contendo a descrição da forma de obtenção do scFv para a resolução da estrutura, por terem os scFvs sido renaturados de corpos de inclusão ou por não terem sido descritos com precisão os métodos de obtenção dos scFvs. Uma pequena parte foi excluída por possuírem ou serem outros tipos de fragmentos de anticorpos.

Para as análises a seguir, além das estruturas selecionadas do PDB, sequências de scFvs expressos na fração solúvel presentes no laboratório também foram incluídas nas análises, sendo eles: 146, 148, 149,1410, CLAU, ANTIEPR, além das sequências dos scFvs MM e OC desenvolvidos neste trabalho. As sequências foram separadas em cadeias pesadas e leves. Em seguida, foram alinhadas utilizando alinhamento estrutural pela plataforma virtual T-coffee expresso. Modelos dos scFvs presentes no laboratório foram gerados pela plataforma Kotai Antibody Builder a fim de realizar a comparação estrutural. Todas as estruturas e os modelos foram analisadas manualmente em busca de características em comum ou singulares.

Os agrupamentos podem ser vistos na FIGURA 18, enquanto o nome de cada estrutura de cada grupo no Quadro 7.

**FIGURA 18.** AGRUPAMENTO DE ESTRUTURAS DE SCFVS



Fonte: O próprio. Em vermelho, as cadeias leves e pesadas. Em roxo, as CDR-H3 e em verde as CDR-L1. Os modelos foram gerados pela plataforma online Kotai Antibody Builder e coloridos no programa PyMol v 4.5.0.

Verificou-se uma diferença de posicionamento das CDRs na região entre as CDRs CDRL1 e CDRH3, sendo esta última a mais importante para reconhecimento dos ligantes de modo geral em anticorpos. A classificação de diferentes estruturas de CDRH3 já foi proposto (KURODA et al., 2008), no entanto, a inclusão da CDRL1 como fator de influência nas estruturas resolvidas por cristalografia nunca foi proposto antes. Os tamanhos destas duas CDRs parecem possuir grupos diferentes de conformidade estrutural, visto que uma parece influenciar a outra na região entre as cadeias leve e pesada. Verificamos também que nosso modelo de scFv OC que é derivado de anticorpo leporínico apresentou estrutura semelhante a outro scFv de origem leporínica, se enquadrando ambos em um grupo separado. Isso pode ser explicado pelo fato de coelhos possuírem anticorpos com características evolutivas singulares quando comparados aos humanos e camundongos. Suas CDRs possuem tamanhos mais diversos e sua conservação do arcabouço diferir mais do que humanos e camundongos entre si (WEBER; PENG; RADER, 2017).

**QUADRO 7. CLASSIFICAÇÃO DE ESTRUTURAS CRISTALOGRÁFICAS E MODELOS EM GRUPOS**

<b>Grupo 1:</b>	<b>Grupo 2:</b>	<b>Grupo 3:</b>	<b>Grupo 4:</b>	<b>Grupo 5:</b>
<b>MM</b>	<b>2GHW</b>	<b>2KH2</b>	<b>2A9M</b>	<b>OC</b>
<b>4NIK</b>	<b>2YBR</b>	<b>4P49</b>	<b>4P48</b>	<b>6CVK</b>
<b>148</b>		<b>4KV5</b>	<b>146</b>	
<b>149</b>		<b>CLAU</b>		
<b>1410</b>				
<b>1KTR</b>				

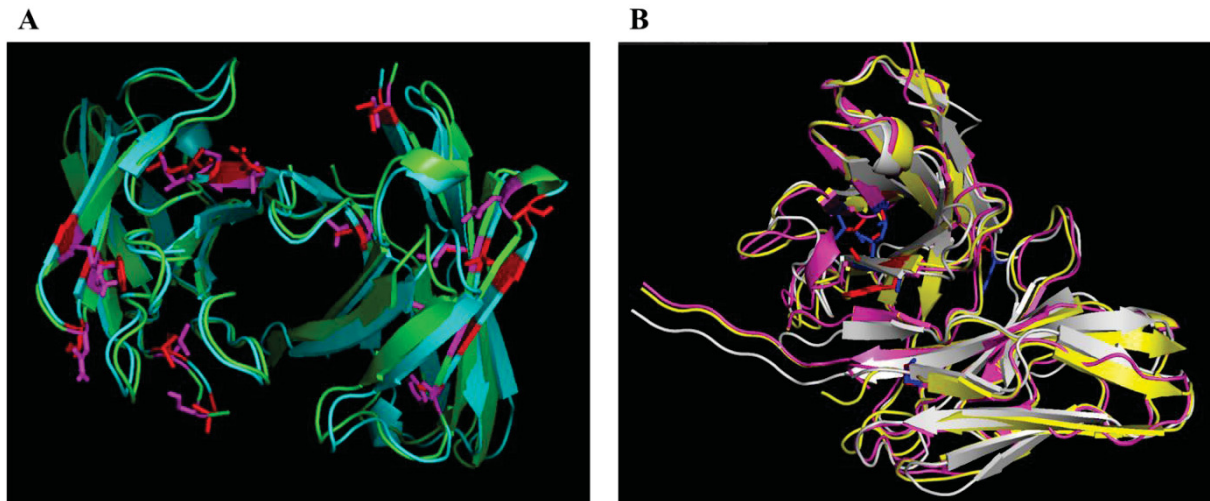
Fonte: O próprio.

Os outros grupos de diferentes posicionamentos das CDRs variaram entre CDRL1/H3 curta, intermediária ou longa em termos de conformação na estrutura cristalizada na região entre as cadeias, não sendo necessariamente relacionado ao tamanho da sequência. Verificou-se a compatibilidade do arcabouço em termos de conservação em sequência dentro das construções agrupadas. Algumas estruturas não conseguiram ser agrupadas com mais de uma estrutura, devido a diferença do tamanho da CDR, sequência ou devido a estrutura em si ser particularmente distinta, assim não foram incluídas em nenhum dos grupos e excluídas da próxima etapa as seguintes estruturas: 1LMK, 3JUY e 4VID.

### **5.2.6. Desenho das construções híbridas**

Decidiu-se então testar construções levando em conta várias características distintas. CDRs das estruturas 148, 146 e 4NIK foram substituídas pelas CDRs do scFv MM. Além disso, uma construção extra foi encomendada substituindo as CDRs da estrutura 4NIK pelas do scFv MM e mudando-se os resíduos da zona de Vernier da estrutura 4NIK pelos do scFv MM. A sobreposição dos modelos pode ser vista na FIGURA 19, enquanto que os alinhamentos das sequências podem ser vistos nas FIGURA 20 e FIGURA 21. Também foi incluída a estrutura do scFv OC para demonstrar a similaridade entre os scFvs de coelho entre si.

**FIGURA 19.** SOBREPOSIÇÃO DOS MODELOS DOS SCFVS COM E SEM ALTERAÇÃO NOS RESÍDUOS DE VERNIER



Fonte: O próprio. A: Sobreposição do scFv 4NIKMM (azul claro) com o scFv 4NIKMM com alterações nos resíduos de vernier (verde), ora nomeado VERMM. Em rosa, os resíduos originais; em vermelho, os alterados. B: Sobreposição do scFv OC (branco), 6CVKOC (rosa) e o scFv 6CVKOC (amarelo) com alterações nos resíduos de vernier, aqui nomeado VEROOC. Os resíduos azuis são os resíduos originais e, em vermelho, os alterados. Os modelos foram gerados pela plataforma online Kotai Antibody Builder e coloridos no programa PyMol v 4.5.0.

A estrutura 4NIK tem origem humana e pertence ao grupo 1, assim como o scFv MM. A alteração dos resíduos da zona de Vernier nessa estrutura se torna compatível com o que já é descrito na literatura (AUBREY; BILLIALD, 2019; KHANTASUP et al., 2015), pois parte da mesma lógica já aplicada a humanização de anticorpos, só que nesse caso, murinizando o anticorpo. Na análise dos resíduos da zona de Vernier, no entanto, o F68 que se torna F68A aparenta fazer interação hidrofóbica com um resíduo fora destes escolhidos para serem alterados. No entanto, visando uma construção solúvel, decidiu-se modificar esse resíduo também pelo originário da estrutura do scFv MM, gerando a modificação M85L.





**FIGURA 21.** ALINHAMENTO ESTRUTURAL DAS SEQUÊNCIAS DOS scFvs DO GRUPO OC



Fonte: O próprio. Triângulos pretos demonstram os resíduos da zona de Vernier que foram alterados entre a construção 6CVKOC e VERO. Asteriscos demonstram resíduos altamente conservados entre as 4 sequências, dois pontos demonstram resíduos com alguma conversação em características físico-químicas próximas e um ponto demonstram resíduos com alguma conservação em termos de tamanho. Em azul, resíduos carregados negativamente. Em rosa, resíduos carregados positivamente. Em vermelho, resíduos apolares. Em verde, resíduos polares.

O modelo 148 tem origem murina e pertence ao grupo 1, sendo o mais próximo do scFv MM. Na experiência no laboratório, ele é um scFv com baixos níveis de expressão na fração solúvel, assim se pode verificar se alterações nas CDRs teriam algum papel no nível de expressão, mesmo em arcabouços de mesmo organismo e tamanhos de CDRs próximos.

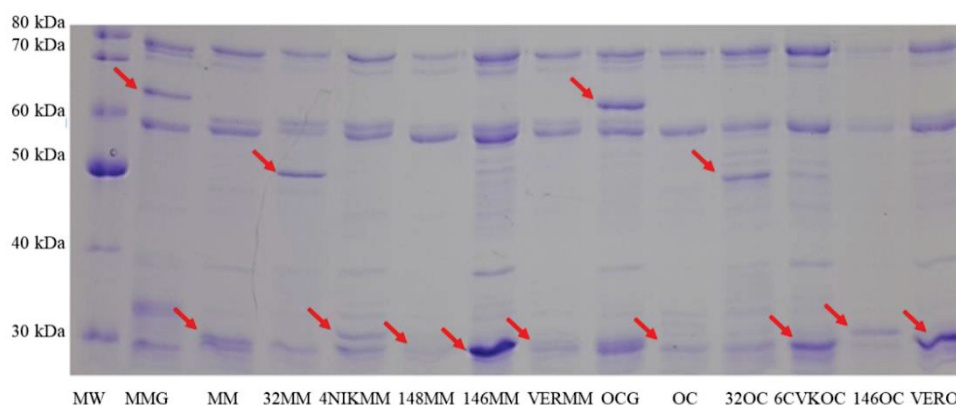
Para o scFv OC, as CDRs das estruturas 146 e 6CVK foram substituídas pelas suas gerando 146OC e 6CVKOC. Além disso, uma construção extra foi encomendada substituindo as CDRs da estrutura 6CVK pelas do scFv OC e mudando-se os resíduos da zona de Vernier da estrutura 6CVK pelos do scFv OC, gerando o scFv VERO.

### 5.2.7. Expressão final dos scFvs

As construções foram, então, testadas na condição padrão (Meio LB. D.O. no momento da indução: 0,6. Quantidade de IPTG: 0,5 mM. Temperatura de expressão: 16 °C. Tempo de expressão: 20 h). No entanto, a cepa de *Escherichia coli* Shuffle, diferentemente da

cepa BL21 (DE3)  $\Delta$ slyD pRARE2, expressa constitutivamente a proteína DsbC, a qual atua como isomerase revertendo oxidações incorretas em pontes dissulfeto e como chaperona, ajudando no enovelamento proteico, já tendo sido demonstrado como capaz de ajudar a expressar scFvs na fração solúvel (AHMADZADEH et al., 2020). O resultado do teste de expressão pode ser visto na **FIGURA 22** e a expressão também foi confirmada por *western blot*, como visto na **FIGURA 23**.

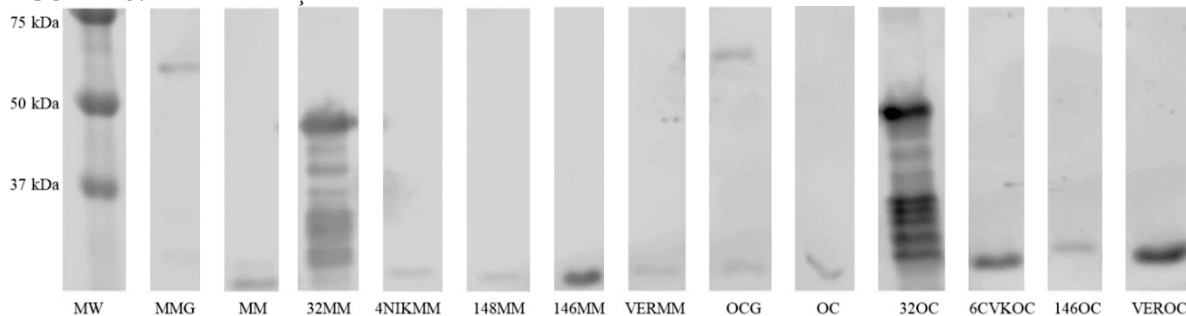
**FIGURA 22. EXPRESSÃO FINAL DOS SCFVS**



Fonte: O próprio. MW: Marcador de massa molecular. Setas vermelhas indicam as proteínas de interesse.

Todas as construções foram expressas na fração solúvel, em diferentes níveis, diferentemente da cepa de *Escherichia coli* BL21 (DE3)  $\Delta$ slyD pRARE2, demonstrando assim mais uma variável importante na escolha do protocolo de obtenção de proteínas recombinantes. A construção com menor rendimento foi a 148MM e a maior a 146MM, ambas construções fora do grupo proposto na análise estrutural, no entanto, a construção 4NIKMM e VERMM não tiveram expressões muito superiores em termos de rendimento e elas estavam nos grupos propostos como compatíveis. As construções fusionadas a Trx mostraram diversas bandas no *western blot*, demonstrando já uma instabilidade inicial, sendo assim, as construções MM e OC foram escolhidas para serem comparadas com as construções híbridas nos ensaios futuros.

**FIGURA 23. CONFIRMAÇÃO DA EXPRESSÃO DOS SCFVS**



Fonte: O próprio. MW: Marcador de massa molecular. Setas vermelhas indicam as proteínas de interesse. Imagem adquirida pelo aparelho LiCor Odyssey Infrared Imaging System (LI-COR Biosciences, Nebraska, EUA).

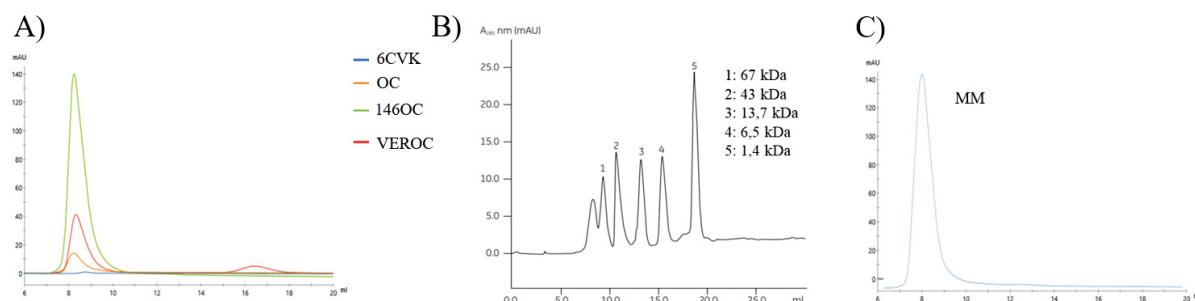


Não houve grande diferença entre as construções com e sem os resíduos da zona de Vernier modificados entre si. O arcabouço 146 teve diferentes níveis de expressão entre as CDRs MM e OC, demonstrando assim realmente uma diferença entre a aceitabilidade com CDRs de tamanhos/espécies diferentes por um mesmo arcabouço, sendo o arcabouço 148 de origem murina assim como as sequências das CDRs MM. O arcabouço 148, assim como com suas CDRs originais, teve um baixo rendimento, se mostrando um arcabouço de pouca ajuda na obtenção de fragmentos de anticorpos na fração solúvel.

### 5.2.8. Purificação dos scFvs

Apenas as construções scFv MM, OC, 146OC, 6CVKOC e VEROOC foram possíveis de serem purificadas. As demais construções se mostraram demasiadamente instáveis por precipitarem ou de baixo rendimento para seguirem até a etapa de cromatografia por exclusão por tamanho. Todas as construções inicialmente eluíram no volume morto da coluna, demonstrando assim serem agregados solúveis (**FIGURA 24**).

**FIGURA 24** CROMATOGRAFIAS DE EXCLUSÃO POR TAMANHO DEMONSTRANDO AGREGADOS SOLÚVEIS



Fonte: Figuras A e C, o próprio. Figura B: Instructions 71-5017-96 AK. Manual da Coluna Superdex 75 10 300 Cytiva. Figura A) Cromatogramas dos scFvs 6CVK, OC, 146OC e VEROOC. Figura B) Padrão de comparação da coluna. Figura C) Cromatograma do scFv MM. Cromatogramas visualizados no programa Unicorn 7.

Testou-se alterar o concentrador utilizado, visando agilizar o processo entre as duas etapas cromatográficas. Sendo assim, foi trocado o concentrador Amicon de 500  $\mu$ L para o de 15 mL. Também se alterou a composição do tampão que inicialmente possuía 50 mM Tris pH 8 e 300 mM NaCl para um tampão com a mesma composição do tampão de cromatografia de exclusão por tamanho dos fatores de crescimento da família EGF, sendo ele composto por 25 mM Tris.HCl pH 8,0 e 150 mM NaCl. Essa escolha foi para simplificar os processos de obtenção dos fragmentos em paralelo com as outras proteínas recombinantes unicamente, visto que tampões com menos sal, em geral, levam a uma maior agregação das proteínas de interesse por permitir uma maior proximidade de regiões hidrofóbicas.

No entanto, nesse caso, a diminuição da quantidade de sais ajudou a solubilização dos scFvs. Duas hipóteses nos ocorreram nesse momento, uma é que a agregação dessas moléculas possa ser mediada pela presença de pontes salinas ou a mudança da quantidade de água livre (GUPTA; DEEP, 2015) ou pela diminuição do tempo entre as duas cromatografias ter permitido ainda a detecção das proteínas na forma solúvel, sugerindo um estado semiestruturado que ao longo do tempo acaba por gerar agregados proteicos. Ambas as hipóteses foram testadas por DLS supersensível.

Todas as construções já demonstraram instabilidade nas frações obtidas pela cromatografia de exclusão de tamanho pelas bandas de menores tamanhos eluídas juntas que foram comprovadas por *western blot* de serem pedaços da proteína inteira (FIGURA 25).

### 5.2.9. Caracterização biofísica dos anticorpos comerciais e sua reatividade

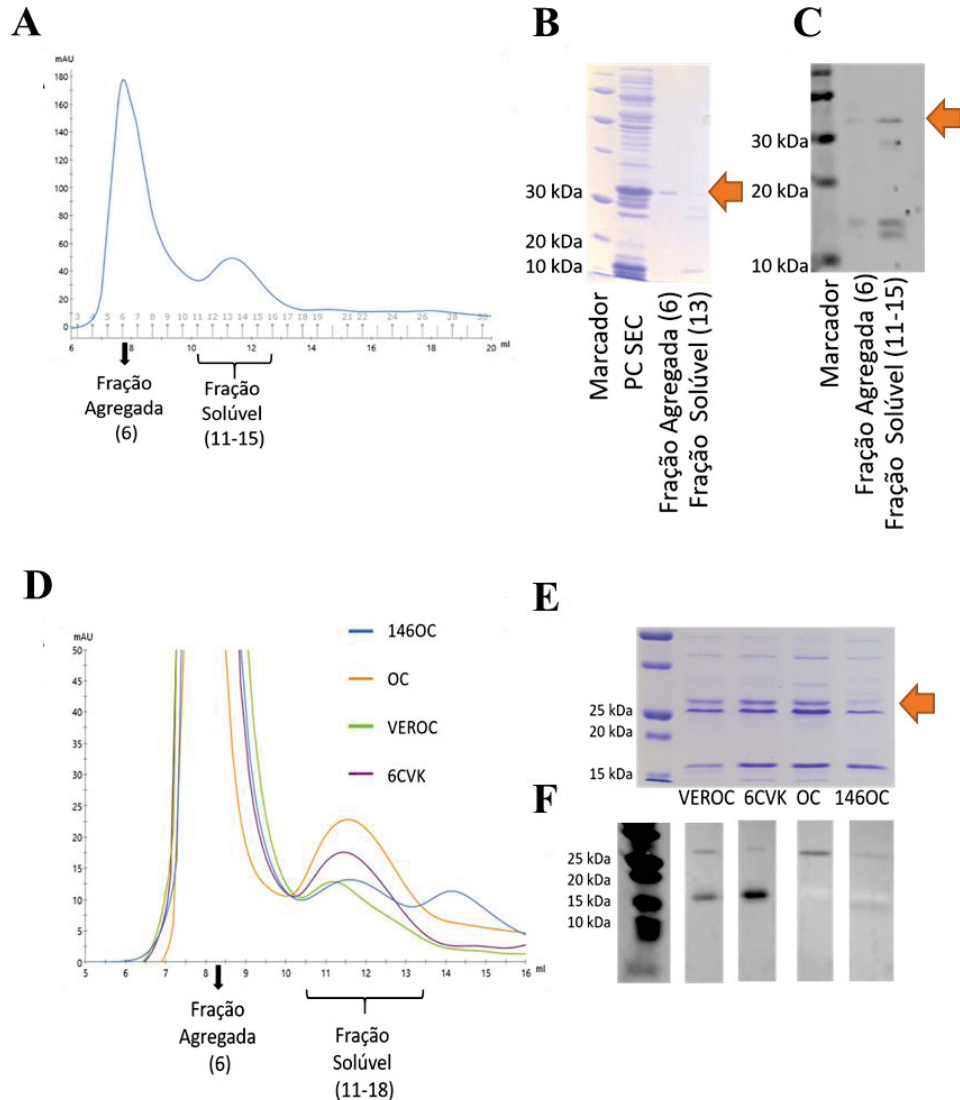
A caracterização dos anticorpos comerciais foi realizada utilizando as técnicas de DLS supersensível, nanofluorimetria de varredura diferencial, *dot blot* e *western blot*. Os anticorpos utilizados são os mesmos sequenciados em trabalho prévio e suas sequencias deram origem aos scFvs encontrados nessa tese.

Na descrição disponível pelos fornecedores, o anti-EGF SC é descrito como uma IgG1  $\kappa$ , produzido em camundongo capaz de reconhecer parte da alça B e a alça C inteira (EALDKYACNCVVG YIGERCQYRDLKWWE) do EGF de rato, sendo capaz de reconhecer essa região também em camundongos e humanos. Ele é recomendado para uso em *western blot*, imunoprecipitação, imunofluorescência e ELISA.

Já o anti-EGF CS é descrito como um IgG neutralizante produzido em camundongo utilizando EGF recombinante humano, sendo recomendado apenas para ensaios de neutralização. Ele é descrito como incapaz de neutralizar o TGF $\alpha$  e a BTC. Para a análise de DLS supersensível foi utilizada a concentração de 1.5 mg/mL de cada anticorpo.

O anticorpo anti-EGF SC demonstrou ser polidisperso, dado que corrobora os dados de checagem de controle de qualidade do Monolith NT.115, o qual acusou agregação do anticorpo em questão. Já o anticorpo anti-EGF CS mostrou um perfil monodisperso com tamanho de raio hidrodinâmico esperado para um anticorpo  $\sim 7$  nm (FIGURA 26).

**FIGURA 25. PURIFICAÇÃO DOS SCFVS E IDENTIFICAÇÃO DE INSTABILIDADE**

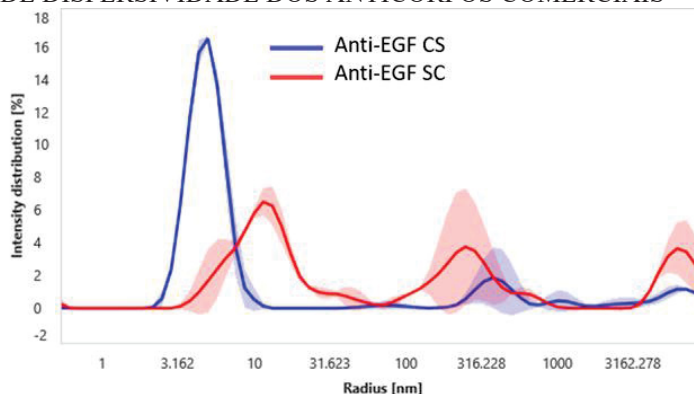


Fonte: O próprio. A. Cromatograma da exclusão por tamanho do scFv MM visualizado no software Unicorn 7. B. SDS-PAGE 13% de uma fração do pico da cromatografia. C. *Western blot* da cromatografia revelado no aparelho LiCor Odyssey Infrared Imaging System (LI-COR Biosciences, Nebraska, EUA). D. Sobreposição dos cromatogramas das cromatografias de exclusão por tamanho dos scFvs OC, VERO, 6CVK e 146OC visualizados no software Unicorn 7. E. SDS-PAGE 13% de uma fração dos picos da cromatografia. F. *Western blot* da cromatografia revelado no aparelho LiCor Odyssey Infrared Imaging System (LI-COR Biosciences, Nebraska, EUA).

Isso é corroborado pelas informações adquiridas durante a análise de nanofluorimetria de varredura diferencial, na qual foi possível obter a temperatura média de transição ( $T_m$ ) do estado enovelado para o desenovelado do anticorpo anti-EGF CS, no entanto, os dados obtidos pelo anticorpo anti-EGF SC demonstram uma população de diversos tamanhos ao longo de todo o gradiente térmico, não havendo determinação clara dos pontos de inflexão. Os valores de  $T_m$  do anticorpo anti-EGF CS foram de 76.73 e 84.34, correspondendo aos valores da cadeia leve e pesada, sendo comumente os valores mais altos de Kd associados a

cadeia pesada (FIGURA 27). Esses valores são compatíveis com os encontrados por outros anticorpos na literatura (IONESCU et al., 2008; KOHLI et al., 2015).

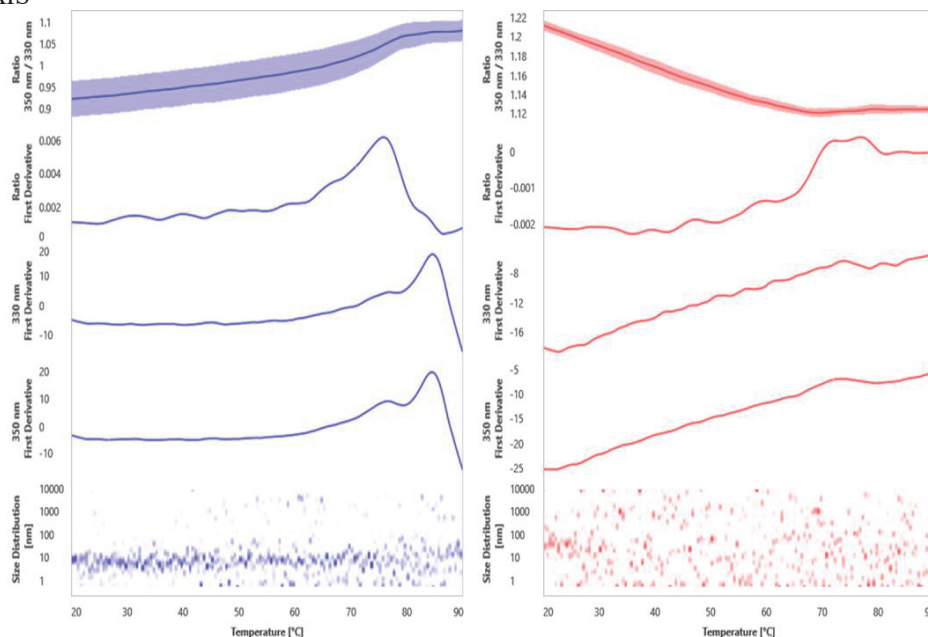
**FIGURA 26.** PERFIL DE DISPERSIVIDADE DOS ANTICORPOS COMERCIAIS



Fonte: O próprio. Imagem do Prometheus Panta (NanoTemper Technologies, Munique, Alemanha).

Para os experimentos de *dot blot* e *western blot* o intuito era avaliar a seletividade de ambos os anticorpos versus diferentes domínios EGF, uma vez que ambos os anticorpos não possuem informações de seletividade dentro da superfamília EGF o suficiente para sua caracterização completa. Os sinais do *dot blot* foram todos normalizados pela quantidade de proteína utilizada por cada gota e a média dos valores de intensidade versus o EGF comercial foi utilizado como valor de referência de 1 para comparação entre as amostras.

**FIGURA 27.** NANOFLUORIMETRIA DE VARREDURA DIFERENCIAL DOS ANTICORPOS COMERCIAIS



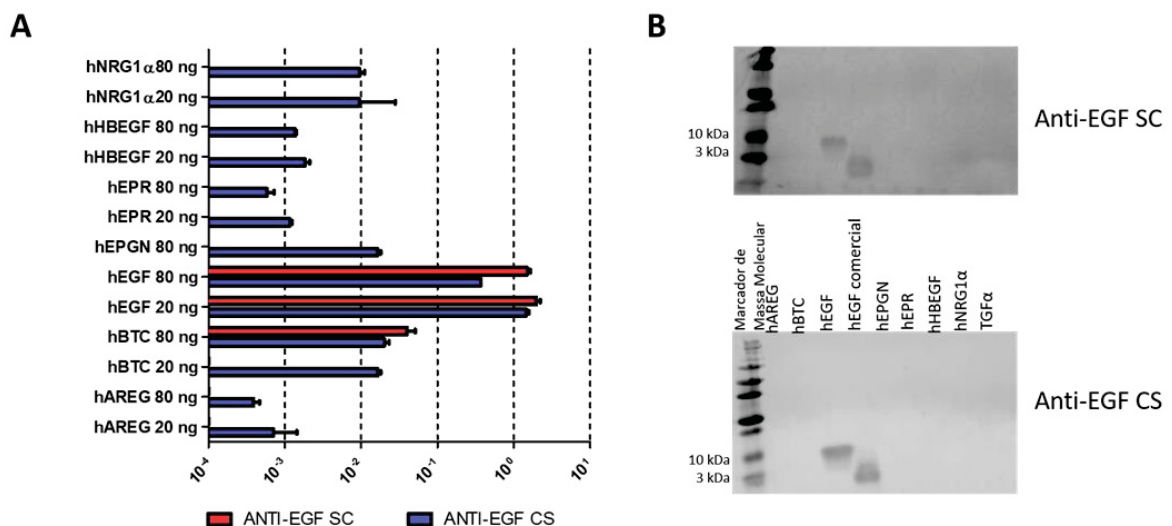
Fonte: O próprio. Em azul, o anticorpo anti-EGF CS e em vermelho o anticorpo anti-EGF SC. Imagem do Prometheus Panta (NanoTemper Technologies, Munique, Alemanha).

Na FIGURA 28A, podemos ver que contra os domínios EGF não desnaturados o anticorpo SC apresentou alta seletividade pelo EGF, havendo, no entanto, uma interação inespecífica 40 vezes mais fraca pela hBTC. Já o anticorpo CS também demonstrou alta seletividade, havendo interações inespecíficas pelo menos 100 vezes mais fracas com todos os domínios EGF, com exceção do hTGF $\alpha$ .

Ambos os valores ficam dentro da margem de diluição de anticorpos para uso em pesquisa. Para o anti-EGF SC vemos que ele é descrito para uso em imunofluorescência dentro da margem de diluição utilizada para esse ensaio, sendo assim, a interação inespecífica com a hBTC preocupante, somada com a baixa qualidade do anticorpo em questão fornecido. Já o anti-EGF CS, o seu uso é recomendado apenas para efeito neutralizante. Considerando que esse tipo de ensaio sempre utiliza um excesso de anticorpo para garantia do efeito, as reações inespecíficas de 100 a 1000 vezes são preocupantes, uma vez que vai atuar em outras proteínas da família se presentes no meio. O fato que esse anticorpo foi descontinuado por motivos não revelados corrobora essa descoberta.

Já na FIGURA 28B, verificamos uma alta especificidade por parte dos dois anticorpos, reconhecendo apenas o EGF recombinante e o EGF de origem comercial. Esse resultado sugere que as ligações inespecíficas ocorrem com os domínios EGF em estado não desnaturado, uma vez que com a desnaturação da proteína esse efeito não é detectado.

**FIGURA 28.** ANÁLISE DE SELETIVIDADE DOS ANTICORPOS ANTI-EGF SC E CS FRENTE AOS DIFERENTES DOMÍNIOS EGF



Fonte: O próprio. A) Valor de fluorescência média das áreas do dot blot. As barras representam a fluorescência normalizada pela ligação dos anticorpos ao EGF comercial contendo o desvio padrão. B) membranas de WB com os anticorpos anti-EGF SC e CS frente os diferentes domínios EGF revelados no aparelho LiCor Odyssey Infrared Imaging System (LI-COR Biosciences, Nebraska, EUA).

A baixa qualidade do anticorpo anti-EGF SC pode ser responsável por ele só ser capaz de reconhecer o seu alvo original, podendo sua afinidade estar diminuída devido a presença de agregados, por outro lado, o anticorpo anti-EGF SC pode ser tão seletivo que mesmo agregado, ele ainda apresente seletividade pelo seu alvo. Seria necessário a aquisição do anticorpo em boas condições biofísicas para se elucidar esse ponto. Já o anticorpo anti-EGF CS, apesar da boa qualidade, demonstra uma menor seletividade pelo seu alvo.

#### **5.2.10. Caracterização biofísica dos scFvs**

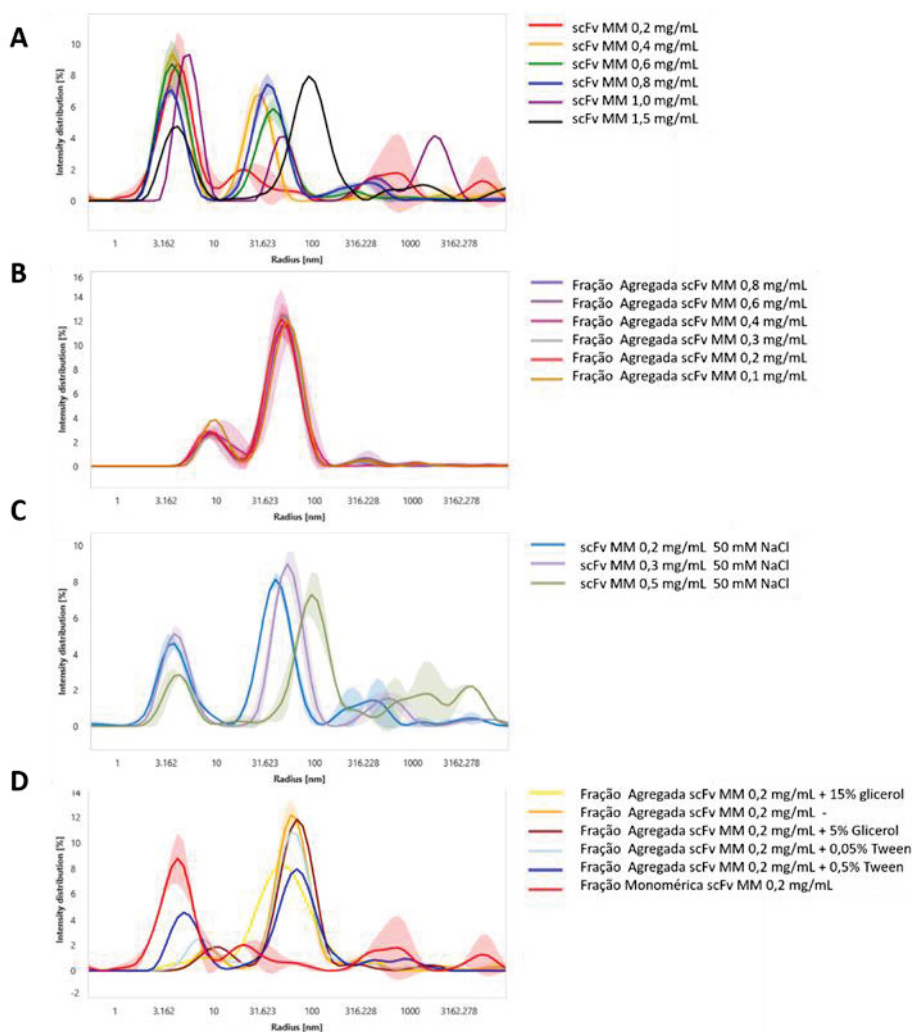
Ao analisar os scFvs no DLS supersensível verificamos uma população agregada. Uma vez que a cromatografia de exclusão por tamanho havia separado as formas agregadas, suspeitamos que o processo de concentração pudesse estar gerando ou acelerando a agregação. Sendo assim, a amostra foi diluída para as concentrações de 1,0 mg/mL, 0,8 mg/mL, 0,6 mg/mL, 0,4 mg/mL e 0,2 mg/mL e seus perfis analisados (FIGURA 29A) e a função de auto-interação do Prometheus Panta foi utilizada para analisar auto interação (FIGURA 30A). A concentração com o melhor perfil monodisperso no DLS supersensível é de 0,2 mg/mL, quase dez vezes menor da concentração inicial. Já a função de auto-interação calcula o parâmetro de difusão de interação com base no coeficiente de difusão e diferentes concentrações da molécula.

Um parâmetro de difusão de interação positivo indica propensão das moléculas de se afastarem enquanto valores negativos indicam propensão a auto interação. O valor obtido do scFv MM foi de -373,83, demonstrando uma alta propensão de auto interação. O mesmo experimento foi realizado diluindo-se a fração agregada e medindo-se seus perfis de dispersão e auto interação (FIGURA 29B e FIGURA 30B). A diluição da fração agregada não levou a desfazer a agregação e seu parâmetro de difusão de interação foi de -31,38, bastante abaixo da fração solúvel, uma vez que já não existem muitas moléculas livres para auto interagirem.

Tentou-se diminuir a quantidade de cloreto de sódio para 50 mM para verificar se modificações adicionais no tampão de cromatografia poderiam ajudar a promover maior estabilidade ou diminuir a auto interação (FIGURA 29C e FIGURA 30C), no entanto, não houve mudança no perfil de dispersão e houve uma diminuição do parâmetro de difusão de interação para -1285,96. Apesar da mudança da quantidade de sal do tampão cromatográfico ter possibilitado o surgimento da população monomérica, o decréscimo do sal nos nossos testes parece ter voltado a favorecer interações hidrofóbicas entre as moléculas fazendo assim o parâmetro de difusão de interação diminuir. Tentou-se então adicionar diferentes concentrações de glicerol (5% e 15%) ou Tween-20 (0,05% e 0,5%).



**FIGURA 29.** PERFIS DE DISPERSÃO DAS DIFERENTES FRAÇÕES CROMATOGRÁFICAS DO SCFV MM



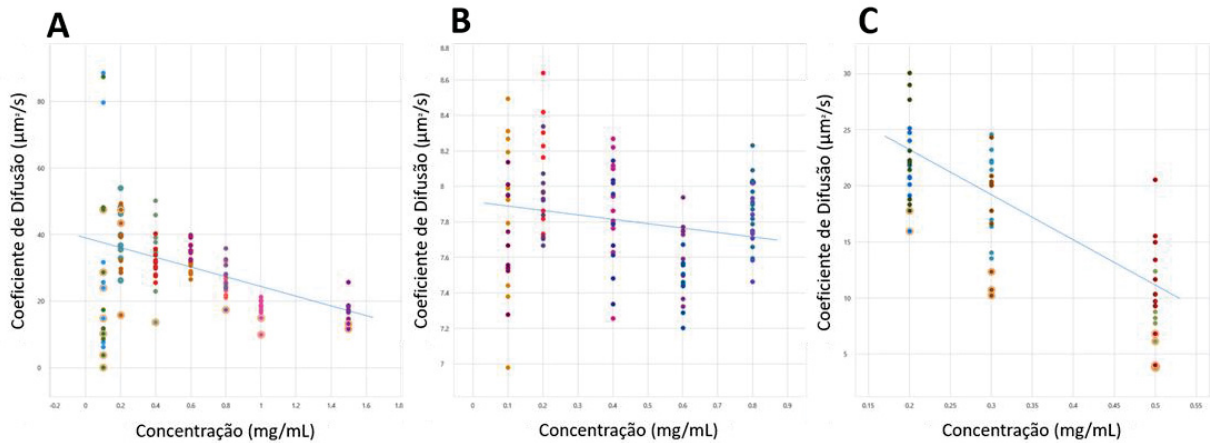
Fonte: O próprio. Gráficos de distribuição de intensidade (%) por raio hidrodinâmico relativo (nm). Imagem do Prometheus Panta (NanoTemper Technologies, Munique, Alemanha).

Apenas com adição de 0,5% de Tween-20 tem-se uma população com raio hidrodinâmico próximo ao do tamanho esperado. Anticorpos em geral são armazenados com no máximo 0,1% de Tween 20, uma vez que ele pode inibir a interação do anticorpo com seu alvo. Já para os testes de nanofluorimetria de varredura diferencial foi possível calcular a  $T_m$  da fração solúvel utilizando-se da derivada da curva de 330 nm, onde as  $T_m$  calculadas foram 50,84 e 61,77, tendo início o processo de agregação  $>40$  °C (FIGURA 31A).

A escolha da utilização da curva da derivada da razão ou dos comprimentos individuais é feita com base em qual tem uma melhor definição dos pontos para a  $T_m$ . Para as frações agregadas verificamos um menor raio cumulativo relativo para as amostras com adição de Tween em ambas as concentrações, tendo a concentração de 0,5% diminuído o raio para o tamanho esperado em entre 70 e 90 graus, no entanto, esse processo se reverteu (FIGURA 31B).



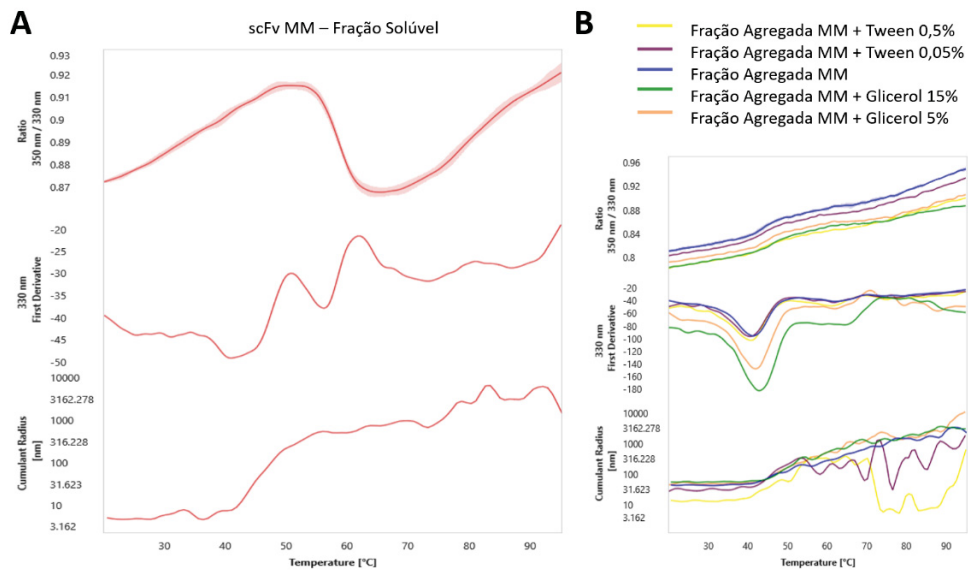
**FIGURA 30. COEFICIENTE DE DIFUSÃO EM RELAÇÃO A CONCENTRAÇÃO**



Fonte: O próprio. Cada círculo é uma medida de DLS supersensível diferente. Halos ao redor dos círculos demonstram pontos fora da curva. Imagem do Prometheus Panta (NanoTemper Technologies, Munique, Alemanha).

Assim como a amostra MM, as amostras do scFv OC e suas construções híbridas também mostram sinais de agregação nas frações solúveis, mesmo tendo sido concentradas para 0,1 mg/mL (FIGURA 32). As frações agregadas sofreram diferentes aquecimentos e tempos. Para as temperaturas 20 °C e 50 °C foram incubadas por 1h, para a temperatura de 90 °C, as proteínas foram incubadas por 10 minutos. Todas foram resfriadas gradativamente à temperatura ambiente e então analisadas. Frações agregadas no gelo e a fração solúvel também foram analisadas e todas comparadas entre si.

**FIGURA 31. NANOFLUORIMETRIA DE VARREDURA DIFERENCIAL DO scFv MM**



Fonte: O próprio. A. Curvas da razão 350/330 nm da fração solúvel, a derivada da curva de 330 nm e raio cumulativo relativo. B da fração solúvel. B. Curvas da razão 350/330 nm, derivada da curva de 330 nm e raio cumulativo relativo. B da fração agregada em diferentes tampões. Imagem do Prometheus Panta (NanoTemper Technologies, Munique, Alemanha).

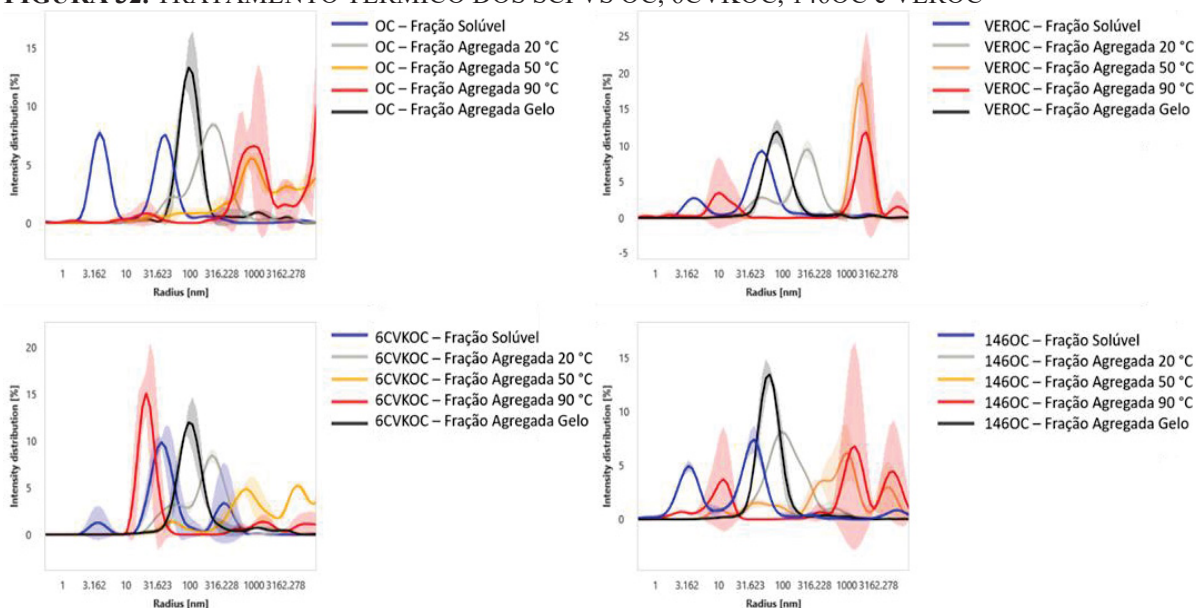
**FIGURA 32. TRATAMENTO TÉRMICO DOS SCFVS OC, 6CVKOC, 146OC e VERO**

Imagem: O próprio. Dados adquiridos no Nanotemper Imagem do Prometheus Panta (NanoTemper Technologies, Munique, Alemanha).

O tratamento térmico a 90 °C demonstrou surgir populações de tamanhos menores para os scFvs 6CVKOC, VERO e 146OC, tendo um efeito mais pronunciado no 6CVKOC, no entanto, nenhuma população com o raio relativo esperado, demonstrando que esses arcabouços não conseguem se renovar após o aquecimento.

### 5.3. SISTEMA DE EXPRESSÃO EM SUPERFÍCIE BACTERIANA

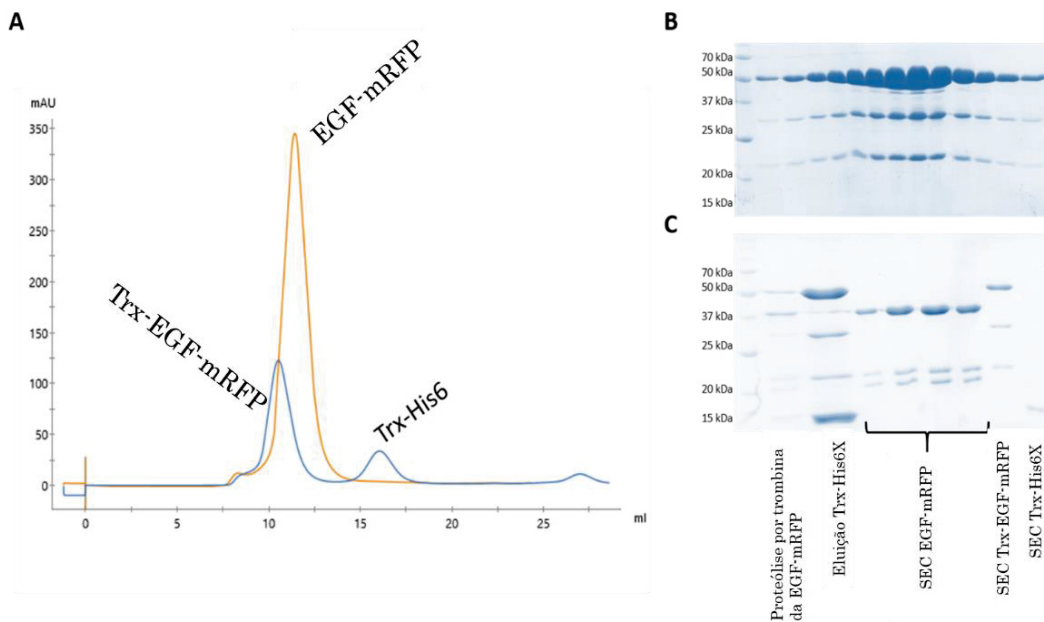
Para testar o sistema de apresentação em superfície bacteriana como uma possibilidade para geração de novos anticorpos do tipo VHH utilizando-se bibliotecas sintéticas, geraram-se seis novas construções. A primeira foi o hEGF fusionado a Trx no N-terminal e a mRFP no seu C-terminal. Seu protocolo de expressão e purificação foi igual ao do EGF fusionado a Trx.

#### 5.3.1. Expressão e purificação do hEGF-mRFP

O Trx-EGF-mRFP foi purificado de maneira semelhante por cromatografia de afinidade ao níquel. Em seguida, as frações foram concentradas e adicionou-se trombina para a retirada da Trx-His6 do N-terminal. O produto da proteólise foi então injetado novamente na coluna de afinidade ao níquel e as frações não ligadas a coluna foram coletadas e concentradas. A fração ligada a coluna também foi eluída para critério de comparação. Ambas as amostras foram utilizadas para a cromatografia por exclusão por tamanho (SEC).

O EGF-mRFP saiu no volume esperado, no entanto, no SDS-PAGE foram identificadas bandas menores em conjunto, o que provavelmente é um produto de degradação. Houve banda de degradação semelhante na proteína inteira, a qual saiu no volume esperado também pela SEC, assim como a Trx livre (FIGURA 33).

FIGURA 33. PURIFICAÇÃO DO EGF-MRFP



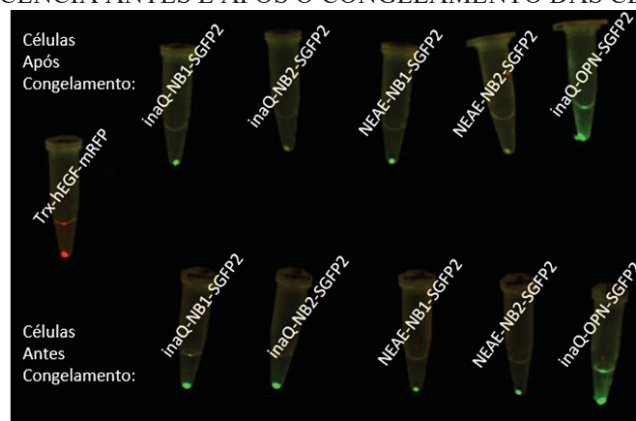
Fonte: O próprio. A. Comparação dos perfis cromatográficos do EGF-mRFP, Trx-EGF-mRFP e Trx-His6. B. SDS-PAGE 13% da cromatografia por afinidade ao níquel do Trx-EGF-Mrpf no software Unicorn 7. A banda de interesse é próxima de 50 kDa. C. SDS-PAGE 13% da amostra digerida por trombina antes da purificação, a fração contendo Trx eluída (15 kDa) da segunda afinidade ao níquel, as frações da SEC do EGF-Mrpf (37 kDa), a fração não digerida na SEC e a fração da Trx na SEC.

As outras cinco construções foram a fusão dos VHHs NB1 e NB6 as proteínas de membrana inaQ e NEAE no N-terminal e a SGFP2 no C-terminal, além da fusão da OPN humana a inaQ e SGFP2, como um controle negativo de interação. Suas expressões foram comparadas em células BL21 (DE3)  $\Delta$ slyD pRARE2 antes e após o congelamento. Também foi verificada a fluorescência da Trx-EGF-mRFP dentro da BL21 (DE3)  $\Delta$ slyD pRARE2 após o congelamento (FIGURA 34).

A escolha das duas proteínas de membrana é a diferença estrutural entre elas, sendo a inaQ intrinsecamente desestruturada e a NEAE ser uma um barril de folhas beta de membrana externa.

As construções com o NB2 parecem perder fluorescência no processo de congelamento, independente da proteína de membrana a que está ancorado, enquanto as outras construções não parecem serem tão afetadas. A fluorescência das construções NEAE antes do congelamento parecem ser menos intensas do que as construções inaQ.

**FIGURA 34.** FLUORESCÊNCIA ANTES E APÓS O CONGELAMENTO DAS CÉLULAS.

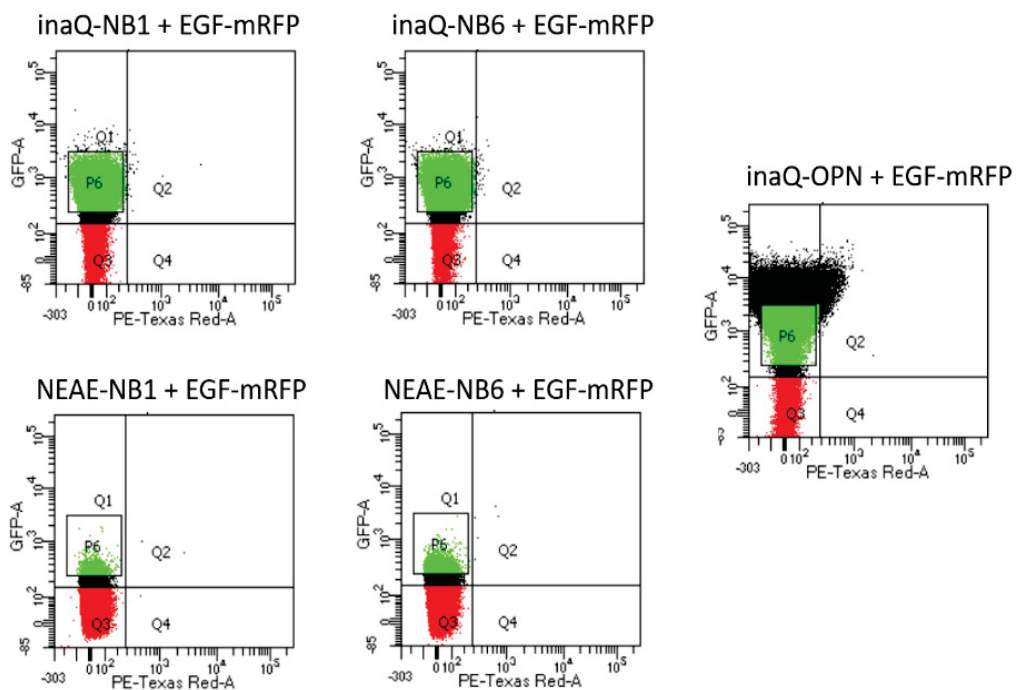


Fonte: O próprio. Imagem adquirida no Invitrogen™ iBright™ CL1500 (Waltham, Massachusetts, EUA).

### 5.3.2. Testes de interação por citometria de fluxo

Os testes de interação por citometria de fluxo foram realizados com as quatro construções incubadas com a proteína EGF-mRFP a fim de recuperar as células duplo-positivas, aquelas que teriam a marcação da proteína de membrana e o ligante (FIGURA 35). Porém, na concentração testada não foi possível verificar população com fluorescência vermelha, seja isolada ou junto à fluorescência verde. A fluorescência das construções NEAE nesse método também se mostrou menos intensa do que das construções inaQ.

**FIGURA 35.** CITOMETRIA DE FLUXO DAS CONSTRUÇÕES DE SUPERFÍCIE BACTERIANA INCUBADAS COM EGF-MRFP



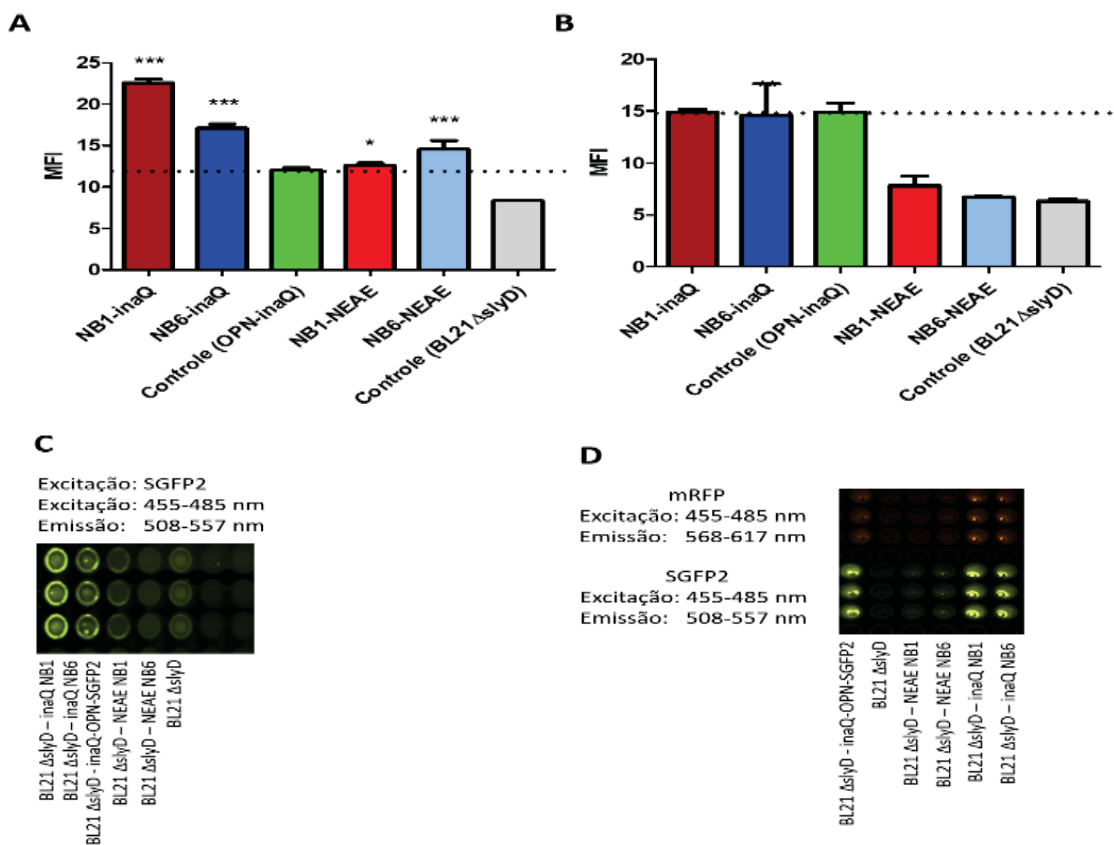
Fonte: O próprio. Histogramas de citometria fluxo. Quadrante Q1 = presença de fluorescência verde. Q2 = presença de fluorescência verde e vermelha. Q3 = ausência de fluorescência verde e vermelha. Q4 = presença de fluorescência vermelha. Dados adquiridos no citômetro FACs Canto II.

Nossa hipótese é que a concentração de EGF-mRFP não foi suficiente para um sinal forte o suficiente para ser detectadas as células duplo positivas, uma vez que a estequiometria proteínas na membrana, células e EGF-mRFP tem que ser considerada.

### 5.3.3. Testes de interação em placa

Para testar a interação entre NB1 e NB6 com o EGF dois desenhos experimentais foram testados. O primeiro, colocando as células com as proteínas expressas na membrana externa no fundo da placa, incubando-se assim elas com o EGF-mRFP e verificando a interação pela dupla fluorescência, ou sensibilizando uma placa com EGF e incubando-se as células na placa, revelando-se pela fluorescência da SGFP2, sem a necessidade de uma segunda molécula fluorescente (FIGURA 36).

FIGURA 36. TESTE DE INTERAÇÃO ENTRE NB1 E NB6 COM EGF



Fonte: O próprio. A. Quantificação da placa sensibilizada com o EGF e incubada com as células. B. Quantificação da placa contendo as células no fundo incubada com o EGF-mRFP. C. Placa sensibilizada com o EGF e incubada com as células. D. Placa contendo as células no fundo incubada com o EGF-mRFP. Imagens de fluorescência adquiridas no Invitrogen™ iBright™ CL1500 (Waltham, Massachusetts, EUA).

No teste da placa sensibilizada com o EGF foi quantificada a fluorescência das construções na superfície das bactérias e foi encontrada significância estatística entre as 4

construções e as células contendo a construção controle e as células não transformadas, sendo novamente as construções inaQ com maior fluorescência detectada.

Já o experimento com as células no fundo da placa incubadas com o EGF-mRFP tiveram a fluorescência vermelha quantificada e não houve diferença entre as construções e o controle, havendo inclusive menos fluorescência nas construções NEAE. Uma possibilidade é que a proteína mRFP interfira com a ligação do EGF aos VHHs, impossibilitando sua utilização em ensaios de interação com NB1 e NB2.

Sendo assim, a estratégia de utilização de VHHs na superfície de membrana externa de bactérias é promissora para o desenvolvimento de futuras bibliotecas ou estudos com anticorpos ligados a membrana. É importante salientar a necessidade de se avaliar a presença de proteínas de fusão seja no VHH na membrana, quanto no ligante, para não haver interferências no desenvolvimento e aquisição de dados.

## 6. CONCLUSÃO

Em resumo, neste trabalho desenvolvemos um protocolo unificado para expressão e purificação para a aquisição dos sete membros da família de fatores de crescimento EGF e NRG1 $\alpha$  de maneira pura e estável, tendo todas as moléculas sido caracterizadas de maneira biofísica e funcional, verificando-se que se encontram funcionais, com a conformação certa e com alta estabilidade, além de ter seus perfis de dispersão, seus espectros de dicroísmo circular e seus padrões de proteólise descritos. Isso é importante para o desenvolvimento de tecnologias nacionais e ampliação dos conhecimentos biofísicos para caracterização visando um controle de qualidade de biomoléculas, tendo várias das características biofísicas aqui descritas pela primeira vez em artigo publicado na revista *Scientific Reports* e no artigo submetido a *Molecular Biology Reports*.

Também foi possível explorar diferentes protocolos de investigação de variáveis importantes para serem alteradas e levadas em conta no momento de se padronizar um protocolo de expressão de proteínas recombinantes de maneira solúvel, seja a composição de tampões, meios, temperatura, tempo, quantidade de agente indutor ou cepa.

Ademais, também conseguimos descrever anticorpos comerciais que são adquiridos comumente no ramo de pesquisa científica e possíveis problemas que estão ligados ao baixo critério das empresas ao fornecer tais produtos, apesar de funcionarem em certas técnicas, suas falhas e problemas devem ser melhor estudadas a fim de garantir também a qualidade da pesquisa produzida com tais insumos.

Por fim, descrevemos diversas características e dificuldades na hora de obter fragmentos de anticorpo do tipo scFv e diversas estratégias para sobrepor os problemas descritos, a fim de não cometermos o mesmo erro dos insumos comercializados como descrito acima.

Nesse sentido, essas descobertas nos levaram a desistir dos scFvs desenvolvidos nesse trabalho e começar a trabalhar em outras maneiras de adquirir fragmentos de anticorpos, como o sistema de expressão em superfície bacteriana, a qual se mostra uma plataforma interessante para ser utilizada, mas que deve ser ainda padronizada e desenvolvida por mais tempo até ser consolidada em nosso grupo.



## 7. REFERÊNCIAS BIBLIOGRÁFICAS

- ABBAS, A. K.; LICHTMAN, A. H.; PILLAI, S. **Imunologia Celular e Molecular**. 7<sup>a</sup> ed. Rio de Janeiro: Elsevier, 2011.
- ABDULL RAZIS, A. F. et al. Expression of recombinant human epidermal growth factor in *Escherichia coli* and characterization of its biological activity. **Applied Biochemistry and Biotechnology**, v. 144, n. 3, p. 249–261, 2008.
- AHMAD, Z. A. et al. ScFv antibody: Principles and clinical application. **Clinical and Developmental Immunology**, v. 2012, p. 1–15, 2012.
- AHMADZADEH, M. et al. Anti-HER2 scFv Expression in *Escherichia coli* SHuffle®T7 Express Cells: Effects on Solubility and Biological Activity. **Molecular Biotechnology**, v. 62, n. 1, p. 18–30, 2020.
- AKBARI, V. et al. Optimization of a single-chain antibody fragment overexpression in *Escherichia coli* using response surface methodology. **Research in Pharmaceutical Sciences**, v. 10, n. 1, p. 81–89, 2015.
- ALIMANDI, M. et al. Epidermal growth factor and betacellulin mediate signal transduction through co-expressed ErbB2 and ErbB3 receptors. **The EMBO Journal**, v. 16, n. 18, p. 5608–5617, 1997.
- AMEZAGA, M. R.; BOOTH, I. R. Osmoprotection of *Escherichia coli* by peptone is mediated by the uptake and accumulation of free proline but not of proline-containing peptides. **Applied and Environmental Microbiology**, v. 65, n. 12, p. 5272–5278, 1999.
- ANVISA. **Resolução da diretoria colegiada- Guia para validação de métodos analíticos e bioanalíticos. Resolução - RE nº 899, de 29 de maio de 2003**. Disponível em: <[www.anvisa.gov.br/legis](http://www.anvisa.gov.br/legis)>. Acesso em: 28 set. 2017.
- ASCOLI, C. A.; BIRABAHARAN, J. **The Antibody Dilemma**. Disponível em: <<https://www.genengnews.com/magazine/253/the-antibody-dilemma/>>. Acesso em: 13 dez. 2021.
- AUBREY, N.; BILLIALD, P. Antibody Fragments Humanization: Beginning with the End in Mind. In: [s.l: s.n.]. v. 1904p. 231–252.
- BAEZ, W. D. et al. Global analysis of protein synthesis in *Flavobacterium johnsoniae* reveals the use of Kozak-like sequences in diverse bacteria. **Nucleic Acids Research**, v. 47, n. 20, p. 10477–10488, 2019.
- BECK, A. et al. Strategies and challenges for the next generation of therapeutic antibodies. **Nature reviews. Immunology**, v. 10, n. 5, p. 345–352, 2010.
- BIENERT, S. et al. The SWISS-MODEL Repository-new features and functionality. **Nucleic Acids Research**, v. 45, n. D1, p. D313–D319, 2017.
- BJO, K. et al. Domain Rearrangements in Protein Evolution. **J. Mol. Biol.**, v. 353, p. 911–923, 2005.
- BOSSI, S. et al. Antibody-mediated purification of co-expressed antigen-antibody complexes. **Protein Expression and Purification**, v. 72, n. 1, p. 55–58, 2010.
- BUSSER, B. et al. The multiple roles of amphiregulin in human cancer. **Biochimica et Biophysica Acta - Reviews on Cancer**, v. 1816, n. 2, p. 119–131, 2011.
- CARRAWAY, J. K. et al. Neuregulin-2, a new ligand of ErbB3/ErbB4-receptor tyrosine kinases. **Nature**, v. 387, n. 6632, p. 512–516, 1997.
- CARVALHO, S. et al. An antibody to amphiregulin, an abundant growth factor in patients' fluids, inhibits ovarian tumors. **Oncogene**, v. 35, n. 4, p. 438–447, 2016.
- CHANG, A. Y. et al. Preparation of calcium competent *Escherichia coli* and heat-shock transformation. **Journal of Experimental Microbiology and Immunology (JEMI)**, v. 1, n. June, p. 22–25, 2017.
- CHEN, J. et al. Expression and function of the epidermal growth factor receptor in

physiology and disease. **Physiological Reviews**, v. 96, n. 3, p. 1025–1069, 2016.

CHEN, R.; JIN, G.; MCINTYRE, T. M. The soluble protease ADAMDEC1 released from activated platelets hydrolyzes platelet membrane pro-epidermal growth factor (EGF) to active high-molecular-weight EGF. **Journal of Biological Chemistry**, v. 292, n. 24, p. 10112–10122, 2017.

COTE, G. M. et al. Neuregulin-1 $\alpha$  and  $\beta$  isoform expression in cardiac microvascular endothelial cells and function in cardiac myocytes in vitro. **Experimental Cell Research**, v. 311, n. 1, p. 135–146, 2005.

CUI, F. J. et al. Optimization of the medium composition for production of mycelial biomass and exo-polymer by *Grifola frondosa* GF9801 using response surface methodology. **Bioresource Technology**, v. 97, n. 10, p. 1209–1216, 2006.

DUCANTEL, F.; MULLER, B. H. Molecular engineering of antibodies for therapeutic and diagnostic purposes. **mAbs**, n. August, p. 445–457, 2012.

DVORAK, P. et al. Exacerbation of substrate toxicity by IPTG in *Escherichia coli* BL21(DE3) carrying a synthetic metabolic pathway. **Microbial Cell Factories**, v. 14, n. 1, p. 1–15, 2015.

ECCLES, S. A. The epidermal growth factor receptor / Erb-B / HER family in normal and malignant breast biology. **Int. J. Dev. Biol.**, v. 44, n. November, p. 685–696, 2011.

EMBL. **SMART: EGF domain annotation**. Disponível em: <[http://smart.embl.de/smart/do\\_annotation.pl?DOMAIN=SM00114](http://smart.embl.de/smart/do_annotation.pl?DOMAIN=SM00114)>.

FARSHDARI, F. et al. The improvement of anti-HER2 scFv soluble expression in *Escherichia coli*. **Brazilian Journal of Pharmaceutical Sciences**, v. 56, p. 2–9, 2020.

FDA. **Analytical Procedures and Methods Validation for Drugs and Biologics**. Disponível em: <<https://www.fda.gov/downloads/drugs/guidances/ucm386366.pdf>>. Acesso em: 19 fev. 2017.

FERREIRA, A. S. **ISOLAMENTO DE FRAGMENTOS DE ANTICORPOS PARA QUANTIFICAÇÃO DE EGF HUMANO**. [s.l.] Universidade Federal do Paraná, 2018.

FERREIRA, A. S. et al. A toolkit for recombinant production of seven human EGF family growth factors in active conformation. **Scientific Reports**, v. 12, n. 1, p. 1–14, 2022.

FRAGOSO-JIMÉNEZ, J. C. et al. Growth-dependent recombinant product formation kinetics can be reproduced through engineering of glucose transport and is prone to phenotypic heterogeneity. **Microbial Cell Factories**, v. 18, n. 1, p. 1–16, 2019.

FRANK, S. A. Specificity and Cross-Reactivity. In: **Immunology and Evolution of Infectious Disease**. 1st. ed. [s.l.] Princeton University Press, 2002. v. 300p. 33–35.

FRANKEL, S.; SOHN, R.; LEINWAND, L. The use of sarkosyl in generating soluble protein after bacterial expression. **Proceedings of the National Academy of Sciences of the United States of America**, v. 88, n. 4, p. 1192–1196, 1991.

FREED, D. M. et al. EGFR Ligands Differentially Stabilize Receptor Dimers to Specify Signaling Kinetics. **Cell**, v. 171, n. 3, p. 683–695.e18, 2017.

FRIEDEL, T. et al. Receptor-targeted lentiviral vectors are exceptionally sensitive toward the biophysical properties of the displayed single-chain Fv. **Protein Engineering, Design and Selection**, v. 28, n. 4, p. 93–105, 2015.

GARRETT, T. P. J. et al. Crystal Structure of a Truncated Epidermal Growth Factor Receptor Extracellular Domain Bound to Transforming Growth Factor  $\alpha$ . **Cell**, v. 110, n. 6, p. 763–773, 2002.

GRILO, A. L.; MANTALARIS, A. The Increasingly Human and Profitable Monoclonal Antibody Market. **Trends in Biotechnology**, v. 37, n. 1, p. 9–16, 2019.

GUARDIOLA, S. et al. Blocking EGFR Activation with Anti-EGF Nanobodies via Two Distinct Molecular Recognition Mechanisms. **Angewandte Chemie - International**

**Edition**, v. 57, n. 42, p. 13843–13847, 2018.

GUMÀ, A. et al. Emerging role of neuregulin as a modulator of muscle metabolism.

**American Journal of Physiology - Endocrinology and Metabolism**, v. 298, n. 4, 2010.

GUPTA, P.; DEEP, S. Salt mediated unusual switching in the aggregation kinetic profile of human carbonic anhydrase. **RSC Advances**, v. 5, n. 116, p. 95717–95726, 2015.

HANG, Q. et al. Integrin  $\alpha 5$  suppresses the phosphorylation of epidermal growth factor receptor and its cellular signaling of cell proliferation via N-glycosylation. **Journal of Biological Chemistry**, v. 290, n. 49, p. 29345–29360, 2015.

HAO, S. et al. Transcriptional Regulation of ompC by Calcium Chloride May Involve envZ. **Journal of Experimental Microbiology and Immunology**, v. 20, n. April, p. 43–46, 2016.

HARRIS, R. C.; CHUNG, E.; COFFEY, R. J. EGF receptor ligands. **Experimental Cell Research**, v. 284, n. 1, p. 2–13, 2003.

HELMA, J. et al. Nanobodies and recombinant binders in cell biology. **Journal of Cell Biology**, v. 209, n. 5, p. 633–644, 2015.

HENRIKSEN, L. et al. Internalization Mechanisms of the Epidermal Growth Factor Receptor after Activation with Different Ligands. **PLoS ONE**, v. 8, n. 3, 2013.

HERRERA, M. et al. Lack of Specificity of Commercial Antibodies Leads to Misidentification of Angiotensin Type 1 Receptor Protein. **Hypertension**, v. 61, n. 1, p. 253–258, 1 jan. 2013.

HOU, H.; HE, H.; WANG, Y. Effects of SDS on the activity and conformation of protein tyrosine phosphatase from thermus thermophilus HB27. **Scientific reports**, v. 10, n. 1, p. 3195, 2020.

HUANG, L. et al. E. coli can eat DNA as excellent nitrogen source to grow quickly. **bioRxiv**, p. 2022.03.07.483256, 2022.

HUDSON, P. J.; SOURIAU, C. Engineered antibodies. **Nature Medicine**, v. 9, n. 1, p. 129–134, 2003.

IONESCU, R. M. et al. Contribution of variable domains to the stability of humanized IgG1 monoclonal antibodies. **Journal of Pharmaceutical Sciences**, v. 97, n. 4, p. 1414–1426, 2008.

JAMES, J. et al. Protein over-expression in Escherichia coli triggers adaptation analogous to antimicrobial resistance. **Microbial Cell Factories**, v. 20, n. 1, p. 1–11, 2021.

JAY, F. **Role of amphiregulin in mediating the bone anabolic actions of parathyroid hormone**. [s.l.] Ludwig-Maximilians-Universität München, 2016.

JL, H. Lysis of cultured cells for immunoprecipitation. **Cold Spring Harbor Protocols**, v. 5, n. 8, 2010.

KADO, Y. et al. Epiregulin Recognition Mechanisms by 9E5(Fab) Antibody. **Journal of Biological Chemistry**, v. 5, p. 1–30, 2015.

KADO, Y. et al. Epiregulin recognition mechanisms by anti-epiregulin antibody 9E5: Structural, functional, and molecular dynamics simulation analyses. **Journal of Biological Chemistry**, v. 291, n. 5, p. 2319–2330, 2016.

KAO, W. T. et al. Common genetic variation in neuregulin 3 (NRG3) influences risk for schizophrenia and impacts NRG3 expression in human brain. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 35, p. 15619–15624, 2010.

KHANTASUP, K. et al. Design and generation of humanized single-chain Fv derived from mouse hybridoma for potential targeting application. **Monoclonal Antibodies in Immunodiagnosis and Immunotherapy**, v. 34, n. 6, p. 404–417, 2015.

KIM, J. A. et al. Identification of Neuregulin-2 as a novel stress granule component. **BMB Reports**, v. 49, n. 8, p. 449–454, 2016.

KÖHLER, G.; MILSTEIN, C. Continuous cultures of fused cells secreting antibody of predefined specificity. **Nature**, v. 256, p. 495–497, 1975.

KOHLI, N. et al. A novel screening method to assess developability of antibody-like molecules. **mAbs**, v. 7, n. 4, p. 752–758, 2015.

KOPP, J. et al. Impact of glycerol as carbon source onto specific sugar and inducer uptake rates and inclusion body productivity in E. Coli BL21(DE3). **Bioengineering**, v. 5, n. 1, 2018.

KRAMER, W. et al. Kinetic studies for the optimization of recombinant protein formation. **Annals of the New York Academy of Sciences**, v. 782, p. 323–333, 1996.

KUMAR, D. et al. QbD Based Media Development for the Production of Fab Fragments in E. coli. **Bioengineering**, v. 6, n. 2, p. 29, 2019a.

KUMAR, D. et al. QbD Based Media Development for the Production of Fab Fragments in E. coli. **Bioengineering**, v. 6, n. 29, p. 1–17, 2019b.

KURODA, D. et al. Structural classification of CDR-H3 revisited: A lesson in antibody modeling. **Proteins: Structure, Function and Genetics**, v. 73, n. 3, p. 608–620, 2008.

LEAHY, D. J. STRUCTURE AND FUNCTION OF THE EPIDERMAL GROWTH FACTOR (EGF / ERBB) FAMILY OF RECEPTORS. **Advances in Protein Chemistry**, v. 68, p. 1–27, 2004.

LEITH, E. M. et al. Characterization of the internal translation initiation region in monoclonal antibodies expressed in Escherichia coli. **Journal of Biological Chemistry**, v. 294, n. 48, p. 18046–18056, 2019.

LIPNIČANOVÁ, S. et al. Optimization of medium composition for propagation of recombinant Escherichia coli. **Nova Biotechnologica et Chimica**, v. 18, n. 2, p. 84–93, 2019.

LONG, J. et al. Optimization of fermentation conditions for an Escherichia coli strain engineered using the response surface method to produce a novel therapeutic DNA vaccine for rheumatoid arthritis. **Journal of Biological Engineering**, v. 12, n. 1, p. 1–10, 2018.

LU, W. et al. High-level expression and purification of heparin-binding epidermal growth factor (HB-EGF) with SUMO fusion. **Molecular Biotechnology**, v. 44, n. 3, p. 198–203, 2010.

MAKABE, K. et al. Thermodynamic consequences of mutations in vernier zone residues of a humanized anti-human epidermal growth factor receptor murine antibody, 528. **Journal of Biological Chemistry**, v. 283, n. 2, p. 1156–1166, 2008.

MARINI, G. et al. Experimental design approach in recombinant protein expression: Determining medium composition and induction conditions for expression of pneumolysin from Streptococcus pneumoniae in Escherichia coli and preliminary purification process. **BMC Biotechnology**, v. 14, n. 1, 2014.

MARTINS, A. et al. Role of calcium in the efflux system of Escherichia coli. **International Journal of Antimicrobial Agents**, v. 37, n. 5, p. 410–414, 2011.

MAUTINO, B. et al. Bioactive recombinant neuregulin-1, -2, and -3 expressed in Escherichia coli. **Protein Expression and Purification**, v. 35, n. 1, p. 25–31, 2004.

MCPHERSON, A.; CUDNEY, B. Optimization of crystallization conditions for biological macromolecules. **Acta Crystallographica Section F: Structural Biology Communications**, v. 70, p. 1445–1467, 2014.

MEI, L.; NAVE, K. A. Neuregulin-ERBB signaling in the nervous system and neuropsychiatric diseases. **Neuron**, v. 83, n. 1, p. 27–49, 2014.

MEI, L.; XIONG, W. C. Neuregulin 1 in neural development, synaptic plasticity and schizophrenia. **Nature Reviews Neuroscience**, v. 9, n. 6, p. 437–452, 2008.

MUYLDERMANS, S. Single domain camel antibodies: Current status. **Reviews in Molecular Biotechnology**, v. 74, n. 4, p. 277–302, 2001.



MUYLDERMANS, S. Nanobodies: natural single-domain antibodies. **Annual Review of Biochemistry**, v. 82, p. 775–797, 2013.

NARHI, L. O. et al. Circular dichroism of reduced and oxidized recombinant human epidermal growth factor. **International Journal of Peptide and Protein Research**, v. 39, n. 2, p. 182–187, 1992.

NELSON, A. L. Antibody fragments: hope and hype. **Landes Bioscience**, v. 2, n. 1, p. 77–83, 2010.

OGISO, H. et al. Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. **Cell**, v. 110, n. 6, p. 775–787, 2002.

OKAMOTO, A. et al. Systemic administration of siRNA with anti-HB-EGF antibody-modified lipid nanoparticles for the treatment of triple-negative breast cancer. **Molecular Pharmaceutics**, 2018.

OKINES, A.; CUNNINGHAM, D.; CHAU, I. Targeting the human EGFR family in esophagogastric cancer. **Nature Reviews Clinical Oncology**, v. 8, n. 8, p. 492–503, 2011.

PARK, D. J. et al. Prognostic Significance of Targetable Angiogenic and Growth Factors in Patients Undergoing Resection for Gastric and Gastroesophageal Junction Cancers. **Annals of Surgical Oncology**, v. 21, p. 1130–1137, 2014.

PINKAS-KRAMARSKI, R. et al. Neu differentiation factor/neuregulin isoforms activate distinct receptor combinations. **Journal of Biological Chemistry**, v. 271, n. 32, p. 19029–19032, 1996.

RAJPAL, A. et al. A general method for greatly improving the affinity of antibodies by using combinatorial libraries. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 24, p. 8466–8471, 2005.

RAZAI, A. et al. Molecular evolution of antibody affinity for sensitive detection of botulinum neurotoxin type A. **Journal of Molecular Biology**, v. 351, n. 1, p. 158–169, 2005.

ROSANO, G. L.; CECCARELLI, E. A. Recombinant protein expression in *Escherichia coli*: Advances and challenges. **Frontiers in Microbiology**, v. 5, n. APR, p. 1–17, 2014.

ROSKOSKI, R. The ErbB/HER family of protein-tyrosine kinases and cancer. **Pharmacological Research**, v. 79, p. 34–74, 2014a.

ROSKOSKI, R. ErbB/HER protein-tyrosine kinases: Structures and small molecule inhibitors. **Pharmacological Research**, v. 87, p. 42–59, 2014b.

ROVIS, T. L.; LEGNAME, G. Prion protein-specific antibodies-development, modes of action and therapeutics application. **Viruses**, v. 6, n. 10, p. 3719–3737, 2014.

RYAN, W. et al. Growth kinetics of *Escherichia coli* and expression of a recombinant protein and its isoforms under heat shock conditions. **Biotechnology Progress**, v. 12, n. 5, p. 596–601, 1996.

SAERENS, D. et al. Identification of a universal VHH framework to graft non-canonical antigen-binding loops of camel single-domain antibodies. **Journal of Molecular Biology**, v. 352, n. 3, p. 597–607, 2005.

SALEMA, V. et al. Selection of Single Domain Antibodies from Immune Libraries Displayed on the Surface of *E. coli* Cells with Two ??-Domains of Opposite Topologies. **PLoS ONE**, v. 8, n. 9, 2013.

SALEMA, V.; FERNÁNDEZ, L. Á. *Escherichia coli* surface display for the selection of nanobodies. **Microbial Biotechnology**, v. 10, n. 6, p. 1468–1484, 2017.

SATO, K. et al. Solution structure of epiregulin and the effect of its C-terminal domain for receptor binding affinity. **FEBS Letters**, v. 553, n. 3, p. 232–238, 2003.

SCHAEFER, G.; FITZPATRICK, V. D.; SLIWKOWSKI, M. X.  $\gamma$ -heregulin: A novel heregulin isoform that is an autocrine growth factor for the human breast cancer cell line, MDA-MB-175. **Oncogene**, v. 15, n. 12, p. 1385–1394, 1997.

SCHMID, M. et al. Effect of sodium sulfite, sodium dodecyl sulfate, and urea on the molecular interactions and properties of whey protein isolate-based films. **Frontiers in Chemistry**, v. 5, n. JAN, p. 1–15, 2017.

SCHNEIDER, M. R.; WOLF, E. The epidermal growth factor receptor ligands at a glance. **Journal of Cellular Physiology**, v. 218, n. 3, p. 460–466, mar. 2009.

SCHNEIDER, M. R.; YARDEN, Y. Structure and function of epigen, the last EGFR ligand. **Seminars in Cell and Developmental Biology**, v. 28, p. 57–61, 2014.

SCHONBRUNN, A. Editorial: Antibody Can Get It Right: Confronting Problems of Antibody Specificity and Irreproducibility. **Molecular Endocrinology**, v. 28, n. 9, p. 1403–1407, 2014.

SELA-CULANG, I.; KUNIK, V.; OFRAN, Y. The structural basis of antibody-antigen recognition. **Frontiers in Immunology**, v. 4, n. OCT, p. 1–13, 2013.

SINGH, B.; CARPENTER, G.; COFFEY, R. J. EGF receptor ligands: recent advances. **F1000Research**, v. 5, n. 0, p. 2270, 8 set. 2016.

SINGH, V. et al. Strategies for fermentation medium optimization: An in-depth review. **Frontiers in Microbiology**, v. 7, n. JAN, 2017.

SONG, I. J. et al. Molecular characterization and expression analysis of mouse epidermal growth factor-like domain 8. **International Journal of Molecular Medicine**, v. 36, n. 2, p. 541–550, 2015.

TAABODI, M. et al. *Aeromonas hydrophila*, *Bacillus thuringiensis*, *Escherichia coli* and *Pseudomonas aeruginosa* utilization of Ammonium-N, Nitrate-N and Urea-N in culture. **Heliyon**, v. 6, n. 4, p. e03711, 2020.

TANHA, J. et al. Selection by phage display of llama conventional VH fragments with heavy chain antibody VHH properties. **Journal of Immunological Methods**, v. 263, p. 97–109, 2002.

TEN-CATEN, F. et al. Internal RNAs overlapping coding sequences can drive the production of alternative proteins in archaea. **RNA Biology**, v. 15, n. 8, p. 1119–1132, 2018.

TICKLE, S. et al. High-Throughput Screening for High Affinity Antibodies. **Journal of the Association for Laboratory Automation**, v. 14, n. 5, p. 303–307, 2009.

TSENG, C. L.; LENG, C. H. Influence of medium components on the expression of recombinant lipoproteins in *Escherichia coli*. **Applied Microbiology and Biotechnology**, v. 93, n. 4, p. 1539–1552, 2012.

VAGENENDE, V.; YAP, M. G. S.; TROUT, B. L. Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. **Biochemistry**, v. 48, n. 46, p. 11084–11096, 2009.

VAZ, M. R. F. et al. Influence of culture medium on the production of EIF antigen from *leishmania chagasi* in recombinant *Escherichia coli*. **Brazilian Journal of Microbiology**, v. 42, n. 4, p. 1390–1396, 2011.

VOSKUIL, J. Commercial antibodies and their validation. **F1000Research**, v. 3, n. 0, p. 232, 2014.

WANG, H. et al. Improving the expression of recombinant proteins in *E. coli* BL21 (DE3) under acetate stress: An alkaline pH shift approach. **PLoS ONE**, v. 9, n. 11, p. 1–11, 2014.

WANG, Y. C. et al. Neuregulin 3 Genetic Variations and Susceptibility to Schizophrenia in a Chinese Population. **Biological Psychiatry**, v. 64, n. 12, p. 1093–1096, 2008.

WEAVER-FELDHAUS, J. et al. Yeast mating for combinatorial Fab library generation and surface display. **FEBS Letters**, v. 564, p. 24–34, 2004.

WEBER, J.; PENG, H.; RADER, C. From rabbit antibody repertoires to rabbit monoclonal antibodies. **Experimental and Molecular Medicine**, v. 49, n. 3, p. e305-12, 2017.

WELLER, M. G. Quality Issues of Research Antibodies. **Analytical Chemistry Insights**, v. 2016, n. 11, p. 21–27, 2016.

WOUTERS, M. A. et al. Evolution of distinct EGF domains with specific functions. **Protein Science**, v. 14, n. 4, p. 1091–1103, 2005a.

WOUTERS, M. A. et al. Evolution of distinct EGF domains with specific functions. **Protein Science**, v. 14, n. 4, p. 1091–1103, 2005b.

XU, Q. et al. A bacterial surface display system expressing cleavable capsid proteins of human norovirus: A novel system to discover candidate receptors. **Frontiers in Microbiology**, v. 8, n. DEC, p. 1–9, 2017.

YAN, J. et al. Construction of a synthetic phage-displayed Nanobody library with CDR3 regions randomized by trinucleotide cassettes for diagnostic applications. **Journal of translational medicine**, v. 12, n. 343, p. 1–12, 2014.

ZARSCHLER, K. et al. High-yield production of functional soluble single-domain antibodies in the cytoplasm of Escherichia coli. **Microbial cell factories**, v. 12, n. 1, p. 97, 2013.

ZENG, F.; HARRIS, R. C. Epidermal growth factor, from gene organization to bedside. **Seminars in Cell & Developmental Biology**, v. 28, p. 2–11, 2014.

ZHANG, H. et al. Metabolic Engineering of Escherichia coli for Ectoine Production With a Fermentation Strategy of Supplementing the Amino Donor. **Frontiers in Bioengineering and Biotechnology**, v. 10, n. January, p. 1–11, 2022.

ZOELEN, E. J. J. V. et al. The EGF Domain: Requirements for Binding to Receptors of the ErbB Family. **Vitamins and Hormones**, v. 59, p. 99–131, 2000.



## 8. APÊNDICE I – SEQUÊNCIAS DAS PROTEÍNAS EM NUCLEOTÍDEOS E AMINOÁCIDOS

As sequências de nucleotídeos indicam as sequências sintetizadas conforme as especificações aos fabricantes e leva em conta seus métodos de otimização de sequência para os organismos alvos ou a sequência final após a amplificação por PCR. As sequências de aminoácidos dizem respeito ao ponto de início da tradução da proteína no vetor em que elas foram subclonadas. Nas proteínas no pET32a foi adicionado na sequência proteica o símbolo // no ponto de proteólise pela trombina. As sequências da TRX foram coloridas de cinza, os domínios EGF foram coloridos de azul claro, as sequências da SGFP2 foram coloridas de verde claro, a sequência da mRFP de vermelho, as sequências das proteínas de superfície bacteriana foram coloridas de rosa e todas as sequências de fragmentos de anticorpos foram coloridas de amarelo.

AREG: Sequência de nucleotídeos

CCATGGCAAGTGTGCGTGTGGAACAGGTGGTTAAACCGCCGCAGAATAAAACCG  
AAAGTGAAAATACCAGCGATAAACCGAAACGTAAAAAAAAAAGGCGGTAAAAAT  
GGTAAAAACCGCCGTAATCGTAAAAAAAAAGAATCCGTGCAATGCAGAATTTTCAG  
AATTTTTGTATCCACGGCGAATGTAAATATATTGAACATCTGGAAGCAGTGACCT  
GCAAATGTCAGCAGGAATATTTTGGCGAACGTTGCGGCGAAAAAAGCATGAAAA  
CCCATAGTATGATTGATAGTAGCCTGAGTAAGCTTTAACTCGAG

TRX-AREG: Sequência de aminoácidos (pET32a)

MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVA  
KLNIDQNPGTAPKYGIRGIPTLLLKNGEVAATKVGALSKGQLKEFLDANLAGSGSG  
HMHSHHHHSSGLVPR//GSGMKETAALKFERQHMDSPDLGTDDEKAMASVRVEQ  
VVKPPQNKTESENTSDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCHGECKYI  
EHLEAVTCKCQEQYFGERCGEKSMKTHSMIDSSLSKL-

TRX-BTC: Sequência de nucleotídeos

CCATGGCAGATGGCAATAGCACCCGCAGCCCGGAAACCAACGGTCTGCTGTGTG  
GCGATCCGGAGGAAAATTGCGCAGCAACCACAACCCAGAGTAAACGCAAAGGCC  
ATTTTCAGTCGTTGCCCGAAACAGTACAAACATTATTGTATTAAGGGCCGCTGCCG

TTTTGTGGTTGCCGAACAGACCCCTAGTTGTGTGTGTGATGAAGGTTATATTGGCG  
CGCGTTGCGAACGTGTTGATCTGTTTTACAAGCTTTAACTCGAG

TRX-BTC: Sequência de aminoácidos (pEt32a)

MSDKIIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVA  
KLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSQKQLKEFLDANLAGSGSG  
HMHSHHHHSSGLVPR//GSGMKETAALKFERQHMDSPDLGTDDDDKAMADGNSTRS  
PETNGLLCGDPEENCAATTTQSKRKGHFSRCPKQYKHYCIKGRCRFVVAEQTPSCVC  
DEGYIGARCERVDLFYKL-

TRX-EGF: Sequência de nucleotídeos

CCATGGCTAATAGCGATAGTGAATGTCCGCTGAGCCATGATGGTTATTGCCTGCA  
TGATGGTGTGTGTATGTATATTGAAGCCCTGGATAAATATGCATGTAATTGCGTG  
GTTGGCTATATTGGTGAACGCTGCCAGTATCGTGATCTGAAATGGTGGGAACTGC  
GTAAGCTTTAACTCGAG

TRX-EGF: Sequência de aminoácidos (pEt32a)

MSDKIIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVA  
KLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSQKQLKEFLDANLAGSGSG  
HMHSHHHHSSGLVPR//GSGMKETAALKFERQHMDSPDLGTDDDDKAMANSSECP  
LSHDGYCLHDGVCMYIEALDKYACNCVVGYIGERCQYRDLKWWELRKL-

TRX-EGF-mRFP: Sequência de nucleotídeos

CCATGGCTAATAGCGATAGTGAATGTCCGCTGAGCCATGATGGTTATTGCCTGCA  
TGATGGTGTGTGTATGTATATTGAAGCCCTGGATAAATATGCATGTAATTGCGTG  
GTTGGCTATATTGGTGAACGCTGCCAGTATCGTGATCTGAAATGGTGGGAACTGC  
GTAAGCTTGGCGGTGGTAGCGACAACACCGAGGACGTCATCAAGGAGTTCATGC  
AGTTCAAGGTGCGCATGGAGGGCTCCGTGAACGGCCACTACTTCGAGATCGAGG  
GCGAGGGGCGAGGGCAAGCCCTACGAGGGCACCCAGACCGCCAAGCTGCAGGTGA  
CCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCAGTTCCAGTA

CGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACATGAAGCTG  
 TCCTTCCCCGAGGGCTTCACCTGGGAGCGCTCCATGAACTTCGAGGACGGCGGCG  
 TGGTGGAGGTGCAGCAGGACTCCTCCCTGCAGGACGGCACCTTCATCTACAAGGT  
 GAAGTTCAAGGGCGTGAACCTCCCCGCCGACGGCCCCGTAATGCAGAAGAAGAC  
 TGCCGGCTGGGAGCCCTCCACCGAGAAGCTGTACCCCCAGGACGGCGTGCTGAA  
 GGGCGAGATCTCCACGCCCTGAAGCTGAAGGACGGCGGCCACTACACCTGCGA  
 CTTCAAGACCGTGTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCAACCACTA  
 CGTGGACTCCAAGCTGGACATCACCAACCACAACGAGGACTACACCGTGGTGGGA  
 GCAGTACGAGCACGCCGAGGCCCGCCACTCCGGCTCCCAGTGACTCGAG

TRX-EGF-mRFP: Sequência de aminoácidos (pEt32a)

MSDKIIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVA  
 KLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSG  
 HMMMMHHSSGLVPR//GSGMKETAATAKFERQHMDSPDLGTDDDDKAMANSDSECP  
 LSHDGYCLHDGVCMYIEALDKYACNCVVG YIGERCQYRDLKWWELRKLGGGS DNT  
 EDVIKEFMQFKVRMEGSVNGHYFEIEGEGEGKPYEGTQTAKLQVTKGGPLPFAWDIL  
 SPQFQYGSKA YVKHPADIPDYMKLSFPEGFTWERSMNFEDGGVVEVQQDSSLQDGT  
 FIYKVKFKGVNFPADGPMQKKTAGWEPSTEKLYPQDGV LKGEISHALKLKDGGHY  
 TCDFKT VYKAKKPVQLPGNH YVDSKLDITNHNEDYTVVEQYEHAEARHSGSQ-

EPGN: Sequência de nucleotídeos

CCATGGCTGAAGGCCCGATTGCCCTGAAATTTAGTCATCTGTGTCTGGAAGATCA  
 TAATAGCTATTGTATTAACGGTGCATGTGCCTTTCATCATGAACTGGAAAAAGCC  
 ATTTGTCGCTGTTTTACCGGCTATAACGGTGAACGCTGCGAACATCTGACCCTGAC  
 CAGTTATGCAGTGGATAGTTATGAAAAGCTTTAACTCGAG

EPGN: Sequência de aminoácidos (pEt32a)

MSDKIIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVA  
 KLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSG

HMHHHHHHSSGLVPR//GSGMKETA AAKFERQHMDSPDLGTDDDDKAMAEGPIALK  
 FSHLCLEDHNSYCINGACAFHHELEKAICRCFTGYTGERCEHLTLTSYAVDSYEK-

EPR: Sequência de nucleotídeos

AGATCTTGTCAGCATCACCAAATGCTCCTCTGACATGAACGGCTATTGTCTGCAC  
 GGCCAGTGTATCTACCTGGTAGACATGTCCCAGA AACTACTGCCGTTGTGAAGTGG  
 GTTATACTGGTGTTCGCTGCGAGCACTTTTTCTGTAAGAATTC

EPR: Sequência de aminoácidos (pEt32a)

MSDKIIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVA  
 KLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSG  
 HMHHHHHHSSGLVPR//GSGMKETA AAKFERQHMDSPDLVSITKCSSDMNGYCLHG  
 QCIYLVDMSQNYCRCEVGYTGVRCEHFFL-

HBEGF: Sequência de nucleotídeos

CCATGGCAGATCTGCAGGAAGCCGATCTGGATCTGCTGCGTGTTACCCTGAGTAG  
 CAAACCGCAGGCCCTGGCAACCCCGAATAAAGAAGAACATGGCAAACGCAAAAA  
 AAAAGGTAAAGGTCTGGGCAAAAAACGTGATCCGTGCCTGCGTAAATATAAAGA  
 TTTTGCATTACGGCGAATGCAAATATGTTAAAGAACTGCGTGCACCGAGTTGC  
 ATTTGTCATCCGGGCTATCATGGCGAACGTTGTCATGGCCTGAGTCTGAAGCTTTA  
 ACTCGAG

HBEGF: Sequência de aminoácidos (pEt32a)

MSDKIIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVA  
 KLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSG  
 HMHHHHHHSSGLVPR//GSGMKETA AAKFERQHMDSPDLGTDDDDKAMADLQEAD

LDLLRVTLSSKPQALATPNKEEHGKRKKKGKGLGKKRDPCLRKYKDFCIHGECKYV  
KELRAPSCICHPGYHGERCHGLSLKL-

NRG1 $\alpha$ : Sequência de nucleotídeos

CCATGGGTACCAGTCATCTGGTGAAATGTGCCGAAAAAGAAAAACCTTTTGCCT  
TAATGGTGGCGAATGCTTTATGGTGAAAGATCTGAGCAATCCGAGCCGCTATCTG  
TGCAAATGCCAGCCGGGCTTTACCGGCGCACGCTGTACCGAAAATGTGCCGATGA  
AAGTTCAGAATCAGGAAAAAGCCGAAGAACTGTATCAGAAGCTTTAACTCGAG

NRG1 $\alpha$ : Sequência de aminoácidos (pEt32a)

MSDKIIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVA  
KLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSCKGQLKEFLDANLAGSGSG  
HMHSHHHSSGLVPR//GSGMKETAATAKFERQHMDSPDLGTDDDDKAMGTSHLVK  
CAEKEKTFCVNGGECFMVKDLSNPSRYLCKCQPGFTGARCTENVPMKVQNQEKAEE  
LYQKL-

TGF $\alpha$ : Sequência de nucleotídeos

CCATGGCTGTTGTGAGTCATTTTAATGATTGTCCGGATAGTCATACCCAGTTTTGC  
TTTCATGGTACCTGTCGTTTTCTGGTGCAGGAAGATAAACCGGCCTGCGTGTGCC  
ATAGTGGCTATGTGGGTGCACGCTGTGAACATGCAGATCTGCTGGCCAAGCTTTA  
ACTCGAG

TGF $\alpha$ : Sequência de aminoácidos (pEt32a)

MSDKIIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVA  
KLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSCKGQLKEFLDANLAGSGSG  
HMHSHHHSSGLVPR//GSGMKETAATAKFERQHMDSPDLGTDDDDKAMAVVSHFN  
DCPDSHTQFCFHGTCTRFLVQEDKPACVCHSGYVGARCEHADLLAKL-

OC: Sequência de nucleotídeos

CCATGGAAATCCAATTGCAACAATCAGGTCCAGAATTGGTGAAACCAGGTGCTA  
GTGTAAAGATAAGTTGTAAGGCTTCTTTGGGTGGATTTGGTTACTTTAAGTCTTCT

GGATACTGGTTGGCATGGGTTAGACAAGCTCCTGGTAAGGGTCTGGAATGGTCTG  
 CTGCTTTGGATGTGAGAGGTGGTTTGTACTCTACTGCTTATGCTTCTTGGGCTAAG  
 GCTTTTGATTTGAGTAAGACTTCTGAGGCTGCTAAGGATATGCATACTTCTTTGGC  
 TGATACTGCTACTTACTTTTGTGCTAGAGATGGTATGCCATCTTCTTCTCAGCCTG  
 CTGGTAAGTTGTGGGGTCCAGGTACTTTGGTTACTGTTTCTTCTGGCGGTGGTAGC  
 GGTGGTGGTGGCGGTGAATTCGGCGGTTCTGGTGGCGGTGGCTCTCAACTTATGA  
 CTAAGACCCCATCCCCAGTCTCCGCTGCTGTCGGAGGTACCGTCACCATCAAATG  
 TCAAGCCTCCGAAGATATCTACAACAGACTTGCTTGGTACCAACAACAACCAGGT  
 AAAAGACCAAAAATTGTTGCTTTACGATGCTTCTAACTTGAATCTGGAGTTCAT  
 CTAGACATCCAGGTCCGGAAGTGGTACTGACTTTTCCGTAAGAATTTCTATGGTC  
 GAAGCTGAAGATTGTGCTACTTACTACGCTCAAGGTGCTGGTCATGTAACCTCTG  
 GTGATAACGCCTTCGGTGGTGGTACTGAAGTTGTTGTTAAGTCGAC

OC: Sequência de aminoácidos (pET28a)

MEIQLQQSGPELVKPGASVKISCKASLGGFGYFKSSGYWLAWVRQAPGKGLEWSAA  
 LDVRGGLYSTAYASWAKAFDLSKTSEAAKDMHTSLADTATYFCARDGMPSSSQPAG  
 KLWGPGLTVTVSSGGSGGGGGEGFGSGGGGSQMLTKTPSPVSAAVGGTVTIKCQA  
 SEDIYNRLAWYQQQPGKRPKLLLYDASNLESGVPSRHPGSGSGTDFSVRISMVEAED  
 CATYYAQGAGHVTSGDNAFGGGTEVVVKLEHHHHHH-

OCG: Sequência de nucleotídeos

CCATGGAAATCCAATTGCAACAATCAGGTCCAGAATTGGTGAAACCAGGTGCTA  
 GTGTAAAGATAAGTTGTAAGGCTTCTTTGGGTGGATTTGGTTACTTTAAGTCTTCT  
 GGATACTGGTTGGCATGGGTTAGACAAGCTCCTGGTAAGGGTCTGGAATGGTCTG  
 CTGCTTTGGATGTGAGAGGTGGTTTGTACTCTACTGCTTATGCTTCTTGGGCTAAG  
 GCTTTTGATTTGAGTAAGACTTCTGAGGCTGCTAAGGATATGCATACTTCTTTGGC  
 TGATACTGCTACTTACTTTTGTGCTAGAGATGGTATGCCATCTTCTTCTCAGCCTG  
 CTGGTAAGTTGTGGGGTCCAGGTACTTTGGTTACTGTTTCTTCTGGCGGTGGTAGC  
 GGTGGTGGTGGCGGTGAATTCGGCGGTTCTGGTGGCGGTGGCTCTCAACTTATGA  
 CTAAGACCCCATCCCCAGTCTCCGCTGCTGTCGGAGGTACCGTCACCATCAAATG  
 TCAAGCCTCCGAAGATATCTACAACAGACTTGCTTGGTACCAACAACAACCAGGT  
 AAAAGACCAAAAATTGTTGCTTTACGATGCTTCTAACTTGAATCTGGAGTTCAT  
 CTAGACATCCAGGTCCGGAAGTGGTACTGACTTTTCCGTAAGAATTTCTATGGTC

GAAGCTGAAGATTGTGCTACTTACTACGCTCAAGGTGCTGGTCATGTAACCTCTG  
 GTGATAACGCCTTCGGTGGTGGTACTGAAGTTGTTGTTAAGGTTCGACAGCGGCTC  
 TGGTATGGTCTCCAAAGGTGAAGAACTGTTACCCGGCGTGGTTCCGATTCTGGTT  
 GAACTGGATGGTGACGTCAACGGCCATAAATTTAGTGTTCGGCGAAGGTGAAG  
 GCGATGCGACCTATGGTAAACTGACGCTGAAGTTCATCTGCACCACGGGTAAACT  
 GCCGGTGCCGTGGCCGACCCTGGTTACCACGCTGACGTATGGTGTGCAGTGT  
 TTTGCACGCTACCCGGATCATATGAAACAACACGACTTTTTCAAGTCAGCGATGCCGG  
 AAGGTTATGTTTCAGGAACGTACCATTTTCTTTAAAGATGACGGCAATTACAAGAC  
 CCGCGCCGAAGTCAAATTTGAAGGTGATACCCTGGTGAACCGTATTGAACTGAAA  
 GGCATCGATTTCAAGGAAGACGGTAATATCCTGGGCCATAAACTGGAATATAACT  
 ACAATAGCCACAACGTGTATATTACCGCTGATAAACAGAAGAACGGCATCAAAG  
 CGAACTTCAAGATCCGCCATAACATCGAAGATGGCGGTGTTCAACTGGCAGACCA  
 CTATCAGCAAAACACCCCGATTGGTGTGGCCCGGTCTGCTGCCGGACAATCAT  
 TACCTGTCAACCCAGTCGAAACTGAGCAAAGATCCGAACGAAAAACGTGACCAC  
 ATGGTCCTGCTGGAATTTGTGACCGCGGCCGGTATCACGCTGGGCATGGATGAAC  
 TGTATAAAGCCCTCGAGCACCACCACCACCACACTGATAACTCGAG

OCG: Sequência de aminoácidos (pET28a)

MEIQLQQSGPELVKPGASVKISCKASLGGFGYFKSSGYWLAWVRQAPGKGLEWSAA  
 LDVRGGLYSTAYASWAKAFDLSKTSEAAKDMHTSLADTATYFCARDGMPSSSQPAG  
 KLWGPGLTVTVSSGGGSGGGGGEGFSGSGGGGSQLMTKTPSPVSAAVGGTVTIKCQA  
 SEDIYNRLAWYQQQPGKRPKLLYDASNLESGVPSRHPGSGSGTDFSVRISMVEAED  
 CATYYAQGAGHVTSGDNAFGGGTEVVVKVESGSGMVSKEELFTGVVPILVELDGD  
 VNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPTLVTTLTYGVCFARYPDH  
 MKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLVNRIELKGIDFKEDG  
 NILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDG  
 PVLLPDNHYLSTQSKLSKDPNEKRDHMLLEFVTAAGITLGMDELYKALEHHHHHHH-

320C: Sequência de nucleotídeos

ATGAGCGATAAAATTATTCACCTGACTGACGACAGTTTTGACACGGATGTACTCA  
 AAGCGGACGGGGCGATCCTCGTCGATTTCTGGGCAGAGTGGTGCGGTCCGTGCAA  
 AATGATCGCCCCGATTCTGGATGAAATCGCTGACGAATATCAGGGCAAACCTGACC  
 GTTGCAAAACTGAACATCGATCAAAACCCTGGCACTGCGCCGAAATATGGCATCC



GTGGTATCCCGACTCTGCTGCTGTTCAAAAACGGTGAAGTGGCGGCAACCAAAGT  
GGGTGCACTGTCTAAAGGTCAGTTGAAAGAGTTCCTCGACGCTAACCTGGCCGGT  
TCTGGTTCTGGCCATATGCACCATCATCATCATTCTTCTGGTCTGGTGCCACG  
CGGTTCTGGTATGAAAGAAACCGCTGCTGCTAAATTCGAACGCCAGCACATGGAC  
AGCCCAGATCTGGGTACCGACGACGACGACAAGGCCATGGAAATCCAATTGCAA  
CAATCAGGTCCAGAATTGGTCAAACCAGGTGCTAGTGTAAGATAAGTTGTAAG  
GCTTCTTTGGGTGGATTTGGTTACTTTAAGTCTTCTGGATACTGGTTGGCATGGGT  
TAGACAAGCTCCTGGTAAGGGTCTGGAATGGTCTGCTGCTTTGGATGTGAGAGGT  
GGTTTGTACTCTACTGCTTATGCTTCTTGGGCTAAGGCTTTTGATTTGAGTAAGAC  
TTCTGAGGCTGCTAAGGATATGCATACTTCTTTGGCTGATACTGCTACTTACTTTT  
GTGCTAGAGATGGTATGCCATCTTCTTCTCAGCCTGCTGGTAAGTTGTGGGGTCCA  
GGTACTTTGGTTACTGTTTCTTCTGGCGGTGGTAGCGGTGGTGGTGGCGGTGAATT  
CGGCGGTTCTGGTGGCGGTGGCTCTCAACTTATGACTAAGACCCCATCCCCAGTC  
TCCGCTGCTGTTCGGAGGTACCGTCACCATCAAATGTCAAGCCTCCGAAGATATCT  
ACAACAGACTTGCTTGGTACCAACAACAACCAGGTAAAAGACCAAAAATTGTTGCT  
TTACGATGCTTCTAACTTGGAAATCTGGAGTTCCATCTAGACATCCAGGTTCCGGA  
AGTGGTACTGACTTTTCCGTAAGAATTTCTATGGTTCGAAGCTGAAGATTGTGCTA  
CTTACTACGCTCAAGGTGCTGGTCATGTAACCTCTGGTGATAACGCCTTCGGTGGT  
GGTACTGAAGTTGTTGTTAAGGTCGACAGCGGCTCTGGTATGGTCTCCAAAGGTG  
AAGAAGTGTTCACCGGCGTGGTTCCGATTCTGGTTGAACTGGATGGTGACGTCAA  
CGGCCATAAATTTAGTGTTTCCGGCGAAGGTGAAGGCGATGCGACCTATGGTAAA  
CTGACGCTGAAGTTCATCTGCACCACGGGTAAACTGCCGGTGCCGTGGCCGACCC  
TGGTTACCACGCTGACGTATGGTGTGCAGTGTTTTGCACGCTACCCGGATCATAT  
GAAACAACACGACTTTTTCAAGTCAGCGATGCCGGAAGGTTATGTTTCAAGAACGT  
ACCATTTTCTTTAAAGATGACGGCAATTACAAGACCCGCGCCGAAGTCAAATTTG  
AAGGTGATACCCTGGTGAACCGTATTGAACTGAAAGGCATCGATTTCAAGGAAG  
ACGGTAATATCCTGGGCCATAAACTGGAATATAACTACAATAGCCACAACGTGTA  
TATTACCGCTGATAAACAGAAGAACGGCATCAAAGCGAACTTCAAGATCCGCCA  
TAACATCGAAGATGGCGGTGTTCAACTGGCAGACCACTATCAGCAAAAACACCCC  
GATTGGTGATGGCCCGGTCTGCTGCCGGACAATCATTACCTGTCAACCCAGTCG  
AAACTGAGCAAAGATCCGAACGAAAAACGTGACCACATGGTCCTGCTGGAATTT

GTGACCGCGGCCGGTATCACGCTGGGCATGGATGAACTGTATAAAGCCCTCGAGC  
ACCACCACCACCACCACTGATAACTCGAG

320C: Sequência de aminoácidos (pET32a)

MSDKIIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVA  
KLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSCKGQLKEFLDANLAGSGSG  
HMHSHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTDDDDKAMEIQLQQSGP  
ELVKPGASVKISCKASLGGFGYFKSSGYWLAWVRQAPGKGLEWSAALDVRGGLYST  
AYASWAKAFDLSKTSEAAKDMHTSLADTATYFCARDGMPSSSQPAGKLWGPGTLV  
TVSSGGGSGGGGGEGFGGSGGGGSQLMTKTPSPVSAAVGGTVTIKCQASEDIYNRLA  
WYQQQPGRPKLLLYDASNLESGVPSRHPGSGSGTDFSVRISMVEAEDCATYYAQQ  
AGHVTSGDNAFGGGTEVVVKVDSGSGMVSKGEELFTGVVPILVELDGDVNGHKFSV  
SGEGEGDATYGKLTCLKFICTTGKLPVPWPTLVTTLTLYGVQCFARYPDHMKQHDFFKS  
AMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEY  
NYNSHNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHY  
LSTQSKLSKDPNEKRDHMLVLEFVTAAGITLGMDELYKALEHHHHHHH-

MM: Sequência de nucleotídeos

CCATGGAAATCCAATTGCAACAATCAGGTCCAGAATTGGTGAAACCAGGTGCTA  
GTGTAAAGGTGAGTTGTAAGGCTTCTGGTTACTCTTTTACTGATTACAACCTTTAT  
TGGGTGAAACAAAGTCATGGTAAGTCCTTGGAATGGATCGGTTACATTGATACTT  
ACGATGGTGGTACAAATTACGTTGATAGATTTAAGGGTAAGGCTACTTTGACTGT  
TGATACTGATAGTTCCACTGCCTTTTTGCATTTGGATTCTTTGACTTCTGAAGATTC  
TGCAGTTTACTATTGTGCCAGACCAGCTGGTTTTGCTTACTGGGGACAAGGTA  
TGGTTACTGTTTCTGCTGGCGGTGGTAGCGGTGGTGGTGGCGGTGAATTCGGCGG  
TTCTGGTGGCGGTGGCTCTGATGTTATGATGACTCAAACCTCCATTGTCTCTTCCAG  
TTTCTCTTGGTGATCAAGCTTCTATTTCTTGTAGAGCCTCTTCTTCTTTACTTCATT  
CTAACGGTAAGACTTACCTTGAATGGTACTTGCAAAAACCCGGACAATCTCCTAG  
ACAATTGCTTTACCAACTTTCTAACAGATTTTCTGGTGTTCCTGATAGATTTTCTG  
GATCTGGTTCTGGTACTGATTTTACTTTGAAGATTAGAAGAGTGGAAGCTGAGGA

TCTTGCCGTTTATTACTGCTTTCAAGGTTCTCATTTTCCATGGACTTTTGGTGGTGG  
TACTAAGTTGGAAATTAAGAGAGTCGAC

MM: Sequência de aminoácidos (pET28a)

MSEIQLQQSGPELVKPGASVKVSKASGYSFTDYNLYWVKQSHGKSLEWIGYIDTYD  
GGTNYVDRFKGKATLTVDTDSSTAFLHLDSLSEDSAVYYCARPAGFAYWGQGLV  
TVSAGGGSGGGGGSGGSGGGGSDVMMTQTPLSLPVSLGDQASISCRASSLLHSNG  
KTYLEWYLQKPGQSPRQLLYQLSNRFSGVPDFRFSGSGSGTDFTLKIRRVEAEDLAVY  
YCFQGSHPWTFGGGKLEIKRLEHHHHHH-

MMG: Sequência de nucleotídeos

CCATGGAAATCCAATTGCAACAATCAGGTCCAGAATTGGTGAAACCAGGTGCTA  
GTGTAAAGGTGAGTTGTAAGGCTTCTGGTTACTCTTTACTGATTACAACCTTTAT  
TGGGTGAAACAAAGTCATGGTAAGTCCTTGGAAATGGATCGGTTACATTGATACTT  
ACGATGGTGGTACAAATTACGTTGATAGATTTAAGGGTAAGGCTACTTTGACTGT  
TGATACTGATAGTTCCACTGCCTTTTTGCATTTGGATTCTTTGACTTCTGAAGATTC  
TGCAGTTTACTATTGTGCCAGACCAGCTGGTTTTGCTTACTGGGGACAAGGTA  
TGGTTACTGTTTCTGCTGGCGGTGGTAGCGGTGGTGGTGGCGGTGAATTCGGCGG  
TTCTGGTGGCGGTGGCTCTGATGTTATGATGACTCAAACCTCATTGTCTCTTCCAG  
TTTCTCTTGGTGATCAAGCTTCTATTTCTTGTAGAGCCTCTTCTTCTTACTTCATT  
CTAACGGTAAGACTTACCTTGAATGGTACTTGCAAAAACCCGGACAATCTCCTAG  
ACAATTGCTTTACCAACTTTCTAACAGATTTTCTGGTGTTCCTGATAGATTTTCTG  
GATCTGGTTCTGGTACTGATTTTACTTTGAAGATTAGAAGAGTGGAAGCTGAGGA  
TCTTGCCGTTTATTACTGCTTTCAAGGTTCTCATTTTCCATGGACTTTTGGTGGTGG  
TACTAAGTTGGAAATTAAGAGAGTCGACAGCGGCTCTGGTATGGTCTCCAAAGGT  
GAAGAACTGTTACCCGGCGTGGTTCCGATTCTGGTTGAACTGGATGGTGACGTCA  
ACGGCCATAAATTTAGTGTTTCCGGCGAAGGTGAAGGCGATGCGACCTATGGTAA  
ACTGACGCTGAAGTTCATCTGCACCACGGGTAAACTGCCGGTGCCGTGGCCGACC  
CTGGTTACCACGCTGACGTATGGTGTGCAGTGTTTTGCACGCTACCCGGATCATAT  
GAAACAACACGACTTTTTCAAGTCAGCGATGCCGGAAGGTTATGTTTCAGGAACGT  
ACCATTTTCTTTAAAGATGACGGCAATTACAAGACCCGCGCCGAAGTCAAATTTG  
AAGGTGATACCCTGGTGAACCGTATTGAACTGAAAGGCATCGATTTCAAGGAAG

ACGGTAATATCCTGGGCCATAAACTGGAATATAACTACAATAGCCACAACGTGTA  
 TATTACCGCTGATAAACAGAAGAACGGCATCAAAGCGAACTTCAAGATCCGCCA  
 TAACATCGAAGATGGCGGTGTTCAACTGGCAGACCACTATCAGCAAAAACACCCC  
 GATTGGTGATGGCCCCGGTCCTGCTGCCGGACAATCATTACCTGTCAACCCAGTCG  
 AAACTGAGCAAAGATCCGAACGAAAAACGTGACCACATGGTCCTGCTGGAATTT  
 GTGACCGCGGCCGGTATCACGCTGGGCATGGATGAACTGTATAAAGCCCTCGAGC  
 ACCACCACCACCACCTAACTCGAG

MMG: Sequência de aminoácidos (pET28a)

MEIQLQQSGPELVKPGASVKVSCKASGYSFTDYNLYWVKQSHGKSLEWIGYIDTYD  
 GGTNYVDRFKGKATLTVDTDSSTAFLHLDSLSEDSAVYYCARPAGFAYWGQGLV  
 TVSAGGGSGGGGGEGFSGGGGSDVMMTQTPLSLPVS LGDQASISCRASSLLHSNG  
 KTYLEWYLQKPGQSPRQLLYQLSNRFSGV PDRFSGSGSGTDFTLKIRRVEAEDLAVY  
 YCFQGSHPWTFGGGKLEIKRVE SSGSMVSKGEELFTGVVPI LVELDGDVNGHKFS  
 VSGEGEGDATYGKLT LKFICTTGKLPVPWPTLVTTLYGVQCFARYPDHMKQHDFE  
 KSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLE  
 YNYSNHNVIYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNH  
 YLSTQSKLSKDPNEKRDHMLLEFVTAAGITLGMDELYKALEHHHHHHH-

32MM: Sequência de nucleotídeos

ATGAGCGATAAAATTATTCACCTGACTGACGACAGTTTTGACACGGATGTACTCA  
 AAGCGGACGGGGCGATCCTCGTCGATTTCTGGGCAGAGTGGTGCGGTCCGTGCAA  
 AATGATCGCCCCGATTCTGGATGAAATCGCTGACGAATATCAGGGCAAACCTGACC  
 GTTGCAAACCTGAACATCGATCAAACCCTGGCACTGCGCCGAAATATGGCATCC  
 GTGGTATCCCGACTCTGCTGCTGTTCAAAAACGGTGAAGTGGCGGCAACCAAAGT  
 GGGTGC ACTGTCTAAAGGTCAGTTGAAAGAGTTCCTCGACGCTAACCTGGCCGGT  
 TCTGGTTCTGGCCATATGCACCATCATCATCATTCTTCTGGTCTGGTGCCACG  
 CGGTTCTGGTATGAAAGAAACCGCTGCTGCTAAATTCGAACGCCAGCACATGGAC  
 AGCCCAGATCTGGGTACCGACGACGACGACAAGGCCATGGAAATCCAATTGCAA  
 CAATCAGGTCCAGAATTGGTGA AACCAGGTGCTAGTGTA AAGGTGAGTTGTAAG  
 GCTTCTGGTTACTCTTTTACTGATTACAACCTTTATTGGGTGAAACAAAGTCATGG  
 TAAGTCCTTGGAATGGATCGGTTACATTGATACTTACGATGGTGGTACAAATTAC  
 GTTGATAGATTTAAGGGTAAGGCTACTTTGACTGTTGATACTGATAGTTCCACTGC

CTTTTGCATTTGGATTCTTTGACTTCTGAAGATTCTGCAGTTTACTATTGTGCCAG  
 ACCAGCTGGTTTTGCTTACTGGGGACAAGGTACTTTGGTTACTGTTTCTGCTGGCG  
 GTGGTAGCGGTGGTGGTGGCGGTGAATTCGGCGGTTCTGGTGGCGGTGGCTCTGA  
 TGTTATGATGACTCAAACCTCATTGTCTCTTCCAGTTTCTCTTGGTGATCAAGCTT  
 CTATTTCTTGTAGAGCCTCTTCTTCTTACTTCATTCTAACGGTAAGACTTACCTTG  
 AATGGTACTTGCAAAAACCCGGACAATCTCCTAGACAATTGCTTTACCAACTTTC  
 TAACAGATTTTCTGGTGTTCCTGATAGATTTTCTGGATCTGGTTCTGGTACTGATT  
 TTACTTTGAAGATTAGAAGAGTGGAAGCTGAGGATCTTGCCGTTTATTACTGCTTT  
 CAAGGTTCTCATTTTCCATGGACTTTTGGTGGTGGTACTAAGTTGGAAATTAAGA  
 GAGTCGAC

32MM: Sequência de aminoácidos (pET32a)

MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVA  
 KLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSG  
 HMHHHHHHSSGLVPRGSGMKETA AAKFERQHMDSPDLGTDDDDKA **MEIQLQSGP**  
**ELVKPGASVKVSCASGYSFTDYNLYWVKQSHGKSLEWIGYIDTYDGGTNYVDRFK**  
**GKATLTVDTDSSTAFLHLDSLSEDSAVYYCARPAGFAYWGQGLVTVSAGGGSGG**  
**GGGGSGGGGGSDVMMTQTPLSLPVSLGDQASISCRASSLLHSNGKTYLEWYLQ**  
**KPGQSPRQLLYQLSNRFSGVPDRFSGSGSDFTLKIRRVEAEDLA VYYCFQGSHPW**  
**TFGGGTKLEIKR**

NB1: Sequência de nucleotídeos

CCATGGGTCAGGTACAGCTGGTAGAATCTGGTGGTGGTTCTGTGCAAGCAGGCGG  
 TTCTCTGCGTCTGTCTTGCACCGCATCTGGTATCATCTTCTCCATCTATGCGTGGTT  
 CCGTCAAGCACCTGGCCAAGAACGTGAAGCTGTTGCGATCACACCGGTGGCTCT  
 CGCTTCACTATCTCCCGTGACAACGCCAAAAACACGGTAACGCTGCAGATGAACC  
 TGAAGCCGGAAGATACTGCAATCTACTACTGCGCGGCTAATGCACTGCCGCTGGT

CGATTACGGTACCGGCTTCGAATATGACTACTGGGGTCAAGGCACTCAGGTGACC  
GTTTCCTCTAAGCTTCATCATCACCATCACCCTAACTCGAG

NB1: Sequência de aminoácidos (pET28a)

MGQVQLVESGGGSVQAGGSLRLSCTASGIIFSIYAWFRQAPGQEREAVAITTGGSRFT  
ISRDNKNTVTLQMNLPEDTAIYYCAANALPLVDYGTGFEYDYWGQGTQVTVSSK  
LHHHHH-

NB6: Sequência de nucleotídeos

CCATGGGTCAAGTACAGCTGGTTGAAAGCGGTGGTGGTTCTGTACAGGCTGGCGG  
TAGCCTGCGTCTGTCTTGTACCGCCTCCGGCTTTACCTTCAGCACCTACGTATGGT  
TCCGCCAGGCTCCGGGTCAGGAACGTGAAGCTGTTGCTATCTCTACCGTTGGCGG  
CCTGCGCTTTACCATCTCTCGTGACAACGCTAAGAATACGGTAACCCTGCAGATG  
AACCTGAAGCCGGAAGATACGGCAATCTACTATTGCGCAGCTAAAGCCGACGGC  
CGTCGCTACGATATCAATTATTGGGGTCAGGGTACCCAAGTTACTGTTTCTAGCA  
AGCTTCATCATCACCATCACCCTAACTCGAG

NB6: Sequência de aminoácidos (pET28a)

MGQVQLVESGGGSVQAGGSLRLSCTASGFTFSTYVWFRQAPGQEREAVAISTVGGGL  
RFTISRDNKNTVTLQMNLPEDTAIYYCAAADGRRYDINYWGQGTQVTVSSKHLH  
HHHHH-

iNAQ: Sequência de nucleotídeos

AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGAT  
AACAAATCCCCTGTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATAT  
GAATTTAGATAAAGCTCTAGTATTGAGGACGTGCGCAAACAACATGGCAGACCA  
TTGTGGTCTTATCTGGCCAGCTAGCGGTACTGTAGAGAGCAAATATTGGCAGAGC  
ACCCGTCGTCATGAAAACGGTCTGGTGGGTTTGCTGTGGGGTGCAGGTTACCTCAG  
CGTTCCTGTCAGTTCACGCAGATGCTCGCTGGATTGTTTCGTGAAGTTGCGGTTGCG  
GACATCATTAGCCTGGAAGAGCCGGGTATGGTTAAGTTCCCGAGAGCTGAGGTG  
GTCCATGTTGGTGACCGTATTAGCGCCTCGCACTTTATTAGCGCGCGCCAGGCGG

ACCCGGCAAGCACGCCGACTCCGACCCCGACGCCGATGACGGCGGCGACCCCGC  
 CACCGACCCCGGCAACCGCTAATGTTACCCTGCCGGTGGCGGAGCAGGCCAGCC  
 ACGAGGTTTTTGTATGTGGCGCTGGTGAGCGCCGCAGCTCCGCCTGTGAATACCCT  
 GCCGGTCACCACCCCGCAGAATCTGCAAACCGCGACCTATGGTTCCACTTTGAGC  
 GCGGACAACCACTCTCGTCTGATCGCGGGCTATGGCTCCAATGAAACCGCCGGTA  
 ACCACAGCGACCTGATCGCCGGCTACGGCTCTACCGGTACCGCCGGGTCCGATTC  
 CAGCCTGGTGGCCGGTTATGGTTCGACCCAAACGGCGGGAGGTGACTCCGCGCTC  
 ACCGCGGGCTACGGTTCTACCCAAACGGCTCGCGAGGGCTCTAACTTGACCGCTG  
 GCTACGGCAGCACAGGCACTGCGGGCTCTGACAGTAGCTTAATCGCGGGTTACGG  
 CAGTACCCAAACCAGCGGTGAAGATAGCTCCTTGACCGCGGGCTACGGAAGCAC  
 GCAAACCGCTCTGGTTCCGCGTGGTAGTCATCACCACCACCACCACATGGCCATG  
 GAGGCCTAATTAATTTAACTCGAG

iNAQ: Sequência de aminoácidos (pET28a)

MNLDKALVLRTCANNMADHCLIWPAVSGTVESKYWQSTRRHENGLVGLLWGAGTS  
 AFLSVHADARWIVREVAVADIISLEPGMVKFPRAEVVHVGDRISASHFISARQADPA  
 STPTPTPTPMTAATPPPTPATANVTLPVAEQASHEVFDVALVSAAAPPVNTLPVTPQ  
 NLQTATYGSTLSGDNHSRLIAGYGSNETAGNHSDLIAGYGSTGTAGSDSSLVAGYGS  
 TQTAGGDSALTAGYGSTQTAREGSNLTAGYGSTGTAGSDSSLIAGYGSTQTSGEDSS  
 LTAGYGSTQTALVPRGSHHHHHHMAMEA-

NEAE: Sequência de nucleotídeos

AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGAT  
 AACCAATCCCCTGTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATAT  
 GATAACACACGGATTTTATGCTAGGACTCGCCATAAACACAACTAAAGAAGAC  
 CTTTATTATGCTGTCTGCGGGCCTCGGCCTGTTCTTCTATGTTAACCAAAATAGCT  
 TCGCGAACGGCGAAAACACTTCAAACCTGGGTTCGGATAGCAAGTTGTTGACCCA  
 TAACTCTTACCAGAATCGTCTGTTTTACTCTGAAGACGGGTGAGACAGTTGCT  
 GACCTGAGCAAAAGCCAAGACATCAACCTGAGCACCATTTGGAGCCTGAATAAA  
 CACCTGTATTCCAGCGAGAGCGAAATGATGAAAGCAGCTCCGGGTCAACAAATC  
 ATTCTGCCGCTGAAGAACTGCCGTTTCGAGTACAGCGCGCTGCCACTGTTGGGTA  
 GCGCGCCTCTGGTTGCCGCGGGTGGAGTGGCTGGTCATACCAATAAGTTAACCAA  
 GATGAGCCCGGATGTTACGAAGTCAAATATGACTGATGACAAGGCGCTCAACTA



CGCAGCCCAACAGGCGGCGAGCTTGGGTAGCCAAGCAGATCCCTTAA  
 TGGCGACTACGCCAAAGATACGGCTCTTGGCATCGCCGGCAATCAGGCCAGCAG  
 CCAACTGCAAGCCTGGCTGCAACACTATGGCACGGCGGAGGTGAATCTGCAATCC  
 GGTAACAATTTTGACGGCTCAAGCCTCGATTTCTGCTGCCATTTTACGACAGCG  
 AGAAGATGTTGGCCTTTGGTCAAGTGGGCGCGCGTTACATCGACAGCCGCTTCAC  
 CGCGAACCTGGGCGCAGGCCAACGTTTCTTTCTGCCGAAAACATGCTGGGCTAC  
 AATGTGTTTCATCGACCAAGACTTCAGCGGCGACAACACCCGTCTGGGTATTGGTG  
 GCGAGTATTGGCGTGATTACTTCAAAGCTCCGTCAACGGTTATTTCCGCATGTCT  
 GGCTGGCATGAATCCTACAACAAAAAGGACTACGACGAACGTCCGGCGAATGGT  
 TTTGACATCCGTTTTAACGGTTACTTGCCGAGCTACCCGGCGCTCGGTGCAAACT  
 GATGTATGAACAGTATTACGGCGATAACGTCGCGTTGTTCAACAGCGACAAGCTG  
 CAAAGCAATCCGGGTGCTGCAACCGTTGGTGTTAACTATACCCCGATTCCGCTTG  
 TTACGATGGGCATCGACTACCGTCATGGTACGGGTAACGAGAACGATCTGTTATA  
 TAGCATGCAGTTTCGCTACCAGTTCGACAAGCCGTGGTCTCAACAGATTGAACCG  
 CAGTATGTGAACGAACACTGCGTACCCTATCGGGCTCTCGCTACGACCTGGTGCAGC  
 GCAACAACAATATTATCCTGGAGTACAAAAAGCAGGACATCCTGTGCTTGAACAT  
 CCCGCATGATATCAATGGCACGGAGCGTTCCACCCAGAAAATTCAGCTGATCGTT  
 AAATCGAAATACGGCCTGGATCGTATTGTGTGGGATGATTCGGCCTTAAGAAGCC  
 AAGGTGGTCAGATTCAGCACAGCGGTTCCAGAGCGCGCAAGATTACCAGGCCAA  
 TTTTGCCAGCGTATGTTTCAGGGTGGTTCTAATGTCTATAAAGTTACCGCACGTGCG  
 TACGACCGCAATGGTAATTCGAGCAACAACGTTCTGTTAACCATTACCGTGTTGT  
 CCAACGGTCAGGTCGTTGATCAGGTTGGCGTGACTGATTTACCGCTGACAAAAC  
 CTCTGCCAAGGCCGATGGCACCGAAGCAATCACCTATACCGCGACCGTCAAGAA  
 GAACGGGGTGGCGCAGGCCAACGTGCCGGTTAGCTTTAATATCGTATCTGGTACT  
 GCGGTGTTGTCTGCGAACTCCGCGAACACCAACGGCTCCGGCAAAGCGACGGTG  
 ACCCTTAAGAGTGACAAACCGGGTCAGGTTGTGGTGAGCGCGAAGACTGCTGAG  
 ATGACCAGCGCTCTCAATGCAAACGCTGTGATTTTTGTTGATCAAACCAAAGCGA  
 GCGGCGCTCCCGTACCGTATCCGGATCTGGAACCGGCGCTGGTTCCGCGTGGTAG  
 TCATCACCACCACCACCATGGCCATGGAGGCCTAATTAATTTAACTCGAG

NEAE: Sequência de aminoácidos (pET28a)

MITHGFYARTRHKHKLKKTFFIMLSAGLGLFFYVNQNSFANGENYFKLGSDSKLLTHN  
 SYQNRLFYTLKTGETVADLSKSQDINLSTIWSLNKHLYSSESEMKAAPGQQIILPLK

KLPFEYSALPLLGSAPLVAAGGVAGHTNKLTKMSPDVTKSNMTDDKALNYAAQQA  
 ASLGSQLQSRSLNGDYAKDTALGIAGNQASSQLQAWLQHYGTAEVNLQSGNNFDGS  
 SLDFLLPFYDSEKMLAFGQVGARYIDSRFTANLGAGQRFFLPENMLGYNVFIDQDFS  
 GDNTRLGIGGEYWRDYFKSSVNGYFRMSGWHESYNKKDYDERPANGFDIRFNGYLP  
 SYPALGAKLMYEQYYGDNVALFNSDKLQSNPGAATVGVNYTPIPLVTMGIDYRHGT  
 GNENDLLYSMQFRYQFDKPWSQQIEPQYVNELRTLSGSRYDLVQRNNNIILEYKKQD  
 ILSLNIPHDINGTERSTQKIQLIVKSKYGLDRIVWDDSAALRSQGGQIQHSGSQAQDYQ  
 AILPAYVQGGSNVYKVTARAYDRNGNSSNNVLLTITVLSNGQVVDQVGVTDFTADK  
 TSAKADGTEAITYTATVKKNGVAQANVPVSNIVSGTAVLSANSANTNGSGKATVTL  
 KSDKPGQVVVSAKTAEMTSALNANAVIFVDQTKASGAPVPYPDLEPALVPRGSHHH  
 HHHMAMEA-

SGFP2: Sequência de nucleotídeos

AGGCCTAGCGGCTCTGGTATGGTCTCCAAAGGTGAAGAACTGTTACCGGCGTGG  
 TTCCGATTCTGGTTGAACTGGATGGTGACGTCAACGGCCATAAATTTAGTGTTTCC  
 GGCGAAGGTGAAGGCGATGCGACCTATGGTAAACTGACGCTGAAGTTCATCTGC  
 ACCACGGGTAAACTGCCGGTGCCGTGGCCGACCCTGGTTACCACGCTGACGTATG  
 GTGTGCAGTGTTTTGCACGCTACCCGGATCATATGAAACAACACGACTTTTTCAA  
 GTCAGCGATGCCGGAAGGTTATGTTTCAGGAACGTACCATTTTCTTTAAAGATGAC  
 GGCAATTACAAGACCCGCGCCGAAGTCAAATTTGAAGGTGATACCCTGGTGAAC  
 CGTATTGAACTGAAAGGCATCGATTTCAAGGAAGACGGTAATATCCTGGGCCATA  
 AACTGGAATATAACTACAATAGCCACAACGTGTATATTACCGCTGATAAACAGAA  
 GAACGGCATCAAAGCGAACTTCAAGATCCGCCATAACATCGAAGATGGCGGTGT  
 TCAACTGGCAGACCACTATCAGCAAAACACCCCGATTGGTGATGGCCCGGTCTCTG  
 CTGCCGGACAATCATTACCTGTCAACCCAGTCGAAACTGAGCAAAGATCCGAACG  
 AAAAACGTGACCACATGGTCCTGCTGGAATTTGTGACCGCGGCCGGTATCACGCT  
 GGGCATGGATGAACTGTATAAAGCCCTCGAGCACCACCACCACCACCACTGACTC  
 GAG

SGFP2: Sequência de aminoácidos (pET28a)

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVP  
 WPTLVTTLTYGVCFAFYDPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEV

KFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHN  
 IEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMLLEFVTAA  
 GITLGMDELYKALEHHHHHH-

4NIKMM: Sequência de nucleotídeos

CCATGGCAGAAGTACAGTTAGTTGAATCAGGGGGAAGTTTAGTGAAACCGGGCG  
 GCAGCCTCAGACTGTCGTGCGCTGCCTCTGGTTATTCTTTCACGGACTACAACCTG  
 TACTGGGTCCGCCAGGCCCGGGCAAAGGTTTGAATGGATCTCCTATATTGATA  
 CTTATGATGGTGGCACCAACTACGTTGACCGTTTTAAAGGCCGTTTTACAATCAG  
 CCGTGATAATGCTACCAATTCATTGTACCTGCAGATGAACTCGCTGCGAGCGGAA  
 GATACTGCGGTCTATTATTGTGTGCGTCCTGCAGGGTTTGCTTATTGGGGGCGCGG  
 TACCCTGGTTACGGTCAGCGCGGGGGGGGGCGGTTTCGGGCGGCGGAGGCTCCGG  
 TGGCGGCGGCAGTCAAAGCGTGCTTACCCAGCCAGCCTCTGTATCGGGTAGCCCA  
 GGTCAAGTCAATTACTATCTCTTGTCGGGCCTCATCGAGCCTGTTACACAGTAATGG  
 TAAACATATCTGGAATGGTATCAACAACATCCCGGCAAAGCCCCGAAGCTGAT  
 GATTTACCAGCTCTCTAACCGCTTCAGCGGAGTGTCCAACCGCTTTTCAGGTAGTA  
 AGTCCGGCAATACGGCCTCCCTAACCATAGTGGCCTGCAAGCGGAGGATGAGG  
 CAGACTATTACTGCTTCCAGGGTTCGCATTTCCCGTGGACCTTTGGTGGGGGTACG  
 AACTGGCGGTGCTTGGAGCGGCCGCGTCAGCAAAGCTTCACCACCACCACCACC  
 ACTAAGAATTC

4NIKMM: Sequência de aminoácidos (pET28a)

MEVQLVESGGSLVKPGGSLRLSCAASGYSFTDYNLYWVRQAPGKGLEWISYIDTYD  
 GGTNYVDRFKGRFTISRDNATNSLYLQMNSLRAEDTAVYYCVRPAGFAYWGRGTLV  
 TVSAGGGGSGGGGSGGGGSQSVLTQPASVSGSPGQSITISCRASSLLHSNGKTYLEW  
 YQQHPGKAPKLMYQLSNRFSGVSNRFSGSKSGNTASLTISGLQAEDEADYYCFQGS  
 HFPWTFGGGTKLAVLGAAASAKLHHHHHH-

148MM: Sequência de nucleotídeos

CCATGGAAGTGCAACTGGTGGAGAGTGGTGGGGGTTTAGTGCAACCGAAAGGTA  
 GTCTGAAAATTTTCATGCGCGGCCTCCGGTTATAGCTTTCACGGACTACAACCTTATAC  
 TGGGTCCGTCAGGCTCCGGGCAAAGGGTCTAGAATGGGTTGCGTACATCGATACCT

ATGATGGCGGCACCAATTACGTTGACCGGTTCAAAGGGCGTTTTACGATTTCCCG  
 CGATGACAGCCAGAGTATGTTGTATCTGCAGATGAATAACCTCAAACTGAGGAT  
 ACTGCTATGTATTATTGCGTTCGCCCTGCAGGTTTCGCGTATTGGGGACAGGGGA  
 CGACTGACCGTGTTCGTCTGCAGGAGGAGGCGGCTCAGGTGGCGGGCGGCTCGG  
 GTGGGGGGGGTTCGGCGGTGGCGGCAGTTACATACAAATGACCCAGTCACCAG  
 CCAGTCTGAGCGTCTCCGTAGGGGAGACCGTGACGATTACCTGTCGCGCCTCTAG  
 TTCGTTTTTGCCTCGAACGGAAAACTACCTGGAATGGTATCAGCAGAAACAA  
 GGTAAGAGCCCGCAATTACTTGTCTATCAACTGTCTAACCGCTTTTCCGGCGTACC  
 GTCGCGTTTCTCAGGCAGCGGCTCGGGTACGCAGTTTTCCCTGAAAATCAATTCTC  
 TACAGAGCGAAGATTTCCGGCACATACTATTGTTTTTCAGGGTAGCCATTTCCCCTG  
 GACCTTTGGCAGCGGTACTAACTGGAAATCAAGCGTAGCAAGCTTCACCACCAC  
 CACCACCACTAAGAATTC

148MM: Sequência de aminoácidos (pET28a)

MEVQLVESGGGLVQPKGSLKISCAASGYSFTDYNLYWVRQAPGKGLEWVAYIDTYD  
 GGTNYVDRFKGRFTISRDDSQSMLYLQMNNLKTEDTAMYCYCVRPAGFAYWGQGT  
 LTVSSAGGGGSGGGGSGGGGSGGGGSYIQMTQSPASLSVSVGETVTITCRASSLLHS  
 NGKTYLEWYQKQKSPQLLVYQLSNRFSGVPSRFSGSGSGTQFSLKINSLQSEDFGT  
 YYCFQGSHPWTFGSGTKLEIKRSLHHHHHH-

146MM: Sequência de nucleotídeos

CCATGGCAGAAGTTCAACTCGTTCAATCCGGTGCGGAAGTAGCTAAACCGGGGG  
 CGAGTGTAAGAGTGTCGTGTAAAGCATCTGGGTACAGTTTTACGGACTACAATCT  
 GTATTGGGTCCGCCAGGCCCCAGGACAAGGGTTGGAGTGGATCGGTTATATTGAT  
 ACCTATGATGGTGGCACGAACTACGTGGATCGGTTCAAGGGGAAAGCGACTCTG  
 ACAGCTGACAAGTCGACTTCAACGGCGTATCTGGAAGTGTCAAGCCTTCGTTCCG  
 AAGATACCGCTGTGTACTACTGCGCCCGACCAGCCGGCTTTGCGTACTGGGGACA  
 GGGCACGCTGGTGACCGTTTCATCCTCAGGCGGCGGTTCCGGGGGCGGCGGCACC  
 GGTGGCGGCGGTAGCATAGTCATGACGCAGTCTCCGCTTAGCCTACCGGTTACCC  
 CGGGTGAACCGGCGAGCATCTCCTGCCGTGCAAGCAGCTCCCTGTTACATTCGAA  
 CGGCAAACTTATTTAGAGTGGTATTTGCAAAAACCGGGACAGTCCCCCAGCTG  
 CTGATTTACCAGCTCAGCAATCGTTTTTCAGGTGTGCCTGATCGTTTTAGTGGTAG  
 TGGCTCTGGTACAGATTTTACCCTGAAAATCAGTCGCGTAGAGGCCGAAGACGTC

GGAGTTTATTATTGTTTCCAGGGTTCGCACTTCCCTTGGACCTTCGGGCAGGGCAC  
CAAAGTGGAAATTAAGCGCAAGCTTCACCACCACCACCACCACTAAGAATTC

146MM: Sequência de aminoácidos (pET28a)

MEVQLVQSGAEVAKPGASVKVSKASGYSFTDYNLYWVRQAPGQGLEWIGYIDTY  
DGGTNYVDRFKGKATLTADKSTSTAYLELSSLRSEDVAVYYCARPAGFAYWGQGLT  
VTVSSSGGGSGGGGTGGGGSIIVMTQSPLSLPVTPGEPASISCRASSLLHSNGKTYLE  
WYLQKPGQSPQLLIYQLSNRFSGVPDFRSGSGSDFTLKISRVEAEDVGVYYCFQGS  
HFPWTFGQGTKVEIKRKLHHHHHH-

VERMM: Sequência de nucleotídeos

CCATGGAAATCCAACCTGGTAGAAAGCGGTGGTTCTCTGGTTAAGCCAGGTGGCTC  
CCTGCGTCTGAGCTGCGCAGCAAGCGGTTACAGCTTACCGACTACAATCTGTAT  
TGGGTTTCGTCAGGCACCAGGTAAAGGCCTGGAGTGGATCGGTTACATTGATACCT  
ACGATGGTGGCACTAACTATGTGGATCGCTTCAAAGGTCGTGCCACCCTGTCCGT  
GGACACCGCTACTAACTCTGCGTACCTGCAACTGAACTCTCTGCGTGCTGAAGAC  
ACTGCTGTGTACTACTGTGCACGTCCAGCAGGCTTCGCGTATTGGGGTCGCGGCA  
CTCTGGTGACTGTTTCTGCTGGTGGTGGTGGCTCCGGTGGTGGTGGTCCGGTGGT  
GGTGGTAGCCAGGTGGTTATGACCCAGCCAGCTTCCGTAAGCGGTTCTCCAGGCC  
AGTCTATCACTATCTCTTGCCGTGCTTCTAGCTCCCTGCTGCACTCTAACGGCAAG  
ACCTACCTGGAGTGGTATCAACAGCACCCAGGCAAAGCTCCGAAACAACCTGCTGT  
ACCAACTGTCCAACCGCTTCTCTGGCGTATCTAACCGTTTCAGCGGTTCTGGCTCC  
GGTACTACCTTCTCTCTGACTATCTCCGGTCTGCAAGCCGAAGACGAAGCCGACT  
ACTATTGCTTTCAGGGTTCTCACTTCCCGTGGACCTTCGGTGGTGGTACTAACTG  
GCTGTACTGGGTGCAGCAGCAAGCGCAAAGCTTCATCATCATCATCACCACCTAAG  
AATTC

VERMM: Sequência de aminoácidos (pET28a)

MEIQLVESGGSVLPKPGSLRLSCLASGYSFTDYNLYWVRQAPGKGLEWIGYIDTYDG  
GTNYVDRFKGRATLSVDTATNSAYLQLNSLRAEDVAVYYCARPAGFAYWGRGLV  
TVSAGGGGSGGGGSGGGGSQVVMTPASVSGSPGQSITISCRASSLLHSNGKTYLE

WYQQHPGKAPKQLLYQLSNRFSGVSNRFSGSGSGTTFSLTISGLQAEDEADY YCFQG  
SHFPWTFGGGTKLAVLGAAASAKLHHHHHH-

6CVKOC: Sequência de nucleotídeos

CCATGGCTGCGGAACTGGTACTGACCCAGACTCCGTCTTCCGTTAGCGCAGCGGT  
AGGTGGTACTGTGACCATCAACTGTCAGGCTTCTGAAGACATCTACAACCGTCTG  
GCATGGTATCAGCAGAAACCAGGCCAACACCGAAACTGCTGATCTACGACGCG  
TCCAACCTGGAGAGCGGTGTTCCATCTCGCTTCAAGGGCTCTGGTCCGGTACTG  
AGTTCACCCTGACTATCTCCGGTGTTC AACGTGATGACGCAGCTACTTACTACTGT  
GCAACCTACTACGCTCAGGGTGCAGGTCACGTTACTAGCGGTGACAACGCGTTCG  
GTGGTGGTACTGAAGTGGTTGTGAAAGGTGGTAGCTCCCGTAGCAGCTCTTCTGG  
TGGTGGTGGCTCTGGTGGTGGTGGTCAGGAACA ACTGGTTGAATCTGGTGGTCGT  
CTGGTTCACCAGGTGGCAGCCTGACCCTGACCTGCACCGTATCCGGTGGTTTCG  
GTTACTTCAAATCCTCTGGCTACTGGGTTCGCCAGGCACCAGGTAAAGGTCTGGA  
ATACATCGGTGCTCTGGACGTTCTGGTGGTCTGTATTCTACTGCTTACGCAAGCT  
GGGCTAAAGCGTTCGATCTGTCTCGTTTCACTATCAGCAAGACCAGCACC ACTGT  
GGCTCTGAAGATGAGCACTCTGACTGCAAGCGACACCGCAACTTACTTCTGTGCT  
CGTGACGGTATGCCATCTAGCAGCCAGCCAGCAGGCAAGCTGTGGGGTCTGGGC  
ACCCTGGTAACTATCAGCTCTCTGGTTCACGCGGCTCTaagcttCACCATCACCACC  
ACCACTAAGAATTC

6CVKOC: Sequência de aminoácidos (pET28a)

MELVLTQTPSSVSAAVGGT VTINCQASEDIYNRLAWYQQKPGQPPKLLIYDASNLES  
GVPSRFKGS GSGTEFTLTISGVQRDDAATYYCATYYAQGAGHVTSGDNAFGGGTEV  
VVKGGSSRSSSSGGGGSGGGGQEQLVESGGRLVPPGGS LTLTCTVSGGFYFKSSGY  
WVRQAPGK GLEYIGALDVRGGLYSTAYASWAKAFDLSRFTISK TSTTVALKMSTLTA  
SDTATYFCARDGMPSSSQPAGKLWGLGTLVTISSLVPRGSKLHHHHHH-

146OC: Sequência de nucleotídeos

CCATGGCCGAGGTCCAGCTTGTACAAAGCGGGGCCGAGGTAGCGAAGCCGGGGG  
CGAGCGTCAAAGTTAGTTGTAAAGCAAGTGGGGGATTTGGCTACTTTAAGAGCTC  
GGGTTATTGGGTGCGTCAGGCACCCGGTCAGGGCCTGGAATGGATAGGAGCCCT

AGATGTCAGGGGTGGCCTGTACTCGACCGCGTACGCCTCATGGGCGAAAGCATT  
 GACCTGTCCAAAGCCACCCTGACTGCTGATAAATCTACCTCCACCGCGTACCTGG  
 AATTAAGCTCCCTTCGGTCGGAAGATACGGCGGTGTATTACTGCGCCCGCGACGG  
 CATGCCTTCAAGTTCTCAGCCGGCTGGTAAGCTGTGGGGCCAGGGA ACTCTCGTG  
 ACCGTGTCATCAAGTGGCGGGGGTTCCGGGGGTGGAGGTACAGGTGGGGGCGGG  
 TCCATTGTCATGACACAAAGTCCACTCAGCTTGCCGGTAACCCCGGGTGAACCTG  
 CCAGCATCAGCTGTCAGGCGTCTGAGGATATTTATAACCGCTTAGCCTGGTATCTT  
 CAGAAACCAGGCCAATCACCGCAGCTACTGATCTACGACGCCTCGAATCTGGAA  
 AGTGGTGTTCGGACCGTTTCAGCGGATCTGGCAGCGGGACTGATTTACGCTGA  
 AAATCAGTCGTGTAGAAGCGGAAGATGTCGGCGTGTATTATTGCGCTACGTATTA  
 TGCTCAGGGTGCGGGCCATGTGACGTCGGGTGATAACGCATTCGGCCAGGGCACC  
 AAAGTGGAGATTAACGCAAGCTTCACCACCACCACCACCTAAGAATTC

146OC: Sequência de aminoácidos (pET28a)

MEVQLVQSGAEVAKPGASVKVSKASGGFGYFKSSGYWVRQAPGQGLEWIGALDV  
 RGGLYSTAYASWAKAFDLSKATLTADKSTSTAYLELSSLRSED TAVYYCARDGMPSS  
 SQPAGKLWGQGTLVTVSSSSGGGSGGGGTGGGGSIVMTQSPLSLPVTPEPASISCSQAS  
 EDIYNRLAWYLQKPGQSPQLLIYDASNLESQVDFRFSGSGSGTDFTLKISRVEAEDVG  
 VYYCATYYAQGAGHVTSGDNAFGQGTKVEIKRKLHHHHHHH-

VEROC: Sequência de nucleotídeos

CCATGGA ACTGGTTATGACCCAGACTCCAAGCTCCGTTTCTGCTGCGGTAGGTGG  
 CACCGTTACCATCAACTGCCAGGCGTCTGAAGACATCTACAACCGTCTGGCGTGG  
 TATCAGCAGAAGCCAGGCCAACCACCGAAGCTGCTGCTGTACGACGCGTCTAATC  
 TGGA AAGCGGTGTGCCGTTCCGTTTCAAAGGTTCCGGTTCTGGCACCGAGTTTAC  
 CCTGACCATTTCTGGCGTTCAGCGTGACGACGCAGCCACCTACTACTGCGCAACC  
 TACTACGCTCAGGGTGTGGTACGTAACCTCTGGCGACAACGCTTTCGGTGGTG  
 GTACTGAAGTGGTAGTGAAAGGTGGTAGCTCTCGCTCTAGCTCCTCTGGTGGTGG  
 TGGTTCTGGTGGTGGTGGCCAGATCCA ACTGGTTGAATCTGGTGGTTCGTCTGGTTC  
 CACCAGGTGGCTCTCTGACTCTGACCTGCACCGTTTCTGGTGGCTTCGGCTACTTC  
 AAATCCTCTGGCTATTGGGTGCGTCAGGCACCAGGTAAAGGCCTGGAGTGGTCTG  
 CTGCTCTGGACGTTTCGTGGTGGTCTGTATTCTACCGCATAACGCGAGCTGGGCAAA  
 GCGT TTTGACCTGTCTCGCTTTACTCTGTCCAAGACTTCTACCACTGCTGCACTGA



AGATGTCCACCCTGACTGCTTCTGACACCGCAACCTACTTCTGTGCTCGTGATGGT  
ATGCCGTCCTCTAGCCAACCAGCAGGTAAACTGTGGGGTCTGGGTACTCTGGTGA  
CTATCTCTTCCCTGGTTCCACGTGGTTCTaagcttCACCACCACCACCATCACTAAGA  
ATTC

VEROC: Sequência de aminoácidos (pET28a)

MELVMTQTPSSVSAAVGGTVTINCQASEDIYNRLAWYQQKPGQPPKLLLYDASNLES  
GVPSRFKGSVSGTEFTLTISGVQRDDAATYYCATYYAQGAGHVTSGDNAFGGGTEV  
VVKGGSSRSSSSGGGGSGGGGQIQLVESGGRLVPPGSLTTLCTVSGGFGYFKSSGY  
WVRQAPGKGLEWSAALDVRGGLYSTAYASWAKAFDLSRFTLSKTSTTAALKMSTLT  
ASDTATYFCARDGMPSSSQPAGKLWGLGTLVTISLVPKSKLHHHHHH-