UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES
PROGRAMA DE PÓS GRADUAÇÃO EM BIOQUÍMICA E BIOLOGIA
MOLECULAR

GEISON CASTRO DA SILVEIRA GUELLER

Interação entre o fator de transcrição CG9571 e módulos reguladores do gene *pair-rule even-skipped* da cascata de segmentação de *Drosophila*

São Paulo
2019

GEISON CASTRO DA SILVEIRA GUELLER

Interação entre o fator de transcrição CG9571 e módulos reguladores do gene *pair-rule even-skipped* da cascata de segmentação de *Drosophila*

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção do título de Mestre em Ciências pelo Programa de Pós-Graduação em Bioquímica e Biologia Molecular

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em 22 de maio de 2019. A versão original encontra-se em acervo reservado na Biblioteca da EACH/USP e na Biblioteca Digital de Teses e Dissertações da USP (BDTD), de acordo com a Resolução CoPGr 6018, de 13 de outubro de 2011.

Área de Concentração:
Bioquímica e Biologia Molecular

Orientador:
Prof. Dr. Luiz Paulo de Moura Andrioli

Coorientador:
Prof. Dr. Shaker Chuck Farah

São Paulo
2019

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

CATALOGAÇÃO-NA-PUBLICAÇÃO

(Universidade de São Paulo. Escola de Artes, Ciências e Humanidades. Biblioteca)

CRB-8 4936

Gueller, Geison Castro da Silveira

Interação entre o fator de transcrição CG9571 e módulos reguladores do gene *pair-rule even-skipped* da cascata de segmentação de *Drosophila* / Geison Castro da Silveira Gueller ; orientador, Luiz Paulo de Moura Andrioli ; coorientador, Shaker Chuck Farah. - 2019
105 f. : il.

Dissertação (Mestrado em Ciências) - Programa de Pós-Graduação em Bioquímica e Biologia Molecular, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo
Versão corrigida

1. Bioquímica. 2. Biologia molecular. 3. Genes. 4. Genética molecular. 5. *Drosophila*. I. Andrioli, Luiz Paulo de Moura, orient. II. Farah, Shaker Chuck, coorient. III. Título

CDD 22.ed. – 572

Nome: GUELLER, Geison Castro da Silveira

Título: Interação entre o fator de transcrição CG9571 e módulos reguladores do gene *pair-rule even-skipped* da cascata de segmentação de *Drosophila*

Dissertação apresentada à Escola de Artes,
Ciências e Humanidades da Universidade de
São Paulo para obtenção do título de Mestre
em Ciências do Programa de Pós-Graduação
em Bioquímica e Biologia Molecular

Área de Concentração:
Bioquímica e Biologia Molecular

Aprovado em: 22/05/2019

Banca Examinadora

Profa. Dra. Marcia Aparecida Sperança
Universidade Federal do ABC. Centro de Ciências Naturais e Humanas

Profa. Dra. Regina Lúcia Baldini
Universidade de São Paulo. Instituto de Química

Profa. Dra. Aline Maria da Silva
Universidade de São Paulo. Instituto de Química

AGRADECIMENTOS

Ao Prof. Dr. Andrioli, meu professor na primeira graduação e meu orientador do TCC na segunda graduação e na pós-graduação. Agradeço pela confiança e por ter aberto seu laboratório para que pudesse começar a trabalhar nesta maravilhosa área que é a biologia do desenvolvimento. Por toda cobrança, incentivo e ensino. Pelo exemplo de dedicação e amor à pesquisa científica. E por todo apoio e esforço para que eu pudesse ser transferido para a EACH. Sem esta transferência, nada disso seria possível.

A todos os professores do Programa de Pós-Graduação em Bioquímica e Biologia Molecular. Agradeço por toda contribuição para minha formação através das disciplinas, conversas e experiências nos laboratórios. Em especial, agradeço ao Prof. Dr. Chambergo, pela disponibilização de equipamentos e materiais e auxílio em experimentos, ao Prof. Dr. Ribamar, pelas conversas e referências para a purificação e renaturação de proteínas, e à Prof. Dra. Campana, pelas conversas, sugestões e pelo trabalho inicial de espectroscopia, que não pôde ser desenvolvido para este trabalho, mas que espero em breve poder sê-lo.

A todos os professores da ala de Ecologia, Evolução e Genética – doutores Fernando Carbayo, Helene Ueno, Leonardo Meireles, Marcos Hara, Luís Schiesari, Rodrigo Willemart e Tiago Franco – à ex-diretora Profa. Dra. Maria Cristina de Toledo e às chefias do serviço de laboratórios da EACH – Jefferson Marins e Patrícia Galdino – que apoiaram de diversos modos a minha pós-graduação, principalmente nas liberações para a participação em disciplinas e congressos.

A todos meus colegas do laboratório de Genética do Desenvolvimento e dos demais laboratórios do Conjunto Multidisciplinar de Pesquisa pela colaboração no desenvolvimento deste trabalho. Em especial, ao Filippo, pelo auxílio nos procedimentos com DNA e com o sistema de expressão *in vitro*, Igor e Lucas pelo auxílio nos procedimentos com DNA e à Patrícia e Adriana por todo o imenso auxílio e suporte nos trabalhos com proteínas.

À Fundação de Amparo à Pesquisa do Estado de São Paulo pelo fundamental apoio financeiro concedido ao nosso laboratório que possibilitou a realização deste trabalho.

A todos meus familiares e amigos que através das mais variadas formas possibilitaram a conclusão deste trabalho. Em especial, aos meus sogros, Egle e João, minha irmã, Gesiê, meu pai e madrasta, Saulo e Angélica, e ao meu amigo Jorge, que me deram e continuam a dar suporte e têm ajudado de diversas formas, principalmente nestes últimos dias de maior tribulação. À minha mãe, Gisley, que nunca mediu esforços por nós. É graças a tudo que você fez que pude chegar até aqui.

À minha querida esposa Jessica, por todo apoio e companheirismo nos últimos anos e durante esta árdua rotina de trabalho como técnico e pesquisador em formação. Por toda a paciência e compreensão pelas minhas ausências. Amo você e espero poder lhe compensar por tudo isto em breve.

Meus sinceros agradecimentos a todos que me ajudaram. Nenhum trabalho como este é feito sozinho.

Acima de tudo e todos agradeço a Deus, a quem devo tudo o que tenho e sou. Não fora a completa transformação operada por Ele em mim não sei onde estaria hoje. A Ele toda a honra e glória.

RESUMO

GUELLER, Geison Castro da Silveira. **Interação entre o fator de transcrição CG9571 e módulos reguladores do gene *pair-rule even-skipped* da cascata de segmentação de *Drosophila***. 2019. 105f. Dissertação (Mestrado em Ciências) – Programa de Pós-graduação em Bioquímica e Biologia Molecular. Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2019. Versão corrigida.

O desenvolvimento do padrão de segmentação de *Drosophila* é estabelecido por uma cascata de genes de segmentação zigóticos. Estes genes são divididos em três classes (*gap*, *pair-rule* e *segment-polarity*) e codificam para fatores de transcrição (FT) que se ligam a módulos cis-reguladores (*CRMs*), reprimindo ou ativando genes alvo. A faixa 2 do gene *pair-rule even-skipped* (*eve 2*) é ativada pelos fatores maternos Bicoid e Hunchback e reprimida pelas proteínas *gap* Giant (Gt) e Krüppel. Estudos posteriores mostraram que o FT Sloppy-paired 1 (Slp 1) e provavelmente um outro FT *forkhead* também atuam na repressão de *eve 2*. O gene anotado CG9571 foi isolado em uma varredura como proteína *forkhead* candidata a repressão de *eve 2*. Estudos genéticos confirmaram essa possibilidade e revelaram que *eve 1* também pode ser alvo deste FT. Este trabalho teve como objetivo verificar a interação de CG9571 com os *CRMs eve 1* e *eve 2*. Para tanto, planejamos obter o domínio de ligação da proteína (CG9571 *BD*) e da proteína completa (CG9571 *FL*) e testar suas interações *in vitro* com fragmentos dos *CRMs* por meio da técnica de retardo da mobilidade eletroforética (*EMSA*). Obtivemos a quantidade necessária de DNA para os experimentos através de *PCR* e preparações plasmidiais de versões clonadas destes *CRMs* que já dispúnhamos em laboratório. Realizamos tentativas de obtenção de CG9571 *BD* por transcrição e tradução *in vitro*, mas esta estratégia não foi bem-sucedida e adotamos a estratégia de clonagem em vetor para expressão em células competentes bacterianas. O fragmento de CG9571 *BD* foi clonado com sucesso, mas não conseguimos verificar a expressão do polipeptídeo em duas linhagens de *E. coli*. Alteramos novamente nossa estratégia e clonamos o fragmento correspondente a CG9571 *FL* em vetor de expressão e conseguimos induzir sua expressão em bactéria, embora não tenha sido obter a proteína recombinante em forma solúvel. Prosseguimos para tentativas de recuperação da proteína a partir de corpos de inclusão. Foram empregados diferentes métodos para solubilização, renovelamento e purificação da proteína. Extratos da fração insolúvel solubilizada em diferentes concentrações de ureia foram submetidos a tentativas de purificação e

renaturação por cromatografia de afinidade, mas não houve adsorção significativa de CG9571 *FL* em colunas com Ni^{2+} imobilizado. Preparações não puras contendo CG9571 *FL* foram obtidas através de procedimentos de renaturação destes extratos e foram utilizadas em ensaios de interação com os *CRMs*. Não houve detecção de retardo da mobilidade eletroforética dos fragmentos em gel. Foram observados efeitos de redução da quantidade de DNA detectado com brometo de etídio nas interações, mas este efeito foi considerado produto da ação de possíveis nucleases contaminantes nas preparações após investigação. Preparações de CG9571 *FL* puras foram obtidas por purificações a partir de SDS-PAGE, mas a maioria das interações da proteína solúvel com *eve* 1 e *eve* 2 não indicou formação de complexo. Obtivemos um único resultado positivo para a interação entre CG9571 *FL* e *eve* 2. Por não ter sido reproduzido, consideramos o resultado inconclusivo e novos experimentos serão conduzidos para dar continuidade à investigação da hipótese do trabalho.

Palavras-chave: *Drosophila*. Cascata de Segmentação. Interação DNA-Proteína. *even-skipped*. CG9571

ABSTRACT

GUELLER, Geison Castro da Silveira. **Interaction between the transcription factor CG9571 and cis-regulatory modules of the pair-rule gene *even-skipped* of *Drosophila* segmentation cascade.** 2019. 105f. Dissertation (Master in Science) – Postgraduated Program in Biochemistry and Molecular Biology, School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2019. Corrected version.

The development of *Drosophila* segmentation pattern is established by a cascade of zygotic segmentation genes. The zygotic genes are grouped in three classes (gap, pair-rule and segment-polarity) and code for transcription factors (TF) that bind to cis-regulatory modules (CRMs) with activation or repression roles. The stripe 2 of the pair-rule gene *even-skipped* (*eve* 2) is activated by the maternal factors Bicoid and Hunchback and repressed by the gap proteins Giant and Krüppel. Later studies showed that Sloppy-paired 1 (Slp 1) and probably another forkhead transcription factor also act for *eve* 2 repression. The annotated gene CG9571 was isolated in a search for putative forkhead protein repressors of *eve* 2. Genetic studies confirmed this possibility and revealed that *eve* 1 could also be a target for this TF. The aim of this work was to verify the interaction of CG9571 with the CRMs *eve* 1 and *eve* 2. To reach this aim, we planned to obtain the binding domain of the protein (CG9571 BD) or of the full-length protein (CG9571 FL) and to test their *in vitro* interactions with the *eve* 1 and *eve* 2 fragments by the electrophoretic mobility shift assay (EMSA). We obtained the necessary amount of DNA for the tests by PCR and plasmidial preparations of cloned versions of these CRMs that we already had in our laboratory. We made attempts to obtain CG9571 BD by *in vitro* transcription and translation system, but this strategy did not work and we adopted the cloning strategy to obtain the protein expressed by bacterial competent cells. CG9571 BD was cloned successfully, but we were not able to detect the polypeptide expression in two *E. coli* strains. We then turned to the CG9571 FL protein that we cloned and succeed to express it in bacteria, although we were not able to obtain the soluble recombinant form. We proceed for attempts of protein recovering from inclusion bodies. Different methods for solubilization, refolding and purification of the proteins were used. Extracts of the insoluble fraction solubilized in solutions with different urea concentrations were used in attempts of refolding and purification by affinity chromatography, but there was not significant CG9571 FL adsorption on columns with Ni²⁺ immobilized. We obtained impure preparations with CG9571 FL through procedures of refolding of these extracts and employed them on

binding assays with the CRMs, but there was no gel shift detection. We observed reduction of the amount of DNA present in the binding reaction samples detected by ethidium bromide, but after further investigations this effect was considered the product of contaminant nuclease reaction from bacteria. Pure preparations of CG9571 FL were obtained by purification from SDS-PAGE, but there was no indication of complex formation on the most binding reaction assays with *eve* 1 and *eve* 2. We obtained only one positive result for the interaction between CG9571 FL and *eve* 2. However, the result was considered inconclusive because we were not able to reproduce it and new experiments will be conducted to investigate the hypothesis of this work.

Key-words: *Drosophila*. Segmentation Cascade. Protein-DNA Interaction. *even-skipped*. CG9571

LISTA DE FIGURAS

Figura 1. Ciclo de vida de <i>Drosophila melanogaster</i>	21
Figura 2. Cascata de segmentação em <i>Drosophila</i>	22
Figura 3. Esquema da estrutura do locus de <i>even-skipped</i>	24
Figura 4. Mecanismo proposto para a regulação da expressão de <i>eve 2</i>	25
Figura 5. Domínio ligador de DNA <i>forkhead</i>	26
Figura 6. Domínio de expressão de CG9571 e efeito da expressão ectópica em faixas de <i>eve</i>	27
Figura 7. Sequência de nucleotídeos de <i>eve 2</i> mutada	32
Figura 8. Sequência de nucleotídeos de <i>eve 1</i>	32
Figura 9. Sequência de nucleotídeos de CG9571	33
Figura 10. Mapa do vetor pCR 2.1-TOPO	36
Figura 11. Mapa do vetor pMiniT 2.0	36
Figura 12. Mapa do vetor pET-15b	37
Figura 13. Região de clonagem e expressão do vetor pET-15b	37
Figura 14. Sequência de aminoácidos de CG9571	41
Figura 15. Checagem das construções plasmidiais pTOPO- <i>eve 1</i> , pBS- <i>eve 2</i> rx e pBS- <i>eve 2</i> ΔGTTT com enzimas de restrição	52
Figura 16. Checagem da construção plasmidial pTOPO-CG9571 com enzimas de restrição	53
Figura 17. Detecção de fragmento correspondente a CG9571 BD obtido por PCR para transcrição e tradução <i>in vitro</i>	54
Figura 18. Detecção de polipeptídeo correspondente a CG9571 BD após transcrição e tradução <i>in vitro</i>	55
Figura 19. Detecção de fragmento correspondente a CG9571 BD obtido por PCR para clonagem no vetor de expressão	56
Figura 20. Clonagem de CG9571 BD em pTOPO verificada com enzimas de restrição	57
Figura 21. Clonagem de CG9571 BD em pET-15b verificada com enzimas de restrição	58
Figura 22. Indução da expressão de CG9571 BD em BL21(DE3)	59
Figura 23. Efeito comparativo da indução de CG9571 BD e CG9571 FL sobre a densidade de células bacterianas em culturas celulares.....	61

Figura 24. Detecção de fragmento correspondente a <i>CG9571 FL</i> obtido por <i>PCR</i> para clonagem em vetor carreador	62
Figura 25. Clivagem de DNA plasmidial extraído de colônias geradas na clonagem de <i>CG9571 FL</i> em pTOPO com enzimas de restrição	62
Figura 26. Clivagem de DNA plasmidial extraído de colônias geradas na clonagem de <i>CG9571 FL</i> em pET-15b com enzimas de restrição	63
Figura 27. Indução da expressão de <i>CG9571 FL</i> (1mM IPTG, 37°C).....	64
Figura 28. Indução da expressão de <i>CG9571 FL</i> (0.1/0.2/0.3mM IPTG, 37°C).....	65
Figura 29. Solubilização de <i>CG9571 FL</i> em condições brandas.....	68
Figura 30. Isolamento de corpos de inclusão e solubilização de <i>CG9571 FL</i> em condições desnaturantes	69
Figura 31. Cromatograma da purificação de extrato solubilizado em ureia 4M.....	71
Figura 32. Frações de purificação e renaturação por IMAC de corpos de inclusão solubilizados em ureia 4M.....	71
Figura 33. Frações de purificação por IMAC de corpos de inclusão solubilizados em ureia 8M.....	72
Figura 34. Purificação de <i>CG9571 FL</i> a partir de SDS-PAGE	73
Figura 35. Proteínas obtidas de fração solubilizada em ureia 4M submetida à renaturação por diluição	74
Figura 36. Proteínas obtidas de fração ressolubilizada em ureia 4M submetida à renaturação por diálise.....	76
Figura 37. Interação entre <i>CRMs eve 1</i> e <i>eve 2</i> mt com <i>CG9571 I</i>	79
Figura 38. Interação entre <i>eve 1</i> e <i>CG9571 I</i> e <i>II (I)</i>	80
Figura 39. Interação entre <i>eve 1</i> e <i>CG9571 I</i> e <i>II (II)</i>	81
Figura 40. Interação entre fragmentos de DNA não específicos e <i>CG9571 II</i>	83
Figura 41. Interação entre <i>eve 1</i> e <i>eve 2 wt</i> e <i>CG9571</i> na presença de EDTA e/ou ausência de cloreto de magnésio	84
Figura 42. Interação entre <i>CRMs</i> e DNA não específico e <i>CG9571 VI</i>	85
Figura 43. Interação entre <i>CRMs</i> e <i>CG9571 VI</i>	86
Figura 44. Interação entre DNA não específico e <i>CG9571 VII</i>	87
Figura 45. Alinhamento entre a sequência de <i>CG9571 BD</i> e o resultado de sequenciamento com <i>primer forward</i> de plasmídeo do clone 1 (clonagem do domínio em pTOPO)	101

Figura 46. Alinhamento entre a sequência de <i>CG9571 BD</i> e parte da sequência obtida para plasmídeo do clone 6 (clonagem do domínio em pET) com <i>primer forward</i>	103
Figura 47. Alinhamento entre a sequência de <i>CG9571 BD</i> e parte da sequência obtida para plasmídeo do clone 6 (clonagem do domínio em pET) com <i>primer reverse</i>	103
Figura 48. Alinhamento entre a sequência de <i>CG9571 BD</i> e as sequências obtidas para clone 6 (clonagem do domínio em pET) com <i>primers reverse e forward</i>	104
Figura 49. Trecho da sequência obtida em sequenciamento com <i>primer reverse</i> com identificação de bases e aminoácidos do vetor pET-15b e do início de <i>CG9571 BD</i> ..	105

LISTA DE TABELAS

Tabela 1. <i>Primers</i> sintetizados para amplificação dos fragmentos	34
Tabela 2. Condições dos programas do termociclador para <i>PCR</i>	34
Tabela 3. Soluções para suspensão de células induzidas.....	42
Tabela 4. Tampões para solubilização branda de proteínas de corpos de inclusão.....	43
Tabela 5. Tampões para solubilização de proteínas de corpos de inclusão em condição desnaturante	44
Tabela 6. Soluções tampão para purificação por IMAC	46
Tabela 7. Tampões para solubilização de proteína recombinante precipitada	47
Tabela 8. Composição de tampões para interação DNA-proteína.....	50
Tabela 9. Principais preparações da proteína recombinante CG9571 <i>FL</i> utilizadas em ensaios de interação DNA-proteína (continua)	77

LISTA DE ABREVIATURAS E SIGLAS

aa	Aminoácido
BN-PAGE	Eletroforese em gel de poliacrilamida nativo contendo Coomassie blue G 250
BSA	Albumina bovina sérica
CRM	Módulos cis-reguladores (do inglês, cis-regulatory modules)
DNA	Ácido desoxirribonucleico
dNTP	Desoxirribonucleotídeo
DO	Densidade óptica
DTT	1,4-ditiotreitol
EDTA	Ácido etilenodiamino tetra-acético
EMSA	Ensaio de retardo de mobilidade eletroforética (do inglês, electrophoretic mobility shift assay)
HEPES	Ácido etanosulfônico 4-2 hidroxietil piperazina-1
IMAC	Cromatografia de afinidade por íons metálicos imobilizados (do inglês, immobilized metal affinity chromatography)
IPTG	Isopropil β -D-1-tiogalactopiranosida
Kb	Do inglês, kilobase. Corresponde a 1000pb
KDa	Do inglês, kilodanton. Corresponde a 1000Da
LB	Luria-Bertani
LBA	LB contendo ampicilina
LBC	LB contendo canamicina
mRNA	RNA mensageiro
MWCO	Do inglês, molecular weight cut-off (em dáltons)
NP-40	Nonil fenoxilpolietoxietanol
PAGE	Eletroforese em gel de poliacrilamida
pb	Pares de bases
PCR	Reação em cadeia da polimerase, do inglês
pH	Potencial hidrogeniônico
pI	Ponto isoelétrico
RNA	Ácido ribonucleico
rpm	Rotações por minuto
SDS	Sódio dodecil sulfato, do inglês

SDS-PAGE	Eletroforese em gel de poliacrilamida contendo SDS
SOC	Do inglês, super optimal broth with catabolite repression
TAE	Tampão tris-acetato-edta
TBE	Tampão tris borato edta
Tris	Tris(hidroximetil)aminometano
Tris-HCl	Tampão tris-ácido clorídrico
wt	Selvagem (do inglês, wild type)
x g	Vezeas a aceleração da gravidade
X-gal	5-bromo-4-cloro-3-indoxil β -D-galactopiranosídeo

LISTA DE SÍMBOLOS

Aminoácidos

A	Alanina
C	Cisteína
D	Ácido aspártico
E	Ácido Glutâmico
F	Fenilalanina
G	Glicina
H/His	Histidina
I	Isoleucina
K	Lisina
L	Leucina
M	Metionina
N	Asparagina
P	Prolina
Q	Glutamina
R	Arginina
S	Serina
T	Treonina
V	Valina
W	Triptofano
Y	Tirosina

Nucleótídeos

A/a	Adenina
C/c	Citosina
G/g	Guanina
T/t	Timina
R	A ou G
Y	C ou T
S	G ou C
W	A ou T
K	G ou T
M	A ou C

SUMÁRIO

1 INTRODUÇÃO	19
2 OBJETIVOS	28
3 MATERIAL E MÉTODOS	28
3.1 LINHAGENS E PREPARO DE CÉLULAS BACTERIANAS COMPETENTES	28
3.2 TRANSFORMAÇÃO DE CÉLULAS COMPETENTES E CULTURA DE CÉLULAS	29
3.2.1 Transformação e cultura de células transformadas	29
3.2.2 Culturas para preparação de DNA plasmidial	30
3.2.3 Culturas para expressão de proteína recombinante	31
3.3 SEQUÊNCIAS E FONTE DOS FRAGMENTOS DE DNA	31
3.4 AMPLIFICAÇÃO DE DNA POR <i>PCR</i>	33
3.5 PURIFICAÇÃO DE FRAGMENTOS DE DNA	34
3.6 CLIVAGEM DE FRAGMENTOS DE DNA POR ENZIMAS DE RESTRIÇÃO	35
3.7 VETORES PARA CLONAGEM E EXPRESSÃO	35
3.8 LIGAÇÃO DE <i>CG9571 BD/FL</i> NOS VETORES DE CLONAGEM	37
3.9 EXTRAÇÃO E PURIFICAÇÃO DE DNA PLASMIDIAL	38
3.10 ANÁLISE DE DNA POR ELETROFORESE EM GEL, ESPECTROFOTOMETRIA E SEQUENCIAMENTO DE DNA	39
3.11 SEQUÊNCIA E PARÂMETROS DO FATOR DE TRANSCRIÇÃO	40
3.12 OBTENÇÃO DA PROTEÍNA POR SISTEMA DE EXPRESSÃO <i>IN VITRO</i>	41
3.13 OBTENÇÃO DA PROTEÍNA POR EXPRESSÃO EM BACTÉRIA	41
3.13.1 Indução da expressão	41
3.13.2 Lise celular e separação de frações solúvel e insolúvel	42
3.13.3 Recuperação de proteína recombinante de corpos de inclusão	43
3.13.4 Purificação e renaturação de proteína em coluna de cromatografia por afinidade	44
3.13.5 Purificação de proteína recombinante a partir de SDS-PAGE	46
3.13.6 Renaturação por diluição e diálise	47
3.13.7 Análise de proteína por eletroforese em gel e espectrofotometria	49
3.14 INTERAÇÃO ENTRE DNA-PROTEÍNA	49
3.14.1 Ensaios de interação entre <i>CRMs</i> e amostras de proteína	49
3.14.2 Ensaios de retardo da mobilidade eletroforética	50
4 RESULTADOS E DISCUSSÃO	52

4.1 CONFIRMAÇÃO DAS CONSTRUÇÕES PLASMIDIAIS CARREADORAS DE <i>EVE 1</i> , <i>EVE 2</i> , cDNA DE <i>CG9571</i> E OBTENÇÃO DOS FRAGMENTOS DE DNA RELATIVOS AOS <i>CRMS</i>	52
4.2 EXPRESSÃO DE <i>CG9571 BD IN VITRO</i> E OBTENÇÃO DE EXTRATO PARA TESTE DE INTERAÇÃO COM <i>EVE 1</i>	53
4.3 CLONAGEM DE <i>CG9571 BD</i> E EXPRESSÃO EM BACTÉRIA	56
4.3.1 Clonagens em vetor carreador e de expressão	56
4.3.2 Tentativas de indução de expressão de pET-15b-<i>CG9571 BD</i>	58
4.4 CLONAGEM DE <i>CG9571 FL</i> , EXPRESSÃO EM BACTÉRIA E OBTENÇÃO DA PROTEÍNA RECOMBINANTE PURIFICADA E SOLÚVEL	61
4.4.1 Clonagens em vetor carreador e de expressão	61
4.4.2 Ensaios iniciais de expressão em células competentes e tentativas de obtenção e purificação da proteína recombinante solúvel	63
4.4.3 Obtenção de proteína a partir de corpos de inclusão por solubilização em condições brandas	66
4.4.4 Obtenção de proteína a partir de corpos de inclusão por solubilização em condições desnaturantes severas	68
4.4.5 Tentativas de purificação e renaturação de proteína recombinante por IMAC	70
4.4.6 Purificação de proteína a partir de SDS-PAGE	73
4.4.7 Renaturação de proteína por diluição e diálise	74
4.5 ENSAIOS DE LIGAÇÃO E DETECÇÃO DE INTERAÇÃO ENTRE <i>CG9571 FL</i> E <i>CRMS</i>	78
5. CONCLUSÕES E PERSPECTIVAS FUTURAS	89
REFERÊNCIAS	92
APÊNDICE 1. SEQUENCIAMENTO DE POSSÍVEIS CLONES <i>CG9571 BD</i> EM VETOR PCR 2.1-TOPO	101
APÊNDICE 2. SEQUENCIAMENTO DE POSSÍVEIS CLONES <i>CG9571 BD</i> EM VETOR PET-15B	103

1 INTRODUÇÃO

A mosca-da-fruta *Drosophila melanogaster* teve papel central no desenvolvimento da moderna genética (Waddington, 2016). Essa espécie se tornou um modelo de excelência em genética por uma série de vantagens que apresenta: ciclo de vida curto, grande quantidade de descendentes produzidos, fácil manipulação e manutenção em laboratório e, sobretudo, pela existência de diversos genótipos mutantes (Garcia, Fernández, 2012). Desde a introdução da *Drosophila* em laboratórios de pesquisas no início do séc. XX foram observadas mutações espontâneas e linhagens com algumas destas mutações foram estabelecidas. Posteriormente, os pesquisadores conseguiram incrementar o surgimento de novas mutações com o emprego de agentes químicos ou físicos (Green, 1961; Lewis e Bacher, 1968; St Johnston, 2002).

A *Drosophila* também passou a ser utilizada amplamente nos estudos de biologia do desenvolvimento a partir de 1970. Com o desenvolvimento da biologia molecular e a interface desta com a genética, foi possível isolar e conhecer a identidade de genes responsáveis por mutações que causavam anomalias no padrão corporal ou, mais precisamente, nos genes que controlam a formação do corpo (Lewis, 1978; Nüsslein-Volhard e Wieschaus, 1980; Johnston e Nüsslein-Volhard, 1992).

Os estudos iniciais que investigaram os genes que controlam a formação do corpo na *Drosophila* foram estudos pioneiros que contribuíram para desvendar a estrutura dos genes e reconhecer a existência de duas partes principais. Uma das partes contém a região codificadora, que é transcrita em mRNA e pode ser traduzida em uma (ou mais) proteína(s). A outra parte, denominada região reguladora ou controladora, controla a transcrição do gene e é composta por vários módulos. Um destes módulos é a região promotora onde é montado o complexo de início da transcrição, que é um complexo holoenzimático composto pela RNA polimerase e outras proteínas. Porém, a transcrição também depende de outros módulos da região reguladora, inicialmente denominados por *enhancers* (intensificadores ou realçadores) e das proteínas que neles se associam (Wolpert *et al.*, 2011).

Os *enhancers*, atualmente denominados por módulos cis-reguladores (*CRMs*, do inglês *cis-regulatory modules*. Ochoa-Espinoza *et al.*, 2005), são sequências formadas em geral por 250 a 1000 pares de bases (pb), dentre as quais são encontrados sítios de ligação (motivos), normalmente compostos por 8 a 12 nucleotídeos cada um, reconhecidos por proteínas ligadoras de DNA denominadas por fatores de transcrição

(Gilbert, 2010). Os fatores de transcrição são proteínas nucleares constituídas de pelo menos dois domínios: um envolvido no reconhecimento e ligação ao DNA e outro que interage com um ou mais componentes do complexo de transcrição (Levine e Manley, 1989). Quando ativados, estes fatores se ligam aos *CRMs*, podendo exercer função de ativação ou repressão do gene alvo (Nelson e Cox, 2014). Portanto, a contribuição de um *CRM* para a expressão gênica em dado momento é o resultado da ocupação de várias moléculas de diferentes tipos nos seus sítios de ligação. Os *CRMs* controlam a taxa de transcrição de um promotor específico e a expressão temporal e espacial de genes é possível pela existência de diferentes *CRMs* (Gilbert, 2010).

Estudos do início do desenvolvimento da *Drosophila* ajudaram muito a entender a transcrição diferencial de genes nos eucariotos. O embrião dessa espécie se desenvolve inicialmente como um sincício e assim permanece até o décimo quarto ciclo de mitose (Rivera-Pomar e Jäckle, 1996. Figura 1). Este modo de desenvolvimento facilita a difusão dos produtos da expressão de genes reguladores envolvidos com a formação do corpo ao longo dos eixos anteroposterior e dorsoventral do embrião. Os núcleos do sincício, que inicialmente estão distribuídos por todo o embrião, ficam sujeitos a diferentes concentrações e combinações destes fatores (Rivera-Pomar e Jäckle, 1996; Wolpert *et al.*, 2011). A interação dos fatores com os *CRMs* de genes-alvo dos núcleos de determinada região determina a ativação ou inativação da expressão local de genes específicos (Small *et al.*, 1991; Ochoa-Espinosa *et al.*, 2005). Portanto, o estudo da transcrição durante a blastoderme sincicial possibilita investigar diretamente ação de genes entre si, sem a interferência da membrana e dos mecanismos de sinalização celular.

Figura 1. Ciclo de vida de *Drosophila melanogaster*

Divisões nucleares não acompanhadas de divisões celulares produzem um sincício que persiste até o final do período de clivagem, quando ocorre a celularização da blastoderme, imediatamente antes da gastrulação. Durante a clivagem ocorre a cascata de segmentação. Fonte: Wolpert (2008).

Os genes responsáveis pela especificação do eixo anteroposterior em *Drosophila* fazem parte da cascata de segmentação. Os genes de segmentação foram agrupados em três classes de genes expressos no zigoto: genes *gap*, *pair-rule* e *segment-polarity* (Nüsslein-Volhard e Wieschaus, 1980). Todas as três classes de genes foram identificadas pelas alterações que os fenótipos mutantes apresentavam no padrão de segmentação do embrião ao longo do eixo anteroposterior da cutícula de larvas de primeiro estágio. Aproximadamente três quartos dos genes de segmentação, incluindo todos os genes *gap* e os genes *pair-rule*, possuem como produtos de sua expressão fatores de transcrição (Alberts *et al.*, 2011). Diversos estudos demonstraram que estes fatores codificados por uma classe de genes de segmentação zigóticos interagem dentro de um mesmo nível da cascata, além de regularem a expressão de genes do nível hierárquico seguinte (e.g. Carroll e Scott, 1986; Harding *et al.*, 1986; Small *et al.*, 1991 e 1992). Em outras palavras, o padrão de segmentação de *Drosophila* é estabelecido por uma cascata de genes, na qual fatores de transcrição dos genes *gap* regulam genes *pair-rule* e fatores de genes *pair-rule* regulam genes *segment-polarity* (Figura 2).

Figura 2. Cascata de segmentação em *Drosophila*

(A) Esquema ilustrando as etapas da cascata de segmentação e os grupos de genes envolvidos: produtos de genes de efeito materno e genes de segmentação zigóticos *gap*, *pair-rule* e *segment-polarity*. Genes homeóticos, cuja atuação está relacionada a determinação da identidade dos segmentos, também são mostrados. Fonte: modificado de Gilbert (2003). (B) Embriões submetidos à técnica de hibridização *in situ* mostrando o padrão de expressão de mRNA de genes de cada nível: *bicoid* (*bcd*, efeito materno), *giant* (*gt*, *gap*), *even-skipped* (*eve*) e *fushi tarazu* (*ftz*, *pair-rule*) e *engrailed* (*en*, *segment-polarity*). Fonte: Andrioli (arquivo pessoal).

Os genes de segmentação zigóticos são acionados por produtos de genes maternos (Nüsslein-Volhard e Wieschaus, 1980; St Johnston e Nüsslein-Volhard, 1992; Anderson e Nüsslein-Volhard, 1984. Figura 4). As células nutrízes e foliculares no ovário da fêmea produzem mRNAs e proteínas que são transferidos para o oócito. No oócito, alguns produtos se distribuem de forma homogênea e outros são diferencialmente distribuídos e ficam retidos em redes de microtúbulos gerando uma assimetria inicial (Rongo e Lehmann, 1996).

Os principais genes de segmentação de efeito materno são: *bicoid* (*bcd*), *nanos* (*nos*), *hunchback* (*hb*) e *caudal* (*cad*) (Wolpert et al, 2011). O mRNA de *bcd* é retido na extremidade que se tornará a região da cabeça do embrião. Após a deposição do ovo, o mRNA de *bcd* é traduzido e a proteína se difunde ao longo do embrião, mas sua maior concentração permanece na região anterior. O gradiente da proteína Bicoid (Bcd)¹ é responsável pela ativação diferencial dos fatores de transcrição de genes *gap* da região anterior que formarão a cabeça e o tórax da larva (Johnston e Nüsslein-Volhard, 1992).

¹ Neste trabalho os genes e seus transcritos (mRNAs) estão escritos em itálico para distinguir das proteínas, com inicial maiúscula e escrita regular, de acordo com a convenção utilizada para *Drosophila*.

Na outra extremidade é encontrada alta concentração de mRNA *nos*, que é traduzido, forma um gradiente posterior da mesma maneira e especifica a formação da região do abdômen da larva. Bcd e Nos regulam a síntese de Hb, com Bcd ativando a expressão de *hb* na região anterior e Nos reprimindo a expressão na região posterior (Wharton e Struhl, 1991; Wang *et al.*, 1994). *caudal mRNA*, diferentemente de *bcd* e *nos*, é distribuído uniformemente pelo ovo e é crítico para o estabelecimento de segmentos posteriores, sendo reprimido por Bcd na região anterior (Dubnau e Struhl, 1996).

Os genes *gap* são os primeiros genes zigóticos da cascata de segmentação expressos no embrião. Eles são necessários em regiões amplas e específicas ao longo do embrião para que este se desenvolva normalmente (Nüsslein-Volhard e Wieschaus, 1980). Mutações nos genes *gap* geram larvas com a ausência de segmentos adjacentes, isso porque a expressão dos genes *gap* é contínua ocupando a região destinada a vários segmentos. Alguns exemplos importantes de genes *gap* são: *tailless (tll)* e *huckebein (hkb)* – genes *gap* terminais; *orthodenticle (otd)*, *empty spiracles (ems)* e *buttonhead (btd)* – genes *gap* da cabeça; e *hunchback* (quem também é gene zigótico, além de materno), *Kruppel (Kr)*, *knirps (kni)* e *giant (gt)* – genes *gap* centrais (Rivera-Pomar e Jäckle, 1996). Como já foi dito, a regulação da expressão de genes *gap* da cabeça e da região central depende de fatores maternos e de fatores de outros genes *gap*, sendo a repressão mútua entre os produtos destes genes possibilita precisar os limites da expressão dos domínios *gap*. Já a regulação de genes *gap* terminais é controlada por fatores maternos, independentemente da atividade de outros genes *gap* (para revisão, Pankratz e Jäckle, 1993). Genes *gap* da cabeça são ativados por Bcd, enquanto os genes *gap* da região central são ativados por uma combinação entre Bcd, Hb e Cad. A regulação da expressão de *Kr* na região central, por exemplo, é realizada com Hb materno e Bcd agindo como ativadores e Kni, Gt e Tll como repressores que refinam a expressão (Hoch, *et al.*, 1991).

A segunda classe de genes da cascata, os genes *pair-rule*, já prenuncia o caráter de segmentação futuro do corpo. Cada gene *pair-rule* é expresso em 7 faixas longitudinais alternadas ao longo do eixo anteroposterior, dividindo o embrião em 14 parassegmentos (Small, 1999). Parassegmentos são unidades metaméricas do embrião que mantém relação com os segmentos anatômicos, consistindo do compartimento posterior de um segmento e do compartimento anterior do segmento seguinte (Martinez-Arias e Lawrence, 1985). Mutações nestes genes produzem embriões onde metade dos parassegmentos estão faltando (Nüsslein-Volhard e Wieschaus, 1980). São conhecidos

ao menos oito genes desta classe. O gene *even-skipped* (*eve*) é um dos genes *pair-rule* mais bem estudados no que diz respeito à regulação. Esse gene codifica uma proteína *homeobox* que é expressa em uma série de 7 faixas transversais (Figura 2B) e que realiza um papel regulatório essencial para estabelecer o plano corporal metamérico (Harding et al, 1986). A região reguladora de *eve* contém uma série de *CRMs* que controla a expressão de faixas individuais ou um par de faixas (Harding *et al.*, 1989; Fujioka *et al.*, 1999) (Figura 3). Há um *CRM* responsável pela regulação da faixa 2, um pelas faixas 3 e 7, outro pelas faixas 4 e 6 e um último pelas faixas 1 e 5 (Fujioka *et al.*, 1999). Estudos têm demonstrado que cada faixa possui um mecanismo regulatório distinto (Small *et al.*, 1996; Fujioka et al, 1999). Esta regulação diferencial se dá por causa da combinação de fatores de transcrição diferencialmente distribuídos ao longo do embrião. Por esta razão, cada *CRM* apresenta uma composição de sítios de ligação para ativadores e repressores e o balanço final dos fatores ligados a estes sítios determinará se ocorrerá ou não a transcrição (Small, 1999).

Figura 3. Esquema da estrutura do locus de *even-skipped*

(A) Região codificadora (Coding) do gene e posição dos módulos reguladores da expressão das 7 faixas. (B-D) Em destaque, *CRM* que regulam a expressão das faixas 1 e 5. Hibridizações *in situ* detectando faixas selvagens e gene repórter lac-Z. Fonte: modificado de Gilbert (2006).

A partir destes estudos das faixas de *eve* e de estudos de outro gene *pair-rule*, *hairy* (*h*) (Pankratz *et al.*, 1990; Langeland e Carrol, 1993), foi proposto um modelo geral para o estabelecimento inicial das faixas individuais no qual os fatores de ativação são encontrados amplamente distribuídos pelo embrião, ativando faixas individuais,

enquanto os fatores de genes *gap* agiriam como repressores e seriam responsáveis pelo estabelecimento das bordas das faixas (Small, 1999). Um segundo momento de regulação das faixas de expressão *pair-rule* é o refinamento, quando cada faixa passa a ocupar de 2 a 3 colunas de núcleos. O refinamento das faixas *pair-rule* ocorre pela mútua repressão entre os genes *pair-rule* e pela autorregulação, observada para *eve* (Fujioka *et al.*, 1996; Jiang *et al.*, 1991) e para o *pair-rule fushi tarazu (ftz)* (Schier e Gehring, 1992).

Dentre as sete faixas de *eve*, o mecanismo de controle do CRM responsável pela regulação da faixa 2 (*eve 2*) é o mais bem elucidado. Experimentos genéticos demonstraram que Bcd e Hb são necessários para a ativação, enquanto os fatores *gap* Gt e Kr agem como repressores, definindo as bordas anterior e posterior da faixa, respectivamente (Figura 4). Experimentos de interação *in vitro* confirmaram a interação destas proteínas com sítios de um elemento mínimo para a expressão desta faixa composto de 480pb (Stanojevic *et al.*, 1991; Small *et al.*, 1991; 1992). Há evidências, contudo, que sugerem que outros fatores são necessários para a formação normal desta faixa – principalmente fatores de função repressora (Small *et al.*, 1992; Wu *et al.*, 1998). Um destes fatores foi descoberto posteriormente: Slp 1 (Andrioli *et al.*, 2002).

Figura 4. Mecanismo proposto para a regulação da expressão de *eve 2*

O primeiro mecanismo mais abrangente proposto para a regulação desta faixa compreendia quatro fatores de transcrição (Bcd, Hb, Gt e Kr). A sequência correspondente ao elemento mínimo para a expressão normal da faixa (*minimal stripe element, MSE*) contém sítios reconhecidos por ativadores e repressores muito próximos, ocorrendo sobreposição em dois agrupamentos de sítios. As letras D indicam regiões entre sítios conhecidos que foram removidas no contexto do CRM controlando o repórter *lacZ*. A deleção 2 (D 2) gerou expansão do repórter e experimentos seguintes revelaram a existência de sítio ligado para Slp1 na sequência correspondente a D2. Fonte: modificado de Andrioli *et al.* (2002).

sloppy-paired (*slp*) é um *locus* que foi identificado originalmente em um experimento de varredura para mutações embriônicas letais que afetavam o padrão de formação do corpo (Nüsslein-Volhard *et al.*, 1984). Posteriormente descobriu-se que o *locus* é composto de dois genes estruturalmente relacionados, *slp 1* e *slp 2*, que codificam proteínas estrutural e funcionalmente relacionadas (Grossniklaus *et al.*, 1992). Slp 1 e Slp 2 estão entre as primeiras proteínas descobertas da família de fatores de transcrição *forkhead* (Grossniklaus *et al.*, 1992), um grupo que é caracterizado por uma região com cerca de 100 resíduos de aminoácidos altamente conservada no domínio de ligação ao DNA (domínio *forkhead*) (Lai *et al.*, 1993) com estrutura tridimensional característica (Figura 5).

Figura 5. Domínio ligador de DNA *forkhead*

A estrutura tridimensional do domínio consiste de uma folha-beta antiparalela formada por três fitas-β (S1, S2 e S3), três ou quatro α-hélices (H1, H2, H3 e H4) e três *loops*. Por causa da semelhança de dois *loops* com asas, o domínio também recebeu o nome de motivo hélice alada. Fonte: Obsil e Obsilova (2008).

Slp 1 e Slp 2 são proteínas redundantes que atuam em dois momentos da cascata de segmentação, como *pair-rule* e como *segment-polarity* (Cardigan *et al.*, 1994; Grossniklaus *et al.*, 1992). No entanto, apenas Slp 1 é expresso como um gene *gap* na região anterior do embrião atuando como um repressor para a expressão da faixa *eve 2* (Andrioli *et al.*, 2002). Os autores descobriram que Slp 1 se liga a uma sequência de repetição (GTTT)_{4x} no *CRM* de *eve 2*, e confirmaram a atividade repressora deste fator *in vitro* e *in vivo*. Mais recentemente, verificou-se que Slp1 também era repressor de outras faixas anteriores *pair-rule*, como *eve 1* (Andrioli *et al.*, 2004). No curso dos estudos com *eve 2*, verificou-se que o duplo mutante para *slp* e *gt* apresentou pequena expansão da borda anterior (Andrioli *et al.*, 2002). A expectativa era de que, na presença

de ativadores ao longo de toda porção anterior da blastoderme e na ausência dos repressores, seria detectada ampla expansão do *CRM* responsável para esta faixa. Ademais, a remoção da sequência (GTTT)_{4x} de *eve 2* apresentou maior expansão da borda anterior desta faixa do que a observada para o mutante duplo para *gt* e *slp*. Essa situação foi interpretada como evidência para a existência de um fator repressor redundante de Slp 1 (Andrioli *et al.*, 2002).

Com esta premissa de um parceiro do repressor Slp 1 atuante na região anterior, genes *forkhead* foram vasculhados no genoma da *Drosophila* e se verificou que o gene anotado *CG9571* apresenta um padrão de expressão na região anterior consistente com o de um gene *gap* repressor (Andrioli, comunicação pessoal) (Figura 6A). Embora não tenham sido verificadas alterações no padrão *pair-rule* em uma deficiência envolvendo *CG9571*, a combinação dessa deficiência com alelo mutante nulo para *slp* (*CG9571-; slp1-*) revelou efeitos acentuados no padrão estriado para a faixa *eve 2* e também para a faixa *eve 1*, em comparação com o detectado para *slp-* (Andrioli, comunicação pessoal). Utilizando *CG9571* em um sistema de expressão ectópica foi detectado repressão específica nas faixas anteriores *pair-rule* (Figura 6B), semelhante ao obtido com *slp1* nesse mesmo sistema (Andrioli, comunicação pessoal; Andrioli *et al.*, 2002).

Figura 6. Domínio de expressão de CG9571 e efeito da expressão ectópica em faixas de eve

Hibridizações *in situ* detectando a expressão de mRNA. (A) *CG9571* (em preto) é expresso na extremidade anterior simultaneamente com outros dois genes *gap*: *Kr* e *slp 1* (ambos em vermelho). (B) A expressão ectópica de *CG9571* na região ventral do embrião (em azul, imagem superior) causa repressão local nas faixas 1 e 2 e, em menor escala, na 3 (imagem inferior). Fonte: Andrioli (arquivo pessoal).

O próximo passo é, portanto, verificar se o fator codificado por *CG9571* interage de fato com os *CRMs* de *eve 1* e *eve 2*. Esta investigação fornecerá evidência para a identificação de um possível novo gene na cascata de segmentação de *Drosophila*,

contribuindo na compreensão do complexo fenômeno de regulação gênica nas fases iniciais do desenvolvimento deste modelo animal. Particularmente, a região anterior precursora da cabeça é a menos conhecida no modelo, pelo fato da complexidade de interação entre os fatores já conhecidos e a antecipação da existência de outros fatores ainda não isolados.

2 OBJETIVOS

O objetivo deste estudo foi verificar a interação do fator de transcrição *forkhead* CG9571 com módulos reguladores de faixas de expressão *pair-rule* do gene *even-skipped*.

Para atingir este objetivo, estabelecemos as seguintes etapas:

- (1) Obtenção dos fragmentos de *eve* 1 e *eve* 2 em quantidade suficiente para experimentos de interação DNA-proteína;
- (2a) Obtenção do domínio de ligação de CG9571 por sistema de transcrição e tradução *in vitro* ou por expressão de polipeptídeo recombinante em bactéria;
- (2b) Obtenção da proteína completa de CG9571 por expressão de proteína recombinante em bactéria;
- (3) Testar a interação *in vitro* entre os fragmentos de DNA de *eve* 1 e *eve* 2 e a proteína ou o domínio de ligação de CG9571.

3 MATERIAL E MÉTODOS

3.1 LINHAGENS E PREPARO DE CÉLULAS BACTERIANAS COMPETENTES

Para as subclonagens foram utilizadas células *E. coli* quimiocompetentes da linhagem JM109 (*Promega*) adquiridas com a empresa e preparadas no laboratório. Para as clonagens e expressão de proteína recombinante foram utilizadas células *E. coli* quimiocompetentes da linhagem BL21(DE3)pLysS (*Promega*), também adquiridas com a empresa e preparadas no laboratório, e células eletrocompetentes da linhagem BL21 STAR (DE3)pLysS (*Thermo Fisher Scientific*), gentilmente cedidas pelo laboratório do Prof. Dr. Felipe Chambergo (EACH, USP).

As células quimiocompetentes foram preparadas a partir dos estoques adquiridos com as empresas. Utilizamos o protocolo desenvolvido por Hanahan (1985) com

pequenas modificações, que consistiram na substituição de alguns componentes dos tampões RF1 e RF2 (MOPS foi substituído por HEPES e RbCl por KCl, em algumas preparações).

As células da linhagem BL21(DE3) contêm um gene que codifica para a enzima T7 RNA polimerase que está sob o controle do promotor *lacUV5*, permitindo que sua indução ocorra na presença de IPTG (*Promega*, 2014). A linhagem BL21(DE3)pLysS ainda contém um plasmídeo (pLysS) que codifica para a enzima lisozima T7 e permite regular ainda mais a expressão do gene alvo, clonado em vetor de expressão (item 3.7), diminuindo a expressão de fundo do gene de interesse enquanto não houver adição de IPTG à cultura (*Promega*, 2014). A linhagem BL21 Star (DE3)pLysS contém, adicionalmente, uma mutação no gene que codifica a RNase E, enzima envolvida na degradação de mRNA, o que permite um aumento da estabilidade de mRNAs do gene alvo nestas células e, consequentemente, um aumento na produção de proteína recombinante (*Invitrogen*, 2010).

3.2 TRANSFORMAÇÃO DE CÉLULAS COMPETENTES E CULTURA DE CÉLULAS

3.2.1 Transformação e cultura de células transformadas

As construções plasmidiais utilizadas neste trabalho foram utilizadas para transformar células competentes seguindo os protocolos dos fabricantes das células. Para células quimiocompetentes, em essência, uma alíquota de DNA plasmidial era adicionada a um tubo de microcentrífuga contendo de 50 a 70 µl de células competentes descongeladas em recipiente com gelo. Os tubos contendo a mistura eram incubados por 20 a 30 min em gelo e, em seguida, incubados em banho de aquecimento a 42°C por 40 a 50 seg. e retornados ao gelo por mais 2 min após o choque. Um volume de 250 a 700 µl de meio de cultura LB (triptona 1%, extrato de levedura 0.5%, NaCl 0.5%) ou SOC (triptona 2%, extrato de levedura 0.5%, NaCl 10mM, KCl, 2.5mM, MgCl₂ 10mM, MgSO₄ 10mM, glicose 20mM) era adicionado ao tubo e as células eram crescidas por 1h em incubadora sob agitação (160 a 200 rpm) a 37°C. Uma alíquota de 80 a 200 µl da cultura era aplicada em placa com meio LB seletivo com a finalidade de selecionar eventos positivos de transformação em colônias isoladas. As culturas em placa eram incubadas em estufa bacteriológica a 37°C por 12 a 16h e então armazenadas em refrigerador.

Para células eletrocompetentes, uma alíquota de construção plasmidial era adicionada em 50 µl de suspensão de células em um tubo de microcentrífuga. A mistura era transferida para uma cubeta de eletroporação estéril de modo que todo o volume estivesse em contato com o fundo da cubeta. Em seguida, a cubeta era inserida em um eletroporador e eram aplicados dois pulsos elétricos de 2500V. Após o choque, 1 ml de meio LB era aplicado na cubeta, misturado com a suspensão de células e todo o volume era transferido para tubo de crescimento. As condições de crescimento e o estriamento em placas de meio seletivo são as mesmas descritas para as células quimiocompetentes.

Para células transformadas com construções plasmidiais contendo o vetor pCR 2.1 TOPO foram utilizadas placas de LB contendo de 150 µg/ml de ampicilina (LBA) ou 50 µg/ml de canamicina (LBC) para espalhamento da suspensão. Em transformações utilizando os demais vetores foram usadas somente placas LBA. Para a distinção entre colônias transformadas apenas com o vetor de colônias contendo plasmídeo mais o inserto, nas clonagens em células da linhagem JM109, foi utilizada a técnica de *blue/white screening*, técnica que se baseia no princípio da complementação funcional para a β-galactosidase (Horwitz *et al.*, 1964; Ullman *et al.*, 1967). Para isso, 40 µl de X-gal 40 mg/ml e 40 µl de IPTG 100 mM eram espalhados em placas LBA cerca de 30min antes do estriamento.

3.2.2 Culturas para preparação de DNA plasmidial

Culturas de células para obtenção de DNA plasmidial foram realizadas conforme protocolos largamente utilizados (Sambrook e Russel, 2001) ou de acordo com os *kits* de preparação de DNA plasmidial utilizados (item 3.9). De modo sucinto, para preparações de pequeno volume (*minipreps*), colônias isoladas de placas de transformações eram selecionadas e inoculadas em 2-5 ml de meio LBA líquido mantido por 12 a 16 h em incubadora com agitação (160-200 rpm) a 37°C. Para preparações de maior volume (*midipreps*), o procedimento realizado era o mesmo das *minipreps* com a diferença de que os pré-inóculos eram incubados por 6 a 8h e uma alíquota desta cultura era diluída de 100 a 1000 vezes para um volume de 25 ou 50 ml de meio LBA líquido contido em frasco de Erlenmeyer. A cultura era incubada nas mesmas condições por 12 a 16h e então as células eram coletadas por centrifugação (3500-6000 x g, 2-10min). A extração e purificação de plasmídeos destas preparações é descrita na seção 3.9.

3.2.3 Culturas para expressão de proteína recombinante

As condições de crescimento e volumes das culturas utilizadas para indução da superexpressão da proteína em bactérias foram bastante variadas. As orientações do manual do fornecedor do vetor de expressão (Novagen, 2003) foram usadas como base para os experimentos e modificadas por tentativa e erro visando-se a otimização deste procedimento.

Colônias isoladas de placas com células transformadas na véspera eram inoculadas em 1-5ml de meio LBA e cultivadas por 4-12h, sob agitação (150-200 rpm), à temperatura ambiente (20-22°C) ou 37°C. Uma alíquota do pré-inóculo era diluída de 20 a 1000 vezes em 4-250ml de meio LBA em um frasco de Erlenmeyer. Culturas de menor volume, em geral, foram utilizadas nos ensaios iniciais de expressão – para a verificação da indução do polipeptídeo/proteína recombinante, seleção de clones com melhor nível de expressão e para padronização e otimização da indução da expressão – enquanto as culturas de maior volume foram usadas para obtenção da proteína para tentativas de purificação e demais procedimentos. Estes inóculos foram cultivados sob agitação (150-200 rpm) em temperaturas que variaram de 18 a 37°C até que a DO₆₀₀ da cultura atingisse valores entre 0.4 e 1 de absorbância. A indução dos inóculos será abordada na seção 3.13.1.

Culturas para obtenção de proteína recombinante foram preparadas a partir de placas estriadas com *stabcultures*. Para o preparo de *stabcultures*, alíquotas de 0.5-1ml de cada cultura final eram separadas imediatamente antes da indução. À estas alíquotas, transferidas para microtubos, era adicionado glicerol estéril à concentração final de 15% e, em seguida, os microtubos eram armazenados em ultrafreezer (-80 a -70°C). Após ensaios de indução, as *stabcultures* dos clones que apresentaram maior expressão da proteína eram mantidas para uso posterior e as demais descartadas.

3.3 SEQUÊNCIAS E FONTE DOS FRAGMENTOS DE DNA

As regiões reguladoras *eve 1* e *eve 2* já foram definidas experimentalmente – *eve 1* apresenta 787 pb (Fujioka *et al.*, 1999) e o elemento mínimo suficiente para a expressão de *eve 2* possui 480 pb (Small *et al.*, 1991; 1992). No laboratório dispomos dos fragmentos destes *CRMs* clonados em vetor plasmidial. Fragmentos de *eve 2* selvagem e com a sequência mutada no sítio GTTT (Figura 7) clonados em pBlue-script (pBS, *Stratagene*) foram gentilmente cedidos pelo Prof. Dr. Stephen Small, New York

University, NY. *eve 1* (Figura 8) foi obtido no laboratório por meio de *PCR* a partir de DNA genômico extraído de *Drosophila melanogaster* e da subsequente clonagem em vetor pCR 2.1-TOPO. Fragmentos dos *CRMs* para os ensaios de ligação foram obtidos por *PCR* (*eve 1*) ou por expansão de DNA plasmidial em bactérias (3.2.2) e clivagem por enzimas de restrição.

Figura 7. Sequência de nucleotídeos de *eve 2* mutada

```

1  ggtcagatctggttacccggtaccgaatgacaatggaacccgaac
46  cgtaactgggacagatcgaaaagctggcctggtttctcgctgtgt
91  gtgccgtgttaatccgtttgccatcagcgagattattagtcaatt
136  gcagttgcagcgtttcgtttcgtcctcgtttcactttcgagtta
181  gactttaccgaagcgtcttgaacaatcgctcgacgtctgttagcac
226  gctgtgccatactttcatttagacggaatcgagggaccctggact
271  ataatcgcacacagagaccgggttgcgaagtcagggcattccgcc
316  gatctagccatcgccatcttctgcgggcgtttgtttgtttgtttg
361  ctgggattagccaagggcttgacttggaatccaatcccgatccct
406  agcccgatcccaatcccaatcccaatcccttgtccttttcattag
451  aaagtcataaaaacacataataatgatgtcgaagggattaggcgc
496  gccttat

```

Em vermelho é destacado o sítio de Gt, em verde a mutação e em roxo os sítios de restrição.

Figura 8. Sequência de nucleotídeos de *eve 1*

```

1  ggcctaatacacttccctgaaatgcataattgtgccgcggccttttg
46  atacgctcctggcggagagggagatgaggaaaggatgcacgggaa
91  ccgcagccaagtggcagtcgagattggcaaatccgccagcggaca
136  atgccagagaatgggcaacaagtagcggcggaattagcaatccta
181  tcatgcttttatggccggccaactcttgcccgcgcacatctcagttc
226  atccgaagcgggaccaggtccaggttcaagtcgaggtccagttacc
271  cctgctatcccgctcaaccccttttagggcgataatccttctaaatg
316  tttgcatttaatttcgagggcgtggacggattagggcgtgctggctg
361  ggcggaacccgcagcagaaacccgccgaggacactgcaccgactga
406  cctgcagcctacagatctctgatcttcgatctctaataccttccgc
451  atttgcaactgacttctgcactgggtccgcccctaataccttccgc
496  cgagaaggcggcagagtcgcgaggtactggcccggggtaatgggga
541  ttatctgcgattacccagatgatccgcagaaagtcaatctgggtt
586  caggggctaattgtcagcgaagtcaactaaatccaatcctttcgc
631  gcccccttctgtttatgtttgtttcgtttgtttgagaattt
676  ctggcaattaagttgcccgttttgatgcgcgggggcgggtgcac
721  aaatcctttcggcatacctgtcctgcacaaatgctgaattccgca
766  tcccatggataaccagatattc

```

A sequência foi delimitada em sequência do *locus* do gene *eve* obtida na base de dados FlyBase a partir dos resultados de Fujioka et al, 1999.

A região codificadora da proteína completa de CG9571 (*FL*, do inglês *full length*) é composta por 783 pb e o domínio de ligação ao DNA (*BD*, do inglês *binding domain*), predito através da ferramenta SMART (<http://smart.embl-heidelberg.de>), apresenta 273 pb (Figura 9). O DNA referente a proteína completa de CG9571 foi originalmente clonado em vetor TOPO (*Invitrogen*) a partir do DNA genômico, pois esse gene não apresenta íntron (dado disponível na entrada do gene, fd19B, na base de dados FlyBase).

Figura 9. Sequência de nucleotídeos de CG9571

```

1  atggatacgacaccaatcttccagtcagcttctccatccgatcg
46  ctgctctccgtggacaaaaaggaggagtcctccatctcgaagcac
91  aactcaggaagcagcttcagctcatgctcgagctccagttccaac
136  tccagctcggattcgatggctgcaaagagcaatgcaaagccggct
181  ttcacctacagcgccctcatagtgatggccatctggagcagctcc
226  gagaagcgtttgaccctaagcgggatctgcaagtggatcgcgga
271  aacttcccgtactatcgcaaccgcaagagcgtctggcagaactcg
316  atccggcacaaacctgagctctcaatccgttcttcggtcggtgccg
361  cgagctctcgatgatcctggacgtggccactactgggcactcgat
406  ccgtatgccgaggatctgtccattggcgagacgacgggacgcctg
451  cgccgcagcaattggcagcagaatacgggggcgagaccaaggctc
496  acaggtcatccctatcagcgaatgccatattatgggcacgggcac
541  ggtaacggcccatatatcaaggcacacagcgcctacttccccata
586  atggaccatcaacaccatgccgccatggtgcagcactaccaggcc
631  atgatgcacagataccagatgatgccacatcctcaccatcaccag
676  catcagcatcagcatcagcatcctcattctcatttcattcagcaa
721  tcaaagccccttcatatccaggagccatatcatcatacccgcctac
766  catcttcaccaagagtga

```

Em vermelho é destacada a sequência do domínio *forkhead* da proteína de CG9571. A sequência foi obtida na base de dados FlyBase (<http://flybase.org/>).

3.4 AMPLIFICAÇÃO DE DNA POR PCR

Reações de amplificação por *PCR* foram empregadas para verificação de clonagens, produção de fragmentos para uso em procedimentos utilizados para a obtenção do polipeptídeo/proteína e para obtenção de fragmentos para ensaios de interação de DNA-proteína.

Todas as reações foram realizadas utilizando-se o *kit GoTaq DNA Polymerase* (Promega), que já contém todos os reagentes necessários para a reação (enzima Taq DNA polimerase, dNTPs, $MgCl_2$ e tampão de reação), com exceção dos oligonucleotídeos iniciadores (*primers*). *Primers* específicos foram sintetizados para cada fragmento e técnica utilizados (Tabela 1). As condições do programa do termociclador utilizadas para cada fragmento são apresentadas na Tabela 2. Seguimos o protocolo do fabricante para a elaboração das reações. Nos casos de uso da técnica para confirmação de clones, metade da colônia de bactérias transformadas e cultivadas para preparação de DNA plasmidial ou para indução da expressão de proteína em células foi raspada da placa e suspensa em água ultrapura. A suspensão foi utilizada na reação de *PCR*.

Tabela 1. Primers sintetizados para amplificação dos fragmentos

Fragmento	Código	Sentido	Sequência (5'→3')
<i>eve 1</i>	ST15'	Forward	CATGAATTTCAGGCCTAATCACTTCCCT EcoRI
	eve 1 R	Reverse	TTGCGGCCTTGGGATATGTGAATATCA
<i>CG9571 BD in vitro</i>	CG-BDN	Forward	AGATCAGATTTATTTAGGTGACACTATAGAACAGA Promotor SP6 CCACCATGGAATGCAAAGCCGGCTTTCACCTACA Região Kozak
	CG-BDC	Reverse	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCATCACGT Cauda poli-A CTCGCCAATGGACAGATCCT Stop códon (2x)
<i>CG9571 BD</i>	d5	Forward	GGAATTCCATATGAATGCAAAGCCGGCTTTCAC EcoRI NdeI
	d3	Reverse	CGCGGATCCCTCACGTCTCGCCAATGGACAGAT BamHI Stop
<i>CG9571 FL</i>	pCG9571 5	Forward	GATCATATGGATACGACACCAATCTTC NdeI
	pCG9571 3	Reverse	ATCGGATCCCTCACTCTTGGTGAAGATGGTA BamHI Stop

Região de hibridização não foi destacada.

Tabela 2. Condições dos programas do termociclador para PCR

	<i>eve 1</i>	<i>CG9571 BD/FL</i>
Desnaturação inicial	94°C - 3min	94°C - 3min
32 ciclos	Desnaturação	94°C - 20seg
	Anelamento	52°C - 25seg
	Extensão	72°C - 25seg
Extensão final	72°C - 10min	72°C - 10min

3.5 PURIFICAÇÃO DE FRAGMENTOS DE DNA

Os kits Wizard SV Gel and PCR Clean-UP System (Promega) e PureLink Quick Gel Extraction & PCR Purification Combo (Invitrogen, Thermo Fisher Scientific) foram utilizados para a purificação de fragmentos amplificados ou de fragmentos de restrição. Seguimos o protocolo do fabricante para purificação de DNA diretamente das reações e a partir de gel de agarose de baixo ponto de fusão.

3.6 CLIVAGEM DE FRAGMENTOS DE DNA POR ENZIMAS DE RESTRIÇÃO

Reações de clivagem de DNA com enzimas de restrição foram empregadas para as mesmas finalidades que as reações de *PCR*. As enzimas EcoRI, BamHI, NdeI, XbaI e XhoI (*Promega*), EcoRI (*Cellco*) e BamHI e NdeI do *Anza Restriction Enzyme Cloning System* (*Thermo Fisher Scientific*) foram utilizadas nas reações conduzidas neste trabalho. Seguimos o protocolo de cada fabricante modificando apenas a quantidade de enzima aplicada na reação de acordo com a eficiência da catálise observada nos resultados das primeiras reações de clivagem com cada enzima. Para os fragmentos que necessitaram de clivagem com duas enzimas cujos tampões para a reação fossem incompatíveis, as reações de clivagem foram realizadas sequencialmente, com uma etapa de purificação entre as reações.

3.7 VETORES PARA CLONAGEM E EXPRESSÃO

Para as tentativas de clonagem de *CG9571 BD/FL* foram empregados os vetores pCR 2.1-TOPO (*Invitrogen*, Figura 10), pMiniT 2.0 (*NEB*, Figura 11) e pET-15b (*Novagen*, Figuras 12 e 13). O vetor pET-15b é um vetor para expressão gênica que permite altos níveis de expressão da proteína recombinante quando clonado em linhagens de *E. coli* que possuam gene para codificação da T7 RNA polimerase, como as linhagens utilizadas neste trabalho (3.1). Este plasmídeo possui um promotor para a enzima T7 RNA polimerase e é construído de modo que a expressão do gene clonado esteja sob o controle deste promotor, o que permite a transcrição do gene após a indução da expressão da T7 RNA polimerase por IPTG (*Novagen*, 2003). pET-15b também possui uma sequência que codifica para seis histidinas (marcador de poli-histidina) que possibilita a purificação por cromatografia de afinidade (3.13.4). Os *kits* com os vetores pCR 2.1-TOPO (pTOPO) e pMiniT 2.0 (pMiniT) linearizados foram adquiridos com os respectivos fabricantes ou com representantes. Uma alíquota de preparação do vetor pET-15B foi gentilmente cedida pelo laboratório do prof. Luis Netto (Instituto de Biociências, USP) e utilizada para preparações plasmidiais no laboratório (seção 3.2).

Figura 10. Mapa do vetor pCR 2.1-TOPO

Destaque para a região contida no gene *lacZα*, múltiplos sítios para enzimas de restrição e sítio para clonagem. *Primers* que podem ser utilizados para sequenciamento anotados. Fonte: manual do fabricante (Invitrogen, 2015). Disponível em: www.lifetechnologies.com.

Figura 11. Mapa do vetor pMiniT 2.0

Destaques para múltiplos sítios para enzimas de restrição e sítio para clonagem. Fonte: manual do fabricante (NEB, 2017). Disponível em: www.international.neb.com.

Figura 12. Mapa do vetor pET-15b

Identificação de principais sítios de enzima de restrição e de elementos importantes. Por seguir a convenção de numeração para o vetor pBR322, a região de expressão de T7 está invertida no mapa circular. Fonte: manual do fabricante (*Novagen*, 1998).

Figura 13. Região de clonagem e expressão do vetor pET-15b

Fonte: manual do fabricante (*Novagen*, 1998).

3.8 LIGAÇÃO DE *CG9571* BD/FL NOS VETORES DE CLONAGEM

Para a clonagem de *CG9571* BD/FL em vetor de expressão foram utilizadas duas estratégias de clonagem direcional: clonagem direta a partir do produto da PCR e clonagem em duas etapas com subclonagem em vetor carreador.

Os vetores pTOPO e pMiniT foram utilizados como vetores carreadores para a segunda estratégia. O kit de clonagem TOPO é um sistema de clonagem com ligação de extremidades coesivas. O vetor pTOPO possui uma timina 3' terminal que permite o

pareamento com a adenina 3' terminal adicionada pela enzima Taq polimerase ao produto de *PCR*. Além disto, a clonagem com este sistema não requer adição de enzima DNA ligase à reação de ligação, pois contém uma enzima topoisomerase I que catalisa a ligação fosfodiéster do inserto ao vetor. Já o sistema de clonagem da *NEB* ocorre por ligação de extremidades cegas. As misturas do *kit* utilizadas para compor as reações de ligação possuem componentes que removem a adenina adicionada à extremidade do produto de *PCR* e também contém a enzima T4 DNA ligase.

Seguimos as orientações de cada fabricante para as reações de ligação dos fragmentos de DNA amplificados e purificados aos vetores pTOPO (*TOPO TA Cloning Kit*) e pMiniT (*NEB PCR Cloning Kit*), incluindo modificações nas condições e na composição das reações quando as clonagens não fossem bem-sucedidas e os controles utilizados no processo de clonagem indicassem problemas nesta etapa. Nas ligações com o vetor pTOPO as modificações consistiram na adição de T4 DNA ligase (*Promega*) à reação, adotando-se o protocolo do fabricante da enzima para a reação, e o aumento de tempo de ligação (de 15-30min) e da relação molar entre inserto/vetor (de 1:1 até 7:1) para o vetor pMiniT. Após a ligação, as reações eram utilizadas para a transformação de células competentes (3.2), que foram cultivadas e utilizadas para preparações de DNA plasmidial (3.12).

Fragmentos de *CG9571 BD/FL* amplificados por *PCR* (primeira estratégia) e fragmentos clonados em vetor carreador (segunda estratégia) foram clivados com BamHI e NdeI seguindo as orientações dos fabricantes das enzimas de restrição. Os fragmentos de restrição purificados foram, então, utilizados em reações de ligação ao vetor pET-15b também clivado pelas mesmas enzimas. As reações de ligação foram conduzidas empregando-se T4 DNA ligase com tampão de ligação comum ou com tampão de ligação rápida (*Promega*). As condições recomendadas pelo fabricante dos reagentes também foram modificadas, elevando-se substancialmente o tempo de reação com o tampão de ligação rápida (de 5min para 2h30min). Todas reações foram conduzidas por 2-3h a temperaturas de 22-26°C.

3.9 EXTRAÇÃO E PURIFICAÇÃO DE DNA PLASMIDIAL

Para a lise celular, extração e purificação dos plasmídeos, foram utilizados os kits *QIAprep Spin Miniprep* (*Qiagen*), *Plasmid Midi Kit* (*Qiagen*) e *ZymoPure Plasmid Midiprep* (*Zymo Research*), seguindo-se o protocolo fornecido por cada fabricante, e o

protocolo de lise alcalina com SDS (Sambrook e Russel, 2001). Todos os métodos se baseiam no método de lise alcalina (Birnboim e Doly, 1979; Ish-Horowicz e Burke, 1981).

Em essência, as células coletadas da cultura por centrifugação (item 3.2.2) eram suspendidas e lisadas em solução contendo detergente e pH elevado. Esta solução também promove a desnaturação de proteínas celulares e do DNA cromossômico e plasmidial. Em seguida, adicionava-se solução para neutralização do pH e precipitação da maior parte dos componentes celulares pela substituição de íons de sódio por íons de potássio (Sambrook e Russel, 2001). As amostras eram centrifugadas em alta velocidade (14 a 15000 rpm) para separação do material desnaturado e o DNA plasmidial era recuperado do sobrenadante. No caso de *midipreps*, a remoção dos materiais celulares desnaturados também era realizada por filtração.

Procedimentos adicionais de purificação do DNA obtido eram realizados promovendo-se a ligação dos plasmídeos à resina da coluna de purificação dos *kits*. Após etapas de lavagem, o DNA plasmidial era eluído da coluna por centrifugação. No caso do protocolo de lise alcalina sem *kit* (Sambrook e Russel, 2001), o DNA plasmidial recuperado era precipitado por etanol e renaturado. Em todos os métodos o RNA contaminante era removido pela adição de RNase no início do processo (*kits*) ou ao final (Sambrook e Russel, 2001), após renaturação dos ácidos nucleicos precipitados. O DNA era mantido em tampão Tris-HCl 10mM pH 8.5.

3.10 ANÁLISE DE DNA POR ELETROFORESE EM GEL, ESPECTROFOTOMETRIA E SEQUENCIAMENTO DE DNA

Amostras de DNA foram analisadas por eletroforese em gel de agarose ou de poliacrilamida não desnaturante (PAGE). Géis de agarose comum foram preparados com tampão TAE 1X em concentrações de 0.6-1.5% e géis de agarose de baixo ponto de fusão em concentração de 1%. Às alíquotas de DNA analisado foram adicionados tampão de amostra Blue/Orange (*Promega*) ou Ez Vizion (*Amresco*) e tampão TAE 1X foi utilizado como tampão de corrida. Para detecção do DNA, os géis foram preparados com brometo de etídio (concentração final de 0.2 µg/ml) ou corados em banho de brometo de etídio após a separação por eletroforese. Quando EZ-VISION foi utilizado como tampão de amostra, o reagente fluorescente do tampão geralmente era suficiente para visualização dos fragmentos. O DNA era detectado por iluminação U.V.

Géis de poliacrilamida nativo foram preparados com tampão TBE 1X em concentrações de 5-8% seguindo protocolo de Sambrook e Russel (2001). Às alíquotas de DNA analisado foi adicionado tampão de amostra Blue/Orange e tampão TBE 1X foi utilizado como tampão de corrida. O gel era corado em solução de brometo de etídio após separação por eletroforese para detecção em iluminação U.V. Foram utilizados marcadores de massa molecular da *Sigma* (1kb) e *Promega* (100pb e 1kb).

As concentrações de DNA de fita simples (*primers*) e fita dupla (fragmentos amplificados, de restrição e construções plasmidiais) foram determinadas por espectrofotometria no Biofotômetro D30 (*Eppendorf*). Foram utilizados os métodos configurados pelo fabricante, que calculam a concentração em massa por volume a partir de uma fórmula que utiliza o valor de absorbância obtido para a varredura entre 225 e 350nm e um fator de conversão (estabelecido para cada tipo de fragmento de DNA).

Amostras de construções plasmidiais das clonagens que apresentaram inserto de tamanho específico em checagens com enzimas de restrição foram selecionadas e enviadas para sequenciamento no Centro de Estudos do Genoma Humano (Instituto de Biociências, USP, SP) para a confirmação dos clones. As sequências foram analisadas através dos programas *DNA Baser Sequence Assembly* e *BioEdit Sequence Alignment Editor* e pela ferramenta online *Clustal Omega Multiple Sequence Alignment*.

3.11 SEQUÊNCIA E PARÂMETROS DO FATOR DE TRANSCRIÇÃO

A proteína completa do fator de transcrição CG9571 é composta por 260 aa (Figura 14) e apresenta massa molecular de, aproximadamente, 30kDa e pI teórico de 9.62. A proteína completa recombinante possui cerca de 21 aa a mais, o que eleva sua massa molecular para cerca de 32.3kDa e seu pI teórico para 9.69. Já o polipeptídeo correspondente ao domínio de ligação de CG9571 (Figura 14) é composto por 91 aa, com massa molecular de aproximadamente 10.5kDa e pI de 9.25, e sua versão recombinante, 12.8kDa. Os parâmetros foram calculados por meio da ferramenta ExPASy ProtParam.

Figura 14. Sequência de aminoácidos de CG9571

```

1  MDTTPIFQSSFSIRSLLSVDKKEESPISKHNSGSSSFSSCSSSSSN
46  SSSDSMAAKSNAKPAFTYSALIVMAIWSSEKRLTLGICKWIAD
91  NFPYYRTRKSVWQNSIRHNLSLNPFFVRVPRALDDPGRGHYWALD
136  PYAEDLSIGETTGRLLRSNWQQNTGARPKVTGHPYQRMPYYGHGH
181  GNGPYIKAHSAYFPIMDHQHHAAAMVQHYQAMMHRYQMMPHPHHHQ
226  HQHQHQHPHSHFIQQSKPLHIQEPYHHTRYHLLHQE

```

Em vermelho é destacada a sequência do domínio *forkhead*. A sequência foi obtida na base de dados FlyBase (<http://flybase.org/>).

3.12 OBTENÇÃO DA PROTEÍNA POR SISTEMA DE EXPRESSÃO *IN VITRO*

Para as tentativas de obtenção direta de CG9571 *BD* foi utilizado o *kit* de expressão *in vitro* *TnT High Yield Wheat Germ Protein Expression System* (Promega). *Primers* específicos para uso neste sistema (Tabela 1) foram sintetizados conforme orientações do fabricante. Utilizamos instruções fornecidas pelo fabricante e obtidas na literatura (Parker *et al.*, 2011) para a elaboração das reações. A reação para transcrição e tradução *in vitro* era composta, essencialmente, por até 8 µl do fragmento da *PCR* purificado e 30 µl do *master mix* e foi incubada à 26°C por 2-3h.

3.13 OBTENÇÃO DA PROTEÍNA POR EXPRESSÃO EM BACTÉRIA

3.13.1 Indução da expressão

Para a indução das culturas de células descritas no item 3.2.3 foram empregadas diferentes concentrações de IPTG (0.1 a 1mM), em variadas temperaturas (18-37°C) por períodos de 30min a 24h. Os protocolos foram elaborados a partir de orientações, recomendações e protocolos da literatura (Sambrook e Russel, 2001; Middelberg, 2002; Novagen, 2003; Vallejo e Rinas, 2004; Gräslund S. *et al.*, 2008).

Para verificação da expressão de CG9571 *BD/FL* (item 3.13.7), diferentes clones selecionados de uma mesma placa eram utilizados para preparação de culturas de pequenos volumes (3 a 10 ml) e realização de ensaios iniciais de expressão. As induções eram realizadas a 37°C, sob agitação (160-180 rpm), com concentrações de IPTG de 0.5 a 1mM. Uma parcela do inóculo era transferida para outro frasco estéril antes da indução e crescida nas mesmas condições para ser empregada como controle negativo. Nos ensaios iniciais de expressão de CG9571 *FL*, os clones que apresentaram maiores níveis de expressão foram armazenados como *stabcultures* (item 3.2.3) e selecionados para ensaios de expressão em larga escala e obtenção da proteína.

Em geral, nas tentativas de obtenção da proteína na forma solúvel, as culturas eram crescidas até as faixas iniciais de crescimento exponencial das bactérias em temperaturas de 30°C e eram induzidas com concentrações mais baixas de IPTG (0.1 a 0.5mM), em temperaturas de 20 a 37°C, por períodos de 30min a 24h. Para as tentativas de obtenção de proteína recombinante a partir de corpos de inclusão foram feitas induções de culturas em condições mais brandas, com 0.5-0.6mM de IPTG, em 20-23°C durante 20 a 24h ou em 30°C por 14h, sob agitação a 150-170 rpm, ou em condições para obtenção de maior nível de expressão, com 0.5-0.8mM IPTG, 37°C, 180-200 rpm, 3-4h.

3.13.2 Lise celular e separação de frações solúvel e insolúvel

Após a indução, as células eram coletadas por centrifugação (3500-6000 x g, 2-10min) e suspensas em diferentes soluções de acordo com os propósitos assinalados na tabela 3. O volume de solução aplicado era equivalente a 10% do volume da cultura induzida.

Tabela 3. Soluções para suspensão de células induzidas

Solução	Aplicação
1 SDS 1%	Verificação do extrato celular total
2 Tris-HCl 20mM, pH 8	Separação de frações de proteínas solúveis e insolúveis
3 Tris-HCl 50mM pH 7.4, NaCl 150mM, EDTA 5mM, NP-40 0.5%	Separação de frações de proteínas solúveis e insolúveis
4 Fosfato de sódio 50mM pH 7.4, NaCl 500mM, imidazol 5mM	Purificação de fração solúvel em cromatografia de afinidade
5 Fosfato de sódio 50mM pH 7.4, NaCl 500mM	Purificação de fração solúvel em cromatografia de afinidade

Para a lise celular, a suspensão de células era submetida de 3 a 5 ciclos de congelamento (em nitrogênio líquido) e descongelamento (em água quente) ou a 2 ciclos de sonicação (10seg pulso, 50seg intervalo, por 2 ou 4min, com intervalo de 10min entre cada sonicação), em banho de gelo, em disruptor ultrassônico com ponteira. A separação das frações solúvel/insolúvel após lise era realizada por centrifugação a 10000 ou 14000 rpm, por 10-20min, a 4°C.

3.13.3 Recuperação de proteína recombinante de corpos de inclusão

Para a recuperação de CG9571 a partir de corpos de inclusão foram estabelecidas duas abordagens. Na primeira abordagem foram utilizados protocolos de solubilização com agentes não desnaturantes ou em condição desnaturante branda desenvolvidos com base na literatura (Jevsevar et al, 2005; Singh et al, 2005; Lu e Lin, 2012; Upadhyay et al, 2014; Singh *et al.*, 2015). Estes métodos foram empregados na tentativa de recuperação de proteínas parcialmente conformadas que podem estar presentes em corpos de inclusão, principalmente nos casos em que a indução é realizada em baixa concentração do indutor (IPTG) e à baixa temperatura. Foram utilizados, portanto, para corpos de inclusão isolados de culturas induzidas nas condições mais brandas (3.13.1). Os inóculos para indução também eram, geralmente, incubados em menores temperaturas (20-23°C ou 30°C). Após a lise celular e centrifugação para separação das frações (3.13.2), o precipitado contendo os corpos de inclusão e outras proteínas insolúveis foi solubilizado em tampão contendo detergente suave ou baixa concentração de ureia (Tabela 4). O volume aplicado era equivalente a 1% do volume da cultura original. A solubilização era conduzida entre 18-22°C, sob agitação, durante a noite. Uma alíquota para análise em gel (3.13.7) era separada e o restante era centrifugado nas mesmas condições já descritas. Após análise em gel, as proteínas eram submetidas a procedimentos de renaturação por diluição e diálise e concentração em tubo de ultrafiltração (3.13.6).

Tabela 4. Tampões para solubilização branda de proteínas de corpos de inclusão

Tampão	Composição
1 Solubilização em condição não-desnaturante	Tris-HCl 40mM pH 8, Triton X-100 0.5%
2 Solubilização em condição desnaturante branda	Tris-HCl 40mM pH 8, ureia 2 M
3 Solubilização em condição desnaturante branda	Tris-HCl 20mM pH 8, NaCl 0.5M, ureia 2 M, imidazol 5mM

Na segunda abordagem, corpos de inclusão foram solubilizados em condições com maior concentração do desnaturante (Tabela 5). Os protocolos foram elaborados a partir de recomendações dos manuais de purificação de proteína da *GE Healthcare* (2007 e 2009) e outras referências (Zaret e Stevens, 1995; Sambrook e Russel, 2001).

Tabela 5. Tampões para solubilização de proteínas de corpos de inclusão em condição desnaturante

Composição
1 Fosfato de sódio 50mM pH 7.2, NaCl 0.5M, ureia 4M
2 Fosfato de sódio 50mM pH 7.2, ureia 4M
3 Fosfato de sódio 50mM pH 7.2, NaCl 0.5M, ureia 8M

Os corpos de inclusão eram solubilizados nas mesmas condições descritas acima ou por agitação vigorosa em vórtex até que o agregado fosse completamente desfeito. Em alguns procedimentos, antes da solubilização em tampão contendo de 4 ou 8 M de ureia, foi acrescentada uma etapa de lavagem dos corpos de inclusão para remoção da maior quantidade possível de proteínas contaminantes. Nestes procedimentos, o precipitado insolúvel era suspenso uma ou duas vezes nos tampões contendo ureia 2 M por agitação vigorosa e centrifugado novamente, sendo suspenso pela última vez em solução contendo maior concentração do desnaturante (4 ou 8 M). Após centrifugação e análise em gel, as proteínas solubilizadas foram submetidas a procedimentos de purificação e renaturação descritos nos próximos itens.

3.13.4 Purificação e renaturação de proteína em coluna de cromatografia por afinidade

Colunas HisTrap HP (*GE Healthcare*) de 1ml e Tricorn (*GE Healthcare*) de 5ml, ambas com matriz carregada com níquel, foram utilizadas em tentativas de purificação da proteína recombinante por cromatografia de afinidade por íons metálicos imobilizados (IMAC). A histidina é um aminoácido que possui afinidade por metais de transição, como o níquel (Bresolin *et al.*, 2009). A existência de uma sequência de seis resíduos de His na proteína recombinante tem como objetivo, portanto, permitir que a proteína de interesse fique retida na matriz da coluna através da interação com Ni^{2+} . Proteínas endógenas do hospedeiro presentes na mesma fração são removidas por etapas de lavagem da coluna e a proteína de interesse é eluída com a aplicação de solução tampão contendo imidazol, substância que também possui afinidade pelo níquel e compete com a interação da proteína de fusão na coluna.

As cromatografias em coluna HisTrap HP de 1ml foram realizadas manualmente, com emprego de seringas e utilizando espectrofotômetro para a detecção de proteína nas frações coletadas no decorrer do procedimento. Já as cromatografias em

colunas Tricorn de 5ml foram realizadas em cromatógrafo ÄKTA Purifier (*GE Healthcare*). Os protocolos utilizados foram elaborados seguindo as instruções e recomendações do manual da coluna e livros técnicos do fabricante (*GE Healthcare*, 2007 e 2009). Em essência, em cada procedimento de cromatografia a coluna era lavada com 5 volumes de água pura ou ultrapura estéril (quando preenchida por etanol 20% para armazenamento ou após uma purificação) e equilibrada com 3-5 volumes de tampão de ligação. A amostra de proteína, previamente filtrada em filtro de seringa (membrana de 0.22 μ m), era aplicada na coluna e, em seguida, a coluna era lavada com 5 volumes de tampão de ligação. Para eluição da proteína, até 5 volumes de tampão de eluição com concentração crescente de imidazol eram aplicados na coluna.

Os tampões utilizados na tentativa de concentração de eventual proteína recombinante na fração solúvel e para purificação de proteínas solubilizadas de corpos de inclusão com ureia são apresentados na Tabela 6. O fluxo de aplicação das soluções e amostra na coluna foi de 0.5-1ml/min, quando o procedimento foi realizado em cromatógrafo, e de aproximadamente 0.1-1ml/min, quando feito manualmente (considerando 30 gotas por minuto equivalentes a 1ml/min).

Colunas para IMAC também podem ser empregadas para renaturação de proteínas. Incluindo-se uma etapa de lavagem com tampão de ligação contendo concentrações decrescentes do agente desnaturante, o remodelamento da proteína pode ser promovido antes da etapa de eluição, enquanto a proteína se encontra imobilizada na matriz, permitindo que a técnica sirva para os dois propósitos, de renovelamento da proteína e purificação. Os protocolos para renaturação em coluna foram elaborados com base em manual e artigos da literatura (*GE Healthcare*, 2007; Rehm *et al.*, 2001; Zouhar *et al.*, 1999) e a cromatografia foi realizada manualmente ou por meio de cromatógrafo.

Os tampões utilizados também estão descritos na Tabela 6. Em essência, a coluna era equilibrada com tampão de ligação contendo ureia. Após injeção da amostra, a coluna era lavada com 2 volumes de coluna com o mesmo tampão de ligação. Na etapa de renaturação, o tampão de ligação com ureia foi diluído gradualmente em tampão de renaturação sem ureia. Foram aplicados 5 (coluna de 5 ml) ou 25 volumes de coluna (coluna de 1 ml). Por fim, para a eluição foi utilizado um gradiente crescente de imidazol com tampão fosfato sem ureia. O fluxo de injeção em todas as etapas foi de 0.5 ml/min.

Tabela 6. Soluções tampão para purificação por IMAC

Aplicação	Tampão	Composição
Purificação de proteína solúvel	Ligação e lavagem	Fosfato de sódio 50mM pH 7.2, NaCl 0.5M
	Eluição	Fosfato de sódio 50mM pH 7.2, NaCl 0.5M, imidazol 0.5M
Purificação e renaturação de proteína insolúvel	Ligação e lavagem	Fosfato de sódio 50mM pH 7.2, NaCl 0.5M, ureia 4M
	Renaturação	Fosfato de sódio 50mM pH 7.2, NaCl 0.5M
	Eluição	Fosfato de sódio 50mM pH 7.2, NaCl 0.5M, imidazol 0.5 M
Purificação de proteína insolúvel	Ligação e lavagem	Fosfato de sódio 50mM pH 7.2, NaCl 0.5M, ureia 8M
	Eluição	Fosfato de sódio 50mM pH 7.2, NaCl 0.5M, ureia 8M, imidazol 0.5M
Purificação e renaturação de proteína insolúvel	Ligação e lavagem	Tris-HCl 20mM pH 8, NaCl 0.5M, ureia 2 M, imidazol 5mM
	Renaturação	Tris-HCl 20mM pH 8, NaCl 0.5 M
	Eluição	Tris-HCl 20mM pH 8, NaCl 0.5M, imidazol 0.5 M

3.13.5 Purificação de proteína recombinante a partir de SDS-PAGE

Como técnica alternativa a purificação por IMAC, proteínas solubilizadas em ureia foram separadas por SDS-PAGE para eluição da proteína recombinante a partir do gel de poliacrilamida. O protocolo foi elaborado a partir de referências na literatura (Hager e Burgess, 1980; Zaret e Stevens, 1995; *Thermo Scientific*, 2009). Espaçadores que delimitam os poços do gel de poliacrilamida foram removidos, após a polimerização, para aplicação do maior volume possível de amostra, deixando apenas um segundo poço, em uma das extremidades do gel, para aplicação de alíquota da proteína recombinante utilizada como referência para a localização de CG9571 no restante do gel. Foram utilizados géis de resolução na concentração de 12.8% e as mesmas condições para SDS-PAGE empregadas na verificação de proteínas (3.13.7).

Ao término da eletroforese, uma faixa vertical correspondente à região da canaleta de aplicação da alíquota para referência era seccionada para detecção da

proteína com Brilliant Blue R (*Sigma*; item 3.13.7). Em seguida, esta canaleta lateral era corada e levada de volta ao seu lugar no gel para delimitar a faixa horizontal em que estava retida a proteína recombinante. A faixa excisada do gel era transferida para um tubo tipo Falcon de 15ml com 1-1.5ml de solução de eluição (TrisCl 50 mM pH 7.9, NaCl 150 mM, EDTA 0.1 mM, SDS 0.03%) e fragmentada com espátula. Para eluição, o tubo era mantido sob agitação, por 10-12h, a 20-22°C. Uma alíquota da eluição era verificada por SDS-PAGE. A eluição era centrifugada a 3000 x g por 5-10 min e todo o sobrenadante era transferido para um novo tubo tipo Falcon.

Para precipitação da proteína e remoção do detergente SDS, 5 volumes de solução de acetona/metanol 1:1 a -20°C eram aplicados na proteína eluída. O tubo era mantido a -20°C durante a noite. O conteúdo do tubo era transferido para um tubo de centrífuga de fundo cônico e centrifugado a 10000 rpm, por 15 min, em centrífuga pré-refrigerada a 4°C. O sobrenadante era descartado e todo resíduo da solução removido cuidadosamente invertendo o tubo em papel toalha por 2-5 min e deixando o tubo secar ao ar por mais 25min à temperatura ambiente ou em gelo. Em alguns procedimentos, antes da remoção completa do resíduo de solução, foi adicionada uma etapa de lavagem com mais 5 volumes da solução de acetona/metanol. Os tubos foram centrifugados novamente e então secos.

Para a solubilização do precipitado de proteína, detectável somente quando amostras de mais de um gel eram reunidas, foram utilizadas diferentes soluções (Tabela 7). A proteína solubilizada em condições desnaturantes foi renaturada por diluição e diálise.

Tabela 7. Tampões para solubilização de proteína recombinante precipitada

Tampão	Composição
1 Solubilização em condição desnaturante	HEPES 20mM, NaCl 100mM, DTT 5mM, MgCl ₂ 1mM, ureia 6M, glicerol 20%
2 Solubilização em condição desnaturante	Fosfato de sódio 50mM pH 7.2, NaCl 0.5M, ureia 4M

3.13.6 Renaturação por diluição e diálise

Na maioria das tentativas de renaturação de CG9571 FL foram empregadas as técnicas de diluição e diálise, geralmente usadas em combinação. Para renovelamento por diluição, as amostras obtidas nos itens 3.13.3, 3.13.4 (apenas purificadas em coluna)

e 3.13.5, com volume máximo de 1ml, foram diluídas de 6 a 15 vezes no mesmo tampão em que estavam solubilizadas, porém sem ureia. As amostras eram aplicadas no tampão de diluição/renaturação em alíquotas, de 15 a 50 µl, com um intervalo de 30 seg. a 5 min entre cada aplicação. Em seguida, as amostras eram dialisadas contra 60-100 ml de tampão fosfato 50 mM pH 7.2 para remoção da ureia diluída e de outros componentes durante 6 a 36 h. No caso das preparações obtidas a partir das amostras do item 3.13.5, tampões fosfato contendo concentração decrescente de NaCl ou tampões HEPES contendo todos os componentes menos ureia foram utilizados para diálise. Menores tempos de diálise foram empregados quando o procedimento era realizado sob agitação em temperaturas de 16-20°C, enquanto tempos maiores foram utilizados para diálise em refrigerador, de 4-7°C.

As diálises eram realizadas através de membrana própria para diálise (*Dialysis tubing cellulose membrane*, Sigma, MWCO 14000 Da), empregada para vedar tubo de ultrafiltração, ou apenas empregando-se tubos de ultrafiltração de proteína (Amicon, Merck, 10000 MWCO e Vivaspin, GE Healthcare, 3000 MWCO), quando a membrana ainda não estava disponível. Os tubos de ultrafiltração possuem uma coluna removível com membrana de corte de massa molecular na base que foi utilizada na tentativa de promover a diálise. Após diálise, as amostras eram concentradas nas mesmas colunas de ultrafiltração descritas a 3900 x g, 4-16°C.

Preparações de proteína para renaturação por diálise eram mantidas em maior volume (até 6 ml) para diminuição da concentração de proteínas em solução. As amostras eram renoveladas nos mesmos sistemas descritos acima. Protocolos de renaturação de proteínas *forkhead* foram utilizados como referência (Pierrou *et al.*, 1994; Yu *et al.*, 1999; Zaret e Stevens, 1995). A concentração de ureia em cada tampão de diálise era reduzida gradualmente. Foram utilizados tampões com reduções de 0.5, 0.3 ou 0.1M a cada troca, sendo cada tampão trocado de 30 em 30 min quando dialisado sob agitação, a 16-20°C, ou a cada 4-6 h quando dialisado em refrigerador. Além da ureia, outros componentes do tampão também eram geralmente removidos ou tinham sua concentração reduzida, como NaCl, imidazol, glicerol e DTT, com exceção de duas preparações em tampão HEPES. As amostras eram concentradas como descrito acima após o término da diálise.

Após os procedimentos de diluição, diálise e ultrafiltração, a concentração de proteína na preparação era determinada por espectrofotometria (item 3.13.7) antes e

após centrifugação a 14000 rpm. Quando possível, as preparações também foram analisadas por SDS-PAGE.

3.13.7 Análise de proteína por eletroforese em gel e espectrofotometria

Amostras de proteína foram analisadas por SDS-PAGE (Laemmli, 1970) no sistema de eletroforese vertical Mini PROTEAN Tetra Cell (*Bio-Rad*). Os géis de poliacrilamida foram preparados com tampão Tris-Glicina 1X em concentrações de 12-14% (para gel de resolução) seguindo protocolo de Sambrook e Russel (2001), a partir de solução de acrilamida/bisacrilamida (29:1) de 30%. Às alíquotas de proteína analisadas era aplicado tampão de amostra 2X (Tris-HCl 100mM pH 6.8, SDS 4%, azul de bromofenol 0.2%, glicerol 20% e DTT 200mM) e tampão Tris Glicina 1X foi utilizado como tampão de corrida. A eletroforese era realizada em 130-150V. Para detecção, o gel era corado com Brilliant Blue R e em seguida tratado com solução descorante de ácido acético 10% e metanol 30%.

A concentração de proteína nas amostras foi calculada através da equação da lei de Beer-Lambert-Bouguer² a partir do coeficiente de extinção molar da proteína recombinante, calculado pela ferramenta online ExPASy em 48360 para CG9571 FL. A absorbância de cada preparação foi determinada por leitura do espectro no comprimento de onda de 280 nm no Biofotômetro D30 (*Eppendorf*).

3.14 INTERAÇÃO ENTRE DNA-PROTEÍNA

3.14.1 Ensaios de interação entre CRMs e amostras de proteína

Foram pesquisadas condições de interação entre fatores *forkhead* e CRMs na literatura e duas condições foram selecionadas (Bao *et al.*, 2011; Julien *et al.*, 2002), incluindo algumas modificações na composição dos tampões (Tabela 8).

² $A_{\lambda} = c\epsilon l$

Onde:

A_{λ} = absorbância

c = concentração (moles/L)

ϵ = coeficiente de absorbância molar ou coeficiente de extinção molar

l = caminho óptico

Tabela 8. Composição de tampões para interação DNA-proteína

Tampão	Composição
1 Tampão I H	HEPES 10mM pH 7.9, KCl 50mM, MgCl ₂ 4mM, BSA 0.5 mg/ml, DTT 1mM, glicerol 10%
2 Tampão I T	TrisHCl 10mM pH 7.5, KCl 50mM, MgCl ₂ 4mM, BSA 0.5 mg/ml, DTT 1mM, glicerol 10%
3 Tampão II H	HEPES 12mM pH 7.9, KCl 100mM, MgCl ₂ 5mM, BSA 10 µg/ml, DTT 0.6mM, glicerol 12%
4 Tampão II T	TrisHCl 10mM pH 7.5, KCl 100mM, MgCl ₂ 5mM, BSA 10 µg/ml, DTT 0.6mM, glicerol 12%

De 25 a 100 ng de DNA eram adicionados à reação com uma quantidade variável da amostra da proteína. Para preparações mais puras, contendo a proteína recombinante isolada ou em maior concentração, de 50 ng a 3 µg de proteína eram utilizados. Para amostras de menor pureza, um volume variável da preparação (3-5 µl) era aplicado. Nas tentativas de interação com amostra do *kit* de expressão *in vitro*, 4µl foram utilizados. O volume final das reações era de 20-30 µl. As reações eram incubadas entre 20 e 27°C por 20-30min.

3.14.2 Ensaios de retardo da mobilidade eletroforética

Para verificação da interação DNA-proteína foi utilizada a técnica de ensaio de retardo de mobilidade eletroforética (*EMSA*) (Fried e Crothers, 1981; Garner e Revzin, 1981). A técnica é baseada na observação de que complexos formados entre proteínas e ácidos nucleicos geralmente apresentam menor mobilidade eletroforética do que ácidos nucleicos livres (Hellman e Fried, 2007), o que resulta na formação de bandas discretas correspondentes a estes complexos (Sambrook e Russe, 2001). Podem ser empregados géis de poliacrilamida ou de agarose nativos nesta técnica, embora géis de poliacrilamida são mais utilizados porque oferecem melhor resolução na detecção do complexo proteína-ácido nucleico (Fried, 1989) e pelo fato de já ter sido observado que alguns complexos apresentam maior estabilidade em matrizes de poliacrilamida do que em agarose (Fried e Bromberg, 1997).

Neste trabalho, as reações de incubação das preparações de proteína com os fragmentos de DNA (3.14.2) e os fragmentos dos *CRMs* livres foram aplicados e separados por eletroforese em géis de poliacrilamida de 5-7%, preparados com solução

de acrilamida/bisacrilamida (29:1) de 45% em TBE 0.5 ou 1X pH 8.4 ou em Tris-Glicina 1X (25mM Tris, 250mM Glicina) pH 8.9, e géis de agarose de 1.5 a 2%, preparados em TBE 0.5X. As eletroforeses eram realizadas em voltagens de 120-200V a temperaturas de 15-20°C ou em temperaturas mais baixas, com a cuba de eletroforese mantida em banho de gelo.

Além da eletroforese vertical convencional em gel de poliacrilamida nativo, também empregamos a técnica Blue-Native-PAGE (BN-PAGE) (Schägger e von Jagow, 1991) em tentativas de análise da interação. A técnica consiste na adição do corante aniônico Coomassie Brilliant Blue G-250 a amostras que contém proteínas básicas para induzir mudança de carga destas proteínas. O corante se liga às proteínas de forma análoga ao SDS, porém sem causar a desnaturação da estrutura proteica, mantendo a conformação e a atividade da proteína intactas e conferindo à proteína carga negativa (Westermeyer, 2016; Wittig *et al.*, 2006). Uma solução de 5% do corante foi preparada com Coomassie Blue G-250 (*Thermo Fisher Scientific*) e alíquotas desta solução eram adicionadas nas reações de interação ao final do tempo de incubação e no tampão do catodo (TBE 0.5 ou 1X) na concentração final de 0.02% do corante. As amostras eram aplicadas em gel de poliacrilamida após 5min de incubação do corante com as reações. Também utilizamos a solução do corante aniônico em reações que foram analisadas por eletroforese em gel de agarose.

Como controle negativo para as interações foram utilizadas sequências de regiões codificantes da proteína completa de outro gene *gap*, *tll* (1356 pb), e do domínio de ligação do próprio gene *CG9571* incubados com amostras de proteína nas mesmas condições. Para a detecção da formação ou não de complexos, os géis foram corados em solução de brometo de etídio (0.2 µg/ml), que já se mostrou suficiente para visualização de fragmentos de DNA nestes complexos (Garner e Revzin, 1986; Förster-Fromme e Jendrosseck, 2010).

4 RESULTADOS E DISCUSSÃO

4.1 CONFIRMAÇÃO DAS CONSTRUÇÕES PLASMIDIAIS CARREADORAS DE *EVE* 1, *EVE* 2, cDNA DE *CG9571* E OBTENÇÃO DOS FRAGMENTOS DE DNA RELATIVOS AOS *CRMS*

Checamos todas as construções plasmidiais por meio de clivagem com enzimas de restrição. As reações liberaram insertos com tamanho esperado para todos os fragmentos: *eve* 1 (Figura 15A), *eve* 2 nas formas selvagem e mutada (Figura 15B) e cDNA completo de *CG9571* (Figura 16).

Figura 15. Checagem das construções plasmidiais pTOPO-*eve* 1, pBS-*eve* 2 rx e pBS-*eve* 2 Δ GTTT com enzimas de restrição

(A) pTOPO-*eve* 1 foi clivado com a enzima EcoRI para liberação do inserto (o vetor apresenta dois sítios para EcoRI, um à jusante e outro à montante do sítio de clonagem). Houve liberação de inserto do tamanho esperado (raia 2). (B) pBS-*eve* 2 rx (forma selvagem de *eve* 2) e pBS-*eve* 2 Δ GTTT (forma mutada) foram clivados com as enzimas XbaI e XhoI e também houve liberação de insertos de tamanho esperado em torno de 600pb (raias 2 e 3). (A e B) Raia 1: MM 100pb *Promega*. Eletroforese em gel de agarose 1%, TAE 1X. Fonte: Gueller (2016).

Figura 16. Checagem da construção plasmidial pTOPO-CG9571 com enzimas de restrição

Clivagem de construção contendo o cDNA com EcoRI (raia 2), para liberação do inserto, e com XbaI (raia 3), para linearização da construção. Fragmento liberado pela clivagem com EcoRI apresenta tamanho esperado (783pb). Raia 1: plasmídeo não clivado (controle). Raia 4: MM de 1kb da *Sigma*. Eletroforese em gel de agarose 0.6%, TAE 1X. Fonte: Gueller (2016).

Para aumentar a quantidade de pTOPO-*eve1*, pBS-*eve 2 rx* (forma selvagem) e pBS-*eve 2 ΔGTTT* (forma mutada), transformamos células competentes e realizamos preparações de DNA plasmidial. Por meio de clivagem dos plasmídeos com enzimas de restrição e purificação dos fragmentos produzidos obtivemos material em quantidade suficiente para os ensaios de interação. Para alguns ensaios, também amplificamos fragmentos de *eve 1* por *PCR* (dado não apresentado). Os fragmentos de *eve 2* foram obtidos por meio de clivagens com EcoRI e XhoI (dado não apresentado) ou com XbaI e XhoI. Os insertos liberados pela última reação apresentam cerca de 30pb a mais.

4.2 EXPRESSÃO DE CG9571 *BD IN VITRO* E OBTENÇÃO DE EXTRATO PARA TESTE DE INTERAÇÃO COM *EVE 1*

Como estratégia inicial, optamos pela obtenção do domínio ligador CG9571 por meio de transcrição e tradução *in vitro*. Verificamos na literatura que o uso do domínio de ligação da proteína é suficiente para a investigação da interação de proteínas com sequências de DNA (Pierrou *et al.*, 1994; Zhu *et al.*, 2012). Optamos, inicialmente, pelo sistema de expressão *in vitro* porque é um método que envolve poucas etapas de manipulação comparativamente ao método clássico de obtenção de formas recombinantes expressas em bactéria.

Após testes realizados para padronizar as condições de *PCR* obtivemos produto do tamanho esperado de 353 pb (Figura 17), considerando-se que as sequências acrescentadas aos *primers* para esta técnica geram um fragmento contendo 80 pb a mais que o domínio *forkhead* (item 3.4).

Figura 17. Detecção de fragmento correspondente a *CG9571 BD* obtido por *PCR* para transcrição e tradução *in vitro*

Amplificação de fragmento com tamanho esperado (aprox. 350pb. Raia 1). Raia 2: MM de 1kb da *Sigma*. Eletroforese em gel de agarose 0.6% TAE 1X. Fonte: Gueller (2016).

O *template* de DNA assim gerado foi purificado de gel e utilizado para reações de transcrição e tradução *in vitro* (item 3.12). Após a reação, alíquotas eram aplicadas em gel SDS-PAGE para verificação da expressão do domínio. Não foi possível identificar em nenhum dos ensaios polipeptídeo do tamanho esperado (10kDa) para CG9571 BD (Figura 18).

Figura 18. Detecção de polipeptídeo correspondente a CG9571 BD após transcrição e tradução *in vitro*

Alíquotas de reações realizadas com dois tempos diferentes de incubação: 2h (raia 3) e 3h (raia 5). Nas raias 2 e 4 foram adicionadas alíquotas de misturas de reação sem fragmentos amplificados de *CG9571 BD*, incubados à mesma temperatura dos experimentos como controle negativo das reações. Não foi possível detectar nenhuma banda de tamanho esperado nas amostras das reações (aprox. 10kDa). SDS-PAGE, gel de resolução 14%, tris-glicina. Fonte: Gueller (2016).

Apesar disto, assumimos ineficiência na geração do polipeptídeo e a possibilidade de existência de uma quantidade mínima suficiente para possibilitar interação com *CRMs* alvo e detecção nos experimentos de EMSA. Volumes variados do extrato foram incubados com *eve 1* nas condições descritas em 3.14.1 com os tampões 1 e 3 (Tabela 8). No entanto, não verificamos retardo na migração eletroforética (dados não mostrados).

Novas tentativas de transcrição e tradução *in vitro* foram repetidas, inclusive com a importação de um novo *kit*. Desconfiamos que a ineficiência dos *kits* possa estar relacionada com a armazenagem dos mesmos por longo no ultrafreezer do laboratório, ou então pelo próprio sistema de transporte do fornecedor. Nossas conversas com fabricante resultaram na necessidade de adquirir outros insumos para verificar o problema e naquele momento, pelo tempo perdido e os gastos já dispendidos com os reagentes, abandonamos essa estratégia.

4.3 CLONAGEM DE *CG9571 BD* E EXPRESSÃO EM BACTÉRIA

4.3.1 Clonagens em vetor carreador e de expressão

A segunda estratégia adotada foi a de construção de DNA recombinante para expressão do domínio da proteína em bactéria. Continuamos optando pela obtenção do domínio *forkhead*, pois fragmentos menores da proteína, como domínios isolados, em geral são mais facilmente expressos na forma solúvel e podem apresentar maior nível de expressão em comparação a proteína completa (Lotti e Pollegioni, 2014; Koth *et al.*, 2003).

Para amplificação do fragmento, incorporamos sítios para as enzimas de restrição BamHI e NdeI nas extremidades dos *primers* utilizados para a *PCR* com o objetivo de direcionar a subclonagem em vetor de expressão pET-15b. Obtivemos fragmentos amplificados de tamanho esperado que foram purificados para serem empregados nas reações de clivagem e ligação ao vetor (Figura 19).

Figura 19. Detecção de fragmento correspondente a *CG9571 BD* obtido por *PCR* para clonagem no vetor de expressão

(A) Amostra da reação de *PCR*. 1) MM de 1kb da *Sigma*; 3) *PCR* de *CG9571 BD*. (B) Amostra de produto da *PCR* após purificação do fragmento (mesma ordem). Amplificação de fragmento de tamanho esperado (aprox. 270pb). Eletroforese em gel de agarose 0.8% TAE 1X. Fonte: Gueller (2016).

Realizamos diversas tentativas de clonagem sem sucesso. Desconfiando da eficiência da clivagem dos fragmentos amplificados, optamos, então, pela tentativa de construção do plasmídeo de expressão através de duas clonagens sequenciais, empregando o vetor pCR 2.1-TOPO como vetor carreador na primeira etapa de clonagem. As tentativas de clonagem utilizando as condições sugeridas pelo protocolo

do fabricante não produziram clones. Levantamos a hipótese de que a enzima responsável pela catalisação da ligação fosfodiéster do *kit* não estava funcional e adicionamos ligase à reação, modificando também as condições da ligação (tempo e temperatura), e obtivemos sucesso na subclonagem. A clonagem foi verificada por *PCR* de colônias selecionadas da placa (dados não mostrados) e, após preparações de DNA plasmidial, por clivagem com enzimas de restrição (Figura 20). Essas checagens confirmaram os tamanhos esperados dos fragmentos clonados e a integridade das extremidades dos fragmentos. O sequenciamento do DNA plasmidial dos clones confirmou a integridade da sequência dos nucleotídeos (e.g. Apêndice 1) e a construção plasmidial do clone 1 foi utilizada para a transferência do fragmento para o vetor de expressão.

Figura 20. Clonagem de *CG9571 BD* em pTOPO verificada com enzimas de restrição

Plasmídeos clivados com BamHI e NdeI. Houve liberação de inserto com tamanho esperado (273pb) de todas as construções verificadas (raias 2-6. Seta vermelha indica os fragmentos). Raia 1: plasmídeo não clivado (controle da reação de restrição). Raia 7: MM de 1kb da *Sigma*. Eletroforese em gel de agarose 1% TAE 1X. Fonte: Gueller (2016).

A construção contendo *CG9571 BD* e o vetor de expressão foram clivados com enzimas de restrição, inserto e vetor foram purificados a partir de gel e utilizados para as reações de ligação. Novamente não obtivemos sucesso nas primeiras tentativas de clonagem. Como o problema parecia residir na ligação do inserto ao vetor, ajustamos as condições da ligação e obtivemos êxito na clonagem alterando significativamente o tempo de reação sugerido para o uso do tampão de ligação rápida para a T4 DNA ligase (3.8). A clonagem também foi verificada por clivagem com enzimas de restrição (Figura 21) e clones positivos foram sequenciados. A integridade da sequência dos nucleotídeos também foi confirmada (Apêndice 2).

Figura 21. Clonagem de *CG9571 BD* em pET-15b verificada com enzimas de restrição

(A) Checagem do DNA plasmidial de dois clones com clivagem pelas enzimas BamHI e NdeI. Inseto com tamanho esperado (273pb) foi liberado para as duas construções (raias 1 e 2). (B) Checagem do DNA plasmidial dos clones com as enzimas BamHI e XbaI. Inseto com tamanho ligeiramente superior ao domínio ligador liberado como esperado para os dois clones (raias 1 e 3). Raia 2: plasmídeo não clivado (controle da reação de restrição). Raias (A) 3 e (B) 4: MM 1kb *Promega*. Eletroforese em gel de agarose 1% TAE 1X. Fonte: Gueller (2016).

4.3.2 Tentativas de indução de expressão de pET-15b-*CG9571 BD*

O plasmídeo construído, pET-15b-*CG9571 BD*, foi utilizado para transformar primeiramente células da linhagem BL21(DE3)pLysS e diferentes clones foram selecionados para ensaios iniciais de expressão do domínio. Conduzimos diversos experimentos de indução variando parâmetros e condições, utilizando tempos de indução de até 15h (26 ou 37°C, 1mM IPTG), mas mesmo assim não foi possível detectar banda de tamanho específico (aproximadamente 12kDa) em análises por SDS-PAGE e coloração por Coomassie Blue.

Além de culturas não induzidas, utilizamos como controle negativo para a indução culturas de células transformadas com o vetor pET-15b vazio, considerando que pudesse ter ocorrido expressão de fundo do domínio, o que teria atrapalhado a identificação de uma possível expressão em menor nível. Descartamos esta hipótese, pois não pudemos identificar nenhuma diferença no perfil de polipeptídeos das amostras induzidas e controle (Figura 22).

Figura 22. Indução da expressão de CG9571 BD em BL21(DE3)

Extrato celular total de células transformadas com vetor pET-15b vazio (controle negativo) e de células transformadas com pET-15b-*CG9571 BD*. Indução da cultura com 1mM IPTG, a 37°C, por 0 (T₀), 2 (T₁), 3 (T₂) ou 4h (T₃). Raia 9: MM BenchMark. Não é possível detectar nenhuma banda de tamanho específico (aprox. 12.5kDa). SDS-PAGE, gel de resolução 14%, tris-glicina. Fonte: Gueller (2017).

Outra tentativa de obtenção do domínio ligador foi realizada utilizando células competentes da linhagem BL21(DE3)Star que permitem aumentar o nível de expressão de proteína recombinante, considerando que o problema pudesse residir na degradação do mRNA de *CG9571 BD* ou do próprio polipeptídeo (item 3.1). No entanto, novamente não fomos capazes de identificar *CG9571 BD* em nenhum nível de expressão ou forma (solúvel ou insolúvel) em nenhum dos ensaios de indução (dados não mostrados).

Realizamos uma tentativa de purificação da fração solúvel do extrato celular de culturas induzidas por IMAC com o intuito de concentrar eventual quantidade do polipeptídeo recombinante expresso em quantidade inferior ao limite de detecção por Coomassie Blue em gel, mas não houve detecção de proteína nas medições em espectrofotômetro de nenhuma fração de eluição (dados não mostrados).

Não conseguimos determinar a razão pela qual o domínio ligador não foi expresso em bactéria – pelo menos não em quantidade mínima para detecção. Proteínas ou polipeptídeos de baixa massa molecular tem a tendência de serem degradados rapidamente em *E. coli*, mas, como regra geral, isto ocorre para moléculas com massa inferior a 5kDa (Bell, 2001), o que não é o caso de *CG9571 BD*. Outra possibilidade para a não detecção ou para detecção em baixos níveis da expressão de proteínas

recombinantes é o efeito tóxico acentuado que a expressão de alguns genes pode exercer nas células (Bell, 2001). Contudo, como o sistema de expressão que utilizamos, que permite reduzir ao máximo a expressão basal do gene alvo durante o crescimento da cultura celular previamente à indução por IPTG, já se mostrou suficiente para a expressão da maioria dos genes potencialmente tóxicos em *E. coli* (Novagen, 2003; Dubendorff e Studier, 1991), deixamos esta possibilidade em aberto e abandonamos a estratégia de obtenção do domínio para tentarmos obter a proteína completa de CG9571.

Após obtermos sucesso na expressão da proteína completa (descrito a partir do item 4.4.2), decidimos realizar um breve experimento para verificarmos a hipótese não descartada da toxicidade de CG9571 *BD* às células hospedeiras. Nós observamos uma aparente redução na massa de células coletadas por centrifugação da cultura induzida em relação a massa coletada da cultura não induzida de mesmo volume diversas vezes durante as tentativas de expressão do domínio, o que consideramos um indicativo de que o polipeptídeo pudesse ter sido expresso e exercido efeito tóxico que produziu a morte das células induzidas. Entretanto, não realizamos nenhum experimento controlado nem utilizamos algum procedimento analítico para análise desta suspeita. Depois de observarmos a expressão de CG9571 *FL*, conduzimos um experimento com o propósito de analisarmos o impacto da indução de culturas transformadas com as construções plasmidiais contendo o domínio ou a proteína completa de CG9571.

Foram preparados e crescidos dois inóculos de 25 e 50 ml de meio LBA para cada construção. Antes da indução, metade do volume de cada inóculo foi transferido para um novo frasco de Erlenmeyer estéril a fim de servir de controle não induzido. Os inóculos foram induzidos com 1 mM de IPTG por 2h30min e ao final da indução a densidade óptica de cada cultura foi medida por espectrofotometria. Os valores obtidos para cada cultura são apresentados na Figura 23.

Figura 23. Efeito comparativo da indução de CG9571 *BD* e CG9571 *FL* sobre a densidade de células bacterianas em culturas celulares

Culturas de células BL21(DE3)pLysS transformadas com pET-15b-CG9571 *BD* e pET-15b-CG9571 *FL*. O primeiro par de colunas de cada construção se refere ao valor de DO₆₀₀ obtido para a cultura de 25 ml e o segundo par de colunas para a cultura de 50 ml.

Os valores de DO₆₀₀ tem relação com a densidade de células em suspensão e, portanto, com o número de células em cada cultura. Verificamos que embora tenha apresentado crescimento em relação ao momento da indução (as culturas foram induzidas em DO₆₀₀ de aproximadamente 1.00), as culturas induzidas de CG9571 *BD* apresentaram uma redução significativa na densidade de células em comparação com as culturas não induzidas, com uma diferença ainda maior na cultura de 50 ml, enquanto não houve nenhuma alteração significativa na densidade de células das culturas induzidas de CG9571 *FL* em comparação com os controles (Figura 23). Estes resultados corroboram a hipótese de que o domínio da proteína tenha sido efetivamente expresso nos ensaios anteriores e que tenha produzido algum efeito tóxico letal às células induzidas.

4.4 CLONAGEM DE CG9571 *FL*, EXPRESSÃO EM BACTÉRIA E OBTENÇÃO DA PROTEÍNA RECOMBINANTE PURIFICADA E SOLÚVEL

4.4.1 Clonagens em vetor carreador e de expressão

Paralelamente aos ensaios das últimas tentativas de expressão do domínio ligador, iniciamos os procedimentos para clonagem da proteína completa de CG9571.

Primers sintetizados para a amplificação de CG9571 FL, também com sítios para as enzimas BamHI e NdeI incorporados, foram utilizados para amplificação do fragmento por *PCR* com sucesso (Figura 24).

Figura 24. Detecção de fragmento correspondente a *CG9571 FL* obtido por *PCR* para clonagem em vetor carreador

Fragmento com tamanho esperado amplificado (783pb). Raia 2: MM 1kb *Promega*. Eletroforese em gel de agarose 0.8% TAE 1X. Fonte: Gueller (2017).

Com um *kit* de clonagem TOPO novo e, em procedimento único para cada etapa, subclonamos CG9571 FL no vetor pTOPO e, em seguida, em vetor pET-15b, seguindo os procedimentos estabelecidos para o domínio e as recomendações do fabricante do novo *kit*. As clonagens foram verificadas por clivagem com enzimas de restrição (Figuras 25 e 26) de plasmídeos obtidos por preparações.

Figura 25. Clivagem de DNA plasmidial extraído de colônias geradas na clonagem de *CG9571 FL* em pTOPO com enzimas de restrição

Plasmídeos linearizados com BamHI (raia 2) e NdeI (raia 3). Clivagem com BamHI e NdeI libera inserto com tamanho específico (raias 4 e 5). Raia 6: MM 1kb *Promega*. Eletroforese em gel de agarose 1% TAE 1X. Fonte: Gueller (2017).

Figura 26. Clivagem de DNA plasmidial extraído de colônias geradas na clonagem de *CG9571 FL* em pET-15b com enzimas de restrição

(A) Plasmídeo clivado com BamHI e XbaI libera inserto de tamanho esperado para a reação (raia 2. Sequência apresenta aprox. 1Kb). Raia 1: plasmídeo não clivado. (B) Plasmídeos clivados com BamHI e NdeI. Insertos com tamanho específico (783pb) podem ser identificados nas raias 1, 5 e 2, este último com intensidade mais fraca. Raias 3 (A e B): MM 1kb *Promega*. Eletroforese em gel de agarose 1%, TAE 1X. Fonte: Gueller (2017).

Estes resultados, juntamente com resultados obtidos na clonagem do domínio, indicam que uma parte significativa dos problemas de clonagem estava relacionada à qualidade dos reagentes dos *kits* de clonagem utilizados. Desconfiamos, tal como mencionado na tentativa de expressão do polipeptídeo *in vitro*, que possam ter ocorrido problemas para a manutenção da temperatura ideal de conservação do reagente durante as etapas de estocagem e transporte destes materiais importados. Além disto, pelo menos uma das enzimas de restrição parece ter dificuldade em clivar fragmentos com sítios de restrição na extremidade (provavelmente a NdeI). As mesmas condições de ligação que resultaram em clonagem dos insertos clivados do vetor carreador não produziram resultado nas tentativas de clonagem em que os fragmentos amplificados foram clivados.

4.4.2 Ensaios iniciais de expressão em células competentes e tentativas de obtenção e purificação da proteína recombinante solúvel

Plasmídeos recombinantes foram utilizados para transformação de bactérias BL21(DE3). No experimento piloto de indução já foi possível verificar em SDS-PAGE banda de tamanho específico (aproximadamente 32 kDa) na amostra do extrato celular total das bactérias da cultura induzida em SDS-PAGE (dados não mostrados). Experimentos seguintes mostram que a proteína recombinante estava presente no pélete

(fração insolúvel) após lise bacteriana. Com frequência, a expressão de proteína recombinante em altos níveis em *E. coli* resulta na formação de agregados da proteína expressa que se depositam no citoplasma da célula e que são denominados corpos de inclusão (Singh *et al.*, 2015). A formação destes agregados pode ser evitada, em muitos casos, reduzindo-se a taxa de expressão da proteína recombinante, o que pode ser conseguido através da indução com menor concentração do indutor e em menor temperatura (Buchner e Rudolph, 1991; Baneyx e Mujacic, 2004). De acordo com estes e outros estudos, elaboramos protocolos de indução buscando expressar a proteína recombinante na forma solúvel. Para isso, selecionamos clones que apresentaram maior nível de expressão e utilizamos diferentes tempos de indução (30min, 1h, 2h, 4h, 12h, 16h, 19h e 24h), concentrações de IPTG (0.1, 0.2, 0.3, 0.5, 0.6, 0.8 e 1mM), temperaturas (23, 30 e 37°C), velocidades de agitação (150 a 200 rpm), densidades de bactérias da cultura para indução (DO₆₀₀ de 0.5 a 0.8) e volumes de cultura (5 a 250ml de meio LB). Conseguimos induzir maior expressão de CG9571 FL, mas não obtivemos sucesso na obtenção da forma solúvel da proteína (Figura 27).

Figura 27. Indução da expressão de CG9571 FL (1mM IPTG, 37°C)

Extrato celular total (ECT) e extrato das frações solúvel (Sol.) e insolúvel (Insol.) de células BL21(DE3)pLysS transformadas com pET-15b-CG9571 FL. Culturas não induzida (controle negativo – 0mM) crescida por 4h (T₂) e induzidas por 2 (T₁) ou 4h (T₂). Raia 6: MM BenchMark. Bandas de tamanho específico (acima do marcador de 30kDa) são observadas nas raias 3 e 5. SDS-PAGE, gel de resolução 14%, tris-glicina. Fonte: Gueller (2018).

Verificamos que mesmo em induções com concentrações baixas de IPTG (0.1 a 0.3mM) e por períodos de tempo curtos (30min, 1 e 2h), as primeiras bandas detectáveis da proteína recombinante apareciam na fração insolúvel (Figura 28).

Figura 28. Indução da expressão de CG9571 FL (0.1/0.2/0.3mM IPTG, 37°C)

Extrato celular total (ECT), extrato da fração solúvel (Sol.) e fração insolúvel (Insol.) de células BL21(DE3)pLysS transformadas com pET-15b-*CG9571 FL* do mesmo clone e induzidas com concentrações variáveis de IPTG. Cultura não induzida (controle negativo - 0mM) e cultura induzida por 30min (T₁), 1 (T₂) ou 2h (T₃). Raia 10: MM BenchMark. Retângulos vermelhos indicam prováveis bandas da proteína recombinante detectada em algumas frações insolúveis [raia 5 (A) e raia 3, 5, 7 e 9 (B)]. SDS-PAGE, gel de resolução 14%, tris-glicina. Fonte: Gueller (2018).

Apesar de não termos identificado expressão de CG9571 *FL* na forma solúvel, realizamos uma tentativa de purificação por IMAC da fração solúvel de células de um dos ensaios de indução (0.5 mM IPTG, 23°C, 24h, 250ml de cultura), tal como realizado para CG9571 *BD*, desta vez empregando cromatógrafo para o procedimento. As frações da purificação foram analisadas por SDS-PAGE, mas não identificamos nenhuma banda de tamanho específico nas frações eluídas (dados não mostrados).

4.4.3 Obtenção de proteína a partir de corpos de inclusão por solubilização em condições brandas

Corpos de inclusão são majoritariamente compostos pela proteína de interesse (Valax e Georgiou, 1993; Carrió *et al.*, 1998). Por esta razão, a expressão de proteína recombinante nesta forma pode permitir a obtenção de grande quantidade de proteína e com algum grau de pureza, além de apresentar outras vantagens, como uma maior proteção contra a ação de proteases endógenas (Clark, 1998; Baneyx e Mujacic, 2004; Jungbauer e Kaar, 2007). Ademais, verificamos na literatura que fatores de transcrição da família *forkhead* geralmente são obtidos por expressão em *E. coli* na forma insolúvel (Kaufmann *et al.*, 1994; Yu *et al.*, 1999; Zaret e Stevens, 1995). Por estes motivos, decidimos prosseguir para tentativas de obtenção de CG9571 *FL* a partir de corpos de inclusão.

A importante desvantagem da utilização de corpos de inclusão é a necessidade de promover o renovelamento *in vitro* da proteína, tarefa que pode se apresentar como um grande desafio. O método mais empregado para obtenção de proteínas com conformação nativa a partir destas estruturas consiste, em essência, na desnaturação da proteína recombinante através de altas concentrações de agentes caotrópicos, como ureia ou hidrocloreto de guanidina, e na renaturação através da remoção do agente desnaturante (Clark, 1998). Existem, contudo, métodos descritos na literatura para recuperação de proteínas a partir de processos que não promovem a completa desnaturação de estruturas secundárias presentes nestas proteínas. Estes métodos foram desenvolvidos a partir de novas informações sobre a estrutura e atividade de proteínas que compõe os corpos de inclusão.

Corpos de inclusão não são homogêneos e suas proteínas podem apresentar diferentes conformações, incluindo proteínas com estrutura nativa ou semelhante a nativa que podem apresentar atividade biológica (García-Fruitós *et al.*, 2005; Ventura e Villaverde, 2006). Corpos de inclusão que apresentam proteínas deste tipo são

denominados corpos não clássicos (Jevševar *et al.*, 2005). Proteínas parcialmente conformadas ou com estruturas semelhantes a nativa exibem alta solubilidade, podendo ser solubilizadas em condições não desnaturantes, através de solventes orgânicos ou detergentes suaves, ou em soluções com baixas concentrações de agentes caotrópicos, o que permite facilitar o processo de renaturação ou, no primeiro caso, a eliminar esta etapa (Jevševar *et al.*, 2005; Singh *et al.*, 2012; Upadhyay *et al.*, 2014).

De acordo com esses estudos, nas nossas abordagens iniciais com corpos de inclusão, exploramos a possibilidade de obter proteína recombinante executando protocolos de solubilização com agentes não desnaturantes ou em condição desnaturante branda. Como qualidade de proteínas recombinantes presentes nos corpos de inclusão depende das condições utilizadas para a indução (Singh *et al.*, 2015), sendo a formação de corpos de inclusão não clássicos favorecida em baixas temperaturas de indução (Jevševar *et al.*, 2005; Peternel *et al.*, 2008), para estes protocolos realizamos induções em temperaturas de 23°C, com 0.5mM de IPTG, por 19-24h.

Os procedimentos de ruptura celular foram realizados por sonicação. Esta técnica se mostrou mais eficiente do que a ruptura por congelamento/fervura empregada em alguns dos experimentos anteriores e por isso a sonicação foi mantida para a realização dos demais ensaios deste trabalho. Após lise celular, proteínas da fração insolúvel, coletada por centrifugação, foram solubilizadas com as soluções tampão 1 ou 2 (Tabela 4) sob agitação, entre 20 e 22°C, durante a noite. Alíquotas de cada amostra bruta (não centrifugada) foram removidas para posterior verificação em gel e o restante foi centrifugado (14000 rpm, 10min) para separação das frações. Alíquotas da fração solúvel foram verificadas em SDS-PAGE juntamente com a amostra bruta (Figura 29).

Figura 29. Solubilização de CG9571 FL em condições brandas

Corpos de inclusão solubilizados em tampão contendo Triton X-100 0.5% (Tr-X) ou ureia 2M (Ur). Alíquotas de amostras bruta e do sobrenadante (após centrifugação do solubilizado). Duplicatas de amostras sendo as pares (2, 4, 6 e 8), respectivamente o dobro do volume das ímpares (1, 3, 5 e 7). Raia 9: MM BenchMark. Retângulos vermelhos indicam diferença de quantidade de CG9571, de uma mesma preparação, apenas suspenso e efetivamente solubilizado em cada processo. SDS-PAGE, gel de resolução 14%, tris-glicina. Fonte: Gueller (2018).

O procedimento de solubilização de CG9571 *FL* com solução contendo 2M de ureia (Tabela 4, tampão 2) permitiu solubilizar maior quantidade da proteína em relação à solubilização com o detergente (Tabela 4, tampão 1) e foi escolhido para a solubilização da proteína recombinante de uma cultura induzida com maior volume (250 ml) nas mesmas condições descritas acima. A fração solubilizada em tampão modificado para preparo de extrato para cromatografia (Tabela 4, tampão 3) foi empregada em procedimento de purificação e renaturação por IMAC (4.4.5).

4.4.4 Obtenção de proteína a partir de corpos de inclusão por solubilização em condições desnaturantes severas

Na segunda abordagem, em condições desnaturantes mais severas (maior concentração do agente caotrópico), utilizamos inicialmente soluções contendo 4 ou 8 M de ureia. Para solubilização em ureia 4M, induzimos uma cultura de mesmo volume e nas mesmas condições descritas para a solubilização em condição desnaturante branda (item 4.4.3) A lise celular foi realizada em tampão Tris-HCl (Tabela 3, tampão 2), a fração insolúvel coletada por centrifugação e solubilizada por agitação vigorosa em vórtex até desagregação do precipitado formado. A solução foi incubada em gelo por aproximadamente 30min antes de nova centrifugação para separação das frações

solubilizadas e não solubilizadas. A preparação obtida foi utilizada para procedimento de purificação e renaturação por IMAC (item 4.4.5).

Para solubilização em ureia 8 M preparamos duas culturas de 250ml que foram induzidas com 0.6mM de IPTG, por 14h a 30°C. Para a primeira destas culturas, seguimos os mesmos procedimentos acima descritos para a obtenção dos corpos de inclusão solubilizados e o extrato foi submetido à purificação por IMAC (item 4.4.5). As células coletadas da segunda cultura induzida foram lisadas no tampão 3 (Tabela 3) e a fração insolúvel coletada foi submetida a duas etapas de lavagem com tampão contendo ureia 2 M (Tabela 4, tampão 2) para remoção do máximo de proteínas contaminantes, separando a cada lavagem a fração solubilizada e não solubilizada por centrifugação. O precipitado obtido após as lavagens foi solubilizado em tampão fosfato contendo ureia 8 M e recuperado do mesmo modo descrito acima. Em análise por SDS-PAGE identificamos CG9571 *FL* nas três etapas de solubilização (Figura 30).

Figura 30. Isolamento de corpos de inclusão e solubilização de CG9571 FL em condições desnaturantes

Frações celulares solúvel (Sol.) e insolúvel (Insol.) após lise celular. Da fração insolúvel foram realizadas centrifugações e solubilizações sequenciais em tampão contendo ureia 2M por duas vezes (1x e 2x) e em tampão contendo ureia 8M. Retângulo vermelho indica as bandas correspondentes a CG9571 *FL* nas três solubilizações. Raia 5: alíquota de CG9571 *FL* I (item 4.4.7) utilizada como parâmetro para a proteína recombinante. SDS-PAGE, gel de resolução 14%, tris-glicina. Fonte: Gueller (2018).

A fração solubilizada em ureia 2 M na segunda lavagem foi submetida a processos de renaturação por diluição e diálise (item 4.4.7) e a fração solubilizada em ureia 8M à purificação a partir de SDS-PAGE (item 4.4.6).

4.4.5 Tentativas de purificação e renaturação de proteína recombinante por IMAC

Na primeira tentativa do uso da técnica para purificação e renaturação de proteínas solubilizadas de corpos de inclusão, realizamos um procedimento manual com amostra solubilizada em tampão contendo 2 M de ureia (item 4.4.3). Tentamos manter um fluxo entre 0.5 e 1ml/min em todas as aplicações de soluções na coluna. Na etapa de renaturação, lavamos a coluna com tampão contendo concentrações decrescentes de ureia a uma redução de 0.1 M do reagente a cada volume de coluna lavado. Para a etapa de eluição, utilizamos 3 volumes de coluna de tampão contendo 50, 250 e 500mM de imidazol. Na análise por SDS-PAGE, contudo, verificamos que a proteína recombinante não ficou retida na coluna, sendo encontrada na fração injetada (dados não mostrados).

Embora os procedimentos de solubilização em condição branda (2 M de ureia) tenham permitido recuperar uma parcela da proteína recombinante expressa, a purificação da proteína por IMAC não foi possível. Decidimos, então, aumentar a concentração de ureia na solução de solubilização, considerando que a sequência de histidinas da proteína recombinante pudesse não estar exposta nas proteínas recuperadas dos corpos de inclusão através deste processo de solubilização. Extratos de proteínas solubilizadas em tampão contendo 4M de ureia (item 4.4.4) foram utilizados para nova tentativa de purificação e renovelamento por IMAC em cromatógrafo (protocolo descrito em 3.13.4). O perfil cromatográfico produzido no processo (Figura 31) foi utilizado para a seleção de amostras do fracionamento por cromatografia. As frações correspondentes aos picos observados no cromatograma nas etapas de injeção (3 a 5), renaturação (13 e 14 e 19 a 21) e eluição (22 e 23) foram concentradas em tubos de ultrafiltração e analisadas por SDS-PAGE (Figura 32).

Figura 31. Cromatograma da purificação de extrato solubilizado em ureia 4M

Picos detectados no início (frações 13 e 14) e no final (19 a 21) da renaturação e pequena ascensão da curva durante a eluição (22 e 23). Curva em azul representa medida de absorbância (em 280 nm) das frações de cromatografia. Em verde, volume aplicado de tampão contendo 0M de ureia para a renaturação. Gradiente formado pela mistura de 0-100% deste tampão com 100-0% do tampão contendo ureia. Em roxo, volume de tampão contendo 0mM de imidazol para a eluição. Gradiente formado pela mistura de 100-0% deste tampão com 0-100% de tampão contendo 500mM de imidazol. Eixo x: em preto, volume em ml de soluções aplicadas; em vermelho, número do tubo de coleta das frações da cromatografia. Fonte: Gueller (2018).

Figura 32. Frações de purificação e renaturação por IMAC de corpos de inclusão solubilizados em ureia 4M

Alíquotas das amostras concentradas correspondentes às frações injetada (Inj.), de renovamento em coluna (Ren.) e eluição (Elu.). Retângulo vermelho indica região do gel correspondente ao tamanho de CG9571 *FL* recombinante. A proteína só é detectada nas frações injetada e de lavagem (raias 2 e 3). Raia 1: MM BenchMark. SDS-PAGE, gel de resolução 14%, tris-glicina. Fonte: Gueller (2018).

A proteína recombinante foi detectada nas frações injetada, novamente, e de lavagem, indicando que a proteína teve fraca adsorção na coluna (Figura 32). Por ter apresentado a proteína de interesse em grande quantidade, ainda que em baixo grau de pureza, decidimos submeter a fração injetada concentrada a processos de renaturação para tentativa de obtenção de CG9571 *FL* solúvel (item 4.4.7).

Uma tentativa de purificação por IMAC também foi realizada com o extrato de proteínas solubilizadas em tampão contendo 8M de ureia (item 4.4.4). Consideramos que mesmo em 4 M de ureia a cauda de histidina poderia não ter ficado exposta – o que explicaria a não adsorção da proteína aos íons metálicos quelatados na coluna – e que o uso do agente em maior concentração poderia romper interações que porventura estivessem impedindo a exposição deste marcador. Com o objetivo de verificar a adsorção da proteína recombinante, empregamos a técnica de IMAC apenas para a purificação do extrato na tentativa de se obter uma fração pura que pudesse ser submetida a processos de renaturação após a purificação. A análise por SDS-PAGE de frações obtidas no ensaio revelou, entretanto, os mesmos resultados obtidos nos ensaios anteriores, ou seja, a não adsorção em coluna (Figura 33).

Figura 33. Frações de purificação por IMAC de corpos de inclusão solubilizados em ureia 8M

Alíquotas de amostras correspondentes a frações (A) de lavagem (Lav.) e eluição (Elu.) e (B) injetada (Inj.). A amostra da etapa de lavagem corresponde a primeira fração de lavagem enquanto a amostra de eluição corresponde a todas as frações coletadas nesta etapa e concentradas por centrifugação em coluna de ultrafiltração. Retângulo vermelho (A) indica CG9571 *FL* recombinante. Raias (A) 3 e (B) 2: alíquota de CG9571 *FL* I utilizada como parâmetro. SDS-PAGE, gel de resolução 14%, tris-glicina. Fonte: Gueller (2018).

Foi possível detectar uma banda de tamanho específico com intensidade muito baixa na fração de eluição (Figura 33A, raia 2). No entanto, acreditamos que isto se deva a contaminação das frações coletadas. Pelo fato deste procedimento de purificação ter sido realizado manualmente, uma pequena parcela da última fração de lavagem foi misturada com a primeira fração de eluição. De qualquer modo, além da baixa concentração de CG9571 *FL* verificada, a amostra também apresentou baixo grau de pureza e o eluato foi descartado.

4.4.6 Purificação de proteína a partir de SDS-PAGE

As proteínas solubilizadas em tampão contendo 8M de ureia (item 4.4.4) foram separadas por SDS-PAGE e eluídas da banda excisada. Alíquotas da eluição foram analisadas por SDS-PAGE e a eluição da proteína de interesse foi confirmada (Figura 34).

Figura 34. Purificação de CG9571 FL a partir de SDS-PAGE

Alíquota da primeira eluição de CG9571 *FL* purificada (Pur.) a partir de gel de poliacrilamida desnaturante. Raia 2: alíquota de amostra bruta (Bru.) aplicada no gel para purificação (fração injetada solubilizada em ureia 8M, item 4.4.5). Raia 3: alíquota da preparação CG9571 *FL* I (item 4.4.7) utilizada como parâmetro. SDS-PAGE, gel de resolução 14%, tris-glicina. Fonte: Gueller (2018).

Em seguida, após a precipitação por acetona/metanol e solubilização em diferentes tampões (item 3.13.5), as proteínas purificadas e solubilizadas foram submetidas a processos de renaturação por diluição ou diálise (item 4.4.7), tendo sido analisadas por espectrofotometria e SDS-PAGE somente após estes procedimentos.

4.4.7 Renaturação de proteína por diluição e diálise

Duas amostras de 1 ml cada da fração injetada da preparação em 4 M de ureia (item 4.4.5, Figura 32, raia 2) foram diluídas 6 vezes para tentativa de renovelamento de CG9571 *FL*. A diluição foi realizada gradualmente, através da aplicação de alíquotas de 50 µl do extrato solubilizado em ureia 4 M por vez no tampão de renaturação durante aproximadamente 3h, em temperatura de 18°C. Em seguida, dialisamos as amostras contra tampão fosfato de sódio pH 7.2 e concentramos por centrifugação em tubo de ultrafiltração até o volume original da amostra antes da diluição. Um precipitado visível se formou nas colunas após estes procedimentos, mas a análise por SDS-PAGE de alíquotas das duas preparações de CG9571 *FL* (CG9571 I e CG9571 II) indicou a presença de grande quantidade da proteína de interesse na forma solúvel (Figura 35).

Figura 35. Proteínas obtidas de fração solubilizada em ureia 4M submetida à renaturação por diluição

Alíquota do sobrenadante das preparações de CG9571 *FL*. Após diluição, diálise e concentração, todo o volume das preparações CG9571 I e CG9571 II foi transferido para microtubo e centrifugado a 14000 rpm para separação das proteínas na forma solúvel das proteínas não renoveladas ou agregadas durante os procedimentos. Raia 3: MM BenchMark. SDS-PAGE, gel de resolução 14%, tris-glicina. Fonte: Gueller (2018).

A concentração total de proteínas das preparações CG9571 I e CG9571 II foi calculada em 10.2 mg/ml e 2.42 mg/ml, respectivamente. Apesar da alta concentração, não foi possível determinar a concentração de CG9571 *FL* nas amostras, pois proteínas contaminantes do hospedeiro também foram recuperadas em forma solúvel após os procedimentos de renaturação, principalmente proteínas de alta massa molecular (Figura

35). Apesar de não termos obtido amostras puras, decidimos utilizar estas preparações para ensaios de interação com os *CRMs* de *eve* (item 4.5).

A fração da segunda lavagem de corpos de inclusão com tampão contendo ureia 2 M (item 4.4.4, Figura 30, raia 3) foi submetida a tentativas de renaturação por diluição e diálise. A primeira metade da fração foi renaturada por diluição, dialisada e concentrada conforme descrito acima, porém em tampão TrisCl 40 mM pH 8. Após a centrifugação, contudo, verificamos, por espectrofotometria e pela formação de agregado no microtubo, que a maior parte da proteína precipitou durante os procedimentos (dados não mostrados) e o sobrenadante da preparação foi desprezado. Para o processo de renaturação do restante do extrato da lavagem, utilizamos procedimento de diálise em membrana. A amostra foi dialisada contra tampões contendo 1.5, 1, 0.5 e 0M (2x) de ureia em agitação e temperatura entre 16 e 17°C. A maior parte das proteínas desta preparação (CG9571 III) também precipitou após centrifugação. A concentração foi calculada em aproximadamente 0.05 mg/ml e não foi possível obter uma detecção clara das proteínas em análise por SDS-PAGE (dados não mostrados). Apesar disto, CG9571 III foi utilizado em reações de interação com DNA imediatamente após o término da preparação (item 4.5).

Por haver grande quantidade de material nos péletes formados nos procedimentos de renaturação da fração solubilizada em ureia 2 M, decidimos realizar tentativas de recuperar as proteínas destas duas amostras. Solubilizamos os precipitados em tampão fosfato contendo 4 M de ureia (Tabela 7, tampão 2) e empregamos um procedimento de diálise mais lenta e gradual para renaturação de ambos. As amostras foram dialisadas contra tampões contendo 3, 2, 1.5, 1.25, 1, 0.7, 0.5, 0.3, 0.15, 0.07 e 0M (2x) de ureia. Após concentração das amostras, a análise por SDS-PAGE revelou que a primeira preparação (pélete de renaturação por diluição), CG9571 IV, era composta principalmente por proteínas de alta massa molecular, embora banda de tamanho específico muito fraca tenha sido detectada (Figura 36A), enquanto a segunda preparação (pélete de renaturação por diálise), CG9571 V, embora pouco concentrada (0.06 mg/ml), apresentou maior quantidade da proteína de interesse em relação a proteínas contaminantes (Figura 36B). Esta preparação também foi empregada logo em seguida em ensaios de interação com DNA (item 4.5).

Figura 36. Proteínas obtidas de fração ressolubilizada em ureia 4M submetida à renaturação por diálise

(A) Alíquota do sobrenadante da preparação CG9571 IV. O retângulo vermelho indica presença de bandas muito fracas correspondentes a CG9571. Foram aplicados dois volumes da mesma preparação no gel. Proteínas de alta massa molecular são vistas próximas aos números das raia. (B) Alíquota da amostra bruta (Bru., não centrifugada) e do sobrenadante (Sob.) da preparação CG9571 V. Retângulo vermelho indica bandas de tamanho específico. Embora maior parte de CG9571 tenha precipitado (comparação entre bandas das raia 1 e 2), a preparação apresenta menor quantidade de proteínas contaminantes. Raia 3 (A e B): alíquota de CG9571 *FL* I utilizada como parâmetro. SDS-PAGE, gel de resolução 14% (A) e 12.8% (B), tris-glicina. Fonte: Gueller (2018).

Três amostras de CG9571 *FL* purificadas a partir de SDS-PAGE (4.4.6) foram solubilizadas, após precipitação, em tampão fosfato contendo ureia (Tabela 7). Na primeira preparação (proteína eluída a partir de um gel preparativo) a amostra foi submetida a processo de renovelamento por diálise contra tampões contendo 3.5, 3, 2.7, 2.4, 2, 1.7, 1.4, 1, 0.7, 0.4, 0.2 e 0M (2x) de ureia. A preparação (CG9571 VI) foi centrifugada em coluna de ultrafiltração e sua concentração calculada em 0.29 mg/ml. Na segunda preparação (CG9571 VII), uma amostra de CG9571 *FL* purificada a partir de quatro géis preparativos foi solubilizada e renaturada de igual forma e sua concentração foi calculada em 0.10 mg/ml. A terceira amostra foi solubilizada em tampão fosfato contendo 8M de ureia (Tabela 5) e diluída 15 vezes para renaturação. Na diálise reduzimos gradualmente a concentração de NaCl até a concentração final de 100 mM. Após a ultrafiltração, a concentração de CG9571 VIII foi calculada em 0.32 mg/ml.

Outras quatro amostras de CG9571 *FL* eluídas de gel foram solubilizadas em tampão HEPES (Tabela 7). Nas três primeiras preparações, o precipitado foi solubilizado em tampão HEPES contendo 6M de ureia e as amostras foram diluídas 15 vezes no mesmo de solubilização sem ureia (Tabela 7, tampão 1). Duas destas preparações foram dialisadas contra este mesmo tampão e uma delas contra tampão que também não continha glicerol e DTT. As preparações foram concentradas em tubos de ultrafiltração e foram obtidas concentrações de 0.09, 0.11 e 0.16 mg/ml. Na quarta preparação, a amostra foi solubilizada no mesmo tampão, mas para renaturação foi utilizado procedimento de diálise contra tampão contendo os componentes do tampão 1 (Tabela 7) com exceção da ureia, glicerol e DTT. Ao final de todos os processos, obtivemos solução com concentração de 0.14 mg/ml. Todas estas preparações foram utilizadas imediatamente em ensaios de interação (item 4.5). A Tabela 9 apresenta um resumo de informações sobre as principais preparações de CG9571 *FL* mencionadas neste trabalho.

Tabela 9. Principais preparações da proteína recombinante CG9571 *FL* utilizadas em ensaios de interação DNA-proteína (continua)

Num.	Origem	Purificação	Renaturação	Tampão final
CG9571 I	Fração injetada de cromatografia	–	Diluição	Fosfato de sódio 50mM
CG9571 II	Fração injetada de cromatografia	–	Diluição	Fosfato de sódio 50mM
CG9571 III	Fração de lavagem de CI	Lavagem de CI com tampão contendo ureia 2M (1X)	Diálise	TrisCl 40mM
CG9571 IV	Precipitado formado durante renaturação	–	Diálise	Fosfato de sódio 50mM
CG9571 V	Precipitado formado durante renaturação	–	Diálise	Fosfato de sódio 50mM

Tabela 9. Principais preparações da proteína recombinante CG9571 *FL* utilizadas em ensaios de interação DNA-proteína (conclusão)

CG9571 VI	Fração insolúvel após lavagens (2X) de CI	Lavagem de CI (2X) e SDS-PAGE	Diálise	Fosfato de sódio 50mM
CG9571 VII	Fração insolúvel após lavagens (2X) de CI	Lavagem de CI (2X) e SDS-PAGE	Diálise	Fosfato de sódio 50mM
CG9571 VIII	Fração insolúvel após lavagens (2X) de CI	Lavagem de CI (2X) e SDS-PAGE	Diluição	Fosfato de sódio 50mM NaCl 100mM

Tampão fosfato de sódio em pH 7.2 e TrisCl em pH 8.

Proteínas das amostras de CG9571 I, II, IV e V solubilizadas em tampão fosfato de sódio 50mM pH 7.2, NaCl 0.5M, ureia 4M; CG9571 VI, VII e VIII solubilizadas inicialmente em tampão fosfato de sódio 50mM pH 7.2, NaCl 0.5M, ureia 8M e, após precipitação por acetona metanol, VIII foi solubilizada no mesmo tampão e VI e VII no tampão contendo ureia 4M; CG9571 III solubilizadas em tampão Tris-HCl 40mM pH 8, ureia 2 M.

CI=corpos de inclusão.

4.5 ENSAIOS DE LIGAÇÃO E DETECÇÃO DE INTERAÇÃO ENTRE CG9571 *FL* E CRMS

Os primeiros ensaios de interação DNA-proteína foram conduzidos com as preparações CG9571 I e II e fragmentos de DNA relativos aos *CRMs eve 1* e *eve 2*. No experimento piloto verificamos que o DNA livre de cada fragmento migrou conforme o esperado, porém não detectamos migração dos fragmentos nas raias em que foram aplicadas as reações de interação (Figura 37).

Figura 37. Interação entre *CRMs eve 1* e *eve 2* mt com CG9571 I

Detecção por brometo de etídio de fragmentos de DNA livre e em reações de interação com proteínas (5µl de CG9571 I) no tampão de reação 3. *eve 1* e *eve 2* mt livres (raias 2, 4, 6 e 7) migraram normalmente durante a eletroforese, mas os fragmentos presentes nas reações de interação não foram detectados ao longo do gel (raias 1, 3 e 5). Setas vermelhas indicam DNA que parece estar retido na base dos poços de aplicação. Seta preta indica fragmento de DNA correspondente ao vetor pTOPO que não foi removido na purificação (*eve 1* foi purificado de gel para novas reações). Raia 9: MM 1kb *Promega*. PAGE, gel 5%, TBE 1X. Fonte: Gueller (2018).

Os fragmentos de *eve 1* e *2* mt presentes nestas reações parecem não ter entrado no gel, pois detectamos DNA na base dos poços em que as reações de interação foram aplicadas (Figura 37, raias 1, 3 e 5). O mesmo fenômeno foi observado no poço de aplicação do padrão de massa molecular, onde bandas de maior massa ficaram retidas.

Estes resultados nos conduziram à hipótese de que CG9571 teria se ligado aos *CRMs*, mas o complexo resultante não conseguiu entrar no gel por algum motivo. Uma possibilidade seria o excesso de proteína empregado nas reações de interação. Em gel de poliacrilamida nativo havíamos estimado a quantidade ideal de DNA a ser utilizada nas reações de interação, mas não quantificamos as amostras de proteína pelo fato das preparações se mostrarem impuras, conforme mencionado anteriormente.

Para os experimentos seguintes, consideramos os valores de concentração obtidos para o total de proteínas nas preparações como a concentração de CG9571 *FL* e ajustamos a quantidade de proteína aplicada nos ensaios de acordo com a quantidade empregada para estes ensaios na literatura (Unezaki *et al.*, 2004; Calzavara-Silva *et al.*, 2004; Cirillo e Zaret, 2007; Andrioli *et al.*, 2004; Fujioka *et al.*, 1996). Além disso, realizamos a separação das amostras por eletroforese durante maior tempo,

considerando que o tempo utilizado no experimento piloto tenha sido insuficiente para permitir que o complexo migrasse no gel.

Em ensaios realizados com baixa quantidade de proteína (50 e 500 ng) das preparações CG9571 I e II, os fragmentos de DNA nas reações de interação migraram a mesma distância que o DNA livre e não foi identificado nenhum efeito adicional (dados não mostrados), mas a análise das interações em que a massa de proteína empregada esteve entre 1 e 5 µg mostrou que a quantidade detectada de DNA livre de *eve 1* diminui conforme a concentração de proteína aumenta, correlação esta que ficou mais evidente para as reações com CG9571 II (Figura 38). No entanto, a partir de então, passamos a não detectar mais retenção de DNA acumulado nos poços com a mesma intensidade que a observada no ensaio anterior (Figura 37).

Figura 38. Interação entre *eve 1* e CG9571 I e II (I)

Deteção por brometo de etídio de fragmentos de DNA livre e em reações de interação com proteínas (1.5, 2, 2.5 e 3µg) no tampão de reação 3. Massa de proteína descrita se refere a massa estimada do total de proteínas de cada preparação. Redução de *eve 1* detectado no gel é verificada com o aumento da quantidade de proteína aplicada na reação de incubação, especialmente no caso de CG9571 II (raias 6 a 9). Raia 10: MM 1kb *Promega*. PAGE, gel 5%, TBE 1X. Fonte: Gueller (2018).

Detectamos resultados semelhantes com a aplicação de amostras das interações em gel de agarose (Figura 39). O gel de agarose possui poros com maior diâmetro em relação aos poros do gel de poliacrilamida e também pode ser empregado em EMSA,

sobretudo em investigações de complexos de maior massa molecular (Ream *et al.*, 2016).

Figura 39. Interação entre eve 1 e CG9571 I e II (II)

Deteção por brometo de etídio de fragmentos de DNA livre e em reações de interação com proteínas (1.5, 2, 2.5 e 3μg) no tampão de reação 3. Diminuição da quantidade de *eve 1* detectada no gel é verificada com o aumento da quantidade de proteína aplicada na reação de incubação, especialmente no caso de CG9571 II (raias 4 a 6). Raia 10: MM 1kb *Promega*. Eletroforese em gel de agarose 1.5%, TBE 0.5X. Fonte: Gueller (2018).

Para explicar o motivo de detectarmos menor quantidade de DNA nas interações contendo maior quantidade de proteína, passamos a considerar um possível problema com a carga do complexo eventualmente formado. CG9571 é uma proteína básica, com pI teórico calculado em 9.69 para a proteína recombinante, o que significa que a proteína apresenta carga líquida positiva em solução que possua pH abaixo deste valor. No sistema que utilizamos para eletroforese em gel nativo, com tampão TBE 1X de pH 8.4 (item 3.14.2), a proteína livre migraria, portanto, na direção do catodo, enquanto o DNA, que apresenta carga negativa, migraria para o anodo. Quando ocorre interação entre proteínas básicas e DNA, entretanto, a carga do complexo formado pode ser negativa, o que permite que o complexo migre na mesma direção que o DNA livre durante a eletroforese. Verificamos que este foi o resultado obtido em alguns trabalhos que utilizaram esta técnica para a investigação da interação de proteínas básicas (proteínas completas ou domínios de ligação, também básicos) com sequências alvo e que empregaram condições em pH abaixo do pI do polipeptídeo para a interação e

eletroforese (tampão para reação, tampão de corrida e gel) (Kaufmann *et al.*, 1994; Häcker *et al.*, 1995; Cirillo *et al.*, 1998).

Por outro lado, também verificamos casos em que pesquisadores relataram fenômenos semelhantes ao que observamos nestes ensaios (retenção de DNA nos poços de aplicação ou redução de DNA detectado no gel) e uma das hipóteses frequentemente levantada durante as discussões destes resultados era a que associava estes efeitos a carga líquida do complexo resultante da ligação de proteínas básicas ao DNA, que poderia apresentar carga líquida próxima a zero ou até mesmo carga líquida positiva (ResearchGate, 2014a, 2014b e 2016). Uma das sugestões fornecidas pelos pesquisadores que discutiam o problema foi o uso da técnica de eletroforese BN-PAGE (item 3.14.2). A técnica foi originalmente desenvolvida para separação de proteínas básicas de membrana (Schägger e von Jagow, 1991) e é usualmente empregada em ensaios de interação proteína-proteína (Krause, 2006; Dresler *et al.*, 2011). Embora não tenhamos encontrado na literatura científica relatos do emprego desta técnica em ensaios de interação de DNA-proteína, o seu uso com sucesso para esta finalidade foi reportado por pesquisadores em pelo menos um dos tópicos do fórum (ResearchGate, 2014a) e decidimos utilizá-la em nossos experimentos.

As reações, conduzidas nas mesmas condições dos ensaios anteriores, com adição do corante (3.14.2), foram separadas por eletroforese em géis de poliacrilamida e também de agarose, mas não verificamos alterações dos resultados (dados não mostrados). Em paralelo, reações de interação sem corante foram aplicadas em gel de agarose preparado com os poços no meio do gel. No entanto, nenhuma banda foi detectada em direção ao polo negativo (dados não mostrados).

No conjunto, esses ensaios não nos permitiram tirar nenhuma conclusão sobre os fenômenos observados. Decidimos, então, conduzir novas reações com controles negativos: sequências de DNA às quais o fator de transcrição não deveria se ligar. Selecionamos as sequências da região codificadora do próprio gene *CG9571* (da proteína completa e domínio de ligação) e de *tailless*, outro gene *gap* que também temos investigado no laboratório. Verificamos que o mesmo efeito obtido para as sequências alvo foi produzido para as sequências não específicas (Figura 40).

Figura 40. Interação entre fragmentos de DNA não específicos e CG9571 II

Detecção de fragmentos de DNA livre e em reações de interação com proteínas (0.5, 1, 2 e 3μg) no tampão de reação 3. Redução da quantidade de *tll* e *CG9571 BD* detectada no gel é verificada com o aumento da quantidade de proteína aplicada na reação de incubação. PAGE, gel 5%, TBE 1X. Fonte: Gueller (2018).

Após este resultado, reavaliamos as imagens obtidas nos numerosos ensaios conduzidos anteriormente e notamos em algumas delas a presença de um arrasto em raias das reações de interação, o que poderia indicar a ocorrência de degradação do DNA por uma nuclease contaminante das preparações. Decidimos, assim, realizar novos ensaios de interação adicionando EDTA, um agente quelante, e removendo o cloreto de magnésio do tampão de ligação tentando interferir na atividade de uma possível enzima nuclease. O resultado do gel mostrou que todo o efeito de redução do DNA livre detectado foi neutralizado, tanto para os *CRMs* alvo quanto para os fragmentos não específicos (Figura 41A).

Figura 41. Interação entre *eve 1* e *eve 2* wt e CG9571 na presença de EDTA e/ou ausência de cloreto de magnésio

Deteção de fragmentos de DNA livre e em reações de interação com 2µg de proteínas no tampão de reação 3. Efeito é inibido em todos os fragmentos na presença de EDTA e ausência de cloreto de magnésio (A, raias 1, 4 e 7) e apenas na ausência de cloreto de magnésio (B, raia 3). PAGE, gel 5%, TBE 1X. Fonte: Gueller (2018).

Em outro ensaio verificamos que apenas a ausência de cloreto de magnésio já era suficiente para inibir a perda de DNA das amostras (Figura 41B). Interpretamos os resultados como uma confirmação da hipótese de atividade de degradação do DNA por nuclease. Por outro lado, quanto a interação de CG9571 com os *CRMs* alvo, a inibição desta degradação de DNA pela adição de EDTA ou pela remoção de cloreto de magnésio não revelou nenhum efeito de retardo na migração do DNA nas raias das reações de interação (Figura 41). Embora a ligação de proteínas ao DNA possa ocorrer na ausência de cátions divalentes como Mg^{2+} (Lane *et al.*, 1992), não consideramos este resultado conclusivo, abandonamos as preparações CG9571 I e II e prosseguimos para novas tentativas de preparação de proteína recombinante.

Novas interações foram realizadas com as preparações CG9571 III, IV e V (item 4.4.7). Nestes experimentos foram incluídas reações com fragmentos não específicos para controle. O resultado para CG9571 III indicou que a preparação era composta principalmente por nuclease bacteriana, pois, mesmo com baixa concentração de proteína, detectamos quantidade de DNA muito baixa para todos os fragmentos testados (dados não mostrados). Já as reações com CG9571 V não apresentaram este mesmo efeito. Aparentemente, a maior parte da proteína precipitada durante os procedimentos de renaturação e concentração da preparação CG9571 III foi a própria proteína

recombinante, enquanto a DNase permaneceu em forma solúvel e ativa. No entanto, os ensaios de CG9571 IV e V com as sequências alvo também não produziram nenhum retardo na migração eletroforética (dados não mostrados).

Decidimos não insistir no uso destas preparações e planejamos a obtenção de CG9571 *FL* através de purificação por SDS-PAGE (item 4.4.6). A primeira preparação de proteína purificada por este método foi utilizada em ensaios de interação logo após o término dos procedimentos de renaturação e concentração e obtivemos resultado positivo para a interação entre CG9571 VI e o fragmento de *eve 2 wt* no primeiro experimento (Figura 42).

Figura 42. Interação entre CRMs e DNA não específico e CG9571 VI.

Deteção de fragmentos de DNA livre e em reações de interação com quantidade variada de CG9571 VI (0.5, 1 e 2 µg) no tampão de reação 3. Segunda banda de DNA foi detectada para a reação com *eve 2 wt* (raia 4, seta vermelha) sugerindo formação de complexo DNA-proteína. PAGE, gel 5%, TBE 1X. Fonte: Gueller (2019).

Imediatamente após este resultado, conduzimos novos experimentos apenas com as sequências alvo, incluindo reações de *eve 2 wt* com quantidade variada da proteína, porém nenhum retardo da mobilidade eletroforética de *eve 2 wt* foi detectado no gel (Figura 43).

Figura 43. Interação entre *CRMs* e CG9571 VI

Deteção de fragmentos de DNA livre e em reações de interação com quantidade variada de CG9571 VI (0.5, 1 e 2 μ g) no tampão de reação 3. Não foi detectado nenhum resultado que indicasse ligação entre a proteína recombinante e *eve 2*. PAGE, gel 5%, TBE 1X. Fonte: Gueller (2019).

Por termos empregado exatamente as mesmas condições de interação e separação por eletroforese, desconfiamos que pudesse ter ocorrido algum problema com a amostra da proteína. A preparação foi centrifugada a 14000 rpm e a determinação de proteína no sobrenadante de CG9571 VI foi calculada em 0.05 mg/ml. A diminuição drástica da concentração de proteína solúvel pode indicar que a maior parte da proteína renovada foi obtida numa conformação ativa, porém instável, o que teria ocasionado a precipitação após poucas horas do término da preparação que esteve armazenada em 4°C.

Pelo fato de ainda haver proteína solúvel na preparação, concentramos CG9571 VI novamente em tubo de ultrafiltração. A preparação, com concentração de 0.17 mg/ml, foi empregada em novos ensaios, nas mesmas condições dos ensaios anteriores ou com pequenas modificações, utilizando outros tampões para incubação (Tabela 8, tampões 1, 2 e 4), aumentando a concentração de cloreto de magnésio nos tampões (de 4 ou 5 para 10mM), o tempo (de 20 para 30min) e a temperatura (de 20-22 para 25°C) de incubação e diminuindo a voltagem da separação por eletroforese (de 200 para 150V), mas não conseguimos reproduzir o resultado anterior.

Concomitantemente a estes últimos experimentos, preparamos uma nova amostra da proteína (CG9571 VII, item 4.4.7). A interação das sequências alvo com CG9571 VII produziu o mesmo fenômeno de redução do DNA detectado com sinais de

degradação dos fragmentos (dados não mostrados). Numa reação com fragmento não específico, o mesmo efeito foi observado (Figura 44).

Figura 44. Interação entre DNA não específico e CG9571 VII

Detecção de fragmentos de *tll* livre e em reações de interação com CG9571 VII no tampão de reação 3. Diminuição de DNA livre detectado nas raia em que foram aplicadas as reações de interação (2, 3 e 4) com arrasto indicando degradação dos fragmentos. PAGE, gel 7%, TBE 1X. Fonte: Gueller (2019).

O resultado sugeriu, novamente, a presença de nuclease contaminante na preparação. Ademais, sugeriu que pelo menos uma nuclease do hospedeiro utilizado para expressão possui tamanho próximo ao do fator de transcrição estudado. A eletroforese para separação de CG9571 *FL* do restante das proteínas do extrato para esta preparação foi conduzida por menor tempo em relação a primeira preparação, o que deve ter permitido que a enzima fosse extraída do gel juntamente com a proteína recombinante. Para as preparações seguintes, utilizamos o mesmo tempo de eletroforese empregado na obtenção de CG9571 VI.

Considerando o provável problema de estabilidade da proteína obtida em CG9571 VI, decidimos fazer preparações com novo tampão. Já havíamos observado que as amostras com tampão Tris-HCl 40 mM, sem nenhum outro aditivo, apresentaram maior precipitação de proteína durante os procedimentos de renovelamento e ultrafiltração. Optamos por utilizar o tampão HEPES, o mesmo empregado em dois tampões de ligação, com os aditivos DTT e glicerol, que podem prevenir a agregação durante o procedimento de renovelamento *in vitro* e conferir maior estabilidade a proteína renovada (Bondos e Bicknell, 2003; Vagenende *et al.*, 2009). As proteínas foram obtidas em menor concentração (item 4.4.7) e nenhum resultado positivo de interação foi detectado (dados não mostrados).

Em nossa última tentativa de obtenção de CG9571 *FL* puro e em forma solúvel e ativa, escolhemos novamente o tampão fosfato para os procedimentos de renaturação e para armazenamento da proteína pelo fato de termos obtido melhor rendimento de proteína solúvel e produção de proteína com capacidade de se ligar em *eve 2* wt na preparação com este tampão. Durante a diálise, reduzimos a concentração de cloreto de sódio para 100 mM e mantivemos esta concentração no tampão final da amostra. Embora tenhamos conseguido, efetivamente, uma preparação (CG9571 VIII) com concentração de proteína maior do que as anteriores com tampão HEPES, não detectamos, novamente, retardo da migração de *eve 2* wt ou dos demais *CRMs* em nenhum dos experimentos de interação com CG9571 VIII (dados não mostrados).

Por fim, realizamos dois últimos ensaios de ligação com as preparações CG9571 VI e CG9571 VIII e os *CRMs* alvo. No primeiro deles, empregamos novamente a técnica de eletroforese BN-PAGE para separação das amostras e no segundo utilizamos tampão Tris-Glicina com pH 8.9, mais alcalino do que o tampão TBE, para produção do gel de poliacrilamida nativo e tampão de corrida. As reações foram conduzidas em temperaturas mais elevadas, entre 25 e 27°C, mas não detectamos resultado positivo de interação para nenhum dos experimentos (dados não mostrados).

Pelos resultados dos experimentos genéticos, era esperado que o fator de transcrição CG9571 se ligasse a *eve 2*, o que foi verificado em um único ensaio (Figura 42). A ausência de interação com a versão mutada deste *CRM*, cuja sequência (GTTT)_{4x} foi deletada, neste mesmo ensaio sugere que o fator reconhece o mesmo sítio que Slp1, algo que também esperávamos devido à similaridade entre a sequência do domínio de ligação das duas proteínas. Contudo, pelo efeito da expressão ectópica de CG9571 em *eve 1* (Figura 6) também supúnhamos verificar resultado positivo para a interação do fator de transcrição com este *CRM*.

Pelo fato de não termos conseguido reproduzir o resultado, entretanto, consideramos que as questões sobre a interação de CG9571 e *eve 2* propostas para este trabalho não foram respondidas de modo satisfatório. Embora tenhamos reputado a não repetição do resultado de interação com CG9571 VI à instabilidade da maior parte das proteínas da preparação, é interessante notar que as proteínas que permaneceram solúveis nesta e nas demais preparações não se ligaram a este *CRM*. Verificamos que a concentração de proteínas em CG9571 VIII foi mantida igual após mais de 6 semanas, indicando que as proteínas solúveis estavam em uma conformação estável.

Como mencionado anteriormente, consideramos a possibilidade de ter obtido, através dos métodos de renovelamento executados, proteínas solúveis com estrutura distinta da conformação nativa da proteína que podem apresentar menor capacidade de ligação ao DNA ou até mesmo nenhuma capacidade. Este não é um fenômeno incomum (Rosano e Ceccarelli, 2014). Mesmo após a otimização dos procedimentos para obtenção de proteínas solúveis a partir de corpos de inclusão, a produção de proteína ativa usualmente varia entre 10 e 50% (Carrió e Villaverde, 2002). Na realidade, uma série de observações independentes mais recentes argumentam contra a ideia de retratar proteínas solúveis como completamente funcionais e estruturalmente homogêneas (Gonzalez-Montalban et al., 2007; Martinez-Alonso et al., 2008). Ao que parece, a fração solúvel de proteínas pode ser melhor compreendida como uma fração composta por populações de proteínas que apresentam características distintas (Martinez-Alonso et al., 2008).

Outra hipótese que estamos estudando se refere ao caráter básico da proteína e sua possível interferência durante a migração do complexo por eletroforese. Embora haja pouca informação a respeito do efeito da carga de uma proteína sobre as propriedades de migração do complexo formado (Lane *et al.*, 1992) e haja exemplos na literatura de ensaios com proteínas básicas que empregaram as mesmas condições que utilizamos para as reações de ligação e eletroforese, como mencionado anteriormente, estamos considerando a possibilidade de que a carga oposta das duas macromoléculas poderia causar a dissociação mediante o campo elétrico formado durante a eletroforese, principalmente se a ligação do fator de transcrição a *eve 2* for de baixa afinidade. Sambrook e Russel (2001) afirmam que o uso de tampões com pH mais elevados para a confecção do gel e eletroforese pode ser necessário para ensaios com proteínas que possuam forte carga positiva.

Os fatores ainda não compreendidos que podem estar por trás da falta de sucesso em repetir a interação com *eve 2* também podem ter afetado a interação de CG9571 com *eve 1*, de modo que também não consideramos esta questão fechada.

5. CONCLUSÕES E PERSPECTIVAS FUTURAS

A realização desse trabalho atendeu outras prioridades do laboratório paralelamente aos objetivos específicos do projeto. A primeira delas foi proporcionar

uma nova área de atuação do laboratório, envolvendo metodologias de manipulação de proteína recombinante. Técnicas foram implementadas possibilitando que alunos ingressantes no grupo agora possam se beneficiar dessa abordagem experimental. A metodologia estabelecida bem como todo conhecimento adquirido ao longo do trabalho será importante para a realização de estudos correlatos com outros fatores de transcrição. Experimentos de interação *in vitro* entre fatores de transcrição e *CRMs* alvo são testes importantes para validar hipóteses de interação direta que experimentos genéticos não são capazes de definir.

O presente estudo reforçou o interesse na caracterização funcional do gene *CG9571* em *Drosophila*. Experimentos desse estudo tiveram como questão central aceitar ou a rejeitar a hipótese de que *CG9571* interaja diretamente com possíveis *CRMs* alvo. Esses experimentos continuam em curso no laboratório e abriram novas possibilidades de investigação. Por exemplo, a obtenção de proteína recombinante é um instrumento que permitirá gerar anticorpos, que por sua vez, são ferramentas que permitirão investigar a proteína *in vivo*. É grande nosso interesse em elucidar a estabilidade e a difusão da proteína no início da blastoderme. A expressão de *CG9571* ocorre na extremidade anterior da blastoderme ao mesmo tempo que outros genes *gap* são expressos, mas ao contrário dos genes *gap* típicos, a transcrição de *CG9571* é transiente. Então, de acordo com nossa hipótese preferencial de trabalho, a atuação direta da proteína *CG9571* em faixas de expressão *pair-rule* dependeria de uma proteína estável e capaz de se difundir. Ainda que a nossa hipótese preferencial de trabalho não seja confirmada, evidências genéticas não deixam dúvidas da contribuição de *CG9571* na regulação das faixas de expressão *pair-rule* ao menos como um fator que potencializa a repressão de *Slp1*, por um mecanismo desconhecido.

Atualmente estamos conduzindo ensaios para a expressão de *Slp1* a partir de construção disponível no laboratório da sequência da proteína completa em vetor de expressão pET-3a. Nosso objetivo é obter a proteína recombinante através de purificação por SDS-PAGE e dos mesmos métodos de renaturação e testar sua interação com *eve 2*. Como a interação de *Slp1* com este *CRM* já foi verificada em outro trabalho, poderemos ter indicativos para avaliar os métodos de renovelamento que foram empregados, bem como as condições de ligação utilizadas, apesar das diferenças entre estas duas proteínas. Também planejamos iniciar em breve novas preparações de *CG9571 FL* para testar a interação de preparações frescas com os *CRMs* utilizando BN-

PAGE e eletroforese em condições mais alcalinas para os ensaios de retardo, a fim de testarmos a hipótese da influência do caráter básico da proteína na interação.

Caso a produção de proteínas instáveis ou solúveis e sem atividade persista como resultado dos métodos empregados, consideramos (1) realizar novas tentativas de obtenção de proteína solúvel através da coexpressão de chaperonas bacterianas (e.g. GroEL) ou do uso de outro vetor de expressão ou (2) produzir anticorpos a partir da proteína recombinante purificada por SDS-PAGE e conduzir experimentos de interação dos *CRMs* com o extrato celular do embrião, utilizando os anticorpos produzidos para detecção da proteína e/ou para ensaios de *supershift*.

REFERÊNCIAS

- Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Biologia molecular da célula. 5. ed. Porto Alegre: Artmed; 2011.
- Anderson KV, Nüsslein-Volhard C. Information for the dorsal--ventral pattern of the *Drosophila* embryo is stored as maternal mRNA. *Nature*. 1984;311(5983):223–7.
- Andrioli LP, Oberstein AL, Corado MSG, Yu D, Small S. Groucho-dependent repression by Sloppy-paired 1 differentially positions anterior pair-rule stripes in the *Drosophila* embryo. *Dev Biol*. 2004;276(2):541–51.
- Andrioli LPM, Vasisht V, Theodosopoulou E, Oberstein A, Small S. Anterior repression of a *Drosophila* stripe enhancer requires three position-specific mechanisms. *Development*. 2002;129(21):4931–40.
- Baneyx F, Mujacic M. Recombinant protein folding and misfolding in *Escherichia coli*. *Nat Biotechnol*. 2004;22(11):1399–408.
- Bao B, Hong B, Feng QL, Xu, WH. Transcription factor fork head regulates the promoter of diapause hormone gene in the cotton bollworm, *Helicoverpa armigera*, and the modification of SUMOylation. *Insect Biochemistry and Molecular Biology*. 2011; 41(9), 670–679.
- Bell, PA. *E. coli* expression systems. In: Gerstein, AS, editor. *Molecular Biology Problem Solver: a laboratory guide*. New York: Wiley-Liss; 2002.
- Bondos SE, Bicknell A. Detection and prevention of protein aggregation before, during, and after purification. *Analytical biochemistry*. 2003;316(2):223-231
- Bresolin ITL, Alves E, Bueno SMA. Revisão. *Quim Nov*. 2009;32(5):1288–96.
- Buchner J, Rudolph R. Routes to active proteins from transformed microorganisms. *Current opinion in biotechnology*. 1991;2(4):532-538.
- Calzavara-Silva CE, Prosdocimi F, Abath FGC, Pena SDJ, Franco GR. Nucleic acid binding properties of SmZF1, a zinc finger protein of *Schistosoma mansoni*. *Int J Parasitol*. 2004;34(11):1211–9.
- Cardigan KM, Grossniklaus U, Gehring WJ. Functional redundancy: the respective roles of the two sloppy paired genes in *Drosophila* segmentation. *Proc. Natl. Acad. Sci. U.S.A.* 1994;91:6324-6328.
- Carrió MM, Corchero JL, Villaverde A. Dynamics of in vivo protein aggregation: building inclusion bodies in recombinant bacteria. *FEMS Microbiol. Lett*. 1998;169:9-15.
- Carrió MM, Villaverde A. Construction and deconstruction of bacterial inclusion bodies. *Journal of biotechnology*. 2002;96(1):3-12.

- Carroll SB, Scott MP. Zygotically active genes that affect the spatial expression of the *fushi tarazu* segmentation gene during early *Drosophila* embryogenesis. *Cell*. 1986;45(1):113–26.
- Cirillo LA, Zaret KS. Specific Interactions of the Wing Domains of FOXA1 Transcription Factor with DNA. *J Mol Biol*. 2007;366(3):720–4.
- Clark EDB. Refolding of recombinant proteins. *Current Opinion in Biotechnology*. 1998;9(2):157-163.
- De Marco A, De Marco V. Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning. *Journal of Biotechnology*. 2004;109(1-2):45-52.
- Dresler J, Klimentova J, Stulik J. Bacterial protein complexes investigation using blue native PAGE. *Microbiol Res* [Internet]. Elsevier; 2011;166(1):47–62.
- Dubendorf JW, Studier FW. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *Journal of molecular biology*. 1991;219(1):45-59.
- Dubnau J, Struhl G. RNA recognition and translational regulation by a homeodomain protein. *Nature*. 1996;379:694–9.
- FLYBASE. Projeto criado e mantido por um consórcio de pesquisadores das universidades de Harvard, Cambridge, Indiana e New Mexico. Identificação do gene no Flybase: FBgn0003720. Disponível em: <<http://flybase.org>>. Acesso em: 16.mar. 2016.
- Förster-Fromme K, Jendrossek D. *AtuR* is a repressor of acyclic terpene utilization (*Atu*) gene cluster expression and specifically binds to two 13 bp inverted repeated sequences of the *atuA-atuR* intergenic region. *FEMS Microbiol. Lett*. 2010;308(2):166-174.
- Fried MG. Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay. *Electrophoresis* 1989;10:366–376.
- Fried MG, Bromberg JL. Factors that affect the stability of protein-DNA complexes during gel electrophoresis. *Electrophoresis* 1997;18:6–11.
- Fried MG, Crothers DM. Equilibria and Kinetics of Lac Repressor-Operator Interactions by Polyacrylamide Gel Electrophoresis. *Nucl Acids Res*. 1981;9:6505–6525.
- Fujioka M, Emi-Sarker Y, Yusibova GL, Goto T, Jaynes JB. Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. *Development*. 1999;126(11):2527–38.

Fujioka M, Miskiewicz P, Raj L, Gulledge AA, Weir M, Goto T. *Drosophila* Paired regulates late even-skipped expression through a composite binding site for the paired domain and the homeodomain. *Development*. 1996;122(9):2697–707.

Galas DJ, Schmitz A. DNase footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucl. Acids Res*. 1978;5(9):3157-70.

Garcia SML, Fernández CG. *Embriologia*. 3. ed. Porto Alegre: Artmed; 2012.

García-Fruitós E, González-Montalbán N, Morell M, Vera A, Ferraz RM, Arís A, Ventura S e Villaverde A. Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microbial cell factories*. 2005;4:27.

Garner MM, Revzin A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon system. *Nucl Acids Res*. 1981;9:3047–3060.

Garner MM, Revzin A. The use of gel electrophoresis to detect and study nucleic acid-protein interactions. *Trends Biol Sci*. 1986;11:395–396

Gatti-lafranconi P, Natalello A, Ami D, Doglia SM, Lotti M. Concepts and tools to exploit the potential of bacterial inclusion bodies in protein science and biotechnology. *FEBS*. 2011;278:2408–18.

GE Healthcare Bio-Sciences. Purifying challenging proteins: principles and methods. Uppsala: Handbook 28-9095-31 AA; 2007. Disponível em: <<http://gelifesciences.com>> Acesso em: jan.2018.

GE Healthcare Bio-Sciences. Recombinant protein purification handbook: principles and methods. Uppsala: Handbook 18-1142-75 AD; 2009. Disponível em: <<http://gelifesciences.com>> Acesso em: mai.2017.

Gilbert SF. *Developmental biology*. 7 ed. Massachusetts: Sinauer Associates; 2003.

Gilbert SF. *Developmental biology*. 9. ed. Massachusetts: Sinauer Associates; 2010.

Gräslund S. *et al*. Protein production and purification. *Nature methods*. 2008;5(2):135-146.

Green MM. Back mutation in *Drosophila melanogaster*. I. X-ray induced back mutation back mutation at the yellow, scute and white loci. *Genetics*. 1961;46:671-682.

Grossniklaus U, Pearson RK, Gehring WJ. The *Drosophila* sloppy paired locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. *Genes Dev*. 1992;6(6):1030–51.

Hager DA, Burgess RR. Elution of proteins from sodium dodecyl sulfate-polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity: results with sigma subunit of *Escherichia coli* RNA polymerase, wheat germ DNA topoisomerase and other enzymes. *Anal. Biochem*. 1980;109:76–86.

- Hanahan D. Techniques for transformation of *E. coli*. In: Glover D M, editor. DNA cloning: a practical approach. Vol. 1. Oxford: IRL Press; 1985. pp. 109–135
- Harding K, Hoey T, Warrior R, Levine M. Autoregulatory and gap gene response elements of the even-skipped promoter of *Drosophila*. EMBO J. 1989;8(4):1205–12.
- Harding K, Rushlow C, Doyle HJ, Hoey T, Levine M. Cross-Regulatory Interactions Among. Science. 1986;233(4767):953–9.
- Hellman LM, Fried MG. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nat. Protoc. 2007;2(8):1849–61.
- Hendrickson W. Protein-DNA interactions studied by the gel electrophoresis-DNA binding assay. BioTechniques. 1985;3(3):198–207.
- Hoch M, Seifert E, Jäckle H. Gene expression mediated by cis-acting sequences of the Krüppel gene in response to the *Drosophila* morphogens bicoid and hunchback. EMBO J. 1991;10(8):2267–78.
- Horwitz JP, Chua J, Curby RJ, Tomson AJ, Da Rooge MA, Fisher BE, Mauricio J, Klundt I. Substrates for Cytochemical Demonstration of Enzyme Activity. I. Some Substituted 3-Indolyl- β -D-glycopyranosides 1a. Journal of medicinal chemistry. 1964;7(4):574–575.
- Jevševar S, Gaberc-Porekar V, Fonda I, Podobnik B, Grdadolnik J, Menart V. Production of nonclassical inclusion bodies from which correctly folded protein can be extracted. Biotechnol. Prog. 2005;21(2):632–39.
- Jiang J, Hoey T, Levine M. Autoregulation of a segmentation gene in *Drosophila*: Combinatorial interaction of the even-skipped homeo box protein with a distal enhancer element. Genes Dev. 1991;5(2):265–77.
- Johnston DS, Nüsslein-Volhard C. The origin of pattern and polarity in the *Drosophila* embryo. Cell. 1992;68(2):201–19.
- Julien E, Bordeaux MC, Garel A, Couble P. Fork head alternative binding drives stage-specific gene expression in the silk gland of *Bombyx mori*. Insect Biochemistry and Molecular Biology. 2002; 32(4), 377–387.
- Jungbauer A, Kaar W. Current status of technical protein refolding. Journal of biotechnology. 2007;128(3):587–596.
- Koth CM, Orlicky SM, Larson SM, Edwards AM. Use of limited proteolysis to identify protein domains suitable for structural analysis. In: Carter Jr CW, Sweet RM (Ed.). Macromolecular Crystallography (Part C). Methods in Enzymology. 2003;368:77–84.
- Krause F. Detection and analysis of protein – protein interactions in organellar and prokaryotic proteomes by native gel electrophoresis: (Membrane) protein complexes and supercomplexes. Electrophoresis. 2006;27:2759–81.

Krystel J, Ayyanathan K. Global analysis of target genes of 21 members of the ZAD transcription factor family in *Drosophila melanogaster*. *Gene*. 2013;512:373-382.

Laemmli, UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 1970;227(5257):680–685.

Lai E, Clark KL, Burley SK, Darnell, JE. Hepatocyte nuclear factor 3/fork head or "winged helix" proteins: a family of transcription factors of diverse biologic function. *Proceedings of the National Academy of Sciences*. 1993;90(22):10421-23.

Lane D, Prentki P, Chandler M. Use of gel retardation to analyze protein-nucleic acid interactions. *Microbiology and Molecular Biology Reviews*. 1992;56(4):509-528.

Langeland JA, Carroll SB. Conservation of regulatory elements controlling hairy pair-rule stripe formation. *Development*. 1993;117(2):585–96.

Levine M, Manley JL. Transcriptional repression of eukaryotic promoters. *Cell*. 1989;59:405-408.

Lewis EB, Bacher F. Methods of feeding ethyl methane sulfonate (EMS) to *Drosophila* males. *Drosophila Information Service*. 1968;43:193.

Lewis EB. A gene complex controlling segmentation in *Drosophila*. *Nature*. 1978;276(5688): 565-70.

Lotti M, Pollegioni L. Aggregation of recombinant proteins: understanding basic issues to overcome production bottlenecks. In: Doglia SM, Lotti M (Ed.). *Protein aggregation in bacteria: functional and structural properties of inclusion bodies in bacterial cells*. New Jersey: Wiley; 2014. p. 221-46.

Martínez-Alonso M, González-Montalbán N, García-Fruitós E, Villaverde A. The functional quality of soluble recombinant polypeptides produced in *Escherichia coli* is defined by a wide conformational spectrum. *Appl. Environ. Microbiol.* 2008;74(23):7431-7433.

Martinez-Arias A, Lawrence PA. Parasegments and compartments in the *Drosophila* embryo. *Nature*. 1985;313:639-642.

Middelberg J. Preparative protein refolding. *TRENDS Biotechnol.* 2002;20(10):437–43.

Novagen. pET system manual. Handbook TB055. Ed. 10; 2003.

Nüsslein-Volhard C, Wieschaus E, Kluding H. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. I. Zygotic loci on the second chromosome. *Wilhelm Roux's Arch. Dev. Biol.* 1984;193:267-282.

Nüsslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in *Drosophila*. *Nature*. 1980;287(5785):795–801.

Obsil T, Obsilova V. Structure/function relationships underlying regulation of FOXO transcription factors. *Oncogene*. 2008;27:2263-75.

Ochoa-Espinosa A, Yucel G, Kaplan L, Pare A, Pura N, Oberstein A, Papatsenko D, Small S. The role of binding site cluster strength in Bicoid-dependent patterning in *Drosophila*. *Proc Natl Acad Sci USA*. 2005;102(14):4960-5.

Pankratz MJ, Jäckle H. Blastoderm segmentation. In: Bate M, Arias AM. *The Development of Drosophila melanogaster*. New York: Cold Spring Harbor Laboratory Press; 1993. p. 467-516.

Pankratz MJ, Seifert E, Gerwin N, Billi B, Nauber U, Jackle H. Gradients of Kruppel and knirps gene products direct pair-rule gene stripe patterning in the posterior region of the *Drosophila* embryo. *Cell*. 1990;61(2):309-17.

Parker DS, White MA, Ramos AI, Cohen BA, Barolo S. The cis-regulatory logic of hedgehog gradient responses: key roles for gli binding affinity, competition and cooperativity. *Sci Signal*. 2011;4(176).

Peternel S, Grdadolnik J, Gaberc-Porekar V, Komel R. Engineering inclusion bodies for non denaturing extraction of functional proteins. *Microb Cell Fact*. 2008;7:34.

Pierrou S, Hellqvist M, Samuelsson L, Enerbäck S, Carlsson P. Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending. *EMBO Journal*. 1994;13(20):5002-5012.

Ream JA, Lewis LK, Lewis KA. Rapid agarose gel electrophoretic mobility shift assay for quantitating protein : RNA interactions. *Anal Biochem*. Elsevier Inc; 2016;511:36-41.

Rehm BHA, Qi Q, Beermann BB, Hinz H-J, Steinbüchel A. Matrix-assisted in vitro refolding of *Pseudomonas aeruginosa* class II polyhydroxyalkanoate synthase from inclusion bodies produced in recombinant *Escherichia coli*. *Biochem J*. 2001;358:263-8.

ResearchGate. Fórum de discussão mantido na plataforma ResearchGate. Tópico: What protocol do you follow for doing native PAGE with proteins with pI higher than 8?, criado em abr. 2014 (2014a) por Zamora-Caballero S (Karolinska Institutet). Uso da técnica BN-PAGE sugerido por Ibelle P (Universidad Nacional Autónoma de México). Disponível em: <https://www.researchgate.net/post/What_protocol_do_you_follow_for_doing_native_PAGE_with_proteins_with_pI_higher_than_8>. Acesso em: 17 out. 2018.

_____. Tópico: Can anyone help with EMSA trouble, as protein/DNA complex trapped in wells?, criado em mai. 2014 (2014b) por Lin TY (Jagiellonian University). Disponível em: <https://www.researchgate.net/post/Can_anyone_help_with_EMSA_trouble_as_protein_DNA_complex_trapped_in_wells>. Acesso em: 17 out. 2018.

_____. Tópico: EMSA problem – no shift/protein stuck in the well?, criado em mar. 2016 por Gart E (Texas A&M University). Disponível em: <<https://www.researchgate>.

net/post/EMSA_problem-no_shift_protein_stuck_in_the_well>. Acesso em: 17 out. 2018.

Rivera-Pomar R, Jäckle H. From gradients to stripes in *Drosophila* embryogenesis: Filling in the gaps. *Trends Genet.* 1996;12(11):478–83.

Rongo C, Lehmann R. Regulated synthesis, transport and assembly of the *Drosophila* germ plasm. *Trends Genet.* 1996;12(3):102–9.

Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in microbiology.* 2014;5:172.

Sambrook J, Russell DW. *Molecular cloning: a laboratory manual.* 3. ed. New York: Cold Spring Harbor Laboratory Press; 2001.

Schägger H, von Jagow G. Blue Native Electrophoresis for Isolation of Membrane Protein Complexes in Enzymatically Active Form. *Anal Biochem.* 1991;199:223–31.

Schier A, Gehring W. Direct homeodomain–DNA interaction in the autoregulation of the fushi tarazu gene. *Nature.* 1992;356(6372):804–7.

Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK. Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microbial Cell Factories.* 2015;14:41.

Singh SM, Sharma A, Upadhyay AK, Singh A, Garg LC, Panda AK. Solubilization of inclusion body proteins using n-propanol and its refolding into bioactive form. *Protein Expr Purif.* 2012;81:75–82.

Small S, Blair A, Levine M. Regulation of even-skipped stripe 2 in the *Drosophila* embryo. *EMBO J.* 1992;11(11):4047–57.

Small S, Blair A, Levine M. Regulation of two pair-rule stripes by a single enhancer in the *Drosophila* embryo. *Dev Biol.* 1996;175(2):314–24.

Small S, Kraut R, Hoey T, Warrior R, Levine M. Transcriptional regulation of a pair-rule stripe in *Drosophila*. *Genes Dev.* 1991;5(5):827–39.

Small S. Pair-rule genes. In: Creighton TE. *Encyclopedia of molecular biology.* New York: Wiley-Interscience Publication; 1999. p. 1754-1759.

St Johnston D. The art and design of genetic screens: *Drosophila melanogaster*. *Nat Rev Genet.* 2002; 3(3):176-88.

Stanojevic D, Small S, Levine M. Regulation of a segmentation stripe by overlapping activators and repressors in the *Drosophila* embryo. *Science.* 1991;254(5036):1385–7.

Thermo Scientific. Purify proteins from polyacrilamide gels. TR0051.1. 2009. Disponível em: <<http://tools.thermofisher.com/content/sfs/brochures/TR0051-Elute-from-polyacrylamide.pdf>>.

Ullman A, Jacob F, Monod J. Characterization by in vitro complementation to an operator proximal segment of the β -galactosidase structural gene of *Escherichia coli*. *J. Mol. Biol.* 1967;24:339.

Unezaki S, Nishizawa M, Okuda-Ashitaka E, Masu Y, Mukai M, Kobayashi S, Sawamoto K, Okano H, Ito S. Characterization of the isoforms of MOVO zinc finger protein, a mouse homologue of *Drosophila* Ovo, as transcription factors. *Gene*. 2004;336(1):47–58.

Vagenende V, Yap MG, Trout BL. Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. *Biochemistry*. 2009;48(46):11084-96.

Valax P, Georgiou G. Molecular Characterization of 8-Lactamase Inclusion Bodies Produced in *Escherichia coli*. 1. Composition. *Biotechnol Prog.* 1993;9:539–47.

Vallejo LP, Rinas U. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microbial Cell Factories*. 2004;3:11.

Ventura S, Villaverde A. Protein quality in bacterial inclusion bodies. *Trends in biotechnology*. 2006;24(4):179-185.

Waddington CH. An introduction to modern genetics. New York: Routledge;2016.

Wang C, Lehmann R. Nanos is the localized posterior determinant in *Drosophila*. *Cell*. 1991;66:637–47.

Westmermeier R. Electrophoresis in practice: a guide to methods and applications of DNA and protein separations. 5 ed. Weinheim: Wiley-VCH;2016.

Wharton RP, Struhl G. RNA regulatory elements mediate control of *Drosophila* body pattern by the posterior morphogen nanos. *Cell*. 1991;67(5):955–67.

Wittig I, Braun H, Scha H. Blue native PAGE. *Nat Protoc*. 2006;1(1):418–28.

Wolpert L, Jessell T, Lawrence P, Meyerowitz E, Robertson E, Smith J. Principles of development. 4. ed. Oxford: Oxford University Press; 2011.

Wu X, Vakani R, Small S. Two distinct mechanisms for differential positioning of gene expression borders involving the *Drosophila* gap protein giant. *Development*. 1998;125:3765–74.

Yu Y, Yussa M, Song J, Hirsch J, Pick L. A double interaction screen identifies positive and negative *ftz* gene regulators and Ftz-interacting proteins. *Mechanisms of Development*. 1999;83:95-105.

Zaret KS, Stevens K. Expression of a highly unstable and insoluble transcription factor in *Escherichia coli*: purification and characterization of the *fork head* homolog HNF3 α . *Protein Expression and Purification*. 1995;6:821-25.

Zhu X, Ahmad S, Aboukhalil A, Busser BW, Kim Y, Tansey TR, Haimovich A, Jeffries N, Bulyk ML, Michelson AM. Differential regulation of mesodermal gene expression by *Drosophila* cell type-specific Forkhead transcription factors. *Development*. 2012;139:1457-1466.

Zouhar J, Nanak E, Brzobohatý B. Expression , single-step purification, and matrix-assisted refolding of a maize cytokinin glucoside-specific β -glucosidase. *Protein Expr Purif*. 1999;17:153–62.

APÊNDICE 1. SEQUENCIAMENTO DE POSSÍVEIS CLONES *CG9571 BD* EM VETOR PCR 2.1-TOPO

As amostras foram sequenciadas com o *primer forward* M13 fornecido junto com o *kit* de clonagem TOPO (Figura 10). Fizemos a complementação de bases e a inversão das sequências obtidas e as alinhamos com a sequência de *CG9571 BD* (Figura 45).

Figura 45. Alinhamento entre a sequência de *CG9571 BD* e o resultado de sequenciamento com *primer forward* de plasmídeo do clone 1 (clonagem do domínio em pTOPO)

TOPO+CG9571_C11_IC CG9571BD	GCASGAGGGAGGCTCCAGGGGAACGCCTGTATCTTATAGTCKGTCGATTGCGCACTCGAC -----
TOPO+CG9571_C11_IC CG9571BD	TGAGCGTCGAATTGTGATGCTCGTCAGGGGGCGGAGCCTATGAAAACGCAGCAACGCG -----
TOPO+CG9571_C11_IC CG9571BD	CTTTTACGGTCCTGCCTTTGCTGGCCTTTGCTCACATGTCCTTCTGCGTTATCCCTGATT -----
TOPO+CG9571_C11_IC CG9571BD	CTGTGGATACCGTATTACCGCCTTGAGTGAGCTGATACCGCTCGCCGCGAGCCGAACGACC -----
TOPO+CG9571_C11_IC CG9571BD	GAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACGCCTCTC -----
TOPO+CG9571_C11_IC CG9571BD	CCCGCGCGTTGGCCGATTCAATATGCAGCTGGCACGACAGTTTCCCGACTGGAAAGCG -----
TOPO+CG9571_C11_IC CG9571BD	GGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTTA -----
TOPO+CG9571_C11_IC CG9571BD	CACTTTATGCTTCCGGCTCGTATGTTGTGTGAATTGTGAGCGGATAACAATTTACACA -----
TOPO+CG9571_C11_IC CG9571BD	GGAAACAGCTATGACCATGATTACGCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAAC -----
TOPO+CG9571_C11_IC CG9571BD	GGCCGCCAGTGTGCTGGAATTCGCCCTTGGAAATTCATATCAATGCAAAGCCGGCTTTCA -----AATGCAAAGCCGGCTTTCA *****
TOPO+CG9571_C11_IC CG9571BD	CCTACAGCGCCCTCATAGTGATGGCCATCTGGAGCAGCTCCGAGAAGCGTTTGACCCCTAA CCTACAGCGCCCTCATAGTGATGGCCATCTGGAGCAGCTCCGAGAAGCGTTTGACCCCTAA *****
TOPO+CG9571_C11_IC CG9571BD	GCGGGATCTGCAAGTGGATCGCGGACAACTTCCCGTACTATCGCACCCGCAAGAGCGTCT GCGGGATCTGCAAGTGGATCGCGGACAACTTCCCGTACTATCGCACCCGCAAGAGCGTCT *****
TOPO+CG9571_C11_IC CG9571BD	GGCAGAACTCGATCCGGCACAACCTGAGTCTCAATCCGTTCTTCGTTCCGGTGCCGCGAG GGCAGAACTCGATCCGGCACAACCTGAGTCTCAATCCGTTCTTCGTTCCGGTGCCGCGAG *****
TOPO+CG9571_C11_IC CG9571BD	CTCTCGATGATCCTGGACGTGGCCACTACTGGGCACTCGATCCGATGCCGAGGATCTGT CTCTCGATGATCCTGGACGTGGCCACTACTGGGCACTCGATCCGATGCCGAGGATCTGT

```

*****
TOPO+CG9571_C11_IC  CCATTGGCGAGACG TGAGGATCCGCG AGGGCGAATTCTGCAGATATCCATCACACTGGC
CG9571BD             CCATTGGCGAGACG-----
*****

TOPO+CG9571_C11_IC  GGCCGCTCGAGCATGCATCTAGAGCCCAATCGCCWTAGAGASYCCGAAGCCT
CG9571BD             -----

```

Bases coincidentes marcadas com asterisco. Em realce verde, a sequência inicial do *primer* d5 anterior à região de hibridização contendo o sítio de restrição para NdeI (fonte em vermelho), que também possui códon para metionina (iniciador da tradução). Em realce amarelo, a sequência (invertida e complementar) do *primer* d3 contendo códon de terminação (fonte em verde) e sítio de restrição para BamHI (fonte em roxo). Em realce vermelho a adenina adicionada pela enzima Taq polimerase que faz pareamento com a desoxitimidina do vetor no sistema AT *cloning*.

APÊNDICE 2. SEQUENCIAMENTO DE POSSÍVEIS CLONES *CG9571 BD* EM VETOR PET-15B

Como não dispúnhamos, num primeiro momento, de *primers* para hibridização com regiões do vetor pET-15b, sequenciamos as amostras com *primers* para *CG9571 BD*. Fizemos alinhamentos entre a sequência de *CG9571 BD*, a sequência obtida para sequenciamento com *primer forward* e a sequência complementar e invertida para o sequenciamento com *primer reverse* (Figuras 46 e 47).

Figura 46. Alinhamento entre a sequência de *CG9571 BD* e parte da sequência obtida para plasmídeo do clone 6 (clonagem do domínio em pET) com *primer forward*

```

CG9571      AATGCAAAGCCGGCTTTCACCTACAGCGCCCTCATAGTGATGGCCATCTGGAGCAGCTCC
Clone6_P5'  -----GATAWKWYGTTCATCYAGTGATGGCCATCTGGAGCAGCTCC
                                     *      *****

CG9571      GAGAAGCGTTTGACCCTAAGCGGGATCTGCAAGTGGATCGCGGACAACTTCCCGTACTAT
Clone6_P5'  GAGAAGCGTTTGACCCTAAGCGGGATCTGCAAGTGGATCGCGGACAACTTCCCGTACTAT
          *****

CG9571      CGCACCCGCAAGAGCGTCTGGCAGAACTCGATCCGGCACAACCTGAGTCTCAATCCGTTT
Clone6_P5'  CGCACCCGCAMGAGCGTCTGGCAGAACTCGATCCGGCACAACCTGAGTCTCAATCCGTTT
          *****

CG9571      TTCGTTTCGGGTGCCGCGAGCTCTCGATGATCCTGGACGTGGCCACTACTGGGCACTCGAT
Clone6_P5'  TTCGTTTCGGGTGCCGCGAGCTCTCGATGATCCTGGACGTGGCCACTACTGGGCACTCGAT
          *****

CG9571      CCGTATGCCGAGGATCTGTCCATTGGCGAGACG-----
Clone6_P5'  CCGTATGCCGAGGATCTGTCCATTGGCGAGACGTGAGGATCCGGCTGCTAACAAAGCCCG
          *****

```

Bases coincidentes marcadas com asterisco.

Figura 47. Alinhamento entre a sequência de *CG9571 BD* e parte da sequência obtida para plasmídeo do clone 6 (clonagem do domínio em pET) com *primer reverse*

```

CG9571      -----AATGCAAAGCCGGC
Clone6_P3'  TCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGAATGCAAAGCCGGC
          *****

CG9571      TTTCACCTACAGCGCCCTCATAGTGATGGCCATCTGGAGCAGCTCCGAGAAGCGTTTGAC
Clone6_P3'  TTTCACCTACAGCGCCCTCATAGTGATGGCCATCTGGAGCAGCTCCGGAAGCGTTTGAC
          *****

CG9571      CCTAAGCGGGATCTGCAAGTGGATCGCGGACAACTTCCCGTACTATCGCACCCGCAAGAG
Clone6_P3'  CCTAAGCGGGATCTGCAAGTGGATCGCGGACAACTTCCCGTACTATCGCACCCGCAAGAG
          *****

CG9571      CGTCTGGCAGAACTCGATCCGGCACAACCTGAGTCTCAATCCGTTCTTCGTTTCGGGTGCC
Clone6_P3'  CGTCTGGCAGAACTCGATCCGGCACAACCTGAGTCTCAATCCGTTCTTCGTTTCGGGTGCC
          *****

CG9571      GCGAGCTCTCGATGATCCTGGACGTGGCCACTACTGGGCACTCGATCCGTATGCCGAGGA
Clone6_P3'  GCGAGCTCTCGATGATCCTGGACGTGGCCACTACTGGCACTCGATCCGTATGCCGAGGA

```



```

*****
CG9571      TCTGTCCATTGGCGAGACG
Clone6_P3'  C-TAACTC-----
              *  *

```

Bases coincidentes marcadas com asterisco.

As regiões adjacentes aos *primers* utilizados para o sequenciamento apresentam diversos mal pareamentos e até inclusões de sequências de bases no caso da sequência obtida com *primer reverse* (Figuras 46 e 47). Tal resultado era esperado, já que não tínhamos *primers* que anelassem com sequências do vetor à montante do sítio de clonagem. As demais bases em direção à extremidade oposta da extremidade hibridizada pelo *primer*, entretanto, coincidiram com CG9571 (Figura 48).

Figura 48. Alinhamento entre a sequência de CG9571 BD e as sequências obtidas para clone 6 (clonagem do domínio em pET) com *primers reverse* e *forward*

```

Clone6_P3 (IC)  AATGCAAAGCCGGCTTTACCTACAGCGCCCTCATAGTGATGGCCATCTGGAGCAGCTCC
CG9571BD       AATGCAAAGCCGGCTTTACCTACAGCGCCCTCATAGTGATGGCCATCTGGAGCAGCTCC
Clone6_P5'     -----BATAWKWYGTCTATCYAGTGATGGCCATCTGGAGCAGCTCC
                  AGAC      C
                  TTTT      T

Clone6_P3 (IC)  CTGAAGCGTTTGACCCTAAGCGGGATCTGCAAGTGGATCGCGGACAACCTCCCGTACTAT
CG9571BD       GAGAAGCGTTTGACCCTAAGCGGGATCTGCAAGTGGATCGCGGACAACCTCCCGTACTAT
Clone6_P5'     GAGAAGCGTTTGACCCTAAGCGGGATCTGCAAGTGGATCGCGGACAACCTCCCGTACTAT

Clone6_P3 (IC)  CGCACCCGCAAGAGCGTCTGGCAGAACTCGATCCGGCACAACCTGAGTCTCAATCCGTTT
CG9571BD       CGCACCCGCAAGAGCGTCTGGCAGAACTCGATCCGGCACAACCTGAGTCTCAATCCGTTT
Clone6_P5'     CGCACCCGCAMGAGCGTCTGGCAGAACTCGATCCGGCACAACCTGAGTCTCAATCCGTTT
                  C

Clone6_P3 (IC)  TTTCGTTTCGGGTGCCGCGAGCTCTCGATGATCCTGGACGTGGCCACTACTGGCACTCGATC
CG9571BD       TTCGTTTCGGGTGCCGCGAGCTCTCGATGATCCTGGACGTGGCCACTACTGGCCACTCGAT
Clone6_P5'     TTCGTTTCGGGTGCCGCGAGCTCTCGATGATCCTGGACGTGGCCACTACTGGCCACTCGAT

Clone6_P3 (IC)  CTTTCCCAWWTM-----
CG9571BD       CCGTATGCCGAGGATCTGTCCATTGGCGAGACG
Clone6_P5'     CCGTATGCCGAGGATCTGTCCATTGGCGAGACG
                  AA A
                  T C

```

Nas linhas Clone6_P3(IC) (sequenciamento *reverse*) e Clone6_P5' (*forward*) são destacadas em verde as bases coincidentes com a sequência do domínio (linha CG9571) e em vermelho as bases não coincidentes. Em azul claro são destacadas as bases indefinidas. Acima ou abaixo destas são apresentadas as bases a que se referem o código, em contorno preto. Quando uma das bases coincide com a sequência, é destacada em verde.

Além da confirmação da identidade do inserto, buscamos verificar se o plasmídeo se tratava do vetor utilizado para subclonagem e se o inserto foi clonado em fase. Analisamos a sequência obtida para o *primer reverse* (complementar e invertida) e encontramos as regiões correspondentes ao vetor pET-15b (Figura 49), confirmando também que o inserto foi clonado em fase.

Figura 49. Trecho da sequência obtida em sequenciamento com *primer reverse* com identificação de bases e aminoácidos do vetor pET-15b e do início de CG9571 BD

```

AGCCCGAAGTGGCGAGCCCGATCTTCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCG
GTGATGCCGGCCAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACC
ATGGGCAGCAGC CATCATCATCATCATCAC AGCAGCGGC CTGGTGCCGCGCGGCAGC CATATGA ATGCAAAGCCGGCT
MetGlySerSer HisHisHisHisHisHis SerSerGly LeuValProArgGlySer HisMet AsnAlaLysProAla
..... poli Histidina ..... Trombina NdeI CG9571 BD
TTCACCTACAGCGCCCTCATAGTGATGGCCATCTGGAGCAGCTCCGRGAAGCGTTTGACCCTAAGCGGGATCTGCAAG
GCGGACAACCTCCCGTACTATCGCACCCGCAAGAGCGTCTGGCAGAACTCGATCCGGCACAACCTGAGTCTCAATCCG

```

Sequência complementar e invertida à sequência obtida no sequenciamento. Em cinza sequência referente ao vetor, com destaques para a sequência codificante de 6 histidinas (vermelho), sequência codificante para trombina (lilás) e sítio para enzima NdeI (azul turquesa – apenas as duas primeiras bases pertencem originalmente ao vetor; as demais fazem parte do inserto). Em verde é destacada parte da sequência do domínio.