

**UNIVERSIDADE FEDERAL DO AMAZONAS – UFAM**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA**  
**– PPGBIOTEC**



**DESENVOLVIMENTO DE NOVOS VETORES BEM  
REGULADOS PARA *Escherichia coli***

Msc. Diego da Silva Moreira

**MANAUS – AM**

**2023**

**UNIVERSIDADE FEDERAL DO AMAZONAS – UFAM**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA**  
**– PPGBIOTEC**

Msc. Diego da Silva Moreira

**DESENVOLVIMENTO DE NOVOS VETORES BEM**  
**REGULADOS PARA *Escherichia coli***

Documento apresentado ao Programa de Pós-graduação em Biotecnologia (PPGBIOTEC) da Universidade Federal do Amazonas (UFAM) como requisito para Defesa de Tese do Curso de Doutorado em Biotecnologia, área de concentração Engenharia Genética.

**Orientador:** Dr. Spartaco Astolfi Filho

**Manaus – AM**

**2023**

## Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

M838d      Moreira, Diego da Silva  
Desenvolvimento de novos vetores bem regulados para  
Escherichia coli / Diego da Silva Moreira . 2023  
203 f.: il. color; 31 cm.

Orientador: Spartaco Astolfi Filho  
Tese (Doutorado em Biotecnologia) - Universidade Federal do  
Amazonas.

1. Promotor sintético regulado. 2. Regulação do operon lac. 3.  
Vetores de expressão. 4. Hormônio de crescimento de tabaqui. I.  
Astolfi Filho, Spartaco. II. Universidade Federal do Amazonas III.  
Título

“E sabemos que todas as coisas contribuem juntamente para o bem daqueles que amam a Deus, daqueles que são chamados segundo o seu propósito”.

Romanos 8 :28

## AGRADECIMENTOS

A Deus Pai Todo Poderoso que ajudou com todo conhecimento, sabedoria e inteligência e por me ajudar nesta grande aventura e jornada fascinante de 15 anos. Pelo carinho fornecido e por sempre estar comigo. Se não fosse pelo seu amor infinito, este trabalho não teria sido concretizado. Seja a glória e glórias para sempre a Ele e sei que seu amor por mim é eterno. Obrigado Deus !!!

Ao meu Pai, Orientador, chefe (Big Boss) e grande amigo Spartaco Astolfi Filho por todos os conselhos pessoais e conhecimentos compartilhados para realização deste novo trabalho; e pela importância/influência que exerce na minha vida profissional. Pela humildade, pela paciência, dedicação e pelo amor em me ensinar a fazer Biotecnologia/Engenharia Genética. Obrigado por toda ajuda professor, eu te amo muito, se sou hoje parte da Biotecnologia é devido ao senhor. Minha dedicação foi sempre extrema para lhe entregar o melhor dos trabalhos.

Dedico também aos meus pais Olival Moreira e Dilce Monteiro, ao meu Tio Raimundo Moreira, meus irmãos Jardeson e Daiana, minha sobrinha e princesa Yanka, por me entender nos momentos de estudo, por toda força e incentivo me concedidos até aqui. Aos pets Bolinha, Lola e o falecido Tom que sempre me deram muito amor.

Dedico a toda minha família, minha avó Neli a grande coluna da família, tias e tios, primos e primas. Para a família do ouro verde e ao tio Irailton, grande diretor.

Aos pilares do CAM: Enedina, Ellen, Edson, CG e Edmar por todo o conhecimento teórico/prático de laboratório, pelos ensinamentos e paciência comigo por todos os anos. Sou muito grato a todos vocês. Dedico também ao professor Mota por toda ajuda na pandemia na parte de PCR em tempo real.

A CAPES pela bolsa concedida nos anos em que estive a frente do trabalho de Doutorado, que foram essenciais para minha formação acadêmica e me proporcionar muitos momentos maravilhosos através do dinheiro.

A Dona Elza, Patrícia, Carmen, Paulo e Júlio pelo companheirismo nos dias da semana, pelas ajudas nas festas e churras no CAM ou bloco M, por toda ajuda na parte teórica e

prática do laboratório e por sempre estarem dispostos a me ajudar com as ministrações das aulas práticas de TDR da graduação e pós-graduação.

A UFAM pela oportunidade de ingressar desde o início da faculdade, Mestrado e agora o Doutorado na área da Biotecnologia, e ao PPGBIOTEC por todos os anos em que me ajudou para este momento ser realizado.

A todos os companheiros do CAM e de laboratório: Kerollen, Joyce, Leissandra, Felipe, Thatiana, Minerval, Lanna, Samantha, Genilton, Wendel, Roberto, Michelle e a todos da turma do IGEM que puderam compartilhar seus projetos comigo e me ensinar tanto. Por sempre estarem no laboratório nos momentos de risadas e de tristezas quando os experimentos não davam certos.

As minhas amigas Amanda, Carleandra e Suellen por todo companheirismo nesses anos de Doutorado e por estarem comigo na pandemia, em momentos difíceis e alegres. Por sempre me apoiarem nos estudos e pela força concedida em todos os momentos. Pelas festinhas em casa e risadas maravilhosas (sempre altas).

Para todas as minhas amizades do sítio do 60. Obrigado por toda ajuda nos anos em que estudei com todo apoio moral para concretização deste trabalho e sempre me ajudando a estar perto de vocês para apreciarmos nossos fins de semana maravilhosos na chácara. Sempre foram os melhores fins de semana com a “turma do funil”.

Agradeço imensamente também a muitos que ajudaram e contribuíram diretamente e indiretamente para realização deste trabalho. Obrigado.

## SUMÁRIO

LISTA DE TABELAS .....	xi
LISTA DE QUADROS .....	xii
LISTA DE FIGURAS .....	xiii
LISTA DE ABREVIATURAS, SÍMBOLOS E UNIDADES .....	xx
RESUMO .....	xxii
ABSTRACT .....	xxiii
1 CONTEXTUALIZAÇÃO .....	1
2 REVISÃO BIBLIOGRÁFICA .....	4
2.1 O PROCESSO DE TRANSCRIÇÃO .....	4
2.2 A RNA POLIMERASE EM PROCARIOTOS .....	6
2.3 FATOR SIGMA ( $\sigma$ ) .....	7
2.4 PROMOTORES PROCARIÓTICOS E SEUS ELEMENTOS .....	10
2.4.1 Regiões -35 e -10.....	13
2.4.2 A região espaçadora.....	14
2.4.3 Sítio de início da transcrição ou +1 .....	14
2.4.4 O elemento UP .....	15
2.4.5 Elemento -10 estendido .....	16
2.4.6 Região discriminadora.....	16
2.5 TERMINADORES DE TRANSCRIÇÃO .....	17
2.6 SEQUÊNCIA SHINE-DALGARNO .....	19
2.7 REGULAÇÃO DA EXPRESSÃO GÊNICA EM PROCARIOTOS BASEADO NO OPERON LAC .....	22
2.8 EXPRESSÃO DE GENES HETERÓLOGOS EM <i>E. coli</i> .....	27
2.9 EXPRESSÃO DE GENES HETERÓLOGOS EM <i>E. coli</i> UTILIZANDO O SISTEMA DE REGULAÇÃO DO OPERON LAC .....	29
2.10 VETORES DE EXPRESSÃO PARA <i>E. coli</i> .....	32
2.11 USO DE CAIXA DE FERRAMENTAS GENÉTICAS .....	33

2.12 HORMÔNIO DE CRESCIMENTO (GH) .....	34
3 OBJETIVOS GERAIS .....	36
3.1 OBJETIVOS ESPECÍFICOS .....	36
4 METODOLOGIA.....	37
4.1 ETAPAS DA CONSTRUÇÃO DO CONJUNTO DE VETORES REGULADOS.....	37
4.2 MATERIAL .....	37
4.2.1 Linhagens Bacterianas .....	37
4.2.2 Plasmídeos utilizados neste trabalho .....	38
4.2.2.1 pUNA .....	38
4.2.2.2 pRIQT.....	39
4.2.2.3 Vetor pGSM-tGH .....	39
4.2.3 Oligonucleotídeos utilizados neste trabalho.....	40
4.2.4 Meios de cultura para bactéria.....	41
4.2.5 Reagentes, tampões e soluções.....	41
4.3 MÉTODOS.....	41
4.3.1 Clonagem molecular.....	42
4.3.2 Transformação genética.....	42
4.3.3 Extrações de plasmídeos e purificação.....	42
4.3.4 SDS-PAGE.....	43
4.3.5 Design do promotor DM e síntese do pDM-GFP.....	43
4.3.6 Transformação genética do pDM-GFP em <i>E. coli</i> , extração plasmidial e análise com endonucleases de restrição.....	44
4.3.7 Desenvolvimento de novos vetores bem regulados baseado no promotor DM para <i>E. coli</i> .....	45
4.3.7.1 Clonagem do cassete de expressão no vetor pUNA (pDMU01).....	45
4.3.7.2 Vetor pCDM (baixa cópia).....	48
4.3.7.3 O vetor pDM02 ( <i>polinlynker</i> ).....	50
4.3.7.4 O vetor pDMU02 (alta cópia).....	53

4.3.8	Análise da regulação e expressão do gene da GFP através do conjunto de vetores regulados.....	55
4.3.9	Análise da regulação e expressão da GFP em <i>E. coli</i> por SDS-PAGE .....	56
4.3.9.1	Avaliação da expressão e regulação da GFP pelo vetor pDM-GFP.....	56
4.3.9.2	Avaliação da expressão e regulação do gene da GFP pelo vetor pDMU01 .....	57
4.3.10	Construção do plasmídeo pDM-TEK e expressão do gene híbrido da tiorredoxina humana e enteroquinase bovina em <i>E. coli</i> .....	58
4.3.11	Expressão do rtGH em <i>E. coli</i> utilizando o vetor pDMU01.....	59
4.3.11.1	Clonagem do gene tGH no vetor pDMU01 .....	59
4.3.11.2	Expressão em <i>E. coli</i> e análise da proteína heteróloga tGH.....	61
4.3.12	Melhoramento do promotor DM por síntese química de DNA randômica.....	62
4.3.12.1	Construção da biblioteca de novos promotores sintéticos baseados no promotor DM .....	63
4.3.12.2	Clonagem dos novos promotores no vetor pCDM e Transformação em <i>E. coli</i> JM110 .....	63
4.3.12.3	Análise da regulação e expressão do gene da GFP através dos diferentes promotores .....	67
4.3.12.4	Sequenciamento e Bioinformática das novas sequências promotoras.....	68
5	RESULTADOS E DISCUSSÃO .....	70
5.1	DESIGN DO PROMOTOR SINTÉTICO, CLONAGEM NO CASSETE DE EXPRESSÃO E SÍNTESE DO pDM-GFP.....	70
5.2	ANÁLISE DO VETOR pDM-GFP COM ENDONUCLEASES DE RESTRIÇÃO .....	73
5.3	CONSTRUÇÃO DE UM CONJUNTO DE VETORES BASEADO NO PROMOTOR DM PARA <i>E. coli</i> .....	76
5.3.1	pDMU01 .....	76
5.3.2	pCDM .....	78
5.3.3	pDM02.....	81
5.3.4	pDMU02.....	84
5.4	ANÁLISE DA EXPRESSÃO DA GFP PELOS PLASMÍDEOS RECOMBINANTES CONSTRUÍDOS: pDM-GFP, pDMU01, pCDM e pDMU02 .....	87
5.5	ANÁLISE DA REGULAÇÃO E EXPRESSÃO DA GFP PELO VETOR pDM-GFP.....	92

5.6 ANÁLISE DA EXPRESSÃO E REGULAÇÃO DO pDM-TEK.....	94
5.7 ANÁLISE DA REGULAÇÃO E EXPRESSÃO DA GFP PELO VETOR pDMU01 .....	96
5.8 ANÁLISE DA REGULAÇÃO E EXPRESSÃO DO tGH PELO VETOR pDMU01 .....	99
5.8.1 Construção do vetor de expressão pDMU-tGH.....	99
5.8.2 Expressão e regulação do gene tGH em <i>E. coli</i> JM110 e análise da expressão em SDS-PAGE.....	103
5.9 MELHORAMENTO DO PROMOTOR DM POR SÍNTESE QUÍMICA DE DNA RANDÔMICA .....	105
5.9.1 Construção da biblioteca dos novos promotores .....	105
5.9.1.1 Clonagem dos novos promotores no vetor pCDM e Transformação Genética em <i>E. coli</i> JM110.....	105
5.9.2 Análise da regulação e expressão do gene da GFP pelos diferentes clones recombinantes .....	107
5.9.3 Sequenciamento e análise das novas sequências promotoras dos clones.....	109
6 CONSIDERAÇÕES E PERSPECTIVAS .....	119
7 REFERENCIAL TEÓRICO.....	120
8 APÊNDICES .....	133
9 ANEXOS.....	134

## LISTA DE TABELAS

Tabela 1: Fatores sigma ( $\sigma$ ) de <i>Escherichia coli</i> . .....	8
Tabela 2: Meio de cultura Luria-Bertani. ....	41
Tabela 3: Reação dos vetores pDM-GFP e pUNA com a endonuclease <i>NotI</i> . ....	46
Tabela 4: Reação de desfosforilação do vetor pUNA com a enzima SAP. ....	47
Tabela 5: Reação do vetor pDM-GFP com a endonuclease <i>SpeI</i> . ....	48
Tabela 6: Reação com <i>SpeI</i> e <i>EcoRI</i> dos DNAs plasmidiais extraídos dos novos clones selecionados. ....	50
Tabela 7: Reação de hibridização dos oligos pKL01 e pKL02. ....	51
Tabela 8: Reação com as endonucleases <i>NdeI</i> e <i>BamHI</i> do vetor pDM-GFP. ....	52
Tabela 9: Reação de restrição com as endonucleases <i>EcoRI</i> e <i>PstI</i> dos vetores pCDM e pRIQT. ....	54
Tabela 10: Reação dos vetores pDMU01 e pGSM-tGH com as endonucleases <i>NdeI</i> e <i>BamHI</i> . ....	59
Tabela 11: Reação do vetor pDMU-tGH com as endonucleases <i>NdeI</i> e <i>BamHI</i> . ....	61
Tabela 12: Reação de hibridização dos oligonucleotídeos pDSM-1 e o pDSM-2. ....	64
Tabela 13: Reação do vetor pCDM e dos oligonucleotídeos utilizando as endonucleases <i>EcoRI</i> e <i>SpeI</i> . ....	66

## LISTA DE QUADROS

Quadro 1: Algumas sequências nucleotídicas de promotores utilizados em vetores de expressão. ....	14
Quadro 2: Oligonucleotídeos utilizados neste trabalho. ....	40
Quadro 3: Reação de preenchimento dos oligonucleotídeos pDSM-1 e o pDSM-2 (hibridizados) utilizando a enzima <i>Klenow</i> . ....	65
Quadro 4: Sequências do UP elemento. ....	112

## LISTA DE FIGURAS

Figura 1: Transcrição de genes bacterianos. O ciclo de transcrição bacteriano. A RNAP holoenzima, que compreende o núcleo da enzima e um fator sigma, interage com o promotor de DNA para formar o complexo fechado (RPc). Após o processo de reconhecimento do promotor ocorre uma isomerização em dois ou mais passos para formação do complexo fechado (RPc) para o complexo aberto (RPo), desenrolando a dupla fita de DNA na região do local de início da transcrição e permitindo que encontre o sítio ativo da holoenzima. A adição de nucleosídeos trifosfatos (NTPs) permite uma transição adicional para o complexo de iniciação que sintetiza o transcrito do RNA mensageiro (RNAm). O processo pode ser mantido no promotor, o que resulta em ciclos de iniciação abortiva em que são produzidos apenas pequenos fragmentos de RNA. Alternativamente, a RNAP pode escapar do promotor para entrar na fase de alongação, levando à liberação do fator sigma e ao alongamento do transcrito de RNA usando NTPs e fatores de transcrição (TFs) (não mostrados). A transcrição prossegue até que a RNAP encontre um terminador de transcrição (TT), fazendo com que o transcrito de RNA seja liberado e a RNAP possa dissociar-se do molde de DNA para voltar a se envolver com um fator sigma para uma nova rodada de transcrição (Fonte: Adaptado de BROWNING & BUSBY, 2016). .....6

Figura 2: Interações específicas das sequências entre a RNA polimerase, sigma 70 e regiões do promotor. Representações esquemáticas das subunidades do núcleo da RNA polimerase,  $\sigma 70$  e o DNA do promotor. RNAP:  $\alpha 2$ : azul-claro;  $\beta$  e  $\beta'$ : cinza;  $\omega$ : preto; sub-regiões do  $\sigma$ : estão mostradas. Promotor: Elemento UP: azul claro; região -35: roxo; elemento -10 estendido (EXT): vermelho; região -10: amarelo; discriminador (DISC): laranja; sítio de início da transcrição ou +1: verde. DS DNA: DNA de dupla fita a jusante do +1: cinza. Os conectores que ligam as subunidades  $\alpha C$  e  $\alpha N$  e subdomínios do  $\sigma$  são mostrados como molas. Parte das regiões de um promotor consenso (CONS) é mostrado e os promotores  $P_R$ , T7A1 e o  $P_1$  do *rrnB* são mostrados abaixo. As bases ausentes são mostradas por traços (Fonte: RUFF et al., 2015)..... 10

Figura 3: Regiões consensuais dos promotores de *E. coli*. Na parte A da figura e destacado em vermelho, mostra os nucleotídeos preferenciais das regiões -35 e -10, no elemento -10 estendido TG (-10 EXT) e no sítio inicial da transcrição ou +1 de promotores de *E. coli*. Na parte B mostra o número de promotores existentes com a sequência espaçadora e sequência discriminadoras preferenciais em número de nucleotídeos (Fonte: modificado de ROSINSKI-CHUPIN et al., 2015). ..... 12

Figura 4: Esquema de um terminador de transcrição intrínseco. Os terminadores no DNA intrínsecos consistem em uma sequência rica em nucleotídeos G e C seguido por uma região rica em Ts (cauda poli-T) em seu lado 3' e que são transcritos em um grampo com uma pequena alça com uracilas no RNA a montante (cauda U) (adaptado de KINGSFORD et al., 2007). ..... 19

Figura 5: O sítio de ligação ao ribossomo e a sequência codificante de proteínas controlam a taxa de início da tradução de um mRNA. A região RBS é subdividida em um local de pausa, um sítio de ligação ao rRNA 16S, um códon de início e uma sequência espaçadora. O reconhecimento de uma subunidade 30S estende-se desde a montante do local de pausa até cerca de 22 bp a jusante do códon de início (Adaptado de SALIS, 2011). ..... 20

- Figura 6: Sítios de ligação ao ribossomo (sequência de Shine-Dalgarno). Algumas sequências de Shine-Dalgarno e sua relação ao códon iniciador. As sequências do RBS de cada mRNA estão sublinhadas (Fonte: com modificação de SCHNEIDER & STEPHENS, 1990). ..... 22
- Figura 7: O operon lac é regulado pelo repressor LacI e pela atividade de CAP/cAMP. O repressor LacI é codificado pelo gene lacI, que se liga ao operador impedindo que ocorra a transcrição pela RNA polimerase. Na ausência de lactose, o repressor LacI está ligado ao operador (A). Na presença de lactose, o repressor LacI desliga-se do operador e assim ocorre a transcrição (B). A proteína CAP ligada com a cAMP reconhece uma sequência específica na região regulatória do gene que se localiza perto do promotor lac denominado de sítio de ligação a CAP e conseqüentemente estimulando a ligação da RNA polimerase ao promotor, ocasionando a ativação máxima da transcrição do operon lac (C)..... 26
- Figura 8: Esquema representando o fluxograma desenvolvido neste trabalho. .... 37
- Figura 9: Mapa físico do vetor pUNA. O vetor possui o gene que confere resistência a ampicilina e uma Ori de alto número de cópias. .... 38
- Figura 10: Vetor pRIQT. O vetor possui o gene que confere resistência a ampicilina e uma Ori de alto número de cópias. Possui o gene lacIqTS que codifica para expressão em alto nível do repressor LacI termo sensível entre um prefixo e sufixo..... 39
- Figura 11: Vetor pGSM-tGH. O vetor possui o gene que confere resistência a ampicilina e uma Ori de alto número de cópias. Possui o gene lacI que codifica para expressão do repressor LacI e o gene tGH entre os sítios das endonucleases *NdeI* e *BamHI*. .... 40
- Figura 12: Representação esquemática de hibridização dos oligonucleotídeos através da região espaçadora e do elemento -10 estendido. .... 64
- Figura 13: Representação esquemática do preenchimento dos oligonucleotídeos pela enzima *Klenow*. .... 65
- Figura 14: Sequência do promotor DM. No sentido da esquerda para direita (5' → 3') estão representadas as sequências de DNA que compõem o promotor, como o elemento UP (Up element), a sequência da região -35 (TTGTGA) sobreposta na sequência completa do operador lacO, a sequência do elemento -10 estendido TGTGTG (-10 EXT), a região -10 (GATAAT) e a região discriminadora GGGGACC (DISC) que antecede o ponto de início de transcrição representado pelo nucleotídeo A (+1)..... 70
- Figura 15: Sequência nucleotídica promotora DM contida no vetor de expressão, o gene GFP e o terminador de transcrição. Da esquerda para a direita (5' → 3') está a região em verde escuro que representa o prefixo dos sítios das endonucleases *EcoRI*, *NotI*, *XbaI*; a região em negro mostra a sequência do elemento UP; a região em azul claro mostra a sequência original do operador lac O1 com a região -35 sobreposta (em amarelo); a sequência elemento -10 estendida em vermelho; a região -10 em rosa, a região em laranja (discriminador) que antecede o ponto de início de transcrição (+1 em verde) seguido de um sítio de *SpeI* em azul escuro e letras brancas. Após marcada em violeta está o Shine-Dalgarno, justaposto com um sítio de *NdeI* (sublinhado) no início da sequência codificante do gene da GFP (em verde); o códon de terminação em vermelho e em verde escuro um sítio de *BamHI*. A sequência marcada de vinho mostra o terminador de transcrição seguido dos sítios das endonucleases *SpeI*, *NotI* e *PstI*. .... 71

- Figura 16: Mapa físico do vetor pDM-GFP portador do promotor DM. A figura mostra os principais componentes do vetor como: os sítios das endonucleases, o promotor DM e o operador lac, o gene da GFP, o terminador de transcrição *rrnB*, o marcador de seleção para canamicina e a origem de replicação..... 72
- Figura 17: Fotografia em gel agarose dos plasmídeos pDM-GFP extraídos. As amostras A, B, C, D e E mostram os perfis eletroforéticos extraídos do vetor pDM-GFP..... 73
- Figura 18: Análise eletroforética do cassete de expressão do vetor pDM02 com endonucleases. A amostra A identifica o plasmídeo pDM-GFP com *EcoRI*; a amostra B está o pDM-GFP com *BamHI*; a amostra C o pDM-GFP com *NdeI*; a amostra D o pDM-GFP com *PstI*; a amostra E mostra o pDM-GFP com *SpeI* e a amostra F identifica o plasmídeo pDM-GFP sem endonuclease..... 74
- Figura 19: Análise eletroforética do perfil de restrição do vetor pDM-GFP com endonucleases. A amostra A está o pDM-GFP sem endonuclease; a amostra B exemplifica o pDM-GFP com *EcoRV*; a amostra C do pDM-GFP com *NotI* e a amostra D o pDM-GFP com *EcoRI* e *PstI*. Na canaleta E, o vetor pDM-GFP com *EcoRI* e *SpeI*. Na amostra F está o vetor pDM-GFP com *EcoRI* e *NdeI* e na amostra G o pDM-GFP com *NdeI* e *BamHI*. ..... 75
- Figura 20: Análise eletroforética do pDM-GFP e do vetor pUNA digeridos com a endonuclease *NotI* e do resultado da purificação dos fragmentos. Na fotografia A, a amostra 1 está o pDM-GFP sem endonuclease e a amostra 2 o plasmídeo pDM-GFP com *NotI*; a amostra 3 está o pUNA sem endonuclease e a amostra 4 o pUNA com *NotI*. Marcado em círculo vermelho na canaleta 2 e 3, estão o fragmento do cassete de expressão e do vetor pUNA linearizado, respectivamente. Na fotografia B, na canaleta 3 está identificado o cassete de expressão e na canaleta 4 exemplifica o vetor pUNA após a purificação..... 77
- Figura 21: O mapa físico do vetor de expressão pDMU01. O vetor possui o gene que confere resistência a ampicilina e uma Ori de alto número de cópias. Possui também o cassete do vetor pDM-GFP clonado entre os sítios das endonucleases *NotI*..... 78
- Figura 22: Análise eletroforética do perfil de restrição do pDM-GFP com a endonuclease *SpeI* e purificação. Na parte A, na canaleta 1 está o pDM-GFP sem tratamento com endonuclease e na canaleta 2 o plasmídeo pDM-GFP digerido parcialmente com *SpeI*. A seta em vermelho mostra o plasmídeo linearizado com *SpeI*. Na canaleta 1 da parte B, está identificado o pDM-GFP sem endonuclease e na canaleta 2 o vetor pDM-GFP digerido parcialmente com *SpeI* após purificação. .... 79
- Figura 23: Análise eletroforética dos DNAs plasmidiais extraídos dos novos clones com as endonucleases *SpeI* e *EcoRI*. As canaletas 1, 2, 3, e 4 estão os DNAs plasmidiais de novos clones tratados com *SpeI* e *EcoRI*. ..... 80
- Figura 24: O mapa físico do vetor pCDM. A figura mostra os principais componentes do vetor como: os sítios das endonucleases, o promotor DM e o operador lac, o gene da GFP, o terminador de transcrição *rrnB*, o marcador de seleção para Canamicina e a origem de replicação. No vetor pCDM existe um único sítio para a enzima *SpeI* como estratégia de clonagem e seleção de novos promotores..... 81
- Figura 25: Fotografia do perfil eletroforético dos oligonucleotídeos pLK1 e pLK2 e do pDM-GFP com as endonucleases *NdeI* e *BamHI*. Na canaleta 1, na parte A da fotografia, mostra

somente o pLK1; na canaleta 2 mostram o pLK1 e pLK2 hibridizados e na canaleta 3 mostra somente o pLK2. Na parte B, a amostra 1 está o pDM-GFP sem digestão com endonucleases e a amostra 2 mostra o plasmídeo pDM-GFP tratado com *NdeI* e *BamHI*. O círculo em vermelho mostra o fragmento que corresponde ao gene da GFP..... 82

Figura 26: Análise eletroforética do vetor pDM-GFP purificado. Na canaleta 1 está identificado o pDM-GFP com *NdeI* e *BamHI* e após purificação..... 83

Figura 27: A figura apresenta o mapa físico do vetor de expressão pDM02. Pode-se identificar os principais componentes do vetor, como: o promotor sintético regulado (DM) e o operador lac; o RBS T7; a região de múltiplos sítios de clonagem composto pelos sítios das endonucleases (*NdeI*, *HpaI*, *AflIII*, *KpnI*, *BglIII*, *PmeI* e *BamHI*); o terminador de transcrição funcional T1, do gene de *rrnB* de *E. coli*; o marcador de seleção que confere resistência ao antibiótico canamicina (CamR); e a origem de replicação de baixo número de cópias – p15A (~10 a 12 cópias). Os sítios das endonucleases estão destacados em negrito. .... 84

Figura 28: Análise eletroforético dos vetores pCDM e pRIQT tratados com as endonucleases *EcoRI* e *PstI*. Na fotografia A, a canaleta 1 está o pCDM digerido com as endonucleases *EcoRI* e *PstI*. O círculo em vermelho mostra o fragmente correspondente ao cassete de expressão. Na fotografia B, a amostra 1 está o pRIQT sem tratamento com endonucleases e a amostra 2 identifica o plasmídeo pRIQT digerido com as endonucleases *EcoRI* e *PstI*. O círculo em vermelho mostra o fragmento correspondente ao vetor linearizado e o círculo em azul corresponde ao fragmento do cassete de expressão que contém o gene *lacIqTS*..... 85

Figura 29: Análise eletroforética do vetor pRIQT digerido com *EcoRI* e *PstI* e purificado. Nas canaletas 1 e 2 está identificado em círculos vermelhos o pRIQT digerido com as endonucleases *EcoRI* e *PstI* e após tratamento de purificação..... 86

Figura 30: Mapa físico do vetor caça promotor pDMU02. A figura mostra os principais componentes do vetor como: os sítios das endonucleases, o promotor DM, o gene da GFP, o terminador de transcrição *rrnB*, o marcador de seleção para ampicilina e a origem de replicação de alto número de cópias. No vetor pDMU02 existe um único sítio para a enzima *SpeI* como estratégia de clonagem e seleção de novos promotores. .... 87

Figura 31: A figura mostra o resultado do crescimento celular dos sistemas do conjunto de vetores. Na parte A, B, C e D da figura, estão o crescimento celular dos sistemas *E. coli* JM110 + pDM-GFP; *E. coli* JM110 + pCDM; *E. coli* DH5 $\alpha$ F'IQ + pDMU01 e *E. coli* DH5 $\alpha$ F'IQ + pDMU02, respectivamente, em função do tempo de cultivo de um clone crescido com o seu antibiótico seletivo, suplementado ou não com glicose e IPTG. O processo de suplementação com glicose ou indução com o IPTG ocorreu a partir do início do cultivo. 88

Figura 32: A figura mostra o perfil de expressão do gene da GFP dos sistemas do conjunto de vetores. Na parte A, B, C e D da figura, estão perfil de fluorescência dos sistemas *E. coli* JM110 + pDM-GFP; *E. coli* JM110 + pCDM; *E. coli* DH5 $\alpha$ F'IQ + pDMU01 e *E. coli* DH5 $\alpha$ F'IQ + pDMU02, respectivamente, em função do tempo de cultivo de um clone crescido com o seu antibiótico seletivo, suplementado ou não com glicose e IPTG. O processo de suplementação com glicose ou indução com o IPTG ocorreu a partir do início do cultivo. 91

Figura 34: O resultado da análise por eletroforese em gel desnaturante de poliacrilamida (SDS-PAGE) e fluorescência da produção de GFP por um clone de *E. coli* JM110 contendo o vetor pDM-GFP em diferentes condições de cultivo. A fotografia A mostra o gel obtido em

transluminador (300 nm) e a B corado com o azul de Coomassie. As amostras nas raias 1, 2 e 3 representam as proteínas de clones crescidos com canamicina e IPTG; as amostras nas raias 4,5 e 6 com canamicina e glicose; e amostras nas raias 7, 8 e 9 somente com canamicina. A raia M mostra o marcador molecular em kDa. A proteína recombinante GFP localiza-se na posição de aproximadamente 25 kDa (seta). ..... 94

Figura 35: Análise do crescimento celular do pDM-TEK. O perfil do crescimento celular (ABS. 600 nm x tempo) de *E. coli* JM110 programada com o plasmídeo pDM-TEK com canamicina e IPTG, com canamicina e glicose e somente com canamicina. O processo de indução com IPTG ocorreu a partir do início do cultivo. .... 95

Figura 36: O perfil eletroforético em gel de poliacrilamida desnaturante de proteínas expressas por um clone de *E. coli* JM110 contendo o vetor pDM-TEK em diferentes condições de cultivo. Nas raias 1 e 4, cultivo com glicose e canamicina; raias 2 e 5, com apenas canamicina; raias 3 e 6, cultivo com canamicina e o indutor IPTG. Nas raias 1, 2 e 3 estão as amostras com 4 horas de cultivo; nas raias 4, 5 e 6 as amostras com 8 horas de cultivo. A raia M, mostra o marcador molecular em kDa. A proteína recombinante TEK localiza-se na posição de 50 kDa (setas). ..... 96

Figura 37: Fotografia dos clones recombinantes programados com o vetor pDMU01. A placa de Petri na esquerda conte um clone isolado e a placa da direita representa os clones crescidos dispostos nas mesmas posições em condições de cultivo. Os clones foram analisados em meio LB: com somente ampicilina, ampicilina e glicose, e ampicilina e IPTG (indutor); iluminados com luz negra (300 nm). ..... 97

Figura 38: O resultado da análise por eletroforese em gel desnaturante de poliacrilamida e fluorescência da produção de GFP por um clone de *E. coli* JM110 contendo o vetor pDMU01 em diferentes condições de cultivo. A fotografia A mostra o gel obtido em transluminador (300 nm) e a B corado com o azul de Coomassie. As amostras nas raias 1, 4 e 7 representam as proteínas de clones crescidos com ampicilina e glicose; as amostras nas raias 2, 5 e 8 com somente ampicilina e amostras nas raias 3, 6 e 9 com ampicilina e IPTG. A raia M mostra o marcador molecular em kDa. A proteína recombinante GFP localiza-se na posição de aproximadamente 25 a 35 kDa (setas) nas raias 3, 6 e 9 na sua estrutura ativa (setas amarelas) e na sua estrutura inativa (setas vermelhas) nos intervalos selecionados de 3, 6 e 8 horas de crescimento celular com 1, 3 e 5 horas de indução, respectivamente. .... 98

Figura 39: Resultado da análise de restrição dos plasmídeos vetores pGSM-tGH e pDMU01 com as endonucleases *NdeI* e *BamHI*. Na figura A, na canaleta 1 está o pGSM-tGH sem tratamento com endonucleases e na canaleta 2 o plasmídeo pGSM-tGH com *NdeI* e *BamHI*. Na figura B, na canaleta 1 está o pDMU01 sem tratamento enzimático e na canaleta 2 o pDMU01 com *NdeI* e *BamHI*. As setas em branco identificam os fragmentos do tGH e da GFP. .... 100

Figura 40: Análise eletroforética dos fragmentos tGH e pDMU purificados. Na parte A, nas canaletas 1 e 2 estão identificados pelas setas brancas o fragmento do tGH. Na parte B da figura, canaleta 1 está identificado o pDMU01. .... 101

Figura 41: Análise eletroforética dos produtos de restrição do novo vetor com *NdeI* e *BamHI*. Na canaleta 1, está o DNA plasmidial do vetor sem tratamento com endonucleases e na canaleta 2 está o DNA plasmidial do novo vetor digerido com *NdeI* e *BamHI*. A seta em branco mostra fragmento correspondente ao gene do tGH. .... 102

Figura 42: Mapa físico do vetor pDMU-tGH. A figura mostra os principais componentes do vetor como: os sítios das endonucleases, o promotor DM e o operador *lac*, o RBS do T7, o gene da recombinante do tGH, o terminador de transcrição *rrnB*, o marcador de seleção para ampicilina e a origem de replicação de alto número de cópias. .... 103

Figura 43: Análise da expressão do tGH por SDS-PAGE por um clone de *E. coli* JM110 contendo o vetor pDMU-tGH em diferentes condições de cultivo. As amostras nas raias 1 e 2 representam a 0 hora com somente ampicilina. As raias 3 e 4 representam as proteínas de clones crescidos com com ampicilina e glicose e ampicilina e IPTG, respectivamente, após 10 horas de crescimento bacteriano e 6 horas de indução. As amostras nas raias 5 e 6 representam as proteínas com ampicilina e glicose e ampicilina e IPTG, respectivamente, em 24 horas de crescimento bacteriano e 20 horas de indução. A raia M mostra o marcador molecular em kDa. A proteína recombinante tGH localiza-se na posição de aproximadamente 24 kDa (setas vermelhas). .... 104

Figura 44: Fotografia da análise do gel de eletroforese da hibridização dos oligonucleotídeos pDSM-1 e pDSM- e tratamento com *Klenow*. Figura A - 1: somente oligonucleotídeo pDSM-1; 2, 3 e 4: oligonucleotídeos pDSM-1 e pDSM-2 anelados; e 5: somente oligonucleotídeo pDSM-2. Figura B - 1: oligonucleotídeos pDSM-1 e pDSM-2 somente anelados; 2: oligonucleotídeos pDSM-1 e pDSM-2 anelados e tratados com *Klenow*. .... 106

Figura 45: Fotografia do perfil eletroforético do vetor pCDM com as endonucleases *SpeI* e *EcoRI* e purificado. Na parte A, na canaleta 1 está o pCDM sem tratamento com endonuclease e na canaleta 2 o plasmídeo pCDM com as endonucleases *EcoRI* e *SpeI*. O círculo em branco mostra o vetor pCDM linearizado e na parte B o vetor pCDM após purificação (marcado em círculo vermelho). .... 107

Figura 46: Fotografia de uma análise dos clones recombinantes em diferentes condições de cultivo e indução. Na fotografia da placa A e B estão alguns diferentes clones em LB com glicose e o antibiótico. Na placa C e D, estão os mesmos clones dispostos em LB com somente o antibiótico canamicina e submetidos com canamicina e o indutor IPTG, respectivamente. Os círculos marcados com as diferentes cores mostram os clones com diferentes níveis de expressão e regulação do gene da GFP em diferentes condições de cultivo. O vetor pCDM está marcado em círculo branco. .... 109

Figura 47: Figura das sequências promotoras sintéticas dos clones da placa 1 (A), placa 2(B) e placa 3(C) que perderam a regulação por glicose, porém mantiveram altos níveis de expressão. Na figura parte A, B e C, se observa os sítios das endonucleases *EcoRI* *NotI* e *XbaI* em vermelho; a região de do elemento UP em quadrado negro; as sequências da regiões -35 em quadrado verde; a região espaçadora com o operador *lacO* sobreposto na região -35 destacada por azul claro; as sequências dos elementos -10 estendidos (-10 ext em chave verde claro); a região -10 está destacada em roxo; as sequências discriminadoras (DISC em chave rosa) e a região do início do gene (+1) em marrom. Após o +1, estão os sítios das endonucleases *PstI*, *NotI* e *SpeI* em amarelo. .... 110

Figura 48: Figura das sequências promotoras sintéticas dos clones da placa 1 (A), placa 2(B) e placa 3(C) que mantiveram a regulação por glicose, porém com nível de expressão maior que o promotor DM. Na figura se observa os sítios das endonucleases *EcoRI* *NotI* e *XbaI* em quadrado vermelho; a regiões de elemento UP em quadrado negro; as sequências da regiões -35 em verde; a região espaçadora com o operador *lacO* sobreposto à região -35 destacada em azul claro; as sequências dos elementos -10 estendidos (-10 ext, em chave verde claro); a

região -10 está destacada em roxo; as sequências discriminadoras (DISC, em chave rosa) e a região do início do gene (+1) em marrom. Após o +1, estão os sítios das endonucleases *Pst*I, *Not*I e *Spe*I em amarelo. .... 111

**LISTA DE ABREVIATURAS, SÍMBOLOS E UNIDADES**

$\alpha$	Alfa
$\beta$	Beta
$\beta'$	Beta linha
$\sigma$	Sigma
$\sigma^{70}$	sigma pertencente à família setenta
$\omega$	Ômega
( $\rho$ )	Rho
<b>ATP</b>	Adenosina trifosfato
<b>cAMP</b>	monofosfato cíclico de adenosina
<b>CAP</b>	<i>Catabolite Activator Protein</i> ou proteína ativadora do catabolismo
<b>CAT</b>	Gene que codifica a proteína cloranfenicol acetil-transferase
<b>C-terminal</b>	Carboxi-terminal
<b>DNA</b>	Ácido desoxirribonucléico
<b>DNTp</b>	Desoxirribonucleotídeo Trifosfato
<b>fMet</b>	<i>N</i> -formil metionina
<b>IFs</b>	Fatores iniciadores
<b>IPTG</b>	isopropil- $\beta$ -D-tiogalactopiranosídeo
<b>ITC</b>	Complexo de transcrição inicial
<b>kb</b>	Kilobases
<b>kDa</b>	QuiloDalton
<b>mV</b>	miliVolts
<i>lac</i>	Lactose

<i>Lac I</i>	Repressor <i>lacI</i>
<b>mRNA</b>	RNA mensageiro
<b>mL</b>	Mililitro
<b>mM</b>	Milimolar
<b>N-terminal</b>	Amino-terminal
<b>ng</b>	Nanograma
<b>O.D.</b>	Densidade óptica
<b>pb</b>	Pares de base
<b>PCR</b>	Reação em cadeia da polimerase
<b>RNA</b>	Ácido ribonucleico
<b>RNAP</b>	RNA polimerase
<b>Rpc</b>	Complexo fechado
<b>Rpo</b>	Complexo aberto
<b>SD</b>	Shine-Dalgarno
<b>UTR</b>	Segmento de DNA que não pode ser traduzido
<b>TEC</b>	Complexo de alongamento da transcrição
<b>tRNA</b>	RNA transportador

## RESUMO

Neste trabalho foi construído um conjunto de vetores de expressão regulados para finalidades de clonagem e expressão de genes heterólogos em *Escherichia coli*. Os promotores são elementos essenciais dos cassetes de expressão utilizados nesse tipo de vetores. Os vetores, derivados desse trabalho, que expressam genes heterólogos em diferentes níveis e de forma bem regulada foram baseados em um promotor sintético denominado de DM. A atividade do promotor DM pode ser modulada tanto por IPTG como por glicose. Os plasmídeos denominados: pDM02 (baixa cópia), pCDM (baixa cópia), pDMU01 (alta cópia) e pDMU02 (alta cópia) foram capazes de expressar em altos níveis e de maneira regulada em *E. coli* as sequências gênicas da GFP (Green Fluorescent Protein), da tiorredoxina humana fusionada com a enteroquinase bovina e do hormônio de crescimento de tabaqui. A sequência promotora DM foi modificada por meio de síntese química randômica de DNA para a construção de uma biblioteca de novos promotores sintéticos e estes foram a seguir ligados ao vetor pCDM em posição adequada para expressar o gene repórter da GFP. Clones de *E. coli* contendo plasmídeos recombinantes com diferentes promotores foram obtidos. Um grupo de clones que possuem novas sequências promotoras mantiveram a regulação por glicose e tiveram maiores níveis de expressão enquanto outro grupo de clones apresentou menor nível de regulação por glicose, porém com níveis bem maiores de expressão de GFP. Os resultados mostram que os novos promotores sintéticos podem ser utilizados para o desenvolvimento de novos, potentes e regulados vetores de expressão para a expressão de genes heterólogos tanto para fins de pesquisa como para bioindustriais.

**Palavras-chave:** Promotor sintético regulado DM, regulação do operon lac, vetores de expressão, hormônio de crescimento de tabaqui.

## ABSTRACT

In this work, a set of regulated expression vectors was constructed for purposes of cloning and expression of heterologous genes in *Escherichia coli*. Promoters are essential elements of the expression cassettes used in this type of vectors. The vectors, derived from this work, which express heterologous genes at different levels and in a well-regulated way were based on a synthetic promoter called DM. DM promoter activity can be modulated by both IPTG and glucose. The plasmids named: pDM02 (low copy), pCDM (low copy), pDMU01 (high copy) and pDMU02 (high copy) were capable of expressing at high levels and in a regulated manner in *E. coli* the gene sequences of GFP (Green Fluorescent Protein), human thioredoxin fused with bovine enterokinase and tambaqui growth hormone. The DM promoter sequence was modified by means of random chemical DNA synthesis to construct a library of new synthetic promoters and these were then ligated into the pCDM vector in an appropriate position to express the GFP reporter gene. *E. coli* clones containing recombinant plasmids with different promoters were obtained. A group of clones that had new promoter sequences maintained glucose regulation and had higher levels of expression, while another group of clones showed a lower level of glucose regulation, but with much higher levels of GFP expression. The results show that the new synthetic promoters can be used for the development of new, potent and regulated expression vectors for the expression of heterologous genes for both research and bioindustrial purposes.

**Keywords:** DM-regulated synthetic promoter, lac operon regulation, expression vectors, tambaqui growth hormone.

## 1 CONTEXTUALIZAÇÃO

O conjunto de ferramentas da Tecnologia do DNA recombinante (rDNA) permite a clonagem de genes heterólogos até a expressão e purificação da proteína isolada para o estudo estrutural e funcional, ou seu uso em aplicações biotecnológicas. A sequência codificante da proteína de interesse pode ser inserida em um vetor ou conjunto de vetores de expressão apropriado e transformada em um hospedeiro procariótico, como a bactéria *Escherichia coli*, que possui longa tradição segura em laboratórios e na produção de proteínas de interesse industrial, como: insulina, hormônios, fatores de crescimento e estimuladores, isso porque é bem caracterizada geneticamente e metabolicamente, reduzindo os custos de produção e melhorando o rendimento (FERRER-MIRALLES et al., 2009; ROSANO et al., 2019).

Um dos principais componentes de um vetor de expressão reconhecido por *E. coli* é o “cassete de expressão gênica”, composto por um conjunto seletivo de sequências de DNA que garantem a produção de proteínas de interesse (STRIEDNER et al., 2010; ZWICK et al., 2012). Um “cassete de expressão” (5' → 3') é composto principalmente de uma sequência de DNA denominada de promotor, capaz de recrutar a RNA polimerase holoenzima ( $\alpha 2 \beta \beta' \omega \sigma$ ) e iniciar a síntese do RNA mensageiro (mRNA); uma região composta pela mensagem genética a ser transcrita, como um gene repórter ou gene heterólogo; um terminador de transcrição – para interromper a síntese do mRNA pela RNA polimerase e o sítio de ligação ao ribossomo (**RBS** – *Ribosome Binding Site*), que recruta o ribossomo e inicia o processo de tradução de um RNA mensageiro em uma proteína (SCHUMANN & FERREIRA, 2004; DADE-ROBERTSON et al., 2018).

A escolha do promotor é muito importante para o bom funcionamento de vetores de expressão em *E. coli* (CHEN et al., 2018). Os promotores rigorosamente regulados são aqueles que possuem um regulador de transcrição no núcleo ou próximo do promotor e são os eleitos para a expressão de genes em altos níveis (BROWNING & BUSBY, 2016; MOREIRA et al., 2018; BROWNING et al., 2019; HOTHERSALL et al., 2021). Moléculas indutoras sintéticas análogas, como isopropil  $\beta$ -D-1-tiogalactopiranosídeo (IPTG) podem ser usadas para regular a expressão dos transgenes em organismos que utilizam um promotor responsivo correspondente, como baseados no próprio promotor lac, e derivados como o promotor Tac, o promotor Trp, LacUV5 e outros sistemas categorizados como promotor

araBAD, promotor L, promotor R e o promotor T7 que combinam diferentes tipos de regulação e força (MOREIRA et al., 2018).

Como os circuitos e vias genéticas existentes são difíceis de controlar devido à falta de sistemas de regulação de genes sem interferência, foram desenvolvidos sistemas com o mínimo de vazamento da expressão da proteína recombinante antes do início do bioprocessamento e assim evitar a morte da hospedeira por toxicidade da expressão pela produção da proteína recombinante. Há também um interesse contínuo em novas abordagens em cassetes mais eficientes com promotores regulados inseridos em diferentes backbones plasmidiais que aperfeiçoem a clonagem e produção de proteínas heterólogas em *E. coli* (SAÏDA, 2007; FERRER-MIRALLES et al., 2009; CHEN et al., 2018; MCCARTY & LEDESMA-AMARO, 2019; ROSANO et al., 2019).

Com todas estas considerações, os cassetes de expressão podem ser inseridos em diferentes plasmídeos com diferentes estratégias para os estudos de expressão de novos genes e por isto inspirou os autores a inventar um novo promotor, que apresentasse estrita regulação e de força média de expressão de genes heterólogos em *E. coli*. O promotor sintético DM classificado com média expressão e ótima regulação, foi construído por síntese química simultaneamente com um cassete de expressão. O cassete de expressão contém a sequência da região 5' UTR do mRNA do gene da proteína 10 do bacteriófago T7 que contém um eficiente RBS (OLINS et al., 1988), o gene repórter da proteína verde fluorescente - GFP (*Green Protein Fluorescent*), e um terminador de transcrição, o T1 do gene de *rrnB* de *E. coli* (OROSZ et al., 1991).

Após a sua concepção, o cassete de expressão foi inserido em diferentes backbones plasmidiais denominados: pDM-GFP, pDMU01, pDM02, pCDM e pDMU02 para testar suas funcionalidades. Nesse contexto, os resultados nesse estudo de fluorescência e SDS-PAGE referentes à expressão do gene repórter da GFP sob o controle dos plasmídeos pDM-GFP (baixa cópia) e pDMU01 (alta cópia) e dos genes heterólogos: do hormônio de crescimento recombinante de tabaqui (tGH), contido no vetor pDMU-tGH; do vetor pDM-TEK, contendo o gene da tiorredoxina humana e a parte catalítica da enzima enteroquinase bovina sob o controle do promotor DM, demonstraram médios e altos níveis de expressão e alta performance de regulação das proteínas mencionadas sob o controle do promotor DM em *E. coli*.

A presente sequência promotora DM, invenção e motivo de patente, clonado em um versátil conjunto de vetores difere dos outros promotores conhecidos atualmente e apresenta vasta aplicação biotecnológica. Principalmente, a presente invenção abre caminho para a clonagem de proteínas recombinantes de interesse industrial e na engenharia metabólica do maquinário celular envolvendo a expressão simultânea de diferentes genes heterólogos. Pode ser utilizado para expressão em altos níveis das proteínas tóxicas, cuja elevada produção e minuciosa regulação tem sido um desafio.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 O PROCESSO DE TRANSCRIÇÃO

A informação genética contida na molécula de DNA possui relação com a síntese de proteínas necessárias ao metabolismo celular, na qual são expressos constitutivamente ou em certas condições, na forma de mRNAs (contido nos genes) para posterior tradução gênica (em proteínas) (SCHUMANN & FERREIRA, 2004; LEWIN et al., 2005). Essa regulação transcricional da conversão de um molde de DNA para RNA é conhecido como transcrição gênica e possui a RNA polimerase (RNAP) como enzima responsável pela sua execução e, por conseguinte, é um dos processos mais fundamentais para a viabilidade da célula (DRENNAN et al., 2012). Todos os organismos vivos codificam pelo menos uma RNAP sendo que a estrutura do núcleo e muitos elementos atuantes e reguladores são compartilhados por RNAP bacterianas, eucarióticas e por archaea (RUFF et al., 2015).

O primeiro passo na expressão gênica em todos os organismos requer a abertura do duplex de DNA para expor uma fita para a síntese de RNA. O nível de transcrição gênica está relacionado com a frequência com que a RNAP inicia o processo, bem como sua afinidade a sequência promotora (HAUGEN et al., 2006, 2008; VAN HIJUM et al., 2009; SAECKER et al., 2011).

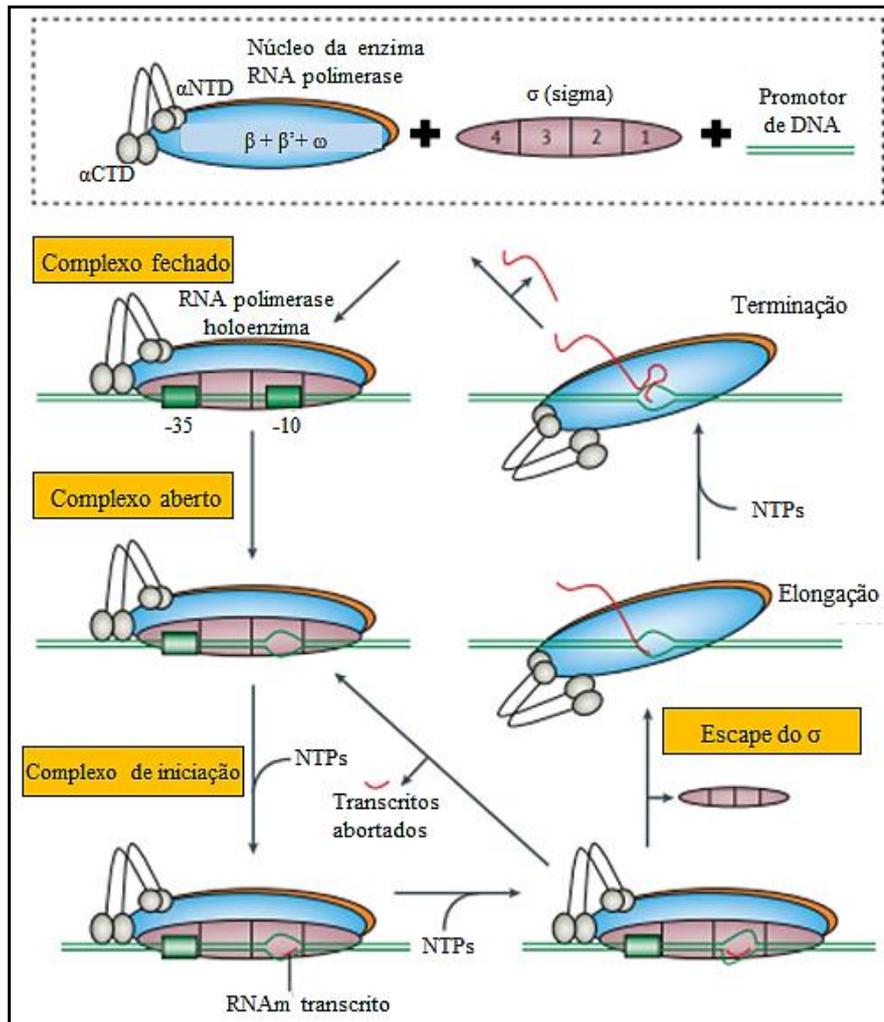
O processo de transcrição gênica ocorre primeiramente quando a RNAP reconhece e se liga a elementos presentes na dupla fita de DNA (interação eletrostática) (FEKLÍSTOV et al., 2014). Logo após, ocorre a formação de um complexo mais estável de reconhecimento formado pelo núcleo da enzima RNAP com a região promotora ou sequência conhecida como promotor (GEZVAIN & LANDICK, 2004). Em bactérias, a RNAP é composta por um núcleo de múltiplas subunidades que é fortemente associado a um fator de especificidade  $\sigma$ . Enquanto o núcleo contém o sítio ativo da polimerase, sendo capaz de sintetizar RNA, o fator sigma especifica quando e onde deve ocorrer o início da transcrição (GRUBER & GROSS, 2003; HOOK-BARNARD et al., 2007). O núcleo da enzima RNAP interagindo com o fator sigma ( $\sigma$ ) forma a holoenzima. Somente a holoenzima é capaz de reconhecer a sequência promotora e ocasionar as mudanças conformacionais no DNA necessárias para separar as duas fitas de DNA para formar uma bolha de transcrição e iniciar de fato iniciar o ciclo da transcrição gênica (LEE et al., 2012).

Em todos os organismos, a transcrição por RNAP dependente de DNA pode ser dividida em quatro etapas mecanicamente e estruturalmente distintas: reconhecimento da região promotora no DNA, iniciação, alongamento e terminação (BORUKHOV & NUDLER, 2003; GEZVAIN & LANDICK, 2004; CAMPBELL et al., 2008; RUFF et al., 2015).

A ligação específica da RNAP holoenzima ao promotor forma um “complexo fechado” (RPc) e põe em movimento uma série de grandes mudanças conformacionais em ambas as biomoléculas. Tais mudanças colocam o duplex de DNA no local do sítio ativo da RNAP, que em seguida, abre o DNA para formar uma bolha de transcrição para a “formação do complexo aberto” (RPo) (SAECKER et al., 2011; RUFF et al., 2015). Isto ocorre devido a uma isomerização, pois supõe que deve resultar em um movimento da cadeia de DNA livre da bolha ao sítio ativo da enzima RNAP de modo que a síntese de RNA possa começar (GEZVAIN & LANDICK, 2004; BROWNING & BUSBY, 2004, 2016).

O complexo aberto é capaz de iniciar a transcrição, mas na maioria dos casos, o promotor mantém-se em um complexo de transcrição inicial (ITC) que sofre rodadas curtas de formação de transcrito e com isso não inicia a transcrição. Após a RNAP deixar o promotor, é formado um complexo de alongamento da transcrição (TEC), onde a subunidade  $\sigma$  se dissocia e o complexo segue de maneira processiva e estável (GEZVAIN & LANDICK, 2004). O fator  $\sigma$  não é necessário para a fase de alongação e é tipicamente liberado do núcleo da RNAP, e apenas esta prossegue polimerizando o RNA (SAECKER et al., 2011).

Ao alcançar o final da região a ser transcrita, sinais presentes na sequência do DNA ou do RNA levam ao término da transcrição, fazendo com que o complexo DNA–RNAP e RNA sejam desfeitos. Tais sequências são chamadas de terminadores de transcrição e são reconhecidas pela RNAP. Este processo de terminação pode ocorrer pela ação de fatores proteicos ou pelo simples desequilíbrio das forças que mantêm o híbrido DNA-RNA juntos (Figura 1) (ROBERTS et al., 2008; VAN HIJUM et al., 2009).



**Figura 1: Transcrição de genes bacterianos.** O ciclo de transcrição bacteriano. A RNAP holoenzima, que compreende o núcleo da enzima e um fator sigma, interage com o promotor de DNA para formar o complexo fechado (RPC). Após o processo de reconhecimento do promotor ocorre uma isomerização em dois ou mais passos para formação do complexo fechado (RPC) para o complexo aberto (RPO), desenrolando a dupla fita de DNA na região do local de início da transcrição e permitindo que encontre o sítio ativo da holoenzima. A adição de nucleosídeos trifosfatos (NTPs) permite uma transição adicional para o complexo de iniciação que sintetiza o transcrito do RNA mensageiro (RNAm). O processo pode ser mantido no promotor, o que resulta em ciclos de iniciação abortiva em que são produzidos apenas pequenos fragmentos de RNA. Alternativamente, a RNAP pode escapar do promotor para entrar na fase de elongação, levando à liberação do fator sigma e ao alongamento do transcrito de RNA usando NTPs e fatores de transcrição (TFs) (não mostrados). A transcrição prossegue até que a RNAP encontre um terminador de transcrição (TT), fazendo com que o transcrito de RNA seja liberado e a RNAP possa dissociar-se do molde de DNA para voltar a se envolver com um fator sigma para uma nova rodada de transcrição (Fonte: Adaptado de BROWNING & BUSBY, 2016).

## 2.2 A RNA POLIMERASE EM PROCARIOTOS

Os procariotos utilizam dois conjuntos distintos de mecanismos para a ativação do início da transcrição, e estes estão focados no promotor ou na RNAP (LEE et al., 2012). A

RNAP bacteriana é composta por pelo menos cinco subunidades constantes e pode ser encontrada em duas formas: a) a apoenzima ou núcleo da enzima que possui todas as subunidades exceto a subunidade  $\sigma$  (ou fator  $\sigma$ ); e b) a holoenzima que possui as subunidades  $\alpha$ ,  $\beta$ ,  $\beta'$ ,  $\omega$ , e  $\sigma$ , capaz de reconhecer promotores gênicos específicos e iniciar a síntese de RNA (BROWNING & BUSBY, 2004; CASES & DE LORENZO, 2005; FEKLISTOV, 2013).

A RNAP da bactéria Gram-negativa *E. coli* é uma das mais estudadas e a mais bem caracterizada dentre os procariotos. As duas subunidades alfa ( $\alpha$ ) possuem 40 kDa cada. A subunidade beta ( $\beta$ ) possui em torno de 155 kDa e a beta linha ( $\beta'$ ) possui 160 kDa (BURGESS, 1969; ISHIHAMA, 2000). A subunidade ômega ( $\omega$ ) possui 10 kDa e os fatores sigma ( $\sigma$ ) possuem massa molecular variando de 25 a 92 kDa. A subunidade  $\alpha$  é formada por dois domínios de aminoácidos independentes e unida por um ligante flexível de 20 aminoácidos: o domínio N-terminal (9 kDa) com resíduos de 1 a 235 aminoácidos e o C-terminal (26 kDa) composto por resíduos de 250 a 329 aminoácidos (MURAKAMI et al., 2002; MURAKAMI & DARST, 2003). As subunidades  $\beta$  e  $\beta'$  possuem 1,342 e 1,407 aminoácidos, respectivamente e a subunidade  $\omega$  (11 kDa) apresenta 91 resíduos de aminoácidos (SAECKER et al., 2011; RUFF et al., 2015).

Cada subunidade tem um papel específico na ligação ao DNA, na interação entre as subunidades da enzima e no processo catalítico. A parte N-terminal de  $\alpha$  está envolvida na formação de dímeros e da ligação de  $\beta$  com  $\beta'$ , e o seu C-terminal está preso por meio de um ligante flexível ao seu N-terminal, é e responsável pela interação com o “elemento UP” presente em alguns promotores que funciona como ativador transcricional (BROWNING & BUSBY, 2004; GEZVAIN & LANDICK, 2004). A subunidade  $\beta$  contém o sítio catalítico ativo da polimerase e juntamente com  $\beta'$  especifica a interação do  $\sigma$  (sigma) com o promotor e a fita de DNA, e também liga os ribonucleotídeos que irão ser polimerizados em RNA (BROWNING & BUSBY, 2004; GEZVAIN & LANDICK, 2004). O papel da subunidade  $\omega$  (ômega) atua como uma “chaperona molecular” para o correto enovelamento da subunidade  $\beta'$  e auxilia na montagem estável da RNAP (SCHUMANN & FERREIRA, 2004).

### 2.3 FATOR SIGMA ( $\sigma$ )

Como mencionado anteriormente, o fator  $\sigma$  está envolvido no reconhecimento dos promotores, sendo que cada um reconhece um conjunto distinto de genes, sendo que

existem, pelos menos, seis tipos diferentes do fator  $\sigma$  em *E. coli*. Cada um desses fatores regula um subconjunto de genes em resposta a diferentes estímulos, como carência nutricional, aumento de temperatura, fase estacionária, estresse oxidativo, estresse osmótico, programa de desenvolvimento, etc (tabela 1) (CHANDRANGSU & HELMANN, 2001). Assim, torna-se possível a um organismo apresentar uma resposta adequada a esses estímulos ambientais e sobreviver a certas condições adversas utilizando a mesma enzima (RNAP) e quando necessário ajustando somente o fator  $\sigma$  (GRUBER & GROSS, 2003; BROWNING & BUSBY, 2004; SAECKER et al., 2011; TRIPATHI et al., 2014).

**Tabela 1:** Fatores sigma ( $\sigma$ ) de *Escherichia coli*.

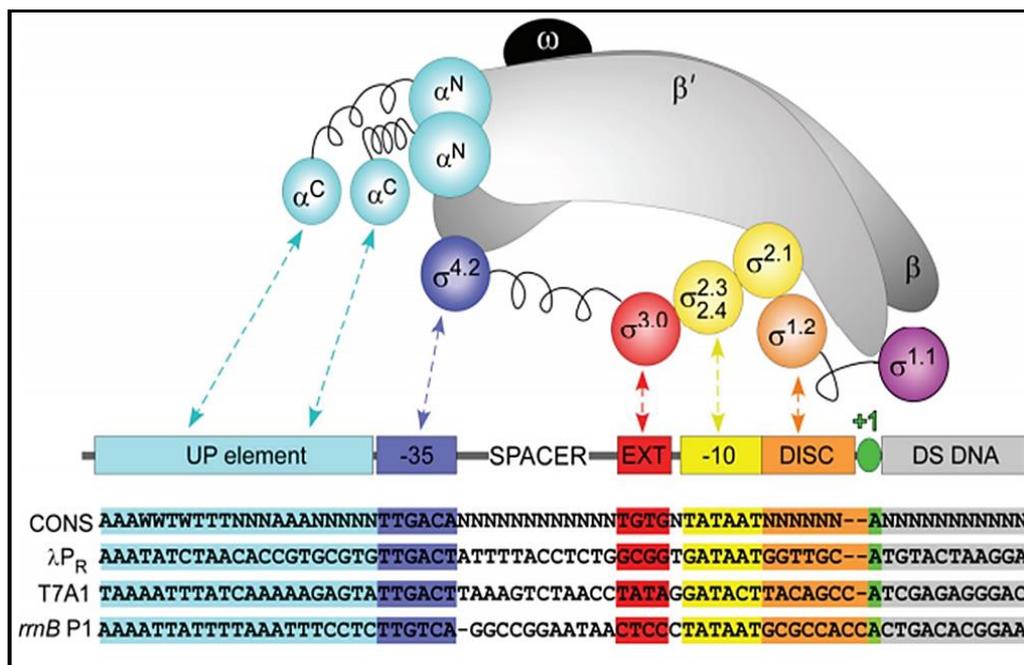
<b>Fator sigma (<math>\sigma</math>)</b>	<b>Gene</b>	<b>Função/Resposta</b>	<b>Seq. Consenso -35 Espaçador -10</b>
<i>Fecl</i>	Fecl	Captação de citrato férrico (ECF)	GAAAAT - 15 bp - TGCCT
24	rpoE	Estresse térmico extra-citoplasmático (ECF)	GGAAGTT- 15 pb- GTCTAA
28	FliA ou rpoF	Mobilidade celular e patogenicidade ou síntese de flagelo	CTAAA-15pb- GCCGATAA
32	rpoH	Choque térmico extremo; renaturação de proteínas	CCCTTGAA -13-15pb - CCCGATNT
38	rpoS	Resposta a estresse e fase estacionária	TTGACA- 16-18 pb- TATACT
54	rpoN	Assimilação ou limitação de nitrogênio (N <sub>2</sub> )	CTGGNA -6pb- TTGCA
70	rpoD	Manutenção basal; crescimento celular, sigma constitutivo	TTGACA - 16-18pb- TATAAT

Fonte: Adaptado de HELMANN, 2001.

Em contraste, todas as espécies de bactérias possuem pelo menos um fator  $\sigma$  primário. Em *E. coli*, o sigma primário é o 70 ( $\sigma$ 70) (codificado pelo gene *rpoD*), que ativa a transcrição principalmente dos genes constitutivos (*housekeeping*) ou transcrição dos genes de manutenção basal da bactéria. Este fator  $\sigma$ 70 é o principal, sendo o fator sigma mais abundante na célula (MERRICK, 1993; HELMANN, 2001). Em *E. coli*, o  $\sigma$ 70 tem uma

maior afinidade para o núcleo da RNAP do que qualquer um dos seus fatores  $\sigma$  alternativos que são utilizados sob certas condições de crescimento ou estresse. Outros fatores pertencem à família sigma 54 ( $\sigma^{54}$ ), responsável pela expressão de genes envolvidos na utilização de nitrogênio (WIGNESHWERARAJ et al., 2008; SACHDEVA et al., 2009; GHOSH et al., 2010; SACHDEVA et al., 2010; LEE et al., 2012).

Como parte da holoenzima funcional,  $\sigma^{70}$  sigma reconhece um promotor constituído por dois elementos bem conservados na sequência, os hexâmeros -35 e -10, que são separados por uma região espaçadora de 16-19 pb (RUFF et al., 2015). O fator  $\sigma^{70}$  pode ser dividido em quatro domínios proteicos peculiares unidos por ligantes flexíveis (regiões 1, 2, 3 e 4) e subdomínios (2.1, 2.2, 2.3 etc) com base na função, estrutura e conservação da sequência (HOOK-BARNARD & HINTON, 2007). Cada domínio de  $\sigma^{70}$  interage com ambos: RNAP e DNA. Os domínios 2 e 4 fazem contatos específicos com as regiões -10 e -35 respectivamente, ao passo que o domínio 3 interage com a região -10 estendida presente em alguns promotores (CAMPBELL et al., 2008). Embora os domínios 2, 3 e 4 são conhecidos por estarem envolvidos no reconhecimento do promotor, a função do domínio 1 não é bem compreendida, está ausente em muitos fatores  $\sigma$  e supõe-se que esteja envolvido na regulação do complexo aberto e no início da transcrição na sequência de DNA do promotor (Figura 2) (MURAKAMI & DARST, 2003; RODRIGUE et al., 2006; SACHDEVA et al., 2010).



**Figura 2: Interações específicas das sequências entre a RNA polimerase, sigma 70 e regiões do promotor.** Representações esquemáticas das subunidades do núcleo da RNA polimerase,  $\sigma$ 70 e o DNA do promotor. RNAP:  $\alpha$ 2: azul-claro;  $\beta$  e  $\beta'$ : cinza;  $\omega$ : preto; sub-regiões do  $\sigma$ : estão mostradas. Promotor: Elemento UP: azul claro; região -35: roxo; elemento -10 estendido (EXT): vermelho; região -10: amarelo; discriminador (DISC): laranja; sítio de início da transcrição ou +1: verde. DS DNA: DNA de dupla fita a jusante do +1: cinza. Os conectores que ligam as subunidades  $\alpha$ C e  $\alpha$ N e subdomínios do  $\sigma$  são mostrados como molas. Parte das regiões de um promotor consenso (CONS) é mostrado e os promotores P<sub>R</sub>, T7A1 e o P<sub>1</sub> do *rrnB* são mostrados abaixo. As bases ausentes são mostradas por traços (Fonte: RUFF et al., 2015).

## 2.4 PROMOTORES PROCARIÓTICOS E SEUS ELEMENTOS

Os promotores procarióticos reconhecidos apresentam tamanho aproximado de 60~80 pares de bases (pb). Está localizada a montante do ponto de início de transcrição (+1) e existe um dado promotor para um ou seletos conjunto de genes. Estes promotores possuem regiões específicas que atuam como sítios de reconhecimento e ancoragem capazes de recrutar a RNAP holoenzima e iniciar a síntese de RNA no sentido 5' para 3' (LI et al., 2015). Desta maneira, controlam os níveis celulares das proteínas codificadas por estes genes através de sua taxa de transcrição. Entretanto, embora cada conjunto de gene(s) possua um promotor, nem todos são idênticos entre si, pois nas células bacterianas a RNAP é confrontada com um conjunto com cerca de pelo menos 2.000 sequências promotoras diferentes para cada 3.000 genes (BROWNING & BUSBY, 2004).

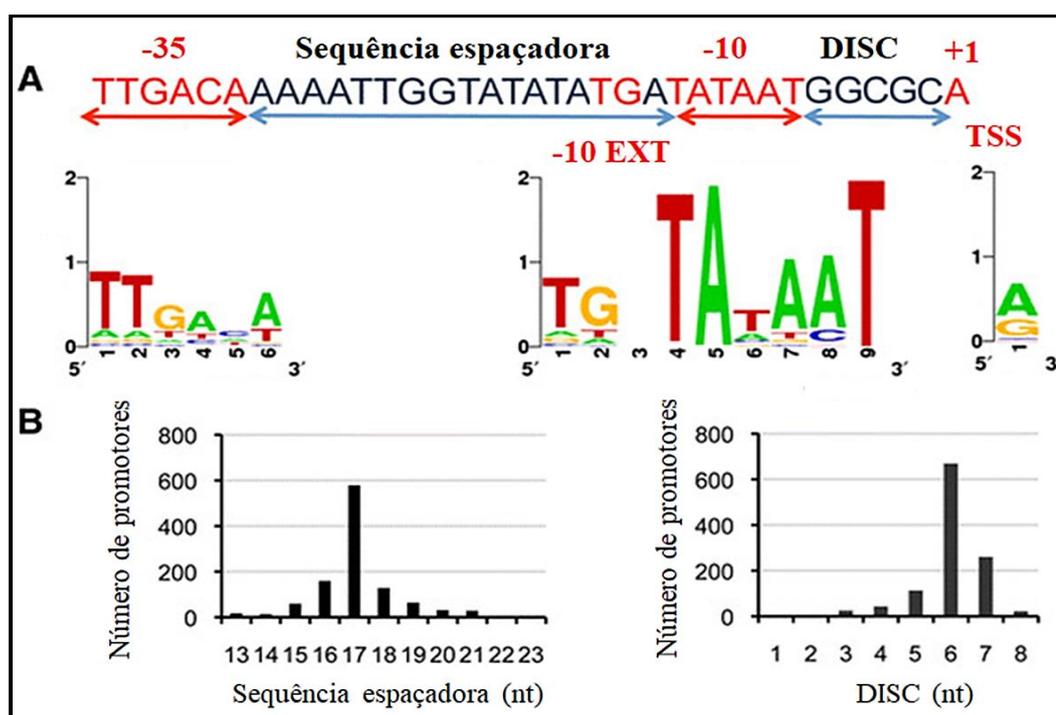
O passo principal na iniciação da transcrição é o reconhecimento do promotor pela RNAP holoenzima. Os estudos cristalográficos e estudos de mobilidade eletroforética

(*footprints*, EMSAs - *Electrophoretic Mobility Shift Assays* e mais recente Microscopia Crioeletrônica), diversos trabalhos de codificação e análise de promotores de procariotos *in silico* através de programas de computador utilizando algoritmos levaram a geração de modelos que explicam como determinadas regiões da sequência de DNA, que compõem o promotor, são reconhecidos por ambos: subunidade  $\sigma$  e RNAP (SAECKER et al., 2021). Foram observadas que em certas regiões dos promotores haviam sequências conservadas e por isso reforçam um conceito existente de uma “sequência promotora consenso” que apresenta identidade entre algumas bases dos promotores nas mesmas posições da sequência e que tais são altamente conservadas entre elas ((HAWLEY & MCCLURE, 1983; HARLEY & REYNOLDS, 1987; HIGUCHI et al., 1988; BEREZHNOY & SHCKORBATOV, 2005; BURDEN et al., 2005; HELLMAN & FRIED, 2007; HOOK-BARNARD & HINTON, 2007; RUFF et al., 2015).

Desde essas análises iniciais considerava-se que, por serem conservadas, essas regiões deveriam ter papéis relevantes no início do processo de transcrição (HARLEY & REYNOLDS, 1987; BEREZHNOY & SHCKORBATOV, 2005). Com a evolução dos procedimentos de sequenciamento de DNA, houve a descoberta de muitas sequências promotoras dentro do genoma de *E. coli* reconhecidas pela holoenzima (RNAP $\sigma$ 70) contendo sequências similares de DNA e que foi descoberta a importância funcional destas sequências em muitas investigações (BROWNING & BUSBY, 2004). Dado que se estabeleceram duas regiões nos promotores que formam o núcleo do representado pelos hexâmeros -35 (<sup>-35</sup> TTGACA <sup>-30</sup>) e -10 (<sup>-12</sup> TATAAT <sup>-7</sup>) e geralmente separado entre si por 17 pb, sendo a transcrição iniciada em uma purina (sítio de início da transcrição ou +1) localizada aproximadamente a 7 pb após a região -10, bem como outras regiões definidas denominadas de: o elemento UP, “elemento -10 estendido e o discriminador (BROWNING & BUSBY, 2004). Essas sequências tornaram-se conhecidas como regiões de identidade sendo localizadas nos procariotos a aproximadamente próximos do núcleo do promotor a montante do início da transcrição (+1) (PRIBNOW, 1975; HAWLEY & MCCLURE, 1983; BEREZHNOY & SHCKORBATOV, 2005; FEKLISTOV, 2013; RUFF et al., 2015).

Em média, promotores de bactérias contêm entre oito e nove bases iguais às do consenso, sendo raro ocorrer promotores com todas as sequências iguais às sequências consenso (nas seis posições da região -35 e -10) (Figura 3) (BROWNING & BUSBY, 2004). Ou seja, foi elucidado que existem variações nas sequências nucleotídicas de

promotores e que estas variações afetam a ligação da RNAP aos promotores, e também que existem regiões mais importantes do que outras para o funcionamento do promotor, como a região -10, que onde é altamente conservada e essencial para o reconhecimento e abertura das fitas pela holoenzima, independente da região -35 estar presente (HAWLEY & MCCLURE, 1983; BROWNING & BUSBY, 2004; BEREZHNOY & SHCKORBATOV, 2005; FEKLISTOV, 2013).



**Figura 3: Regiões consensuais dos promotores de *E. coli*.** Na parte A da figura e destacado em vermelho, mostra os nucleotídeos preferenciais das regiões -35 e -10, no elemento -10 estendido TG (-10 EXT) e no sítio inicial da transcrição ou +1 de promotores de *E. coli*. Na parte B mostra o número de promotores existentes com a sequência espaçadora e sequência discriminadora preferenciais em número de nucleotídeos (Fonte: modificado de ROSINSKI-CHUPIN et al., 2015).

Sabe-se que as sequências promotoras que apresentam elementos mais parecidos com as da sequência consenso funcionam mais eficientemente, tendem a ser mais ativos e possuem uma maior atividade de ligação com a RNAP do que promotores cujas sequências nucleotídicas difere da sequência consenso preferencial de  $\sigma 70$  (5' TTGACA – 17 pb – TATAAT 3') (SCHUMANN & FERREIRA, 2004). Isso implica que, promotores com maior afinidade pela RNAP, resultam em uma maior frequência na transcrição dos genes. Ao contrário, promotores com baixa afinidade pela RNAP ocasionam em uma menor transcrição de genes (MURAKAMI & DARST, 2003b; SAECKER et al., 2011).

### 2.4.1 Regiões -35 e -10

As regiões de hexâmeros -35 e -10 desempenham um papel muito importante na seletividade e na frequência da iniciação da transcrição, contudo formam o núcleo do promotor. A região -35 possui a sequência de consenso 5' - TTGACA - 3' com o -35 = T; -34 = T; e -33 = G sendo estas bases as mais altamente conservadas. Estas bases interagem com o subdomínio  $\sigma$ 4.2 em promotores e podem recrutar ativadores como  $\lambda$ CI e CAP influenciando em diversas etapas da transcrição. Também pode interagir com o núcleo da RNAP para mediar o recrutamento de fatores de alongamento (RUFF et al., 2015).

A região -10 apresenta uma região consenso 5' - TATAAT - 3' onde os pb desta região geralmente são ricos em A e T. As bases -11 = A e -7 = T são as bases mais altamente conservadas do hexâmero. O hexâmero ou região -10 desempenha um papel fundamental de modo a formar o complexo aberto (abertura das fitas), enquanto a região ou hexâmetro -35 serve como um alvo para ativadores de transcrição (HOOK-BARNARD & HINTON, 2007; RUFF et al., 2015).

Muitas mutações ocorrem dentro das regiões -35 e -10, em geral, muitas dessas mutações podem fugir da região consensual e conseqüentemente podem diminuir ou aumentar a frequência de iniciação da transcrição do(s) gene(s) (CAMPBELL et al., 2008; RUFF et al., 2015; HEYDUK & HEYDUK, 2018). Um promotor historicamente importante, caso à parte, é o promotor do operon *lac* de Jacob e Monod de 1961 (JACOB & MONOD, 1961). O promotor *lac* não é tão semelhante à sequência de nucleotídeos presente na região -35 e -10 conhecida como consenso. Possui uma base na região -35 e duas na região -10 que difere da sequência consenso e a região espaçadora entre estas duas regiões é de 18 pb (CZARNIECKI et al., 1997).

Entretanto, considerando que promotores que apresentam as sequências mais próximas às do consenso são mais fortes, construíram outro promotor, o promotor *LacUV5* a partir do promotor *lac*. O promotor *LacUV5* difere em apenas uma base da região -35 consenso, e apresenta a região -10 idêntica à da região consenso. Com essa mudança, a atividade promotora aumentou cerca de 2,5 vezes (Tabela 2) (FULLER, 1982; SCHUMANN & FERREIRA, 2004).

Quadro 1: Algumas sequências nucleotídicas de promotores utilizados em vetores de expressão.

Promotor	Região		
	Região -35	espaçadora	Região -10
$P_{Lac}$	TTtACA	18 pb	TATgtT
$P_{LacUV5}$	TTtACA	18 pb	TATAAT
$P_{Trp}$	TTGACA	17 pb	TtaAcT
$P_{Tac}$	TTGACA	16 pb	TATAAT
$\lambda P_L$	TTGACA	17 pb	gATAcT
$\lambda P_R$	TTGACT	17 pb	gATAAT
<b>Consenso</b>	<b>TTGACA</b>	<b>17 pb</b>	<b>TATAAT</b>

Os nucleotídeos presentes na sequência consenso são demonstrados em negrito. Nas sequências dos promotores, os nucleotídeos idênticos à sequência consenso são mostrados em letras maiúsculas, e os que não são idênticos, são mostrados em nucleotídeos minúsculos (SCHUMANN & FERREIRA, 2004).

#### 2.4.2 A região espaçadora

A região espaçadora é um segmento de DNA que não apresenta nenhuma sequência de consenso de nucleotídeos e está localizada entre as regiões -35 e -10. Mas existe um comprimento de consenso, que é o espaçamento entre os subdomínios  $\sigma_{4,2}$  e  $\sigma_{2,3}$ . Sendo que, há autores que afirmam que esta região é a mais importante que a sequência como um todo para a determinação do promotor, isso porque esse espaço seria o mais preferencial do fator sigma 70 (MITCHELL et al., 2003; KIRYU et al., 2005; SHIMADA et al., 2014).

Foi observado em promotores fortes que o espaçamento de 17 nucleotídeos era preferencial entre os hexâmeros consensuais -35 e -10. Posteriormente, foi confirmado por cristalografia de Raios-X que o espaçamento de 17 nucleotídeos era o ideal para o acoplamento da subunidade sigma 70 e mutações que alterem o tamanho do espaçador de 17 pb existentes entre as sequências -35 e -10 também modificam o nível da transcrição dirigida pelo promotor mutado (BROWNING & BUSBY, 2004; SCHUMANN & FERREIRA, 2004; KIRYU et al., 2005; RUFF et al., 2015).

#### 2.4.3 Sítio de início da transcrição ou +1

Os “sítios de início de transcrição” (*Transcription Start Site* - TSS) é o nucleotídeo onde a transcrição se inicia e geralmente é numerado de +1. Quando a RNA polimerase está

posicionada na região promotora, ela começa a polimerizar os ribonucleotídeos da fita de RNA antiparalela nesta posição, ou seja, o início de gene dá se nesta posição (RANGEL-CHAVEZ et al., 2017).

A base do sítio de início da transcrição preferencial na posição +1 é Adenina (43,88%), seguido por Guanina (24,31%), Uracila (18,03%) e Citosina (13,76%) (RUFF et al., 2015). Esta distribuição foi obtida a partir de uma população de 8.500 iniciações de transcrição mapeadas. Não é bem entendido como o nucleotídeo é escolhido para iniciar a transcrição e parece ser meramente baseado na existência de uma distância adequada do promotor para onde a RNA polimerase se liga para iniciar a transcrição (RANGEL-CHAVEZ et al., 2017).

#### 2.4.4 O elemento UP

Uma lista inicial de sequências promotoras dependentes de  $\sigma 70$  revelou uma sequência rica em bases AT a montante do elemento -35 que apresentava identidade e verificou que as sequências dentro desta região aumentavam significativamente a transcrição de alguns promotores e, portanto atribuiu-se à função *enhancer* transcricional para esta sequência (ROSS et al., 1993; RAO et al., 1994; ESTREM et al., 1998; GOURSE et al., 2000; ROSS & GOURSE, 2005; RHODIUS et al., 2012; LEE & BORUKHOV, 2016).

Trabalhos subsequentes elucidaram dois subsítios de sequências ricas em A/T, um elemento UP proximal (-45 a -37) e um elemento UP distal (-58 a -45). Cada um dos dois elementos pode ser reconhecido por uma das duas subunidades  $\alpha$  presentes na RNAP, e o contato dos elementos proximal, distal ou ambos pode afetar a atividade do promotor (HOOK-BARNARD & HINTON, 2007). Embora uma sequência específica seja preferida para a ligação, as subunidades  $\alpha$  também podem interagir de forma não específica com esta região do DNA (se houver um período prolongado de interação entre ambos) e que pode levar a tempos necessários para formar complexos abertos entre RNAP $\alpha$ CTDs e elemento UP nesses promotores gerando sequências funcionais para a transcrição. Os promotores podem ter um ou ambos, embora o elemento UP distal seja mais funcional sozinho, uma sequência completa do elemento UP completo é mais eficiente (RUFF et al., 2015).

Os elementos UP foram notados pela primeira vez por sua capacidade de aumentar a transcrição dos promotores de rRNA, e usando o promotor ribossomal *rrnB* P1, Ross e colaboradores demonstraram que a região de -60 a -38 do ponto de início da transcrição

(+1), denominada elemento UP, realmente fazia contato com o domínio C-terminal das subunidades  $\alpha$  da RNAP e uma sequência de elemento UP de consenso (5' NNNAAAWWTWTTTTNNNAAANN3' (W = A ou T e N = qualquer base) foi determinada usando análise mutacional aumentando a transcrição 330 vezes *in vivo* (ROSS et al., 1993; GAAL et al., 1994; RAO et al., 1994; ESTREM et al., 1998, 1999; GOURSE et al., 2000; ; ROSS & GOURSE, 2005; HOOK-BARNARD & HINTON, 2007; RHODIUS et al., 2012).

#### 2.4.5 Elemento -10 estendido

Uma compilação inicial e análise de promotores indicaram que havia alguma preferência pela sequência TG a montante de -10 (HAWLEY & MCCLURE, 1983). Dados genéticos e bioquímicos também sugeriram a importância do contato entre RNAP nesta região do DNA. Através destes estudos verificaram uma sequência consenso reconhecido por  $\sigma^{70}$  como motivo -15 TGn -13, também chamado de sequência estendida -10 (BERMAN & LANDY, 1979; SIEBENLIST & GILBERT, 1980; BUSBY et al., 1984; GRANA et al., 1985; PONNAMBALAM et al., 1986).

Investigações sobre a importância do motivo -15 TGn -13 mostraram que esta sequência existe quando há a ausência de elemento -35 ou -10 ou que diferente de suas sequências consensos (5' TTGACA - 17pb - TATAAT 3') ou em espaçadores mais longos (18 pb ou mais). Promotores que se desviam significativamente do consenso nesses elementos são mais propensos a ter um motivo TGn (HOOK-BARNARD & HINTON, 2007).

Muitos promotores de *E. coli* foram analisados e identificados por sequenciamento de DNA, e os resultados revelaram que o motivo TGn da posição -15 a posição -13 está presente em cerca de 20% dos promotores de *E. coli*. A atividade ideal requer a presença de Timina na posição -15 e Guanina na posição -14, embora dependendo do promotor, a presença de apenas um desses nucleotídeos pode ser vantajoso para a atividade do promotor. O elemento -10 estendido foi determinado como um motivo de 3-6 nucleotídeos (5' -17 TRTGn -13, onde R = A ou G; N = qualquer nucleotídeo 3') localizado logo a montante da região -10 (5' TATAAT 3') e que pode afetar o nível de transcrição através de contatos específicos com domínio  $\sigma^3$  (HOOK-BARNARD; HINTON, 2007).

#### 2.4.6 Região discriminadora

Muitas hipóteses das taxas da transcrição abortiva e de iniciação pela RNAP ( $\alpha 2\beta\beta'\omega\sigma 70$ ) de *E. coli* foram discutidos e elucidadas por microscopia crio eletrônica nas sequências nucleotídicas posicionadas entre as regiões - 10 e o sítio de início da transcrição de promotores (PEMBERTON et al., 2000; HAUGEN et al., 2006, 2008; RUFF et al., 2015; HENDERSON et al., 2017). A principal função desta sequência discriminadora é na regulação do tempo de vida do complexo aberto (RPo) do DNA pela RNA polimerase. Estudos cristalográficos do sítio ativo da RNAP mostraram que os nucleotídeos presentes à jusante da região -10, juntamente com o sítio de início da transcrição (+1) são reconhecidos pelo  $\sigma 70$  subdomínio  $\sigma 1.2$  (HAUGEN et al., 2006; BARINOVA et al., 2008; RUFF et al., 2015).

Análises de mutagêneses e de bioinformática de promotores nessa região, mostraram que o tempo de vida diminui à medida que o comprimento do discriminador aumenta de 6 para 8 bases. Estas observações indicam que muitos promotores de *E. coli* formam complexos abertos de longa duração e que os locais de início da transcrição *in vivo* são de grande importância (RUFF et al., 2015).

O discriminador é uma região com cerca de 5 a 8 pb e extremamente rica em bases nucleotídicas G e C. Possui 5' GGG 3' como sequência consenso nas posições -7, -6 e -5, pois são muito importantes para o aumento e estabilidade do complexo aberto e fazem uma interação mais forte com  $\sigma 1.2$  (FEKLISTOV et al., 2006; BARINOVA et al., 2008; LEE & BORUKHOV, 2016). Abordagens bioquímicas mapearam esses grandes efeitos na estabilidade de RPo para interações específicas com a região conservada  $\sigma 1.2$ . Discriminadores de comprimentos de DNA menores em ITCs são propostos como instáveis e a troca de nucleotídeos GC para AT podem implicar em desestabilidade dos complexos (RPc e RPo) levando a curtos tempos de vida. O impacto do conteúdo de G e C em relação a interação do discriminador com o subdomínio  $\sigma 1.2$  pode variar em promotores e diferentes purinas podem ser favorecidas para uma interação melhor com  $\sigma 1.2$  (BARINOVA et al., 2008; RUFF et al., 2015; HENDERSON et al., 2017).

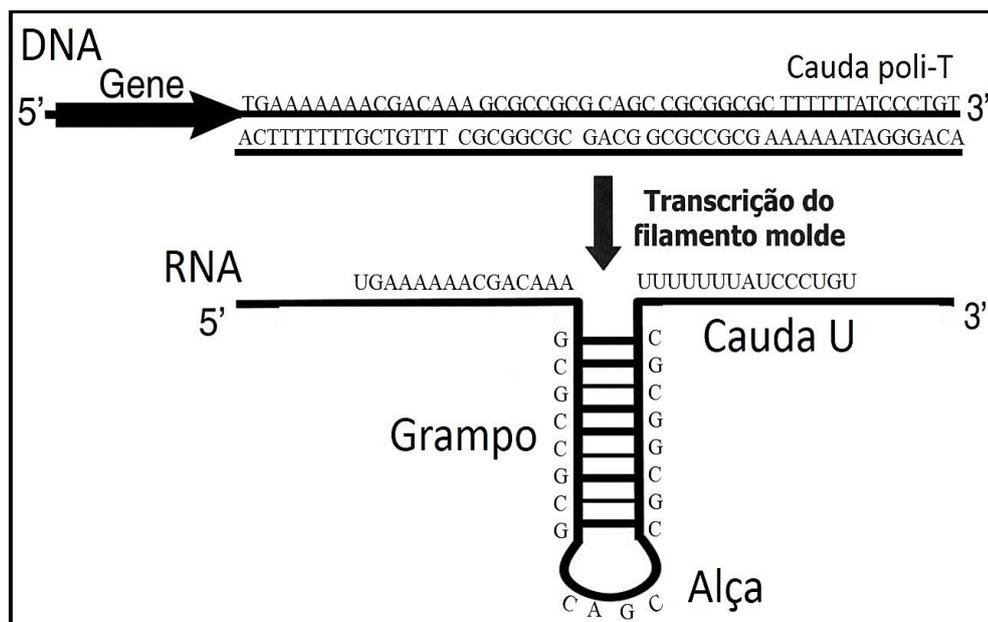
## 2.5 TERMINADORES DE TRANSCRIÇÃO

Além das fases de iniciação e do alongamento, a terminação da transcrição representa outro fator essencial para a transcrição do DNA, e, portanto, é outra fase regulatória muito importante na etapa da expressão gênica. O sistema de paradigma no qual

a antiterminação da transcrição foi descrita pela primeira vez foi o sistema mediado por bacteriófago  $\lambda$ -N em *E. coli*. Estudos subsequentes revelaram um conjunto crescente de sistemas de controle da terminação de transcrição, utilizando uma variedade de mecanismos para controlar a formação de sinais de terminação no RNA nascente, ou a capacidade da RNAP de reconhecer esses sinais. A investigação desses mecanismos reguladores forneceu informações valiosas sobre a operação da maquinaria de transcrição em procariontes (HENKIN, 1996, 2000; CAMBRAY et al., 2013).

As bactérias usam dois mecanismos alternativos para encerrar a transcrição: a terminação dependente do fator de transcrição ou proteína *Rho*, que depende da interação com a RNAP e ocorre em locais que não podem ser facilmente caracterizados por sua estrutura ou sequência e a terminação independente de *Rho* ou terminação intrínseca que envolve sinais de sequência e estrutura que podem ser identificados computacionalmente. Ambos os terminadores, dependentes de *Rho* e intrínsecos são afetados por sistemas de controle de terminação, embora alguns sistemas de término sejam específicos do tipo de terminador (HENKIN, 1996; GALE et al., 2021). No entanto, as características da sequência exata e as etapas de recrutamento e função de *Rho* não são tão bem compreendidas o suficiente para uso em sistemas genéticos sintéticos (HENKIN, 1996; CAMBRAY et al., 2013).

As características dos terminadores independente do fator *Rho* ou terminadores intrínseco foram bem estudadas em *E. coli* e é responsável pelos 80% dos eventos de terminação de transcrição neste organismo (CAMBRAY et al., 2013). Os terminadores intrínsecos são funcionais na ausência de fatores adicionais em reações de transcrição realizadas *in vitro*, portanto podem terminar a transcrição sem a ajuda de fatores auxiliares, consistem em um grampo (*hairpin*) de comprimento 6-12 pb fechado por uma alça curta de 3-5 nucleotídeos, rico em bases G e C e seguido a jusante por um trecho de 7 a 12 nucleotídeos de U na transcrição de RNA, que corresponde no DNA a um palíndromo invertido seguido por uma cauda poli-T (figura 4) (HENKIN, 1996; CAMBRAY et al., 2013; GALE et al., 2021).



**Figura 4: Esquema de um terminador de transcrição intrínseco.** Os terminadores no DNA intrínsecos consistem em uma sequência rica em nucleotídeos G e C seguido por uma região rica em Ts (cauda poli-T) em seu lado 3' e que são transcritos em um grampo com uma pequena alça com uracilas no RNA a montante (cauda U) (adaptado de KINGSFORD et al., 2007).

Como a terminação intrínseca depende das afinidades de ligação diferencial entre os nucleotídeos do DNA e RNA, a RNAP reconhece a sequência terminadora e realiza uma pausa na transcrição para a formação da cauda poli-U e do grampo com a alça curta. A interação entre as bases U do mRNA e A do DNA é fraca resultando na desestabilização do mRNA e deslocando-o para fora do sítio ativo da RNAP. A presença da alça em grampo (rico em G e C) no sítio ativo da RNAP interrompe a ligação mRNA-DNA devido ao alongamento mais lento existente da cauda rica em U que permite a o dobramento da estrutura do grampo desestabilizando o complexo de transcrição (LI et al., 2015; GALE et al., 2021). O dobramento do grampo é um passo fundamental no processo de terminação, e experimentos moleculares mostraram que se não existir a estrutura do dobramento em grampo, permite que a RNAP continue realizando a transcrição através do terminador (LI et al., 2015).

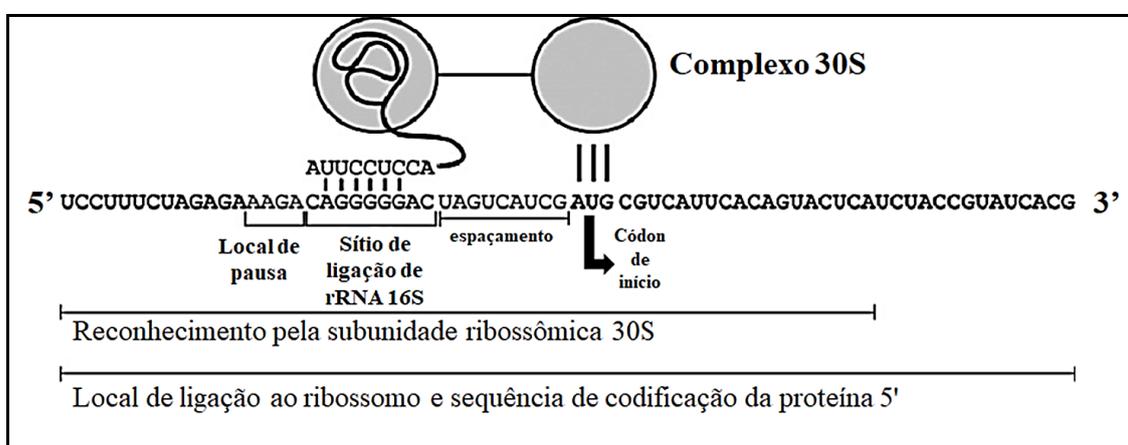
## 2.6 SEQUÊNCIA SHINE-DALGARNO

Tradução é o processo de síntese de proteínas, e ocorre em uma estrutura de várias subunidades chamado de ribossomo. Em procariotos, a transcrição e a tradução ocorrem praticamente juntas, ou seja, os ribossomos iniciam a tradução no RNA nascente

simultaneamente com a transcrição (KACZANOWSKA & RYDÉN-AULIN, 2007; LI et al., 2015). Assim, o acoplamento transcrição-tradução é uma característica específica da expressão gênica procariótica e desempenha papéis essenciais em muitos aspectos da regulação (KACZANOWSKA & RYDÉN-AULIN, 2007; LI et al., 2015).

A sequência de Shine e Dalgarno (SD), baseada nos sobrenomes dos cientistas que a descreveram pela primeira vez, ou simplesmente “sítio de ligação ao ribossomo” (**RBS** - *Ribosome Binding Site*) foram descritos primeiramente em mRNAs de bacteriófagos (SHINE & DALGARNO, 1974). Estes pesquisadores sugeriram que esta região interagia com uma extremidade rica em pirimidinas, a 3' anti-SD (que possui a sequência complementar UCCUCC) do rRNA 16S, presente na subunidade ribossômica 30S. Estas regiões unidas auxiliam na formação do complexo de iniciação da tradução para facilitar a ancoragem do mRNA e a adaptação ao ribossomo 30S (SHINE & DALGARNO, 1974).

A região de iniciação da tradução compreende três sequências diferentes: (1) a sequência de Shine-Dalgarno, (2) o códon de início, (3) a região espaçadora entre a sequência de Shine-Dalgarno e o códon de início, e às vezes (4) intensificadores de tradução (OMOTAJO et al., 2015). Neste sentido, o ribossomo liga-se ao local e sinaliza ao ribossomo a orientação correta em relação ao códon de início e, assim, permite que a tradução comece (figura 5) (CHEN et al., 1994; THIEL et al., 2018).



**Figura 5: O sítio de ligação ao ribossomo e a sequência codificante de proteínas controlam a taxa de início da tradução de um mRNA.** A região RBS é subdividida em um local de pausa, um sítio de ligação ao rRNA 16S, um códon de início e uma sequência espaçadora. O reconhecimento de uma subunidade 30S estende-se desde a montante do local de pausa até cerca de 22 bp a jusante do códon de início (Adaptado de SALIS, 2011).

Tal como na iniciação da transcrição, em que um dos quatro nucleotídeos é mais preferencial que o outro, na iniciação da tradução alguns códon também são mais preferenciais do que outros (RANGEL-CHAVEZ et al., 2017). Em *E. coli*, o códon preferencial de início nos genes para a iniciação da tradução é AUG (91%), seguido por GUG (8%) e UUG (1%) de acordo com a eficiência de tradução, onde AUG é predominante encontrado (SHINE & DALGARNO, 1974; RINGQUIST et al., 1992; BARRICK et al., 1994; SCHUMANN & FERREIRA, 2004; GUALERZI & PON, 2015). Esta preferência nos códon utilizados para o início da tradução é conservado em outras bactérias (RANGEL-CHAVEZ et al., 2017).

Muitos pesquisadores descobriram em RBS de genes de *E. coli* com o auxílio de programas de bioinformática (utilizando algoritmos de busca) que existem espaçamentos preferencias entre a sequência de SD e o códon de início (AUG) em uma variedade de mRNAs que variavam de 5 a 13 nucleotídeos com parte ou uma totalidade de uma sequência rica de várias purinas (UAAGGAGGU) e que culminou para os seguintes resultados: (1) A sequência de SD 5' UAAGGAGG 3' permite uma produção de proteína 3 a 6 vezes maior do que 5' AAGGA 3'; (2) o espaçamento ótimo UAAGGAGG foi determinado como sendo 4 a 8 nucleotídeos e 5 a 7 para AGGAGG encontrado em muitos genes predominantemente usados para armazenamento e processamento de informações e tradução, e biogênese de ribossomos (figura 6) (SHINE & DALGARNO, 1975; RINGQUIST et al., 1992; BARRICK et al., 1994; MA et al., 2002; JIN et al., 2006; KACZANOWSKA & RYDÉN-AULIN, 2007; VIMBERG et al., 2007; PASOTTI & ZUCCA, 2014; OMOTAJÓ et al., 2015; THIEL et al., 2018).

mRNA	RBS	Códon de início	Sequência do gene
<b>Proteína A do fago Q<math>\beta</math></b>	CUG AGU AUA <u>AGA GGA</u> CAU	AUG	CCU AAA
<b>Capsídeo A do fago Q<math>\beta</math></b>	CUU UGG <u>GUC</u> AAU UUG AUC	AUG	GCA AAA
<b>Fago <math>\gamma</math>Cro</b>	AUG UAC UAA <u>GGA GGU</u> UGU	AUG	GAA CAA
<b>Capsídeo do fago <math>\phi</math>I</b>	UUU AAU <u>GGA</u> AAC UUC CUC	AUG	AAA AAG
<b>Fago <math>\Phi</math>X174 A*</b>	CUU UGG <u>GUC</u> AAU UUG AUC	AUG	GCA AAA
<b>Lipoproteína</b>	AUC UAG <u>AGG</u> GUA UUA AUA	AUG	AAA GCU
<b>RecA</b>	GGC AUG ACA <u>GGA</u> GUA AAA	AUG	GCU AUC
<b>GalE</b>	AGC CUA AUG <u>GAG</u> CGA AUU	AUG	AGA GUU
<b>GalT</b>	CCC GAU UAA <u>GGA</u> ACG ACC	AUG	ACG CAA
<b>Ribossomo L10</b>	CAU CAA <u>GGA</u> GCA AAG CUA	AUG	GCU UUA
<b>T7 gene 10</b>	UUA AGA <u>AGGAGAU</u> AUACA U	AUG	GCU AGC

**Figura 6: Sítios de ligação ao ribossomo (sequência de Shine-Dalgarno).** Algumas sequências de Shine-Dalgarno e sua relação ao códon iniciador. As sequências do RBS de cada mRNA estão sublinhadas (Fonte: com modificação de SCHNEIDER & STEPHENS, 1990).

A eficácia da interação primária entre o RBS e o ribossomo depende do (i) grau de complementaridade entre a sequência do núcleo da SD e da sequência do anti-SD do rRNA 16S, (ii) a distância do SD do códon de iniciação da tradução e (iii) as sequências nucleotídicas circundantes, que podem formar estruturas secundárias que interferem na ligação de ambos (MA et al., 2002; OMOTAJÓ et al., 2015; THIEL et al., 2018).

A estrutura secundária na região de iniciação da tradução do mRNA desempenha um papel importante na eficiência da expressão gênica. Aparentemente, a sequência SD não é essencial para a ligação do mRNA aos ribossomos durante a tradução, embora a falta da sequência SD e/ou do códon de início afeta a acessibilidade da subunidade ribossômica 30S ao mRNA impedindo a tradução ou a eficiência da iniciação é diminuída na sua ausência (MELANÇON et al., 1990; SCHUMANN & FERREIRA, 2004). Entretanto, a força de interação da SD, um espaço longo excessivo ou curto entre o SD e a escolha do códon de início pode acarretar um efeito direto sobre a produção de proteínas, assim como na eficiência da iniciação da tradução, em até 600 vezes, sendo AUG o mais eficaz do que os outros códons (MA et al., 2002; MUTALIK et al., 2013).

## 2.7 REGULAÇÃO DA EXPRESSÃO GÊNICA EM PROCARIOTOS BASEADO NO OPERON LAC

Os procariotos monitoram constantemente as condições físico-químicas extracelulares, para que possam responder modificando seu padrão de expressão do genoma

(ISHIHAMA, 2012; BICH et al., 2016). A iniciação da transcrição é a principal etapa de regulação da expressão do genoma, mesmo que a síntese de mRNA também seja regulada na etapa de alongamento e término da transcrição (ISHIHAMA, 2012). Apenas uma parte de genes estimados é transcritos em um dado momento e para economizar na síntese proteica alguns produtos gênicos são transcritos em larga escala, enquanto outros são necessários em menor escala. O princípio que norteia a regulação gênica é o princípio da economia de energia (BOSDRIESZ et al., 2015).

Os primeiros estudos de expressão gênica em procariotos foram do metabolismo da lactose em *E. coli*. O operon lac foi o primeiro sistema de regulação gênica descrito, resultante de pesquisa realizada por François Jacob e Jacques Monod em 1961, tendo como a ação de um repressor controlando a transcrição em resposta a um indutor (alolactose) (JACOB & MONOD, 1961; LEWIS, 2013). O operon lac é o conjunto de genes responsável pela síntese de três enzimas envolvidas no metabolismo da lactose, que se ativa na presença desse açúcar, mas na ausência da glicose, substrato inicial preferido. Devido à este sistema, acabou se tornando o paradigma no estudo da regulação de outros genes em bactérias para a biologia de sistemas atuais (GILBERT & MAXAM, 1973; PENUMETCHA et al., 2010; BICH et al., 2016; PHILLIPS et al., 2019).

O operon lac consiste em uma região regulatória constituída como: (i) o promotor (P); (ii) um operador lac (lacO), onde um repressor (LacI) se liga; e uma proteína ativadora do catabolismo – CAP, também conhecida como proteína receptora de cAMP (CRP) (PENUMETCHA et al., 2010; BICH et al., 2016). Essas regiões antecedem as três sequências estruturais dos genes lacZ, lacY e lacA, respectivamente, que codificam as proteínas responsáveis para a utilização do açúcar lactose apenas quando a mesma está disponível (ULLMANN, 2009; PHILLIPS et al., 2019).

Na ausência de lactose, o operon lac é bloqueado por uma proteína repressora LacI, codificada pelo gene lacI, que se encontra geralmente em plasmídeos ou a montante do promotor lac na forma natural. Esta proteína é um regulador negativo da transcrição do operon lac reprimindo a transcrição do operon lacZYA através da ligação no operador que está a jusante do promotor lac, sobreposto no sítio de início da transcrição (figura 7-A) (LEWIS et al., 1996; OEHLER et al., 2006; PENUMETCHA et al., 2010; BICH et al., 2016; SCHULLER et al., 2020). O operador consiste em uma sequência de nucleotídeos específica que é reconhecida pelo repressor LacI que se liga fortemente, bloqueando fisicamente

(estrangulando) o início da transcrição. O operador O1 e repressor LacI ligado foram isolados como um oligonucleotídeo de 27 pares de bases protegido da digestão por DNase. A estrutura tridimensional do repressor LacI ligado a um DNA operador de 21 pb (5'-AATTGTGAGCGGATAACAATT-3') também foi determinada por difração de raios-X (OEHLER et al., 1990).

O repressor LacI tem alta afinidade pela lactose. Na presença de lactose, o repressor é inibido alostericamente por alolactose que se liga ao repressor LacI, induz uma mudança conformacional na estrutura da proteína e o LacI não pode mais se ligar ao sítio lacO. Isso deixa o sítio lacO desobstruído para a ligação de RNAP e, assim, a transcrição gênica do operon ocorre (figura 7-B) (LEWIS et al., 1996; OEHLER et al., 2006; PENUMETCHA et al., 2010; BICH et al., 2016; SCHULLER et al., 2020).

Quando a bactéria *E. coli* é cultivada no meio contendo glicose e lactose, a expressão do operon lac permanece inativada, ou seja, enquanto houver glicose disponível, a bactéria irá consumi-la, e não utilizará a lactose devido a glicose ser a fonte de carbono principal, coerente com o fato de que o açúcar é mais energeticamente eficiente a ser metabolizado, portanto, faz sentido evitar a expressão do operon lac quando a glicose está presente. Embora a lactose possa induzir a expressão do operon lac, o nível de expressão é muito baixo. A razão para isso é que o operon lac está sujeito à repressão catabólica ou à expressão reduzida de genes provocada pelo crescimento na presença de glicose (LEWIS, 2013).

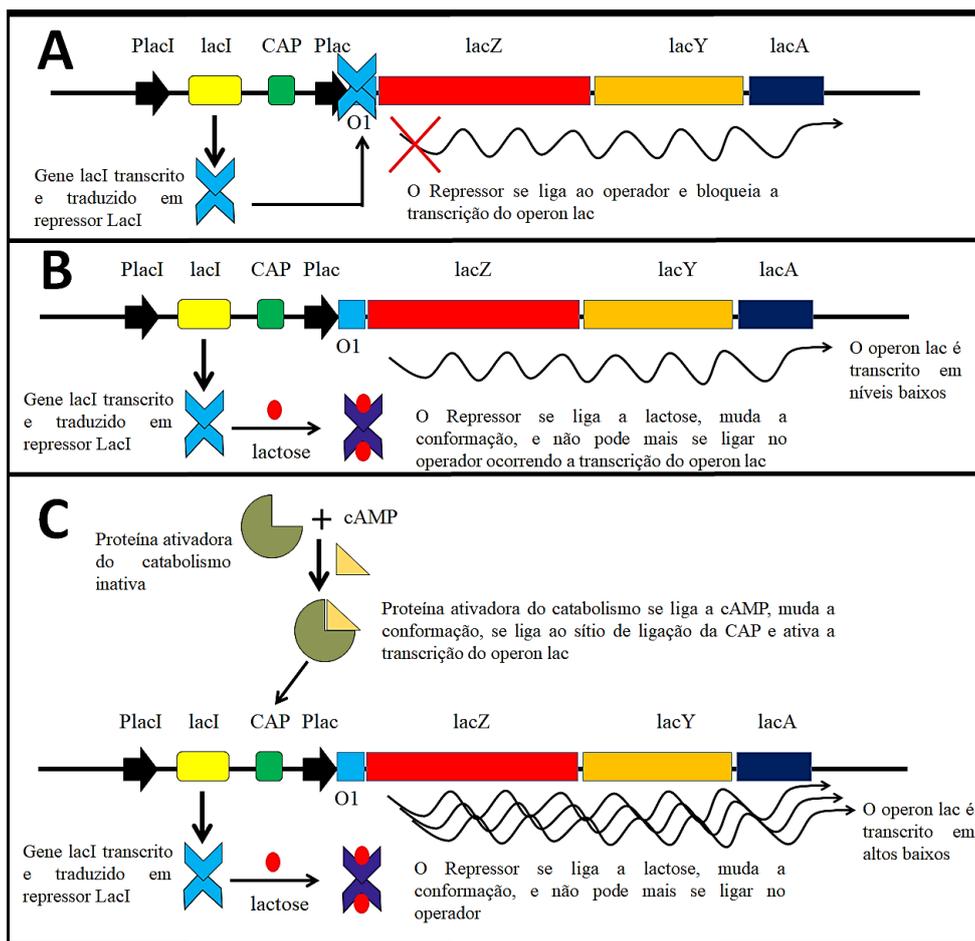
O promotor lac devido a sua sequência não ser tão idêntica a consensual, apresenta pouca capacidade de se ligar à RNA polimerase e formar um Rpo, e por isso é pouco transcrito mesmo quando induzido. Existe um sítio de ligação a montante do promotor lac para uma proteína denominada de “proteína ativadora do catabolismo” (CAP). A CAP ativa a transcrição de mais de 100 promotores e controla positivamente a atividade da RNAP aumentando a capacidade dela para transcrever vários genes ou operons catabólicos (BUSBY & EBRIGHT, 1999; LAWSON et al., 2004; ULLMANN, 2009; LEWIS, 2013; BICH et al., 2016).

CAP é um dímero de duas subunidades idênticas, cada uma com 209 resíduos de aminoácidos e possui massa molecular de 45 kDa capaz de forte interação com sítio de DNA de 22 pares de bases com simetria dupla, 5'- AATGTGATCTAGATCACATTT-3' (BUSBY & EBRIGHT, 1999; LAWSON et al., 2004; ULLMANN, 2009). Uma subunidade

da CAP interage com a primeira metade do sítio de DNA e a outra subunidade de CAP interage com a metade restante do sítio de DNA (EBRIGHT et al., 1989; LAWSON et al., 2004).

Para se ligar ao seu sítio de ligação, a proteína CAP é alostericamente regulada por monofosfato de adenosina cíclico ou AMP cíclico (cAMP), um pequeno sinalizador intracelular da ausência de glicose. Na presença de concentrações elevadas de cAMP, devido à ausência de glicose, a CAP se liga a cAMP, funciona como um ativador de transcrição cooperativamente com a RNAP para reconhecer seu sítio de ligação de DNA particular nas regiões regulatórias de vários genes, inclusive do operon lac. Em células com pouca oferta de glicose e a alta oferta de lactose, o nível de cAMP é elevado e os genes necessários para o metabolismo da lactose são expressos ao máximo (LEWIS, 2013; BICH et al., 2016). Na presença de glicose diminui o nível de proteínas CAP e não há síntese de cAMP em altos níveis, com isso a transcrição do operon lac enfraquece (BICH et al., 2016).

O sítio de ligação de CAP/cAMP no operon lac a montante do Plac faz com que a polimerase aumente a afinidade pelo promotor. A presença deste ativador ligado (CAP/cAMP) juntamente ligado ao seu sítio de DNA favorece a interação e ligação da RNAP ao promotor lac para produzir o RPo e com isso ativar a transcrição aumentando em torno de 50 vezes a sua atividade (figura 7-C) (LAWSON et al., 2004; BICH et al., 2016).



**Figura 7: O operon lac é regulado pelo repressor LacI e pela atividade de CAP/cAMP.** O repressor LacI é codificado pelo gene lacI, que se liga ao operador impedindo que ocorra a transcrição pela RNA polimerase. Na ausência de lactose, o repressor LacI está ligado ao operador (A). Na presença de lactose, o repressor LacI desliga-se do operador e assim ocorre a transcrição (B). A proteína CAP ligada com a cAMP reconhece uma sequência específica na região regulatória do gene que se localiza perto do promotor lac denominado de sítio de ligação a CAP e consequentemente estimulando a ligação da RNA polimerase ao promotor, ocasionando a ativação máxima da transcrição do operon lac (C).

Promotores dependentes de CAP para ativação da transcrição podem ser agrupados em duas classes com base na posição do sítio de DNA para CAP e o mecanismo correspondente para ativação da transcrição (BUSBY & EBRIGHT, 1999; LAWSON et al., 2004).

Nos promotores dependentes de CAP Classe I, o sítio de DNA para CAP está localizado a montante do promotor. Os promotores dependentes de CAP classe I mais bem caracterizados são o promotor lac e o promotor artificial CC que possuem um sítio de DNA para CAP posicionado na região -61,5. A ativação da transcrição em promotores dependentes de CAP Classe I envolve uma única interação proteína-proteína entre CAP e

RNAP holoenzima, e prossegue através de um mecanismo simples de "recrutamento", pelo qual CAP facilita a ligação de RNAP ao promotor para produzir o complexo fechado RNAP-promotor (LAWSON et al., 2004).

Nos promotores dependentes de CAP de Classe II, o sítio de DNA para CAP se sobrepõe ao elemento  $-35$  no núcleo do promotor. Os promotores dependentes de CAP de classe II mais bem caracterizados são o promotor galP1 e o promotor artificial CC que possuem um sítio de DNA para CAP centrado na posição  $-41,5$ . A ativação da transcrição em promotores dependentes de CAP de classe II envolve três conjuntos de interações proteína-proteína entre CAP e RNAP e prossegue através de mecanismos de "recrutamento" e "pós-recrutamento", pelo qual o CAP facilita a ligação de RNAP ao promotor para produzir o R<sub>P</sub>c e facilitar a isomerização para R<sub>po</sub> (LAWSON et al., 2004).

O controle positivo mediado pelo cAMP–CAP tem sido frequentemente associado ao efeito do cAMP no sistema da repressão catabólica. Este modelo, no entanto, foi contestado após vários resultados que mostram que a repressão catabólica do operon lac (e de outros sistemas) pode ser modulada independentemente do cAMP. Estudos adicionais cristalográficos e biofísicos *in vitro* revelaram que a ligação do CAP também estabiliza a ligação do repressor LacI ao operador. Portanto, a CAP atua não simplesmente ativando a transcrição, mas também aumentando a repressão dos genes catabólicos (LAWSON et al., 2004). Os mecanismos responsáveis pela repressão catabólica, ativação, repressão e o seu sítio de reconhecimento ainda precisam de mais estudos (CROOKS et al., 2004; LAWSON et al., 2004; ULLMANN, 2009; GUAZZARONI & SILVA-ROCHA, 2014; AMORES et al., 2015).

## 2.8 EXPRESSÃO DE GENES HETERÓLOGOS EM *E. coli*

A bactéria mais comumente utilizada para a expressão de genes heterólogos é a *Escherichia coli*, uma enterobactéria Gram-negativa que tem sido caracterizada ao longo de décadas, de utilização segura nos laboratórios de pesquisa (utilizada principalmente para clonagem, modificação genética e produção de proteínas em pequena escala para fins de pesquisa) e na indústria (em bioprocessos biotecnológicos) (HUANG et al., 2012). Essa bactéria tem atraído interesse, devido à habilidade de crescer rapidamente atingindo altas concentrações celulares, permitir o uso de substratos baratos e possuir várias cepas com

genótipo conhecido (CHOI et al., 2006; SIVASHANMUGAM et al., 2009; SANCHEZ-GARCIA et al., 2016).

A rDNA desenvolvida no final dos anos 70 usando a *E. coli* como plataforma biológica permitiu a aprovação do uso clínico pelo FDA (Food and Drug Administration) da insulina humana recombinante produzida em *E. coli* para o tratamento de diabetes em humanos, sendo o primeiro fármaco recombinante a entrar no mercado (FERRER-MIRALLES et al., 2009; HUANG et al., 2012; BAESHEN et al., 2014; OVERTON, 2014; SANCHEZ-GARCIA et al., 2016; SELAS CASTIÑEIRAS et al., 2018).

Assim, muitos esforços recentes de rDNA, engenharia metabólica e biologia sintética têm se centrado no desenvolvimento de ferramentas no ajuste da expressão para facilitar a busca de um equilíbrio ótimo entre a viabilidade celular e o cassete de expressão para a produção de diferentes proteínas recombinantes em *E. coli* (ANDRIANANTOANDRO et al., 2006; FERRER-MIRALLES et al., 2009; OVERTON, 2014; CHEN et al., 2018). O uso da *E. coli* para obtenção de proteínas em quantidade suficiente para o estudo da estrutura e função da mesma ou para diversas aplicações clínicas ou industriais é hoje um marco na história do conhecimento de proteínas (HUANG et al., 2012; OVERTON, 2014; SANCHEZ-GARCIA et al., 2016; CHEN et al., 2018).

Vários são os fatores que ocasionam a escolha preferencial da expressão em *E. coli*: as linhagens mutantes disponíveis; um grande número vetores de clonagem e expressão compatíveis; os marcadores para seleção de clones; a força e a regulação do promotores; controle do número de cópias do gene; a eficiência dos terminadores de transcrição; a estabilidade do RNA mensageiro após a transcrição; a força e estrutura do RBS bem definidos; capacidade de otimização de códons; linhagens com co-expressão de tRNAs raros; tags de afinidade para purificação; e fáceis condições de cultivo, como: tempo; temperatura; pH; características nutricionais e meio de cultura bem definidos (SAHDEV et al., 2008; FERRER-MIRALLES et al., 2009; HUANG et al., 2012; ARPINO et al., 2013; GOPAL & KUMAR, 2013; ROSANO & CECCARELLI, 2014; CHEN et al., 2018).

A *E. coli* é capaz de produzir proteínas em níveis tão altos quanto de 10 a 50% da proteína celular total, além de ser bem caracterizada geneticamente, possui seu metabolismo amplamente conhecido, apresenta rápido crescimento celular, e existem várias revisões descrevendo várias cepas modificadas de *E. coli* para produzir proteínas terapêuticas com

alto rendimento, baixo custo e qualidade (SCHUMANN & FERREIRA, 2004; JANA & DEB, 2005; SØRENSEN et al., 2005; DEMAÏN & VAISHNAV, 2009; SIVASHANMUGAM et al., 2009; ADRIO & DEMAÏN, 2010; HUANG et al., 2012; PACHECO et al., 2012; OVERTON, 2014; ROSANO & CECCARELLI, 2014; BAESHEN et al., 2015; MARSCHALL et al., 2016; SELAS CASTIÑEIRAS et al., 2018).

No entanto, somente alguns genes podem ser expressos neste organismo e as principais desvantagens da *E. coli* como sistema de expressão incluem: a incapacidade de fazer modificações pós-traducionais (MPT) encontradas na maioria das proteínas eucarióticas, como: glicosilação; fosforilação; processamento não proteolítico; a incapacidade de formar pontes dissulfeto e a não secreção da proteína para o meio de cultura (CHOI et al., 2006; FERRER-MIRALLES et al., 2009; SIVASHANMUGAM et al., 2009; CHEN, 2012; HUANG et al., 2012; SELAS CASZTIÑEIRAS et al., 2018). Nesta situação, o processamento de proteínas a estas respostas de estresse celular pode tornar os produtos inúteis, principalmente por falta de solubilidade, e muitas espécies de proteínas se depositam em grandes quantidades como agregados proteicos conhecidos como corpos de inclusão (IBs) (FERRER-MIRALLES et al., 2009; SINGH et al., 2015).

A formação dos IBs pode ser vantajosa, uma vez que se obtêm proteínas recombinantes com rendimentos relativamente elevados e podem ser purificadas após o bioprocessamento (SINGH et al., 2015). São apropriados para a produção de proteínas tóxicas e são principalmente resistentes a ataques de proteases (BERLEC & ŠTRUKELJ, 2013). Muitas proteínas terapêuticas recombinantes aprovadas foram produzidas com êxito na forma solúvel em *E. coli* ou purificadas com sucesso a partir de IBs (CHOI et al., 2006; FERRER-MIRALLES et al., 2009; HUANG et al., 2012; BERLEC & ŠTRUKELJ, 2013; OVERTON, 2014; SANCHEZ-GARCIA et al., 2016; SELAS CASTIÑEIRAS et al., 2018).

## 2.9 EXPRESSÃO DE GENES HETERÓLOGOS EM *E. coli* UTILIZANDO O SISTEMA DE REGULAÇÃO DO OPERON LAC

Para a expressão de proteínas recombinantes em altos níveis em *E. coli* é necessário que o promotor do “cassete de expressão” seja regulado, pois a expressão em altos níveis da proteína “estranha” no interior da célula pode ser prejudicial e muitas vezes letal, principalmente se ela for tóxica à célula hospedeira (SAIDA et al., 2006; SAÏDA, 2007; MARSCHALL et al., 2016).

Um promotor eficiente de *E. coli* adequado para a síntese de proteínas em alto nível para aplicação industrial deve apresentar várias características desejáveis, entre elas: 1) deve ser forte, resultando em um acúmulo de 20 a 40% ou mais da proteína no citoplasma celular; 2) ter um nível mínimo de atividade de expressão basal, sendo rigorosamente regulado; 3) e sua indução deve ser feita de forma simples, efetiva, economicamente viável (MAKRIDES, 1996; SCHUMANN & FERREIRA, 2004; JANA & DEB, 2005; SØRENSEN et al., 2005; BERLEC & ŠTRUKELJ, 2013).

A maioria dos sistemas que empregam promotores reguláveis nos vetores de expressão se baseia na repressão por impedimento estérico, onde reguladores de transcrição (repressores) ligam-se ao seu operador específico, posicionado sobreposto ou justaposto aos hexâmeros consensuais de reconhecimento (-35 e -10) da sequência promotora. Dessa maneira, com a ligação da molécula repressora ao operador, haverá o impedimento do reconhecimento do promotor pelo fator sigma e pela RNA polimerase dificultando ou impedindo o início do processo transcricional do gene heterólogo (ANDRIANANTOANDRO et al., 2006; BROWNING; BUSBY, 2016; RUDGE et al., 2016; ENGSTROM & PFLEGER, 2017).

Não existe um sistema de expressão universal, e a escolha de cada combinação (hospedeira e o vetor de expressão) vai depender de cada proteína a ser expressa. A *E. coli* tem sido utilizado com sucesso como hospedeira para expressar genes heterólogos sob o controle de vários promotores regulados. Dos sistemas de regulação existentes, os promotores atualmente empregam a regulação derivada do operon lac em seus sistemas de expressão, entre eles, destacam-se o promotor lac e seus derivados tac, trp, trc e LacUV5, comumente usados tanto na pesquisa básica quanto na indústria e todos eles são induzidos por IPTG (CHOI et al., 2006; OVERTON, 2014; ROSANO & CECCARELLI, 2014; RAVISHANKAR et al., 2015; MOREIRA et al., 2018).

Normalmente para a expressão de genes heterólogos em *E. coli*, utiliza-se o indutor sintético isopropil- $\beta$ -D-tiogalactopiranosídeo (IPTG), pois é um forte indutor químico análogo da lactose utilizado na indução do promotor lac e derivados (CHOI et al., 2006; OVERTON, 2014; ROSANO & CECCARELLI, 2014; RAVISHANKAR et al., 2015; MOREIRA et al., 2018). Na ausência do indutor como a própria molécula do IPTG, o repressor LacI se liga ao operador situado logo após ao promotor. Quando o indutor IPTG está presente liga-se ao produto gênico do lacI (repressor LacI) e causa a dissociação do

operador ocasionando a transcrição do(s) gene(s) heterólogo(s) (VAN HOEK & HOGEWEG, 2007; FERNÁNDEZ-CASTANÉ et al., 2012).

O IPTG é um indutor sintético que apresenta duas vantagens: não é dependente de lac permease, se difundindo pela membrana celular por difusão e não pode ser clivado pela  $\beta$ -galactosidase e assim não perde sua capacidade indutora, prevenindo o término da transcrição (SCHUMANN & FERREIRA, 2004; VAN HOEK & HOGEWEG, 2007; FERNÁNDEZ-CASTANÉ et al., 2012; ROSANO & CECCARELLI, 2014). Se por um lado ele apresenta efetividade na indução e não sendo consumido durante o processo, por outro lado o uso de promotores induzíveis por IPTG limita sua aplicação na produção em larga escala devido à sua toxicidade e alto custo (CHOI et al., 2006; HUANG et al., 2012; RETNONINGRUM et al., 2019).

Por ser tóxico, o uso de IPTG não é recomendado para a produção de proteínas recombinantes com finalidades terapêuticas em saúde humana ou animal em bioprocessos em escala industrial. Se utilizado nesses casos, procedimentos para remoção das impurezas do IPTG deverão ser incorporados ao processo e será necessário comprovar que a está ausente no produto final (JONASSON et al., 2002; RETNONINGRUM et al., 2019). Para contornar estas desvantagens, o IPTG pode substituir a lactose em concentrações baixas (0,5 a 1mM) para minimizar potenciais problemas de solubilidade de proteínas (YABUTA et al., 1995; ANDREWS et al., 1996; STUDIER, 2005; CHOI et al., 2006; PETI & PAGE, 2007; SHOJAOSADATI et al., 2008; SIVASHANMUGAM et al., 2009; MÜHLMANN et al., 2017).

A purificação é o último passo a serem realizadas após a expressão. Atualmente as construções dos sistemas de expressão são realizadas com a finalidade de facilitar este processo. Entre o grande número de sistemas existentes, várias estratégias baseadas em sequências de aminoácidos manipuladas, fusões de proteínas e domínios de proteína única. O mais utilizado são os que expressam as proteínas de fusão, onde o gene da proteína alvo está ligado a um gene de outra proteína que ficará retida na fase estacionária de uma coluna cromatográfica. Tendo vantagem adicional é que a proteína de fusão pode proteger a proteína de interesse do ataque de proteases, possivelmente estabilizando o mRNA e podem também podem auxiliar na formação de pontes dissulfeto aumentando o rendimento de proteína solúvel (BERLEC & ŠTRUKELJ, 2013; OVERTON, 2014; SANCHEZ-GARCIA et al., 2016; OLIVEIRA & DOMINGUES, 2018).

## 2.10 VETORES DE EXPRESSÃO PARA *E. coli*

Os vetores de expressão são veículos de clonagem utilizados para a introdução de informação genética dentro de um hospedeiro utilizando as ferramentas da rDNA para estudo ou produção de proteínas recombinantes (SMILLIE et al., 2010; BERLEC & ŠTRUKELJ, 2013). Os vetores mais simples e estritamente utilizados são conhecidos como plasmídeos. Plasmídeos são moléculas circulares, fita dupla de DNA e extracromossomais existentes em microrganismos. Com exceção dos vetores que se integram no genoma, que não são comumente utilizados, a grande maioria apresenta: replicação independente; uma marca de seleção; geralmente um ou mais genes e sítios únicos para enzimas de restrição que facilitam a ligação de diferentes genes exógenos (BERLEC & ŠTRUKELJ, 2013).

Os vetores de expressão funcionais em *E. coli* contêm elementos comuns responsáveis pela produção controlada de proteínas recombinantes e possuem uma grande variabilidade de aplicações, como: clonagem de genes; para a seleção de clones recombinantes; seleção de origem de replicação; seleção de promotores; expressão ou ainda expressão e secreção de uma dada proteína (BERLEC & ŠTRUKELJ, 2013; OVERTON, 2014).

Um vetor de clonagem e expressão utilizado em *E. coli* deve possuir as seguintes propriedades: habilidade para se replicar na célula hospedeira, como a presença de uma origem de replicação independente (Ori) que controla a taxa de replicação do plasmídeo e, portanto, o número de cópias; deve acomodar insertos de vários tamanhos sem danificar as funções de replicação; deve possuir a facilidade de inserção na célula hospedeira depois de manipulação *in vitro*; deve conter um ou mais genes marcadores de seleção e permitir uma manutenção estável dos plasmídeos; e conter somente um sítio de múltiplo de clonagem (MCS) para endonucleases de restrição para clonagem de genes ou partes biológicas de DNA (OVERTON, 2014)

As Oris mais comuns são ColE1 (15-20 cópias na célula) derivado do plasmídeo pBR322, p15A (10-12 cópias por célula) e uma versão mutante da Ori do pMB1 que está presente na série pUC (500 a 700 cópias na célula) (HUANG et al., 2012; BERLEC & ŠTRUKELJ, 2013; OVERTON, 2014).

Os marcadores de seleção conferem resistência a antibióticos e permitem a manutenção estável do plasmídeo. Os antibióticos comumente utilizados na rDNA são

Ampicilina, Canamicina, Tetraciclina e Cloranfenicol. A ampicilina pode ser problemática devido à sua suscetibilidade à degradação no meio de crescimento e pode ser substituída por carbenicilina. A tetraciclina mostrou ser altamente estável durante o cultivo (BERLEC & ŠTRUKELJ, 2013; OVERTON, 2014).

## 2.11 USO DE CAIXA DE FERRAMENTAS GENÉTICAS

A padronização das partes biológicas e sua montagem é uma das idéias centrais por trás da biologia sintética para controlar com precisão a expressão gênica (KANG et al., 2014).

No início, sistemas de superexpressão de genes começaram a ser projetados e testados levando à produção industrial de proteínas valiosas como insulina e interferons (ROSANO & CECCARELLI, 2014; MOREIRA et al., 2018). Depois vários plasmídeos sofisticados para superexpressão de proteínas em *E. coli* e cepas hospedeiras apropriadas foram concebidas e construídas geneticamente para melhorar mais ainda a expressão gênica e a biossegurança no manuseio de OGMs (HOCHKOEPLER, 2013; KANG et al., 2014).

Embora o desenvolvimento de uma variedade de promotores e vetores plasmidiais tenha sido relatado, eles muitas vezes carecem da informação dos perfis de expressão gênica e não são otimizados para a expressão gênica de múltiplos genes que codificam uma via heterólogas (KANG et al., 2014). Para minimizar o desequilíbrio metabólico, vários plasmídeos disponíveis comercialmente foram projetados para cada processo biotecnológico, escolhendo entre diferentes promotores e origens de replicação, ambos são responsáveis pelo controle chave no nível de expressão do(s) gene(s) de interesse(s) (IRLA et al., 2016; LOZANO TEROL et al., 2021).

Diferentes promotores induzíveis por IPTG (T7, A3, lpp, tac, Pac, Sp6, lac, npr, trc e sin) têm sido utilizados para estudos de expressão de proteínas recombinantes. Geralmente promotores são introduzidos em backbones vetoriais que possuem diferentes origens de replicação e estratégias de clonagem e expressão de genes heterólogos (HEISS et al., 2016; LOZANO TEROL et al., 2021; YU et al., 2021).

Devido a essa variedade de parâmetros, um manuseio simples e a possibilidade de trocar facilmente elementos genéticos específicos devem ser pré-requisitos no desenvolvimento de uma plataforma de expressão adequada e bem estruturada. Um exemplo

é a caixa de ferramentas BioBrick, que possui partes genéticas padronizadas, como vetores de expressão integrativos e replicativos, origens de replicação, diferentes promotores com diferentes regulação e níveis de força, hospedeiras geneticamente modificadas e genes repórteres diferentes (KRÜGER et al., 2022).

Os genes repórteres desempenham um papel fundamental nos estudos de expressão gênica. Eles são mais comumente usados como indicadores dos níveis de transcrição de muitas variantes de sequência de DNA e de tradução dentro das células (IRLA et al., 2016). As células são testadas medindo diretamente a quantidade da proteína repórter. As proteínas fluorescentes são consideradas como repórteres ideais, pois não requerem substratos. A proteína verde fluorescente (GFP) que originalmente veio de uma água-viva bioluminescente chamada *Aequorea victoria* mostra uma fluorescência verde brilhante quando exposta à luz ultravioleta, possui um único pico de excitação centrado em cerca de 395 nm, com um pico de emissão em 509 nm e é devidamente muito utilizada em estudos laboratoriais para diversas aplicações de estudos, sob o controle de diversos promotores constitutivos e induzíveis em *E. coli* (YANG et al., 2016; DEUTCH, 2018; ABDELAAL & YAZDANI, 2021).

## 2.12 HORMÔNIO DE CRESCIMENTO (GH)

O hormônio do crescimento (GH) é um polipeptídeo globular de aproximadamente 24 kDa composto de 188 até 210 aminoácidos ligados entre si através de uma estrutura de hélice quádrupla com duas pontes dissulfeto entre as cadeias e um segmento aminoterminal invariável (YOUSEFIAN & SHIRZAD, 2011; COSTA et al., 2016; GHAVIM et al., 2017).

A quantidade de GH natural é extremamente baixa e a sua disponibilidade é limitada. O hormônio do crescimento recombinante (rGH) de peixes demonstrou que possui a mesma função do GH natural. Visto que em condições fisiológicas a quantidade endógena do GH é extremamente baixa e sua disponibilidade é limitada, a expressão heteróloga é uma importante alternativa para produção exógena de maiores níveis deste hormônio, o que pode viabilizar a realização de estudos voltados para a fisiologia e desempenho zootécnico das espécies de peixes utilizadas na piscicultura (LI et al., 2003).

Como a forma nativa de GH é uma proteína não glicosilada de cadeia simples e a produção de rGH foi extensivamente realizada utilizando diferentes sistemas de expressão em *E. coli* como preferencial a outros microorganismos ou sistemas celulares mais

complexos. Por sua vez, bibliotecas de cDNAs codificadoras de GH de várias espécies foram construídas utilizando a bactéria *E. coli* para o desenvolvimento de estudos pré-clínicos e clínicos em humanos (em estudos endocrinológicos no tratamento de nanismo, fraturas ósseas nos casos de osteoporose, queimadura de pele, hemorragias por úlcera, doenças cardiovasculares, entre outros) e sido amplamente utilizado eficiência no consumo alimentar, composição nutricional e crescimento de peixes. Algumas dessas sequências são derivadas de humanos, ratos, suínos, bovinos, de galinhas e peixes (VENUGOPAL et al., 2002; CUMMINGS & MERRIAM, 2003; LI et al., 2003; HÖYBYE & CHRISTIANSEN, 2015; COSTA et al., 2016).

O tambaqui tem sido, até o presente, a única espécie de peixe nativa contemplada com pesquisas sobre o gene codificador do GH. O dobramento correto das proteínas terapêuticas do GH é essencial para sua função se for utilizado para terapia gênica. Eficientes métodos para produção, purificação e caracterização de GHs em larga-escala têm sido desenvolvidos, sendo a produção em sistemas de expressão heteróloga em *E. coli* um caminho promissor.

### 3 OBJETIVOS GERAIS

O presente trabalho refere-se ao desenvolvimento de novos vetores bem regulados para clonagem e expressão de genes heterólogos em *E. coli*.

#### 3.1 OBJETIVOS ESPECÍFICOS

- Design do promotor sintético DM, síntese química do cassete de expressão bem regulado e construção do pDM-GFP;
- Clonagem do cassete de expressão contendo o promotor DM em diferentes backbones plasmidiais para construção de um conjunto de vetores;
- Clonagem e expressão de genes heterólogos em *E. coli* utilizando os vetores de expressão construídos;
- Melhoramento do promotor DM por síntese química de DNA randômica.

## 4 METODOLOGIA

### 4.1 ETAPAS DA CONSTRUÇÃO DO CONJUNTO DE VETORES REGULADOS

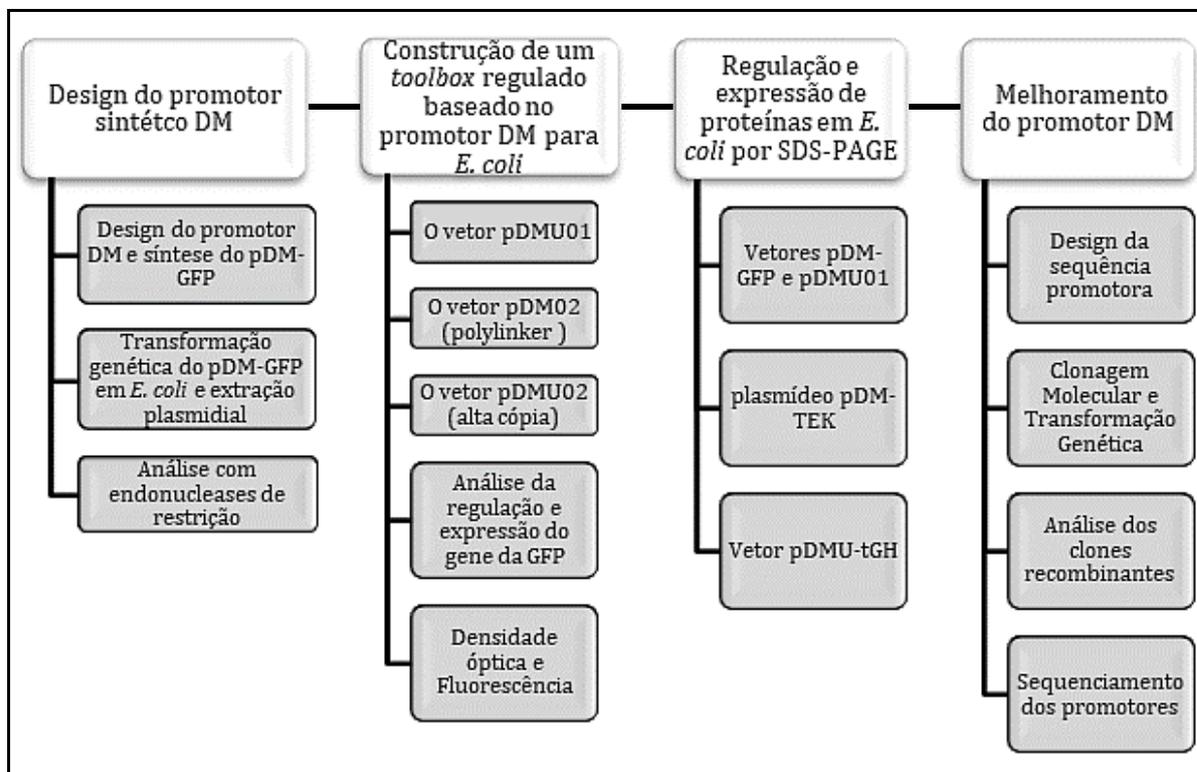


Figura 8: Esquema representando o fluxograma desenvolvido neste trabalho.

### 4.2 MATERIAL

#### 4.2.1 Linhagens Bacterianas

As linhagens hospedeiras de *E. coli* utilizadas neste trabalho foram a: **JM110** e **DH5 $\alpha$ F'1Q**.

A hospedeira **JM110** possui o seguinte genótipo: *rpsL (Strr) thr leu thi-1 lacY galK galT ara tonA tsx dam dcm supE44  $\Delta$  (lac-proAB) [F' traD36 proAB lacIq  $\Delta$ M15]*, confere resistência ao antibiótico estreptomicina e contém o gene *lacI* em um plasmídeo F' que codifica a proteína repressora LacI em alto nível capaz de bloquear a transcrição de promotores baseados no sistema *lac*.

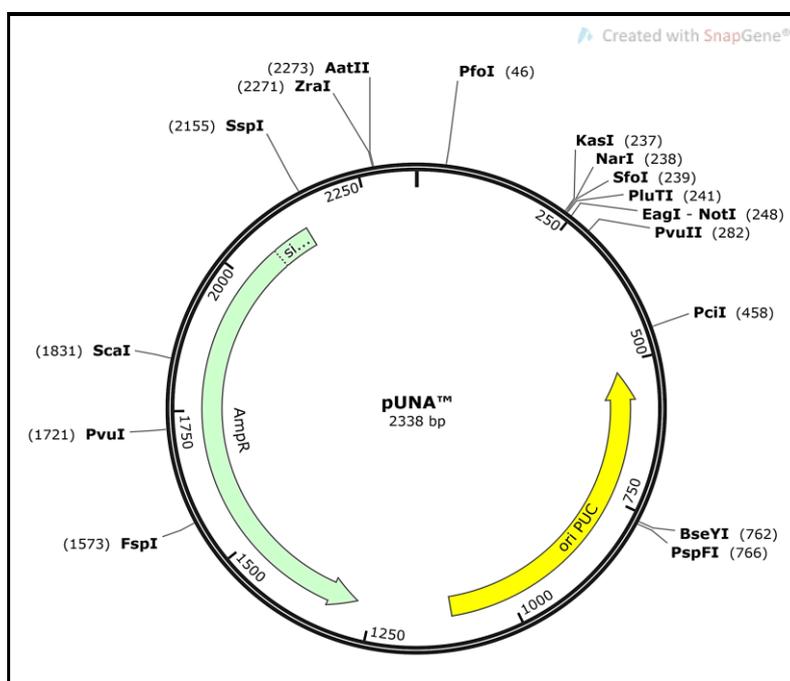
A linhagem **DH5 $\alpha$**  (*F*- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20  $\phi$ 80dlacZ $\Delta$ M15  $\Delta$ (lacZYA-argF) U169, hsdR17(rK-mK+ $\lambda$ -)). Deficiente em recombinação. Deficiente em Endonuclease A. Boa para clonar grandes plasmídeos (DÜRWARD & HOFFMANN-BERLING, 1968; HANAHAAN, 1983; GRANT et al., 1990).*

A linhagem de *E. coli* **DH5 $\alpha$ F'IQ** apresenta o seguinte genótipo: *F*- $\phi$ 80lacZ $\Delta$ M15  $\Delta$ (lacZYA-argF) U169 *recA1 endA1 hsdR17 (rk-, mk+) phoA supE44  $\lambda$ - thi-1 gyrA96 relA1/F' proAB+ lacIqZ $\Delta$ M15 zzf: Tn5 [KmR]. A linhagem DH5 $\alpha$ F'IQ também apresenta expressão para o produto do gene lacIq, conferindo a produção de repressor LacI em alto nível e apresenta resistência ao antibiótico canamicina (JESSE & BLODGETT, 1988). A linhagem **DH5 $\alpha$ F'IQ** foi utilizada neste trabalho para transformação genética do plasmídeo pDMUrGH.*

## 4.2.2 Plasmídeos utilizados neste trabalho

### 4.2.2.1 pUNA

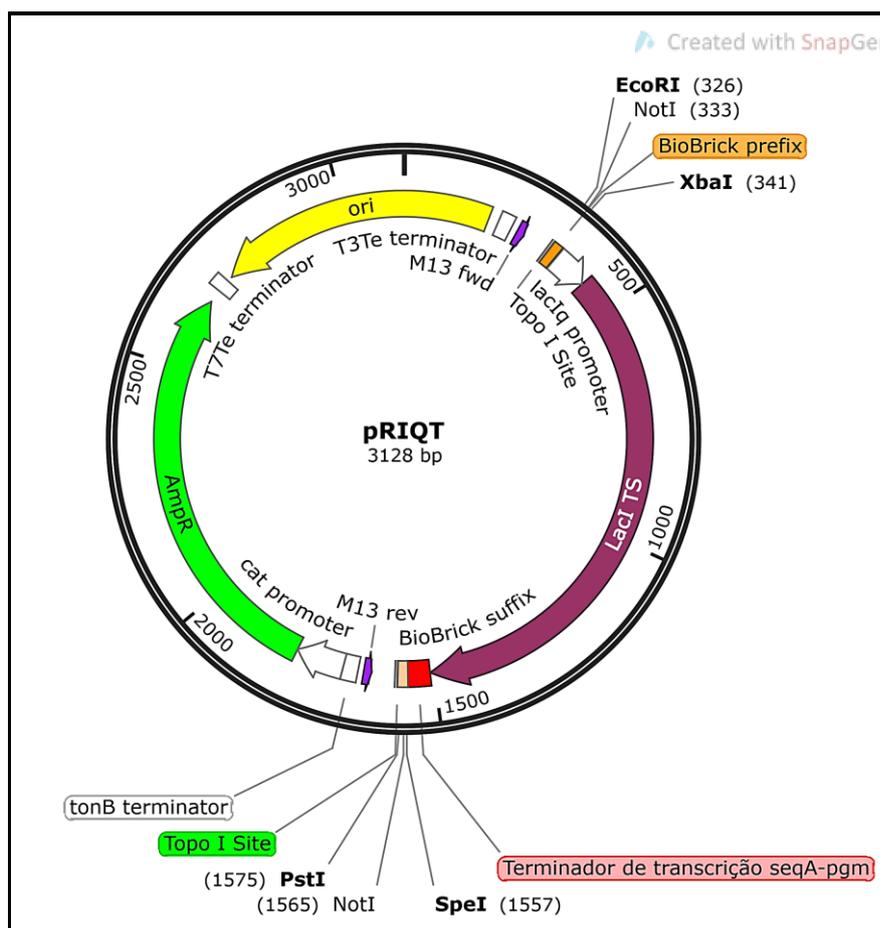
O vetor pUNA apresenta 2338 pares de bases, possui como marcador genético o gene que confere resistência a ampicilina, a Ori do pUC (500-700 cópias na célula) e apresenta um sítio único de restrição da enzima de restrição *NotI* na posição 248.



**Figura 9: Mapa físico do vetor pUNA.** O vetor possui o gene que confere resistência a ampicilina e uma Ori de alto número de cópias.

#### 4.2.2.2 pRIQT

O vetor pRIQT apresenta 3128 pares de bases, possui o gene que confere resistência a ampicilina, a Ori do pUC (500-700 cópias na célula) e apresenta o prefixo de reconhecimento para as endonucleases *EcoRI*, *NotI* e *XbaI* (na posição 326 até 341) e sufixo de reconhecimento para as endonucleases *SpeI*, *NotI* e *PstI*. O vetor possui o gene *lacIqTS* que codifica para expressão em alto nível do repressor LacI termo sensível.

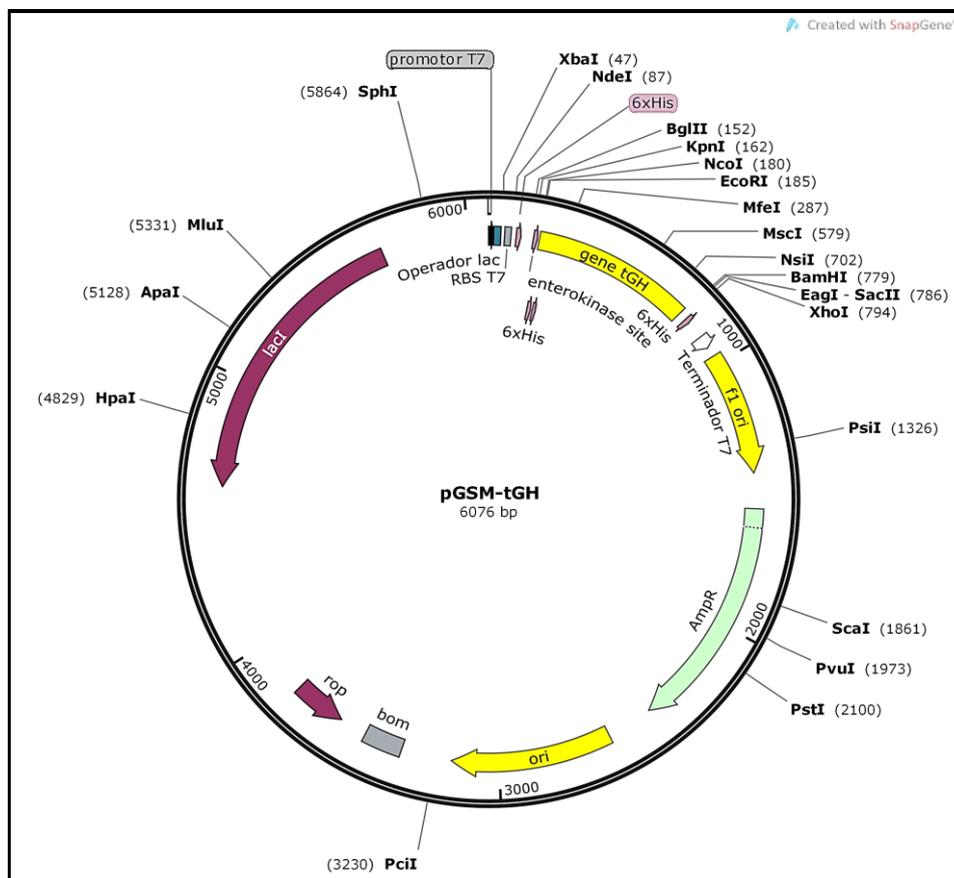


**Figura 10: Vetor pRIQT.** O vetor possui o gene que confere resistência a ampicilina e uma Ori de alto número de cópias. Possui o gene *lacIqTS* que codifica para expressão em alto nível do repressor LacI termo sensível entre um prefixo e sufixo.

#### 4.2.2.3 Vetor pGSM-tGH

O vetor pGSM-tGH possui 6076 pares de bases, possui o gene que confere resistência a ampicilina, a Ori do pUC de alto número de cópias (500-700 cópias na célula). Possui o gene *lacI* que codifica para expressão do repressor LacI para regulação de

promotores baseados no lac e apresenta o gene recombinante tGH entre os sítios das endonucleases *NdeI* e *BamHI*.



**Figura 11: Vetor pGSM-tGH.** O vetor possui o gene que confere resistência a ampicilina e uma Ori de alto número de cópias. Possui o gene *lacI* que codifica para expressão do repressor LacI e o gene *tGH* entre os sítios das endonucleases *NdeI* e *BamHI*.

#### 4.2.3 Oligonucleotídeos utilizados neste trabalho.

Os oligonucleotídeos foram sintetizados quimicamente pela (*EXXTEND* – Campinas/SP) e estão listados no quadro 2.

**Quadro 2: Oligonucleotídeos utilizados neste trabalho.**

Oligonucleotídeos	Sequência 5' → 3'	MER*
pLK1	TATGTTAACTTAAGGTACCAGATCTGTTTAAACG	34
pLK2	ACAATTGAATTCCATGGTCTAGACAAATTTGCCTAG	36

PDSM-1	CGAATTCGCGGCCGCTTCTAGAWWAAAWWTWTTTTW WAAWAKAATTGWSWCGGATAACAATTGTGT	67
PDSM-2	TACTAGTAGCGGCCGCTGCAGA TYSSTCCCATTATMACACAATTGTTATCCGC	53
VF1	TGCCACCTGACGTCTAAGAA	20
RCP1	GAGAAAGTAGTGACAAGTGTTGGCC	25

\**nucleotídeos*

#### 4.2.4 Meios de cultura para bactéria.

Tabela 2: Meio de cultura Luria-Bertani.

	Composição
<b>Líquido</b>	Peptona 10g/L; Extrato de levedura 5g/L; Cloreto de sódio 10g/L; Água destilada 1000 mL
<b>Sólido</b>	Adição de ágar 15g/L
<b>Ajuste para pH 7,5 e esterilização por autoclavagem.</b>	

Todos os meios de cultura bacteriana foram dissolvidos em água destilada ou deionizada e posteriormente autoclavados a 121 °C por 15 min.

O meio seletivo Luria Bertani (LB) sólido e/ou líquido contendo canamicina 50 µg/mL ou ampicilina 100 µg/mL foram usados para seleção de clones recombinantes, crescimento bacteriano, ensaios de fluorescência e produção de proteínas.

#### 4.2.5 Reagentes, tampões e soluções

Todos os reagentes, tampões e soluções estão apresentados no **ANEXO 1**. Os reagentes e tampões utilizados para o preparo das soluções foram de grau analítico. A água utilizada neste trabalho foi destilada e deionizada (água ultrapura tipo 1).

As soluções foram esterilizadas a 120°C por 15 minutos ou, quando necessário, filtradas em membranas do tipo milipore com poro de 0,22µm.

### 4.3 MÉTODOS

### 4.3.1 Clonagem molecular

Todas as técnicas de clonagem e manipulação de DNA recombinantes foram realizadas de acordo com (MANIATIS et al ., 1982), ou de acordo com os fabricantes das enzimas e kits. As endonucleases (enzimas de restrição) e outras enzimas modificadoras de DNA (SAP- Fosfatase Alcalina de Camarão e T4 DNA ligase) foram usadas conforme especificado pelos fornecedores ( *New England Biolabs*, Ipswich, MA, USA). Os DNAs purificados, vetores e fragmentos, quando submetidos a reação de ligação, utilizou-se a enzima T4 DNA ligase (*New England Biolabs*, Ipswich, MA, USA) e a reação foi incubada a 16°C por 16-18 horas (MANIATIS et al ., 1982).

Todos os resultados com endonucleases foram analisados em gel de agarose, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador. Os produtos da digestão quando necessários foram analisados utilizando os marcadores moleculares da GeneRuler™™ 1 kb DNA Ladder (Fermentas), Ladder 1Kb (Quatro G) ou Invitrogen 100 pb DNA ILdder (ANEXO 2).

### 4.3.2 Transformação genética

Os protocolos de preparo de células eletrocompetentes e quimiocompetentes estão listados no ANEXO 3 e 4, respectivamente. O preparo de células competentes para transformações genéticas foram realizadas segundo HANAHAN, 1983.

### 4.3.3 Extrações de plasmídeos e purificação

As extrações de plasmídeos dos clones selecionados foram realizadas com o Kit de extração *Illustra plasmidPrep Mini Spin Kit* (GE Healthcare Bioscience) ou com o kit da Cellco *Fast-n-easy Plasmid Mini-Prep kit (Column based isolation of plasmid DNA)* o qual possuem soluções que atuam nas fases de lise celular, inibição de proteases, lavagem e eluição do DNA plasmidial (ANEXO 5). Os procedimentos foram realizados de acordo com as especificações do fabricante.

A purificação dos fragmentos ou vetores após digestões quando necessárias foram realizadas através do *Kit comercial illustra GFX PCR DNA and Gel Band Purification* (GE Healthcare). Os procedimentos foram realizados de acordo com as especificações do fabricante (ANEXO 6).

O protocolo de purificação de DNA por precipitação está listado no **ANEXO 7**.

#### **4.3.4 SDS-PAGE**

As proteínas recombinantes produzidas por clones de *E. coli* portadores dos plasmídeos foram analisada por eletroforese em gel desnaturante de poliacrilamida (SDS-PAGE) segundo LAEMMLI, 1970. Os produtos da produção de proteínas foram analisados utilizando o Blueye prestained protein ladder (11 to 245 kda) (**ANEXO 2**).

#### **4.3.5 Design do promotor DM e síntese do pDM-GFP**

O promotor DM foi construído para ser reconhecido pela maquinaria RNAP associada com o  $\sigma 70$  de hospedeiras *E. coli*. Foi realizado o design da sequência do promotor DM utilizando as melhores sequências consensos da atualidade moderna de promotores e reconhecimento eficiente pela RNAP ( $\alpha I \alpha II \beta \beta' \omega \sigma 70$ ) como o elemento UP, as regiões -35 e -10, a região espaçadora, o elemento -10 estendido, o discriminador e o sítio de início da transcrição (HOOK-BARNARD & HINTON, 2007; RUFF et al., 2015; RANGEL-CHAVEZ et al., 2017; MAZUMDER & KAPANIDIS, 2019).

No design do promotor DM foi não adicionado o sítio da CAP para tornar o promotor não responsivo à repressão catabólica como encontrado nos promotores dependentes de CAP Classe I e II, nas regiões -61,5 e 41,5, respectivamente (LAWSON et al., 2004).

Para obtermos um promotor mais estritamente regulado, decidiu-se mimetizar o que ocorre no operon MER (que codifica resistência a mercúrio) de *E. coli* sobrepondo a região operadora lacO1 selvagem de 21 pb (5'-AATTGTGAGCGGATAACAATT-3') à região -35.

O promotor DM foi encomendado, construído por síntese química de DNA pela empresa *Genone Biotecnologies Ltda* (Rio de Janeiro) e inserido em um cassete de expressão com cerca de 1015 nucleotídeos para testar sua funcionalidade pela maquinaria de tradução de proteínas (**RBS**mRNA). O cassete de expressão contém a jusante do sítio de início da transcrição ou +1 a sequência da região 5' UTR do mRNA do gene da proteína 10 do bacteriófago T7, que contém um eficiente RBS (OLINS et al., 1988) e o gene repórter da

proteína verde fluorescente - GFP (*green protein fluorescent*) sobreposto com um sítio de restrição da enzima *NdeI* no códon de início da tradução.

Após o códon de término da tradução do gene da GFP foi adicionado um sítio de restrição da enzima *BamHI*, seguido do terminador de transcrição, o T1 do gene de *rrnB* de *E. coli* (OROSZ et al., 1991) para interromper a síntese do mRNA. Nas extremidades 5' e 3' do cassete de expressão foram adicionados os sítios das endonucleases *EcoRI*, *NotI*, *XbaI* e *SpeI*, *NotI* e *PstI*, respectivamente, de modo a permitir a transferência deste cassete de expressão para outros vetores de expressão ou expressão/secreção da série IGEM\_Manauas e outros da série Biobrick.

O cassete de expressão foi subclonado em um plasmídeo denominado pSB3K3-TH2 e o vetor resultante foi denominado pDM-GFP. O vetor pDM-GFP possui Ori de baixo número de cópias (10-12 cópias) proveniente do plasmídeo p15A (CHANG & COHEN, 1978) e o gene de resistência a canamicina originário do transposon Tn903 (GRINDLEY & JOYCE, 1980).

#### **4.3.6 Transformação genética do pDM-GFP em *E. coli*, extração plasmidial e análise com endonucleases de restrição**

O plasmídeo pDM-GFP foi recebido liofilizado (4µg) e redissolvido em 20 µL de água ultrapura estéril. A seguir, para obtenção de massa plasmidial e análise da expressão do gene da GFP, o plasmídeo foi introduzido em células de *E. coli* JM110 (YANISCH-PERRON et al., 1985) com 0,2 µL da solução de DNA plasmidial misturados com 50 µL de células eletrocompetentes e submetidas ao procedimento de eletroporação (ANEXO 3). Em seguida, diferentes alíquotas dos sistemas foram semeadas em placas de Petri com meio LB sólido suplementado com canamicina. As placas foram incubadas invertidas a 37°C durante 16 horas.

Colônias transformantes foram selecionadas para pré-inóculos contendo canamicina. Os pré-inóculos foram incubados sob agitação constante (140 rpm) por um período de 14 horas a 37°C. Clones transformantes de *E. coli* JM110 portadores do vetor pDM-GFP foram selecionados e seus plasmídeos extraídos (ANEXO 5). Os resultados da extração do vetor pDM-GFP foi submetido à eletroforese em gel de agarose 0,8%, corado com brometo de

etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador. Os DNAs dos clones extraídos foram guardados a -20°C até o seu uso novamente.

Para uma melhor caracterização dos sítios únicos endonucleases que existem no mapa físico do plasmídeo pDM-GFP, este foi tratado com endonucleases *EcoRI*, *BamHI*, *NdeI*, *PstI*, *SpeI*, *EcoRV*, *NotI* e simultaneamente com *EcoRI* e *PstI*, *EcoRI* e *SpeI*, *EcoRI* e *NdeI*, *NdeI* e *BamHI*. Os sistemas foram incubados à 37°C por 2 horas e os resultados foram submetidos à eletroforese em gel de agarose 0,8%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador. Os DNAs plasmidiais dos clones extraídos foram estocados à -20°C até o seu uso novamente.

#### **4.3.7 Desenvolvimento de novos vetores bem regulados baseado no promotor DM para *E. coli***

Após sua análise de funcionalidade, para expandir os estudos sobre regulação e expressão de proteínas, o cassete de expressão sob o controle do promotor sintético induzível regulado com alta *performance* denominado de DM (domínio de patente) foi inserido em conjunto de diferentes backbones plasmidiais.

O conjunto de vetores construído neste trabalho consiste em cinco diferentes vetores: o próprio pDM-GFP (baixa cópia), o pDM02 (*polylinker*) e pDMU01 (alta cópia), pCDM (baixa cópia) e pDMU02 (alta cópia). O cassete de expressão presente em todos os vetores é flanqueado por diferentes sítios de endonucleases e oferece a possibilidade de clonagem (utilizando endonucleases e ligação clássica) para a troca de novas partes de DNA com funções biológicas para novos estudos de regulação e expressão de genes heterólogos em *E. coli*. Para a validação do conjunto de vetores, utilizamos nos experimentos deste estudo o pDM-GFP (BR 10 2019 000470 3) e a hospedeira *E. coli*.

##### **4.3.7.1 Clonagem do cassete de expressão no vetor pUNA (pDMU01)**

Para a construção do plasmídeo de alta cópia contendo o promotor DM, o cassete de expressão pertencente ao vetor pDM-GFP de baixa cópia foi clonado no vetor pUNA (de alta cópia) utilizando a enzima de restrição *NotI*. O vetor pUNA (4.2.2.1) possui um único sítio de restrição para a enzima *NotI* no seu backbone plasmidial e o vetor pDM-GFP possui dois sítios para *NotI* nas posições 188 e 1182 do backbone plasmidial (5' - 3') que permite a transferência do cassete de expressão para o vetor pUNA.

A seguir, para obtenção de massa plasmidial, 1  $\mu\text{L}$  do vetor pUNA foi misturado em um microtubo com 50  $\mu\text{L}$  de células *E. coli* DH5 $\alpha$  eletrocompetentes e transformado geneticamente pela técnica de eletroporação (**ANEXO 3**). Em seguida, diferentes alíquotas dos sistemas foram semeadas em placas de Petri com meio LB sólido contendo somente ampicilina. As placas foram incubadas invertidas a 37°C durante 16 horas.

Foram selecionados clones transformantes portadores dos vetores pDM-GFP e pUNA. Os clones foram inoculados em pré-inóculos contendo canamicina e glicose para os clones portadores do pDM-GFP e somente ampicilina para os clones portadores do vetor pUNA. Os pré-inóculos foram incubados durante a noite a 37°C sob agitação de 150 rpm. Os plasmídeos dos clones foram extraídos (**ANEXO 5**), o resultado foi submetido à eletroforese em gel de agarose 0,8%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador. Os DNAs dos clones extraídos foram guardados a -20°C até o seu uso novamente.

O vetor pUNA e o vetor pDM-GFP foram submetidos a tratamento com enzima de restrição *NotI*. Foram montadas as reações em microtubos para um volume final de 50  $\mu\text{L}$ . Os componentes dos sistemas com a endonuclease estão especificados na tabela 3 a seguir.

**Tabela 3: Reação dos vetores pDM-GFP e pUNA com a endonuclease *NotI*.**

<b>Componentes</b>	<b>pDM-GFP sem enzima</b>	<b>pDM-GFP com enzima</b>	<b>pUNA sem enzima</b>	<b>pUNA com enzima</b>
H2O	1	2	4	12
Tampão 3.1 10X	1	5	1	5
DNA plasmidial	8	40	5	30
<i>NotI</i> (10U/ $\mu\text{L}$ )	-	3	-	3
V. F.	10 $\mu\text{L}$	50 $\mu\text{L}$	10 $\mu\text{L}$	50 $\mu\text{L}$

Os sistemas foram incubados por 1 hora a 37°C em banho seco. Após o tempo da reação, foi adicionado 15  $\mu\text{L}$  de TBE 5x azul nas reações com a endonuclease *NotI* e em seguida foram submetidos à eletroforese em gel de agarose 0,8%. Os fragmentos digeridos correspondentes ao inserto do cassete de expressão e do vetor pUNA linearizado foram purificados com kit (**ANEXO 6**). O resultado da purificação foi analisado eletroforese em

gel de agarose 0,8%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador.

O vetor pUNA foi tratado com a enzima SAP para desfosforilação das pontas 5' e 3' (retirada do fosfato) para impedir sua religação na hora da reação com a T4 DNA ligase. Foram adicionados os reagentes em um microtubo para um volume de 60  $\mu$ L. Os componentes da reação de desfosforilação do vetor pUNA com a enzima SAP estão especificados na tabela 4 a seguir.

**Tabela 4: Reação de desfosforilação do vetor pUNA com a enzima SAP.**

<b>Componentes</b>	<b>pUNA com enzima SAP</b>
H2O	7
Tampão 10X SAP	6
Vetor pUNA	45
Enzima SAP (1U/ $\mu$ L)	2
V. F.	60 $\mu$ L

O sistema foi incubado por 30 minutos a 37°C em banho seco, seguido de 60°C por 15 minutos. Após a reação de desfosforilação e inativação da enzima, foi adicionado 15  $\mu$ L de TBE 5x azul no microtubo e a amostra foi submetida a gel de agarose 0,8%. Após a eletroforese, a banda de interesse foi purificada por meio do protocolo de purificação do kit (**ANEXO 6**).

A reação de ligação do vetor pUNA desfosforilado e do inserto do cassete de expressão digeridos com *NotI* e purificados foram submetidos a uma reação de ligação para um volume de 20,5  $\mu$ L. Foi adicionado em um microtubo na ordem: 2  $\mu$ L de tampão 10X de T4 DNA ligase, 12  $\mu$ L de vetor pUNA, 5  $\mu$ L do fragmento do cassete de expressão e 1,5  $\mu$ L da enzima da enzima T4 DNA ligase (200U/ $\mu$ L). A reação foi incubada por 16°C durante a noite (16-18 horas) e após o tempo de reação, o sistema de ligação no microtubo foi submetido a concentração de DNA por precipitação (**ANEXO 7**).

Após a purificação, 2  $\mu\text{L}$  do sistema de ligação foi misturado com 50  $\mu\text{L}$  de células *E. coli* DH5 $\alpha$ F'IQ eletrocompetentes e transformadas geneticamente pela técnica de eletroporação (ANEXO 3). Alíquotas do sistema foram semeadas em placas de Petri com LB sólido contendo ampicilina. As placas foram incubadas invertidas a 37°C durante 16 horas.

Foi selecionado 1 clone de *E. coli* DH5 $\alpha$ F'IQ com o novo vetor e inoculado em pré-inóculo contendo ampicilina. O frasco foi incubado durante a noite a 37°C sob agitação de 150 rpm. O DNA plasmidial foi extraído (ANEXO 5), o resultado foi submetido à eletroforese em gel de agarose 0,8%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador. O novo vetor passou a se chamar de pDMU01. O DNA do clone extraído foi guardado a - 20°C até o seu uso novamente.

#### 4.3.7.2 Vetor pCDM (baixa cópia)

O vetor pDM-GFP (figura 4.2.2.1) possui o sítio de *SpeI* localizados nas posições 262 e 1174 dentro do seu backbone plasmidial (5' - 3'). Quando ocorre a digestão com esta enzima (*SpeI*) o fragmento (912 pb) que correspondente ao RBS e o gene da GFP é liberado do vetor. Através da digestão parcial com a enzima *SpeI* nas condições apropriadas, a enzima pode cortar somente um ou o outro sítio de *SpeI* localizado no vetor pDM-GFP.

Foi selecionado 1 clone transformante portador do vetor pDM-GFP e inoculado em pré-inóculo contendo canamicina e suplementado com glicose. O pré-inóculo foi incubado durante a noite a 37°C sob agitação de 150 rpm. O plasmídeo do clone foi extraído (ANEXO 5), o resultado foi submetido à eletroforese em gel de agarose 0,8%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador.

Foi montada a reação do vetor pDM-GFP com a endonuclease *SpeI* 1U/ $\mu\text{L}$ . Foi montada a reação em microtubo para um volume final de 40  $\mu\text{L}$ . Os componentes da reação estão especificados na tabela 5 a seguir.

**Tabela 5: Reação do vetor pDM-GFP com a endonuclease *SpeI*.**

<b>Componentes</b>	<b>pDM-GFP sem enzima</b>	<b>pDM-GFP com enzima</b>
H2O	5	18
Tampão 10X CutSmart	1	4

DNA plasmidial	4	16
<i>SpeI</i> (1U/ $\mu$ L)	-	2
V. F.	10 $\mu$ L	40 $\mu$ L

Os sistemas foram incubados por 30 minutos a 37°C em banho seco, seguido de inativação da enzima *SpeI* com 10  $\mu$ L de TBE5x azul e 3  $\mu$ L de EDTA 200mM. O resultado da reação foi submetido à eletroforese em gel de agarose 0,8%. Após a eletroforese, a banda de interesse foi purificada por meio do protocolo de purificação do kit (**ANEXO 6**), o resultado foi submetido à eletroforese em gel de agarose 0,8%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador.

O fragmento correspondente ao pDM-GFP digerido com *SpeI* e purificado foi tratado com a enzima T4 DNA polimerase para completar novamente a ponta dupla fita do fragmento digerido pela enzima *SpeI* e inativar um dos dois sítios. A reação foi montada para um volume de 15  $\mu$ L. Foi adicionado em um microtubo: 2,8  $\mu$ L de H<sub>2</sub>O, 1,5  $\mu$ L de tampão 2.1 10X, 10  $\mu$ L de vetor pDM-GFP (60ng) e 0,4  $\mu$ L de dNTPs (2,5 mM). O sistema foi pré incubado a 12°C por 10 minutos em termociclador *BIO RAD T100 Thermal Cycler* e logo em seguida adicionado 0,3  $\mu$ L da enzima T4 DNA polimerase (1U/  $\mu$ L) na reação e o sistema foi novamente incubado a 15 minutos por 12°C em termociclador *BIO RAD T100 Thermal Cycler*. Após o tempo, a enzima foi inativada com 1,5  $\mu$ L de EDTA 200mM seguido de 75°C por 20 minutos.

O volume de 16,5  $\mu$ L no microtubo foi submetido a concentração de DNA por precipitação (**ANEXO 7**), o pellet de DNA foi redissolvido em 8,5  $\mu$ L de H<sub>2</sub>O e a seguir foi montado a reação de ligação para um volume de 10  $\mu$ L. Foi adicionado em um microtubo: 8,5  $\mu$ L do fragmento digerido parcialmente com *SpeI* e purificado, 1  $\mu$ L de tampão 10X de T4 ligase e 0,5  $\mu$ L da enzima T4 DNA ligase (200U/ $\mu$ L). A reação foi incubada por 16°C durante a noite.

Após o tempo da reação, 1  $\mu$ L do sistema de ligação foi misturado com 50  $\mu$ L de células *E. coli* DH5 $\alpha$  eletrocompetentes e transformadas geneticamente pela técnica de eletroporação (**ANEXO 3**). Alíquotas do sistema foram semeados em placas de Petri com

LB sólido contendo canamicina e glicose. As placas foram incubadas invertidas a 37°C durante 16 horas.

Foram selecionados clones transformantes portadores do novo vetor oriundo do pDM-GFP e inoculados em pré-inóculos com canamicina e glicose. Os pré-inóculos foram incubados durante a noite a 37°C sob agitação de 150 rpm. Os DNAs plasmidiais dos novos clones foram extraídos (ANEXO 5), o resultado foi submetido à eletroforese em gel de agarose 0,8%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador.

Para confirmação que o sítio 1174 foi deletado nos DNAs plasmidiais dos novos clones, decidiu realizar a digestão com a enzima *SpeI* e *EcoRI*. As reações analíticas foram montadas em microtubos para um volume de 10 µL e os componentes estão especificados na tabela 6 a seguir.

**Tabela 6: Reação com *SpeI* e *EcoRI* dos DNAs plasmidiais extraídos dos novos clones selecionados.**

Componentes	DNAs plasmidiais com enzima
H2O	3,4
Tampão 10X CutSmart	1
DNA plasmidial	5
<i>SpeI</i> (1U/µL)	0,3
<i>EcoRI</i> (10U/µL)	0,3
V. F.	10 µL

As reações foram incubadas por 2 horas a 37°C em banho seco. O resultado foi submetido à eletroforese em gel de agarose 0,8% e visualizado no transluminador sob luz UV. Os DNAs dos novos clones foram guardados a -20°C até o seu uso novamente. O novo vetor de baixo número de cópias passou a se chamar de pCDM.

#### 4.3.7.3 O vetor pDM02 (*polinlynker*)

Para a construção do vetor pDM02, o vetor pDM-GFP deveria ser digerido com as endonucleases *NdeI* e *BamHI* para retirada da sequência codificante do gene da GFP e inserido um oligonucleotídeo dupla fita com as bordas digeridas pelas mesmas endonucleases. As duas sequências de DNA fita simples denominadas de pLK1 e pLK2, compõem o *polylinker* fita dupla após o anelamento. O *polylinker* é uma sequência com múltiplos sítios de clonagem (MCS) e o gene da GFP no vetor pDM-GFP é flanqueado pelo sítio das endonucleases *NdeI* e *BamHI*.

Como estratégia de clonagem os oligonucleotídeos foram desenhados como se as suas bordas já tivessem sido tratadas com as endonucleases *NdeI* e *BamHI* após o anelamento correto. Por isto não foi realizado o tratamento com as endonucleases de restrição *NdeI* e *BamHI* com os oligos pLK1+pLK2 dupla fita (dPLK).

Foram verificados os sítios das endonucleases existentes no vetor pDM-GFP. Os sítios de endonucleases de restrição de cortes únicos e que não existem no vetor pDM-GFP foram escolhidas para serem adicionadas ao *polylinker*. O *polylinker* foi desenhado com os sítios únicos das endonucleases *NdeI*, *HpaI*, *Afl* II, *KpnI*, *BglII*, *PmeI* e *BamHI* na sua sequência (5' – 3'). Dois oligonucleotídeos pLK1 e pLK2 (Quadro 1) foram sintetizados quimicamente pela empresa *EXXTEND*. Os oligonucleotídeos foram redissolvidos em um microtubo com Tris-HCl 2mM pH 8 para 100pmol/ $\mu$ L.

Foi realizado a hibridização dos oligonucleotídeos pLK1 e pLK2 para formarem dupla fita. Foi adicionado em um microtubo um sistema somente com o oligo pLK1, outro somente com o oligo pLK2 e outro com ambos os oligos pLK1 e pLK2 (dPLK). As reações foram montadas para um volume final de 50  $\mu$ L. A reação de anelamento foi montada de acordo com a tabela 7 a seguir.

**Tabela 7: Reação de hibridização dos oligos pKL01 e pKL02.**

<b>Componentes</b>	<b>Somente com pKL01</b>	<b>pKL01 e pKL02 (anelados)</b>	<b>Somente com pKL02</b>
H2O	42	39,5	42,5
NaCl 0,5M	5	5	5
pKL01 (1,04 $\mu$ g/ $\mu$ l)	3	3	-
pKL02 (1,1 $\mu$ g/ $\mu$ l)	-	2,5	2,5

V. F.	50 $\mu$ L	50 $\mu$ L	50 $\mu$ L
-------	------------	------------	------------

Os sistemas após montados nos microtubos foram colocados em termociclador *BIO RAD T100 Thermal Cycler* em um circuito de: 2 minutos por 100°C, 5 minutos por 80°C, 5 minutos por 70°C, 5 minutos por 50°C seguido de 5 minutos em temperatura ambiente. O resultado foi submetido à eletroforese em gel de agarose 2%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador. Os microtubos foram guardados a -20°C até o seu uso novamente.

Para a clonagem do *polylinker* no vetor pDM-GFP, o vetor pDM-GFP foi tratado com as endonucleases *NdeI* e *BamHI*. Foi montada a reação em um microtubo para um volume final de 50  $\mu$ L e os componentes estão especificados na tabela 8 a seguir.

**Tabela 8: Reação com as endonucleases *NdeI* e *BamHI* do vetor pDM-GFP.**

Componentes	pDM-GFP com enzima
H <sub>2</sub> O	15
Tampão 3.1 10X	5
pDM-GFP	25
<i>NdeI</i> (10U/ $\mu$ L)	2,5
<i>BamHI</i> (10U/ $\mu$ L)	2,5
V. F.	50 $\mu$ L

A reação com as endonucleases foi incubada por 2 horas a 37°C em banho seco. Após o tempo, foi adicionado 15  $\mu$ L de TBE5x axul no volume restante (45  $\mu$ L) e então submetido a eletroforese em gel de agarose 0,8%. A banda de interesse foi extraída por meio do protocolo de purificação (**ANEXO 6**) e o resultado foi novamente submetido à eletroforese em gel de agarose 0,8%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador.

A seguir o oligonucleotídeo sintético dPLK foi ligado com a enzima T4 DNA ligase ao vetor pDM linearizado. A reação de ligação foi montada para um volume de 10  $\mu$ L em

um microtubo com: 0,5 µL (250 ng) do oligonucleotídeo sintético (dPLK); 7,5 µL (~52,5 ng) do vetor pDM (digerido com as endonucleases *NdeI* e *BamHI* e purificado); 1 µL de tampão 10X de T4 ligase e 1 µL da enzima T4 DNA ligase (200U/µL). A reação foi incubada por 16°C durante a noite. Após a reação de ligação, o volume final de 10 µL após no microtubo foi submetido a concentração de DNA por precipitação (ANEXO 7).

O pellet de DNA foi redissolvido em água e 1 µL do sistema de ligação purificado foi misturado com 50 µL de células *E. coli* DH5α eletrocompetentes e transformadas geneticamente pela técnica de eletroporação (ANEXO 3). Alíquotas do sistema foram semeados em placas de Petri com LB sólido contendo canamicina e glicose. As placas foram incubadas invertidas a 37°C durante 16 horas.

Foram selecionados clones de uma placa de Petri com *E. coli* DH5α contendo o novo vetor para extração dos plasmídeos. Os clones foram inoculados em pré-inóculos com canamicina e glicose por 18 horas e seguiu-se para extração plasmidial (ANEXO 5). Os DNAs plasmidiais dos novos clones extraídos foram submetidos à eletroforese em gel de agarose 0,8%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador. Os novos vetores com o *polylinker* passaram a se chamar de pDM02.

#### 4.3.7.4 O vetor pDMU02 (alta cópia)

Para a construção do vetor caça promotor de alta cópia, o cassete de expressão do vetor pCDM deveria ser clonado no vetor pRIQT (4.2.2.3) utilizando as endonucleases *EcoRI* e *PstI*. O vetor pCDM possui os sítios de restrições da *EcoRI* na posição 181 e *PstI* na posição 1196 (5' → 3') e quando é digerido com estas endonucleases libera um fragmento que corresponde ao cassete de expressão (1015 pb). O vetor pRIQT possui os mesmos sítios das endonucleases (*EcoRI* e *PstI*) nas posições 326 e 1565 (5' - 3'). A digestão dos vetores pCDM e pRIQT com as endonucleases *EcoRI* e *PstI* permite a transferência do cassete de expressão gênica (com estratégia para clonagem de promotores) de um vetor de baixa cópia (pCDM) para um backbone de alta cópia (pRIQT).

O plasmídeo pRIQT foi misturado com 50 µL de células *E. coli* JM110 eletrocompetentes e transformado geneticamente pela técnica de eletroporação (ANEXO 3). Em seguida, diferentes alíquotas dos sistemas foram semeadas em placas de Petri com meio LB sólido suplementado somente com ampicilina. As placas foram incubadas invertidas a 37°C durante 16 horas.

Clones recombinantes de *E. coli* JM110 portadores dos vetores de expressão pCDM e pRIQT foram selecionadas para pré-inóculos contendo canamicina e glicose e ampicilina, respectivamente. Os pré-inóculos foram incubados sob agitação constante (140 rpm) por um período de 14 horas a 37°C e após o crescimento, seus DNAs plasmidiais dos clones foram extraídos (ANEXO 5), o resultado foi submetido à eletroforese em gel de agarose 0,8%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador. Os DNAs dos clones extraídos foram guardados a -20°C até o seu uso novamente.

Foram escolhidas amostras de DNAs plasmidiais dos vetores pRIQT e pCDM com as endonucleases *EcoRI* e *PstI*. Foram montadas as reações dos vetores pRIQT e pCDM em microtubos para um volume de 50 µL e 20 µL, respectivamente. Os componentes da reação estão especificados na tabela 9 a seguir.

**Tabela 9: Reação de restrição com as endonucleases *EcoRI* e *PstI* dos vetores pCDM e pRIQT.**

<b>Componentes</b>	<b>pRIQT com enzima</b>	<b>pCDM com enzima</b>
H2O	25	1
Tâmpão 3.1 10X	5	2
DNA plasmidial	15	15
<i>EcoRI</i> (10U/µL)	2,5	1
<i>PstI</i> (10U/µL)	2,5	1
V. F.	50 µL	20 µL

As reações foram incubadas por 3 horas a 37°C em banho seco. Após transpassado o tempo, foi adicionado 15 µL de TBE 5x azul no volume restante (45 µL) e o resultado foi submetido à eletroforese em gel de agarose 0,8%. As bandas de interesse foram purificadas por meio do protocolo de purificação do kit (ANEXO 6) e o resultado desta purificação foi novamente submetido à eletroforese em gel de agarose 0,8%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador. Os DNAs plasmidiais foram guardados a -20°C até o seu uso novamente.

A seguir, o vetor pRIQT purificado e o fragmento correspondente ao cassete de expressão digeridos com *EcoRI* e *PstI* foram submetidos a reação de ligação. A reação foi montada para um volume final de 20  $\mu\text{L}$ . Foi adicionado em um microtubo: 7,5  $\mu\text{L}$  do DNA do inserto do cassete de expressão+backbone do pCDM, 10  $\mu\text{L}$  do vetor pRIQT, 2  $\mu\text{L}$  de tampão 10X de T4 ligase e 0,5  $\mu\text{L}$  da enzima T4 DNA ligase (200U/ $\mu\text{L}$ ). A reação foi incubada por 16°C durante a noite. Após a reação de ligação, o volume final de 10  $\mu\text{L}$  no microtubo foi submetido à concentração de DNA por precipitação (ANEXO 7).

O pellet de DNA foi redissolvido em 4  $\mu\text{L}$  de H<sub>2</sub>O e 2  $\mu\text{L}$  do sistema de ligação purificado foram misturados com 50  $\mu\text{L}$  de células *E. coli* DH5 $\alpha$  eletrocompetentes e transformadas geneticamente pela técnica de eletroporação (ANEXO 3). Alíquotas do sistema foram semeados em placas de Petri com LB sólido contendo ampicilina e glicose. As placas foram incubadas invertidas a 37°C durante 16 horas.

Foram selecionados novos clones *E. coli* DH5 $\alpha$  de uma placa de Petri contendo o novo vetor para extração dos plasmídeos e inoculados em pré-inóculos com ampicilina e glicose por 18 horas. Após, seguiu-se para extração plasmidial (ANEXO 5) e análise em gel de agarose 0,8%. O resultado foi corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador. O novo vetor passou a se chamar de pDMU02.

#### **4.3.8 Análise da regulação e expressão do gene da GFP através do conjunto de vetores regulados**

Para a análise quantitativa da expressão e regulação do gene da GFP com a produção da proteína sob o controle promotor sintético regulado DM, clones recombinantes de *E. coli* JM110 ou DH5 $\alpha$ F'IQ programadas com os plasmídeos recombinantes pDM-GFP, pDMU01, pCDM e pDMU02 foram submetidos a diferentes condições de cultivo e indução em meio LB líquido. Foram inoculados inicialmente em pré-inóculos contendo glicose e o antibiótico seletivo canamicina para os sistemas pDM-GFP+*E. coli* JM110 e pCDM+ *E. coli* JM110, e ampicilina e glicose para os sistemas *E. coli* DH5 $\alpha$ F'IQ+pDMU01 e *E. coli* DH5 $\alpha$ F'IQ+pDMU02. As culturas foram incubadas durante a noite a 37°C sob agitação de 150 rpm.

No dia seguinte, foram inoculados 1000  $\mu\text{L}$  dos pré-inóculos em Erlenmeyers de 250 mL estéreis e fechados contendo 50 mL de meio LB líquido nas seguintes condições: um sistema com somente canamicina ou ampicilina, um sistema com canamicina ou ampicilina

suplementado com glicose e um sistema com canamicina e o indutor IPTG. As culturas foram então incubadas à 37°C com agitação de 150 rpm. Nos intervalos de 2, 4, 6, 8, 10, 12, 24 e 30 horas, alíquotas de 1 mL foram coletadas para mensurar o crescimento celular e fluorescência.

Para mensurar o crescimento celular, uma alíquota de 200 µL de cada cultura dos clones nos sistemas, foi adicionada em microplacas de 96 poços de fundo transparente. Para avaliação da fluorescência, as alíquotas restantes (800 µL) foram submetidas a centrifugação por 3 minutos, o sobrenadante foi descartado e adicionado o mesmo volume de tampão TN. O sedimento celular restante (pellet) foi ressuscitado e dessas suspensões foram retiradas alíquotas de 200 µL e transferidas para placa de culturas de 96 poços de cor negra. As leituras de fluorescência foram realizadas com excitação de 395 nm e emissão a 509 nm e o crescimento bacteriano a partir da absorbância a 600nm em espectrofluorímetro *Hidex Chameleon Plate V*.

Após 8 horas de cultivo foi necessária à diluição a cultivo de células para a leitura de ABS e fluorescência, sendo então diluído 1:9. O resultado foi analisado no software *IBM SPSS*, com significância de 0,05 e os gráficos elaborados.

#### **4.3.9 Análise da regulação e expressão da GFP em *E. coli* por SDS-PAGE**

##### 4.3.9.1 Avaliação da expressão e regulação da GFP pelo vetor pDM-GFP

Para a análise qualitativa da expressão e regulação do gene da GFP com a produção da proteína verde fluorescente (GFP) sob o controle promotor sintético regulado DM, clones de *E. coli* JM110 programadas com o plasmídeo recombinante de baixa cópia pDM-GFP foram submetidos a diferentes condições de cultivo e indução em meio contendo LB sólido em placas de Petri.

Após a análise qualitativa em placas de Petri, prosseguiu-se para a análise quantitativa da expressão e regulação do gene da GFP em diferentes condições de cultivo e indução em meio líquido. Para isso, um clone recombinante portador do vetor pDM-GFP foi inoculado em pré-inóculo contendo glicose e o antibiótico seletivo canamicina. O pré-inóculo foi incubado sob agitação constante (150 rpm) por um período de 14 horas a 37°C.

No dia seguinte, foi inoculado 500  $\mu$ L do pré-inóculo em triplicatas de Erlenmeyers de 250 mL contendo 50 mL de meio LB líquido separadamente nas seguintes condições: três somente com canamicina, três com canamicina e glicose e três com canamicina e o indutor IPTG. As culturas foram então incubadas por 8 horas à 37°C com agitação de 150 rpm. Em intervalos de 2 horas, alíquotas foram coletadas para análise das proteínas por eletroforese em gel desnaturante de poliacrilamida (SDS-PAGE).

Para avaliação da fluorescência emitida pela GFP, alíquotas de 1 mL de cada cultura eram coletadas e centrifugadas a 12.000g por 3 minutos e o sobrenadante descartado. Os sedimentos celulares restantes (pellet) foram ressuspensos em 1 mL de tampão TN e dessa suspensão celular, 100  $\mu$ L foram transferidos para placa de culturas de 96 poços de cor negra. A fluorescência emitida foi medida em espectro fluorímetro *Hidex Chameleon* (comprimento de onda de excitação de 340 nm e de emissão de 535nm). Durante o desenvolver das culturas a produção de GFP era acompanhada também visualizando as células bacterianas por meio de microscopia de fluorescência.

No intervalo de 8 horas, alíquotas de 1 mL foram coletadas para mensurar a produção das proteínas GFP programado vetor pDM-GFP na hospedeira *E. coli* JM110 por eletroforese em gel desnaturante de poliacrilamida 12% (SDS-PAGE).

Como o fluoróforo da GFP é tridimensional, para ser possível sua visualização no gel de poliacrilamida por meio da fluorescência, após a solubilização das proteínas celulares no tampão de amostras para o processo eletroforético, essa solução não foi fervida, para que a GFP não desnaturasse. O gel contendo as proteínas foi visualizado em UV antes de ser corado.

#### 4.3.9.2 Avaliação da expressão e regulação do gene da GFP pelo vetor pDMU01

Para a analisar qualitativamente a expressão e regulação do gene da GFP com a produção da proteína verde fluorescente (GFP) sob o controle promotor sintético regulado DM, um clone de *E. coli* JM110 programada com o plasmídeo recombinante de alta cópia pDMU01 foi submetido a diferentes condições de cultivo e indução em meio contendo LB em placas de Petri.

Após a análise qualitativa em placas de Petri, prosseguiu-se para a análise quantitativa da expressão e regulação do gene da GFP em diferentes condições de cultivo.

Para isso, um clone recombinante portador do vetor pDMU01 foi inoculado em pré-inóculo contendo glicose e o antibiótico seletivo ampicilina. O pré-inóculo foi incubado sob agitação constante (150 rpm) por um período de 14 horas a 37°C.

No dia seguinte, foi inoculado 500 µL do pré-inóculo em triplicatas de Erlenmeyers de 250 mL contendo 50 mL de meio LB líquido separadamente com somente com ampicilina. Após 2 horas de crescimento bacteriano foi adicionada em três frascos a glicose e em outros três frascos o indutor IPTG. As culturas foram então incubadas novamente por 8 horas à 37°C com agitação de 150 rpm.

Para analisar a regulação e a expressão do gene da GFP e produção das proteínas, os pellets (1 mL) coletados nos intervalos selecionados de 3, 6 e 8 horas de crescimento celular e 1, 3 e 5 horas de indução, respectivamente, foram submetidas para análise das proteínas por eletroforese em gel desnaturante de poliacrilamida (SDS-PAGE). Durante o desenvolver das culturas, a produção de GFP era visualizada por meio de microscopia de fluorescência.

#### **4.3.10 Construção do plasmídeo pDM-TEK e expressão do gene híbrido da tiorredoxina humana e enteroquinase bovina em *E. coli***

Visando analisar o funcionamento do promotor DM com a expressão de outra sequência codificante, selecionou-se o gene híbrido da tiorredoxina humana com o gene da enteroquinase bovina (TEK), uma vez que a enteroquinase, além de ser tóxica por ser uma protease, tem diversas aplicações industriais e alto valor de mercado.

A indução da expressão do gene da proteína híbrida TEK por *E. coli* JM110 programada com o plasmídeo pDM-TEK, com adição de IPTG, foi realizada quando as culturas atingiram a absorvância em 600 nm igual a 0,1. A partir deste momento, alíquotas de 2 mLs dos clones de *E. coli* portadores de plasmídeos pDM-TEK foram retiradas a cada 2 horas até o tempo total de 8 horas para as leituras de absorvância a 600nm (elaboração das curvas de crescimento) e também para análise das proteínas recombinantes produzidas em eletroforese em gel de poliacrilamida 12% (SDS-PAGE).

Este trabalho intitulado “EXPRESSÃO HETERÓLOGA DA ENZIMA ENTEROQUINASE EM *Escherichia coli*” foi realizado pela discente Kerollen Runa Pinto (PINTO., 2017).

### 4.3.11 Expressão do rtGH em *E. coli* utilizando o vetor pDMU01

#### 4.3.11.1 Clonagem do gene tGH no vetor pDMU01

O plasmídeo pGSM-tGH (4.2.2.3) possui o gene tGH (584 pb) entre os sítios das endonucleases de restrição *NdeI* e *BamHI* e o vetor pDMU01 possui o gene da GFP (723 pb) flanqueado pelos mesmos sítios de restrição. Para a clonagem do tGH pela troca do gene da GFP no vetor pDMU01, os vetores pDMU01 e pGSM-tGH deveriam ser digeridos com as duas endonucleases *NdeI* e *BamHI* e o gene tGH inserido no vetor pDMU01 desprovido da sequência codificante da GFP.

Alíquotas de 1 µl dos plasmídeos pDMU01 e pGSM-tGH foram misturados separadamente com 50 µL de células *E. coli* JM110 quimiocompetentes e transformadas geneticamente por choque térmico (ANEXO 4). Em seguida, diferentes alíquotas dos sistemas foram semeadas em placas de Petri com meio LB sólido suplementado contendo glicose e ampicilina para seleção de clones portadores dos vetores. As placas foram incubadas invertidas a 37°C durante 16 horas.

Clones recombinantes isolados de *E. coli* JM110 portadores dos vetores pDMU01 e pGSM-tGH foram selecionadas para pré-inóculos contendo glicose e ampicilina. Os pré-inóculos foram incubados sob agitação constante (140 rpm) por um período de 14 horas a 37°C e seus plasmídeos extraídos (ANEXO 5). Os resultados da extração dos vetores pDMU01 e pGSM-tGH foram submetidos à eletroforese em gel de agarose 0,8%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador. Os DNAs plasmidiais dos clones extraídos foram guardados a -20°C até o seu uso novamente.

Para a clonagem do gene tGH ao vetor pDMU01, amostras de DNAs plasmidiais dos vetores pGSM-tGH e pDMU01 foram submetidas com as endonucleases *NdeI* e *BamHI*. Foram montadas as reações em microtubos para um volume de 50 µL. Os componentes da reação estão especificados na tabela 10 a seguir.

**Tabela 10: Reação dos vetores pDMU01 e pGSM-tGH com as endonucleases *NdeI* e *BamHI*.**

Componentes	pGSM-tGH com enzima	pDMU01 com enzima
H2O	-	15

Tampão 3.1 10X	5	5
DNA plasmidial	40	25
<i>Nde</i> I (10U/ $\mu$ L)	2,5	2,5
<i>Bam</i> HI (10U/ $\mu$ L)	2,5	2,5
V. F.	50 $\mu$ L	50 $\mu$ L

A reação foi incubada por 1 hora e 30 minutos a 37°C em banho seco. Após a digestão, o volume foi submetido a eletroforese em gel de agarose 0,8% e as bandas de interesse foram extraídas por meio do protocolo de purificação (**ANEXO 6**). O resultado da purificação foi submetido à eletroforese em gel de agarose 0,8%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador.

A seguir, o vetor pDMU01 linearizado (desprovido do gene da GFP) e o fragmento correspondente ao gene tGH digeridos com *Nde*I e *Bam*HI e purificados, foram submetidos a reação de ligação com a T4 DDNA ligase. A reação foi montada para um volume de 10  $\mu$ L. Foi adicionado em um microtubo: 4,5  $\mu$ L de H<sub>2</sub>O, 2  $\mu$ L (30ng) do DNA do vetor pDMU01, 2  $\mu$ L (15 ng) do fragmento tGH, 1  $\mu$ L de tampão 10X de T4 ligase e 0,5  $\mu$ L da enzima T4 DNA ligase (200U/ $\mu$ L). A reação foi incubada por 16°C durante a noite.

Após a reação de ligação, o volume final de 10  $\mu$ L no microtubo foi submetido à concentração de DNA por precipitação (**ANEXO 7**). Após o tempo da reação, 1,5  $\mu$ L do sistema de ligação purificado foram misturados com 50  $\mu$ L de células *E. coli* JM110 eletrocompetentes e transformadas geneticamente pela técnica de eletroporação (**ANEXO 3**). Alíquotas do sistema foram semeadas em placas de Petri com LB sólido contendo ampicilina e glicose. As placas foram incubadas invertidas a 37°C durante 16 horas.

Foi selecionado um clone recombinante de uma placa de Petri contendo a hospedeira *E. coli* JM110 transformante com o novo vetor. Este clone foi submetido a um pré-inóculo com ampicilina e glicose por 18 horas. Após, seguiu-se para extração plasmidial (**ANEXO 5**) e o resultado foi submetido à eletroforese em gel de agarose 0,8%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador. O novo vetor foi denominado de pDMU-tGH.

Para confirmação da clonagem do gene tGH no vetor pDMU, o DNA plasmidial do vetor pDMU-tGH foi submetido a a análise de restrição com as endonucleases *NdeI* e *BamHI*. As reações foram montadas em microtubos para um volume de 10  $\mu\text{L}$  e os componentes estão especificados na tabela 11 a seguir.

**Tabela 11: Reação do vetor pDMU-tGH com as endonucleases *NdeI* e *BamHI*.**

<b>Componentes</b>	<b>pDMU-tGH sem enzima</b>	<b>pDMU-tGH com enzima</b>
H <sub>2</sub> O	4	3
Tampão 3.1 10X	1	1
DNA plasmidial (~80 ng)	5	5
<i>NdeI</i> (10U/ $\mu\text{L}$ )	-	0,5
<i>BamHI</i> (10U/ $\mu\text{L}$ )	-	0,5
V. F.	10 $\mu\text{L}$	10 $\mu\text{L}$

As reações foram incubadas por 2 horas e 30 minutos a 37°C em banho seco. Após a digestão, o resultado foi submetido à eletroforese em gel de agarose 0,8%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador.

#### 4.3.11.2 Expressão em *E. coli* e análise da proteína heteróloga tGH

Para análise da regulação e expressão do tGH sob o controle do promotor DM, um clone recombinante de *E. coli* JM110 possuindo o vetor pDMU-tGH foi selecionado para análise da regulação e expressão do gene tGH. Uma colônia isolada (JM110+pDMU-tGH) foi inoculada em pré-inóculo contendo glicose e ampicilina. No dia seguinte, foram inoculados 1000  $\mu\text{L}$  de pré-inóculo em dois Erlenmeyers de 250 mL estéreis e fechados contendo 50 mL de meio LB líquido com somente ampicilina. As culturas foram então incubadas à 37°C com agitação de 150 rpm.

O crescimento celular foi acompanhado, e quando a cultura bacteriana atingiu  $\text{OD}_{600\text{nm}} = 0.4$ , o valor foi normalizado para 1,0. Os microtubos foram centrifugados por 1 minuto por 12.000 g e o sobrenadante descartado. Os pellets foram guardados a -80°C até o seu uso

novamente. A partir deste momento considerou-se como tempo zero (0) e após a estocagem dos pellets, imediatamente em um sistema com somente ampicilina foi adicionado a glicose e no outro sistema com ampicilina foi adicionado o indutor IPTG. As culturas foram incubadas à 37°C com agitação de 150 rpm.

Nos intervalos de 10 e 24 horas de crescimento bacteriano e 6 e 20 horas de indução, respectivamente, alíquotas de 1 mL foram coletadas para mensurar o crescimento celular seguido do procedimento de normalização. Os microtubos foram centrifugados por 1 minuto por 12.000 g e o sobrenadante descartado. Os pellets foram estocados a -80°C até o seu uso novamente.

Para analisar a regulação e a produção do tGH, os pellets coletados nos intervalos selecionados de 0, 10 e 24 horas de crescimento celular e 0, 6 e 20 horas de indução foram submetidas para análise das proteínas por eletroforese em gel desnaturante de poliacrilamida (SDS-PAGE).

Foi preparada a solução para o preparo das amostras (pellets) em um falcon estéril. Para tal foi adicionado 5 mL de Tampão A (Tris-HCl pH 8 20mM; NaCl 500mM), 100 µL de Triton100X e 100 µL de PMSF (100mM). O volume da solução foi completado para 10 mL com Tampão A, a solução foi fervida por 10 segundos em forno de micro-ondas e na sequência resfriada em gelo. Em seguida foi adicionado 250 µL da solução em cada amostra (pellet), seguido de homogeneização. As amostras após a ressuspensão foram submetidas à disruptor ultrassônico com potência de 80 watts (1 pulso de 10 seg) em banho de gelo.

Uma alíquota de 40 µL após a sonicação foi transferida para outro microtubo estéril. Neste microtubo foi adicionado 10 µL de tampão de corrida 5X com azul. As amostras foram fervidas por 100°C por 5 minutos. Foram aplicados 15 µL nos poços do gel de acrilamida e o resultado foi analisado por eletroforese SDS-PAGE 12%. A corrida foi realizada em 200 volts, 25 mA e 30 Watts por 40 minutos, seguido de coloração com o corante *Coomassie Blue* (Fisher Scientific, Madrid, Espanha) por 1 hora. Depois foi lavado com água destilada e descolorado por 1 hora. O gel após descoloração foi lavado com H<sub>2</sub>O destilada novamente e procedido à visualização e captura da imagem em software *Labscan Image Master V6*.

#### **4.3.12 Melhoramento do promotor DM por síntese química de DNA randômica**

Existem atualmente muitas técnicas estabelecidas para gerar essas bibliotecas de promotores. O primeiro método desenvolvido a partir de mutações aleatórias, Jensen e Hammer, 1998, construíram uma biblioteca de promotores sintéticos através de alterações das sequências espaçadoras entre as regiões -35 e -10 de promotores de bactérias, mantendo as regiões de consenso intacta. Com isso, construíram promotores artificiais utilizando oligonucleotídeos sintéticos degenerados com o intuito de aumentar o nível de expressão gênica em *Lactococcus lactis*. Como resultado, tiveram sucesso na expressão do gene da luciferase em 25, 50, 200 e 400% acima da expressão normal do gene e começaram a aplicar em outros microrganismos procarióticos, como em *E. coli* e *Lactobacillus plantarum*) e organismos eucarióticos (*Saccharomyces cerevisiae*) (JENSEN & HAMMER, 1998; SANTOS & STEPHANOPOULOS, 2008).

#### 4.3.12.1 Construção da biblioteca de novos promotores sintéticos baseados no promotor DM

Para a obtenção de promotores com a regulação e níveis maiores de expressão, foi realizado o design dos novos promotores sintéticos inspirados na sequência do promotor DM e em promotores fortes de *E. coli* descritas na literatura (MOREIRA et al., 2018). Para a obtenção de regiões mais consensuais no promotor DM, as bases mais eficientemente reconhecidas por RNAP  $\sigma 70$  foram mantidas nos novos promotores sintéticos.

Os oligonucleotídeos foram desenhados conservando as regiões complementares entre si em uma parte da região correspondente ao operador lac e a região do elemento -10 estendido (localizada entre as regiões -35 e -10) e variando-se nucleotídeos de outras posições (vide figuras 12 e 13). A síntese química dos oligonucleotídeos foi realizada pela empresa EXXTEND. O sítio de *EcoRI* junto com *SpeI* foram adicionados nas bordas para ligar os novos promotores com o plasmídeo pCDM, pois eles clivam a região promotora DM do plasmídeo pCDM, permitindo a clonagem de outros novos promotores. O sítio da enzima de restrição *PstI* foi posicionado após o promotor para confirmação da clonagem de novos promotores.

#### 4.3.12.2 Clonagem dos novos promotores no vetor pCDM e Transformação em *E. coli* JM110

Para completar a dupla fita (Figura 12) (anelamento dos oligonucleotídeos através de uma parte da região espaçadora e o elemento - 10 estendido) os oligonucleotídeos pDSM-1

e pDSM-2 foram redissolvidos conforme recomendações do fabricante. Foi realizada a reação de hibridização (anelamento) dos oligonucleotídeos pDSM-1 (2 µg / µL) e o pDSM-2 (1,6 µg / µL) (Figura 12).

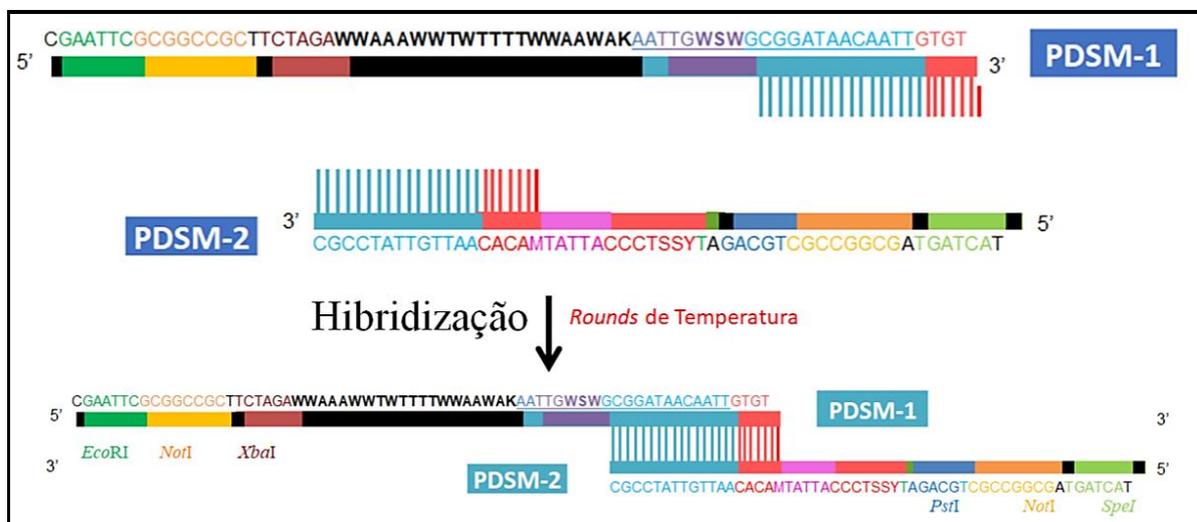


Figura 12: Representação esquemática de hibridização dos oligonucleotídeos através da região espaçadora e do elemento -10 estendido.

Os volumes finais dos sistemas de hibridização foram ajustados para 50 µL. Os componentes da reação de hibridização dos oligonucleotídeos pDSM- 1 e o pDSM-2 estão especificados na tabela 12 a seguir.

Tabela 12: Reação de hibridização dos oligonucleotídeos pDSM-1 e o pDSM-2.

Componentes	pDSM-1 pDSM-2	pDSM-1 pDSM-2	pDSM-1 pDSM-2	pDSM-1 pDSM-2	pDSM-1 pDSM-2
H2O	44	43,5	43	42,5	44
Tampão 2 NEB 10X	5	5	5	5	5
pDSM-1	1	1	1	1	-
pDSM-2	-	0,5	1	1,5	1
V. F.	50 µL				

As reações nos microtubos foram ajustadas nas seguintes condições em termociclador BIO-RAD T100 thermal cycle: 95°C / 2 min., 80 °C / 2 min., 56 °C / 5 min., 45 °C / 2 min., 36 °C / 2 min., 20 °C / 2 min., 4 °C / hold. Os resultados das reações foram

analisados por eletroforese em gel de agarose 2% corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador.

Após a hibridização, realizou-se o preenchimento das regiões fita-simples utilizando a enzima *Klenow* (Figura 13).

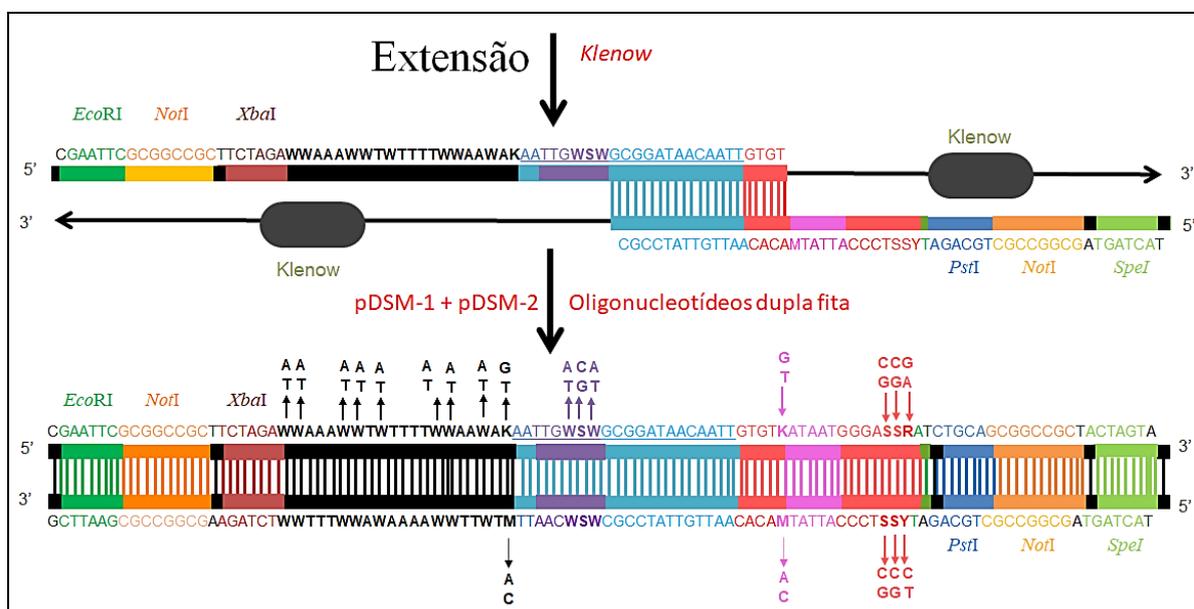


Figura 13: Representação esquemática do preenchimento dos oligonucleotídeos pela enzima *Klenow*.

Os volumes finais dos sistemas foram ajustados para 25  $\mu$ L. Os componentes da reação de preenchimento dos oligonucleotídeos pDSM-1 e o pDSM-2 estão especificados no quadro 3 a seguir:

**Quadro 3:** Reação de preenchimento dos oligonucleotídeos pDSM-1 e o pDSM-2 (hibridizados) utilizando a enzima *Klenow*.

Componentes	Sem Klenow	Com Klenow
Água deionizada filtrada	9,5	7,5
Tampao 2 NEB 10X	2,5	2,5
dNTPs 2,5 mM	3	3
DNA anelados	10	10
Klenow (5U/ $\mu$ L)	-	2
Volume Final	25 $\mu$ l	25 $\mu$ l

As reações nos microtubos foram ajustadas nas seguintes condições em termociclador BIO-RAD T100 thermal cycle: 16°C/ 30 min., 20 °C / 30 min., 75 °C / 20 min., 4 °C / *hold*. Os resultados das reações foram analisados por eletroforese em gel de agarose 2% corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador. A purificação dos oligonucleotídeos dupla fita foi realizada excisando o fragmento do gel com bisturi estéril e adicionando em um microtubo. A seguir, 200µL de água ultrapura estéril foi adicionada no microtubo seguido de maceração do gel de agarose e incubação a 4°C por 24 horas. Após o tempo, o microtubo foi centrifugado por 10 minutos/14.000 rpm, retirou-se o sobrenadante e foi adicionado em outro microtubo estéril. Os DNAs dos oligonucleotídeos foram guardados a -20°C até o seu uso novamente.

Foi escolhida uma amostra do vetor pCDM e os oligonucleotídeos hibridizados e tratados com *Klenow* para análise de restrição com as endonucleases *EcoRI* e *SpeI*. A seguir foi montada a reação do vetor pCDM e dos oligonucleotídeos hibridizados e tratados com *Klenow* em microtubos para um volume de 50 µL. Os componentes estão especificados na tabela 13 a seguir.

**Tabela 13: Reação do vetor pCDM e dos oligonucleotídeos utilizando as endonucleases *EcoRI* e *SpeI*.**

<b>Componentes</b>	<b>pCDM com enzima</b>	<b>Oligonucleotídeos hibridizados e tratados com <i>Klenow</i> com enzima</b>
H2O	33,5	33,5
Tâmpão 2.1 NEB 10X	5	5
DNA (pCDM ou Oligos)	7,5	7,5
<i>EcoRI</i> -HF (20U/µL)	2,5	2,5
<i>SpeI</i> -HF (20U/µL)	1,5	1,5
V. F.	50 µL	50 µL

As reações foram incubadas por 20 minutos a 37°C em banho seco. Após o tempo das reações, foi adicionado 15 µL de TBE 5x azul em cada microtubo e as amostras foram submetidas a eletroforese em gel de agarose 0,8%. A banda de interesse do vetor pCDM foi purificada por meio do protocolo de purificação do kit (**ANEXO 6**) e o resultado foi

novamente submetido à eletroforese em gel de agarose 0,8%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador. Os DNAs plasmidiais linearizados foram guardados a -20°C até o seu uso novamente.

Foi realizada novamente a purificação dos oligonucleotídeos excisando o fragmento do gel de agarose e adicionando em um microtubo com 200µL de água ultrapura estéril seguida de maceração e incubação a 4°C por 24 horas. Após o tempo, o microtubo foi centrifugado por 10 minutos/14.000 rpm, retirou-se o sobrenadante e foi adicionado em outro microtubo estéril. Os DNAs dos oligonucleotídeos foram guardados a -20°C até o seu uso novamente.

Os fragmentos purificados (pCDM e oligonucleotídeos) foram quantificados no Qubit Fluorometric Quantification. A seguir os oligonucleotídeos sintéticos foram ligados aos vetores pCDMs desprovidos dos promotores DM. A reação de ligação foi montada para um volume final de 40 µL em um microtubo com: 15 µL (2,1 ng) do oligonucleotídeos sintéticos; 20 µL (30 ng) do vetor pCDM; 4 µL de tampão 10X de T4 ligase e 1 µL da enzima T4 DNA ligase (200U/µL), e incubada durante a noite a 16°C. Após a reação de ligação, o DNA contido no sistema foi concentrado por precipitação (**ANEXO 7**).

O pellet de DNA foi redissolvido em água deionizada e 1 µL do sistema de ligação purificado foi misturado com 50 µL de células *E. coli* JM110 eletrocompetentes e transformadas geneticamente pela técnica de eletroporação (**ANEXO 3**). Alíquotas do sistema foram semeados em placas de Petri com LB sólido contendo canamicina e glicose. As placas foram incubadas invertidas a 37°C durante 16 horas.

4.3.12.3 Análise da regulação e expressão do gene da GFP através dos diferentes promotores

Para analisar a produção de GFP dos novos clones contendo os diferentes promotores, foram selecionadas as colônias recombinantes e realizado a replica plate em três diferentes condições com: 1) glicose + antibiótico canamicina; 2) somente o antibiótico canamicina e 3) com o antibiótico canamicina + indutor IPTG.

A partir das análises qualitativas das réplicas plate, foram selecionados os clones que expressaram o gene da GFP quando suplementado com glicose e com maior intensidade quando induzido por IPTG; e os clones que mantiveram a regulação em glicose e

expressaram o gene da GFP com mais intensidade que os clones controle com o pCDM. Os clones selecionados inoculados com canamicina e glicose e incubados a 37°C no agitador durante a noite. Seguiu-se para extração plasmidial dos clones e os resultados foram submetidos à eletroforese em gel de agarose 0,8%, corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador. Após a extração plasmidial dos novos clones, estes foram submetidos com a endonuclease *Pst*I para confirmação das clonagens dos novos promotores. A partir deste momento, alguns clones foram selecionados para determinação das novas sequências promotoras.

#### 4.3.12.4 Sequenciamento e Bioinformática das novas sequências promotoras

Para se obter as novas sequências promotoras sob diferentes níveis de regulação e expressão do gene da GFP dos clones, os DNAs plasmidiais extraídos dos diferentes clones foram quantificados em Qubit fluorometric. Foi realizado a reação de PCR das sequências dos promotores dos vetores utilizando o par de primers iniciadores: VF1 e RCP1 (Quadro 2) para amplificar 504 pares, onde inclui a região dos novos promotores. A PCR foi realizada com o kit da Platinum *Taq* DNA Polymerase High Fidelity (*Thermo Fisher Scientific Inc*) conforme orientações do fabricante. A concentração dos DNAs plasmidiais foram ajustados para 6 ng/μl para as reações de PCR. Os resultados das reações foram analisados por eletroforese em gel de agarose 2% corado com brometo de etídio e visualizado sob luz ultravioleta (UV) em fotodocumentador. Após a PCR, os amplicons foram submetidos a reação de sequenciamento.

As reações de sequenciamento foram feitas utilizando o kit Sequencing-Big Dye Terminator Cycle Sequencing Ready ABI versão 3 – Applied Bioscience/Life Technologies, USA, seguindo as recomendações do fabricante. Cada reação de sequenciamento foi ajustada para um volume final de 10 μL, com o DNA amplificado, os oligonucleotídeos iniciadores VF1 ou RCP, e tampão Big-Dye para 1X final e 2 μL Big-Dye. As condições de reação em termociclador foram programadas: 96°C / 1 min.; 25 ciclos (96°C / 15seg, 50 °C / 15seg. 60 °C / 1,30 min.) e as leituras de fluorescência realizadas no sequenciador ABI, modelo 3500XL (Applied Biosystems).

Os dados brutos provenientes do sequenciador de DNA foram submetidos e analisados pelo software Sequencing Analysis versão 5.4. e a seguir, para verificar qualidade das sequências nucleotídicas foi realizado procedimento via PHRED disponível no site da

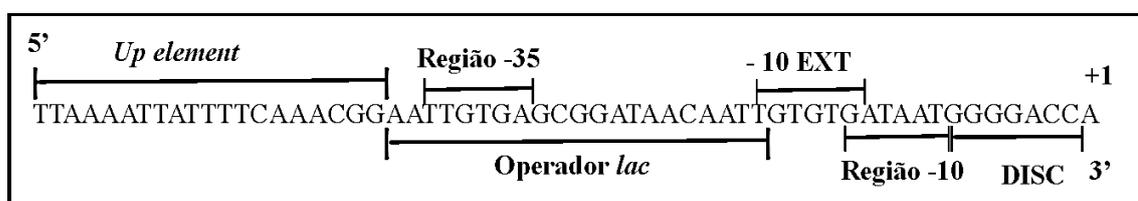
UnB (<http://www.biomol.unb.br/phph/index.html>). As sequências foram alinhadas utilizando o programa *BioEdit* (disponível no site: [www.mbio.ncsu.edu/bioedit/bioedit.html](http://www.mbio.ncsu.edu/bioedit/bioedit.html)) para identificação das regiões com função promotora.

## 5 RESULTADOS E DISCUSSÃO

### 5.1 DESIGN DO PROMOTOR SINTÉTICO, CLONAGEM NO CASSETE DE EXPRESSÃO E SÍNTESE DO pDM-GFP

O promotor DM foi construído para reconhecimento em hospedeiras *E. coli* pela maquinaria RNAP associada com o  $\sigma 70$ . Foi adicionada a sequência do elemento UP distal e proximal na região -58 a -39 (TTAAAATTATTTTCAAACGG) e o operador lac O1 logo após a sequência de elemento UP. Como descrito, o promotor DM possui o operador lac (lacO) do sistema do operon lac sobreposto à região conservada -35 e a região espaçadora preferencial com 17 nucleotídeos completada com o -10 EXT (TGTGTG) entre a região -35 e a -10 (TATA ou *Pribnow box*) do promotor. Portanto, a região -35 no presente design apresenta a sequência TTGTGA, ao invés da sequência consensual TTGACA. Dessa forma, diferentemente do promotor lac natural, no promotor sintético DM a região de reconhecimento -35 da subunidade sigma 70 está sobreposta à região operadora, onde se acopla a proteína repressora LacI, aprimorando consequentemente a eficiência de bloqueio da transcrição deste complexo promotor-operador- regulador.

Essa estratégia faria com que o promotor operasse com bom nível de regulação por meio do gene *lacIq* que codifica a proteína repressora LacI em altos níveis. Para completar os 17 nucleotídeos preferenciais da região espaçadora entre as regiões -35 e -10, foi adicionado o elemento TGTG antes da região -10, considerado uma extensão da mesma, utilizado em promotores que diferente de suas sequências consensos (5' TTGACA - 17pb – TATAAT 3') capaz de torná-la mais eficiente e potencializar a força da expressão. Foi adicionada à sequência discriminadora na posição -7 a -1 (GGGGACC), sendo o -7 GGG -5 consenso importante na estabilidade do Rpo e inserido a Adenina na posição +1 (Figura 14) (HOOK-BARNARD & HINTON, 2007; RUFF et al., 2015).



**Figura 14: Sequência do promotor DM.** No sentido da esquerda para direita (5' → 3') estão representadas as sequências de DNA que compõem o promotor, como o elemento UP (Up element), a sequência da região -35 (TTGTGA) sobreposta na sequência completa do operador lacO, a sequência do

elemento -10 estendido TGTGTG (-10 EXT), a região -10 (GATAAT) e a região discriminadora GGGGACC (DISC) que antecede o ponto de início de transcrição representado pelo nucleotídeo A (+1).

Para a construção de um vetor de expressão para *E. coli*, a montante do sítio de início da transcrição ou +1, foram adicionados a sequência da região 5' UTR do mRNA do gene da proteína 10 do bacteriófago T7, que contém um eficiente RBS (AAGAAGGAGA); o gene repórter da GFP e o terminador de transcrição T1 do gene de *rrnB* de *E. coli* para interromper a síntese do mRNA (CHALFIE et al., 1994; OLINS & RANGWALA, 1989; OROSZ et al., 1991).

O promotor DM foi engenheirado para ser muito bem regulado (com *performance*), posicionado a montante do cassete de expressão a sequência codificadora do RBS, o gene repórter da proteína verde fluorescente - GFP flanqueado por diversos sítios de enzimas de restrição para permitir sua fácil transferência a outros vetores ou para troca de partes genéticas para estudos de expressão e regulação. Após sua concepção, o polinucleotídeo foi desenhado com cerca de 1015 nucleotídeos. A figura 19 mostra a sequência completa de nucleotídeos do cassete de expressão (promotor DM + GFP + terminadores de transcrição) (figura 15).

```

CGAATTCGCGGCCGCTTCTAGATTAAAATTATTTTCAAACGGAAATTGTGAGCGGA
TAACAATGTGTGATAATGGGGACCATACTAGTACTAGAAATAATTTTGTTTAAC
TTTAAGAAGGAGATATACATATGGCTAGCAAAGGAGAAGAACTTTTCACTGGAG
TTGTCCCAATTCTTGTGAATTAGATGGTGTATGTTAATGGGCACAAATTTTCTGTC
AGTGGAGAGGGTGAAGGTGATGCTACATACGGAAAACCTACCCTTAAATTTATTI
GCACTACTGGAAAACCTACCTGTTCCATGGCCAACACTTGTCACTACTTTCTCTTAT
GGTGTTC AATGCTTTTCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAA
GAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATACTTTCAAAGATGAC
GGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATC
GTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTCGGACACAA
ACTGGAGTACA ACTATACTCACACAATGTATACATCACGGCAGACAAACAAA
GAATGGAATCAAAGCTAACTTCAA AATTCGCCACAACATTGAAGATGGCTCCGTT
CAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTT
TACCAGACAACCATTACCTGTCTACACAATCTGCCCTTTCGAAAGATCCCAACGA
AAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACAT
GGCATGGATGAGCTTACAAATAATGGATCCACTAGAGCCAGGCATCAAATAAA
ACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTATCTGTTGTTTGTTCGGTGA
ACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTAT
ATACTAGTAGCGGCCGCTGCAGC

```

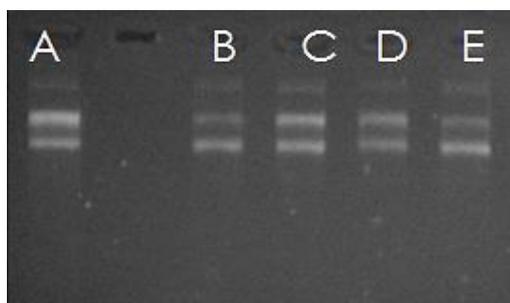
**Figura 15: Sequência nucleotídica promotora DM contida no vetor de expressão, o gene GFP e o terminador de transcrição.** Da esquerda para a direita (5' → 3') está a região em verde escuro que representa o prefixo dos sítios das endonucleases *EcoRI*, *NotI*, *XbaI*; a região em negro mostra a sequência do elemento UP; a região em azul claro mostra a sequência original do operador *lac* O1 com a região -35 sobreposta (em amarelo); a sequência elemento -10 estendida em vermelho; a região -10 em rosa, a região



A presente sequência promotora DM situado no vetor de expressão caracteriza-se por uma invenção e motivo de patente (BR 10 2019 000470 3). O cassette de expressão situado no vetor apresenta: um promotor com média força e regulado com a sequência operadora do sistema lacO sobreposta na região -35; um sítio de ligação do ribossomo (RBS) proveniente do gene 10 do fago T7 (OLINS et al., 1988); a sequência do gene repórter da proteína verde fluorescente (Green Fluorescent Protein – GFP) e o terminador de transcrição forte derivado do gene T1 de *rrnB* de *E. coli* (OROSZ et al., 1991). O vetor ainda contém uma Ori (p15A) de baixo número de cópias (10~12 cópias por célula) (COZZARELLI et al., 1968; CHANG & COHEN, 1978) e o gene de resistência da Canamicina originário do transposon Tn903 (GRINDLEY & JOYCE, 1980).

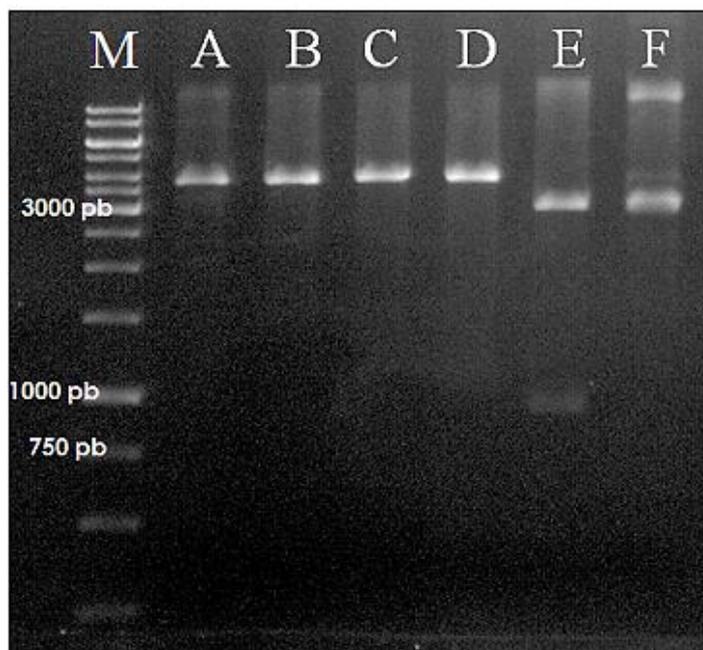
## 5.2 ANÁLISE DO VETOR pDM-GFP COM ENDONUCLEASES DE RESTRIÇÃO

O vetor pDM-GFP foi introduzido em células *E. coli* JM110 eletrocompetentes e transformado pela técnica de eletroporação. Foram extraídos após 16-18 horas de incubação, colônias recombinantes geneticamente isoladas para pré-inóculos com canamicina. Os pré-inóculos foram incubados durante a noite e no outro dia seguiu-se para a extração dos plasmídeos pDM-GFP (ANEXO 5). Foram extraídos DNAs plasmidiais do pDM-GFP de 5 clones selecionados como mostrados na figura 17 a seguir.



**Figura 17: Fotografia em gel agarose dos plasmídeos pDM-GFP extraídos.** As amostras A, B, C, D e E mostram os perfis eletroforéticos extraídos do vetor pDM-GFP.

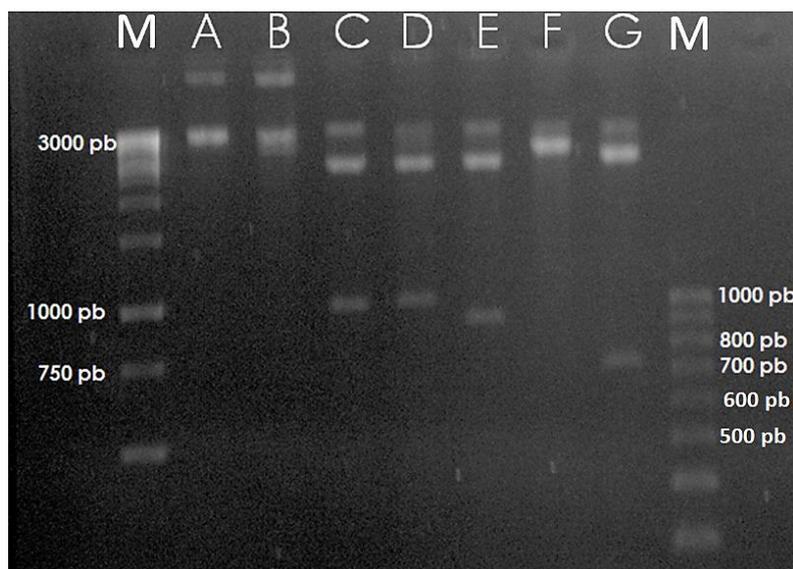
A seguir, uma amostra do plasmídeo pDM-GFP extraído, foi tratado com as endonucleases *EcoRI*, *BamHI*, *NdeI*, *PstI* e *SpeI* para confirmação dos sítios de restrição. O resultado foi analisado em gel de agarose e visualizado sob luz UV como mostra na fotografia 18 a seguir.



**Figura 18: Análise eletroforética do cassete de expressão do vetor pDM02 com endonucleases.** A amostra A identifica o plasmídeo pDM-GFP com *EcoRI*; a amostra B está o pDM-GFP com *BamHI*; a amostra C o pDM-GFP com *NdeI*; a amostra D o pDM-GFP com *PstI*; a amostra E mostra o pDM-GFP com *SpeI* e a amostra F identifica o plasmídeo pDM-GFP sem endonuclease.

Como esperado, não houve a liberação de insertos do pDM-GFP quando este foi tratado separadamente com as endonucleases *EcoRI*, *BamHI*, *NdeI* e *PstI* (canaletas **A**, **B**, **C** e **D**) respectivamente. Foi observada a liberação de um fragmento de aproximadamente 920 nucleotídeos correspondente ao *Shine-Dalgarno* + GFP + terminador de transcrição quando tratado com *SpeI* (canaleta **E**), confirmando a presença de dois sítios da endonuclease no vetor. A canaleta **F** corresponde ao pDM-GPF sem endonuclease.

Os plasmídeos pDM-GFP foram submetidos com as endonucleases: *EcoRV*; *NotI*; e com *EcoRI* e *PstI*; *EcoRI* e *SpeI*; *EcoRI* e *NdeI*; *NdeI* e *BamHI* para confirmar se os sítios desenhados estariam no vetor através dos fragmentos que deveriam ser liberados. Após a digestão, o resultado foi analisado em gel de agarose e visualizado sob luz UV como mostra na fotografia 19 a seguir.



**Figura 19: Análise eletroforética do perfil de restrição do vetor pDM-GFP com endonucleases.** A amostra A está o pDM-GFP sem endonuclease; a amostra B exemplifica o pDM-GFP com *EcoRV*; a amostra C do pDM-GFP com *NotI* e a amostra D o pDM-GFP com *EcoRI* e *PstI*. Na canaleta E, o vetor pDM-GFP com *EcoRI* e *SpeI*. Na amostra F está o vetor pDM-GFP com *EcoRI* e *NdeI* e na amostra G o pDM-GFP com *NdeI* e *BamHI*.

O resultado da análise de restrição mostra que os perfis de restrição identificados no gel são compatíveis com o esperado pelo resultado da figura 16. A canaleta **A** exemplifica o pDM-GFP sem endonuclease (controle). Na canaleta **B**, o plasmídeo pDM-GFP com a *EcoRV* não mostrou o perfil do fragmento do cassete de expressão.

Na canaleta **C** e **D** houve a liberação do fragmento correspondente à sequência completa do cassete de expressão (promotor DM + GFP + terminador de transcrição) quando este tratado com a endonuclease *NotI* e na combinação *EcoRI* e *PstI*. No poço **E**, ocorreu a liberação de dois fragmentos correspondente a sequência RBS + GFP + terminador de transcrição (912 pb) e o backbone plasmidial (2842 pb) quando o pDM-GFP foi tratado com as endonucleases *EcoRI* e *SpeI*. Não foi possível observar o fragmento correspondente ao promotor DM no gel, devido ao tamanho do fragmento ser bem menor e a concentração do gel ser de 0,8%, ocorrendo uma maior migração e difusão para estes tipos de fragmentos pequenos e menor coloração com brometo de etídio.

Na canaleta **F**, correspondente ao pDM-GFP com as endonucleases *EcoRI* e *NdeI*, e foi possível visualizar um fragmento correspondente ao vetor desprovido do promotor DM + RBS. A amostra **G** foi tratada com as endonucleases *NdeI* e *BamHI* apresentando um fragmento de 723 pares de bases (GFP + terminador de transcrição).

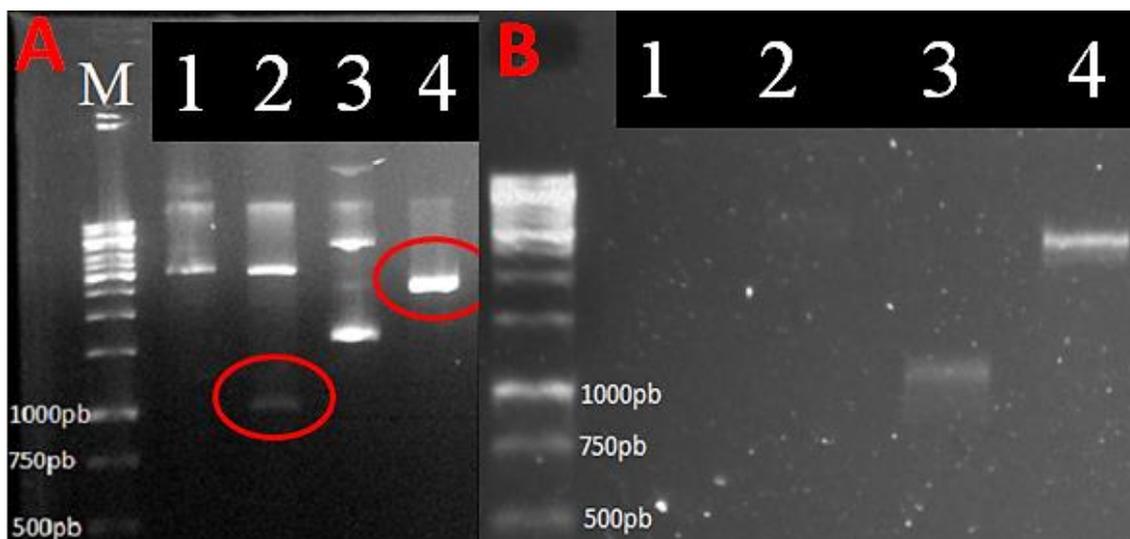
### 5.3 CONSTRUÇÃO DE UM CONJUNTO DE VETORES BASEADO NO PROMOTOR DM PARA *E. coli*

Estudos de expressão do gene da GFP contido no pDM-GFP (baixa cópia) (BR 10 2019 000470 3) e do vetor pDM-TEK (baixa cópia (PINTO, 2017) que possui o gene heterólogo que codifica a proteína híbrida entre tiorredoxina humana e enteroquinase bovina (TEK) clonado no lugar da GFP, demonstram que o promotor DM é capaz de programar a expressão de proteínas heterólogas em *E. coli* JM110 (possui o gene lacIq) de forma bem regulada pelo operador lacO e de resultar em níveis médios e altos de expressão das proteínas mencionadas.

Com base nos resultados, realizamos a construção de um conjunto de vetores sob o controle do promotor DM com novas estratégias para expressão e regulação de proteínas. O cassete de expressão foi inserido em diferentes backbones plasmidiais, são eles: o vetor pDM02 (vetor com *polylinker*), pDMU01 (vetor de alta cópia), pCDM (vetor de baixa cópia) e pDMU02 (vetor de alta cópia).

#### 5.3.1 pDMU01

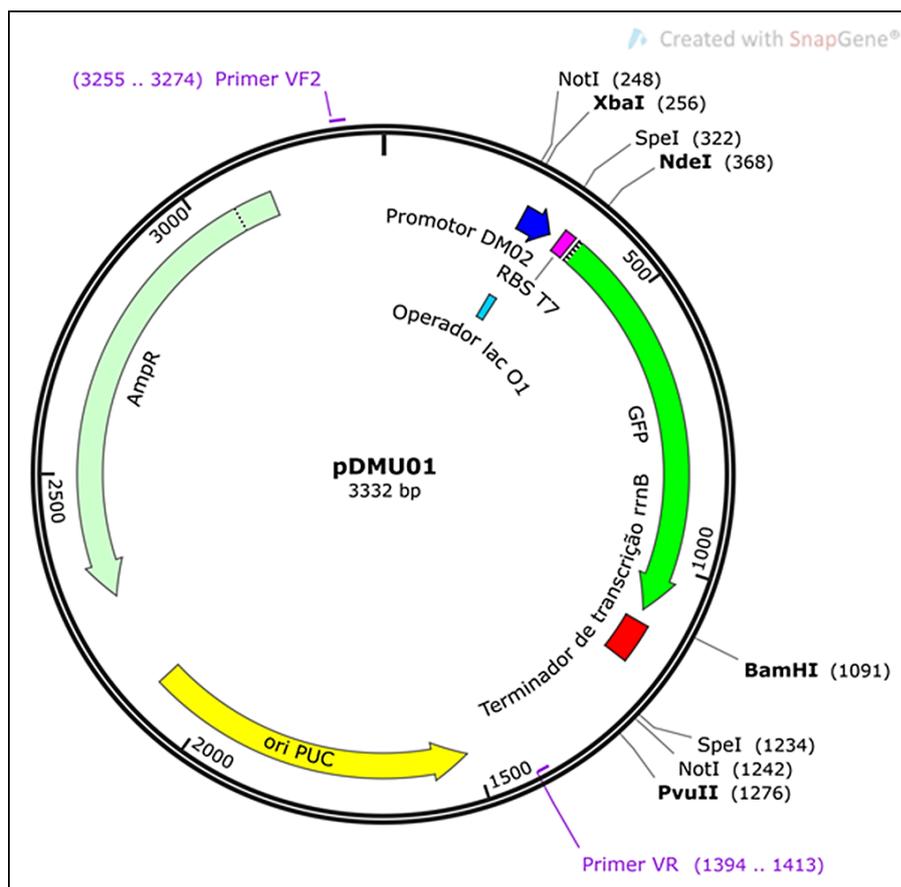
Dois clones de *E. coli* JM110 com os vetores pDM-GFP e pUNA foram escolhidos para pré-inóculos e seus DNAs plasmidiais extraídos. Para a clonagem do cassete de expressão do vetor pDM-GFP no vetor pUNA, uma amostra de DNA plasmidial do vetor pDM-GFP e do vetor pUNA, foram digeridos com a enzima de restrição *NotI* em sistema analítico. Houve a liberação do inserto do cassete de expressão correspondente a 994 pb quando o vetor pDM-GFP e a linearização do vetor pUNA quando tratados com a enzima *NotI*. Após o resultado da análise de restrição e a eletroforese, os fragmentos correspondentes aos DNAs do cassete de expressão e do vetor pUNA linearizado foram purificados. A fotografia 20-A a seguir mostra os DNAs dos plasmídeos pDM-GFP e pUNA digeridos com a endonuclease *NotI*, nas canaletas 2 e 4, respectivamente. A fotografia 20-B, mostra o resultado após a digestão e a purificação dos fragmentos do cassete de expressão (canaleta 3) e vetor pUNA linearizado (canaleta 4).



**Figura 20: Análise eletroforética do pDM-GFP e do vetor pUNA digeridos com a endonuclease *NotI* e do resultado da purificação dos fragmentos.** Na fotografia A, a amostra 1 está o pDM-GFP sem endonuclease e a amostra 2 o plasmídeo pDM-GFP com *NotI*; a amostra 3 está o pUNA sem endonuclease e a amostra 4 o pUNA com *NotI*. Marcado em círculo vermelho na canaleta 2 e 3, estão o fragmento do cassete de expressão e do vetor pUNA linearizado, respectivamente. Na fotografia B, na canaleta 3 está identificado o cassete de expressão e na canaleta 4 exemplifica o vetor pUNA após a purificação.

O cassete de expressão e o vetor pUNA desfosforilado ambos digeridos e purificados, foram submetidos a reação de ligação. A seguir, o sistema de ligação foi misturado com células de *E. coli* DH5 $\alpha$ F'IQ eletrocompetentes e transformadas por eletroporação.

O novo vetor passou a se chamar de pDMU01 com 3332 pares de bases, possuindo o gene de resistência a ampicilina, de alto número de cópias (pUC – 500~700 cópias) e com o cassete de expressão contendo o promotor DM. O seu mapa físico está representado na figura 21 a seguir.



**Figura 21: O mapa físico do vetor de expressão pDMU01.** O vetor possui o gene que confere resistência a ampicilina e uma Ori de alto número de cópias. Possui também o cassete do vetor pDM-GFP clonado entre os sítios da endonuclease *NotI*.

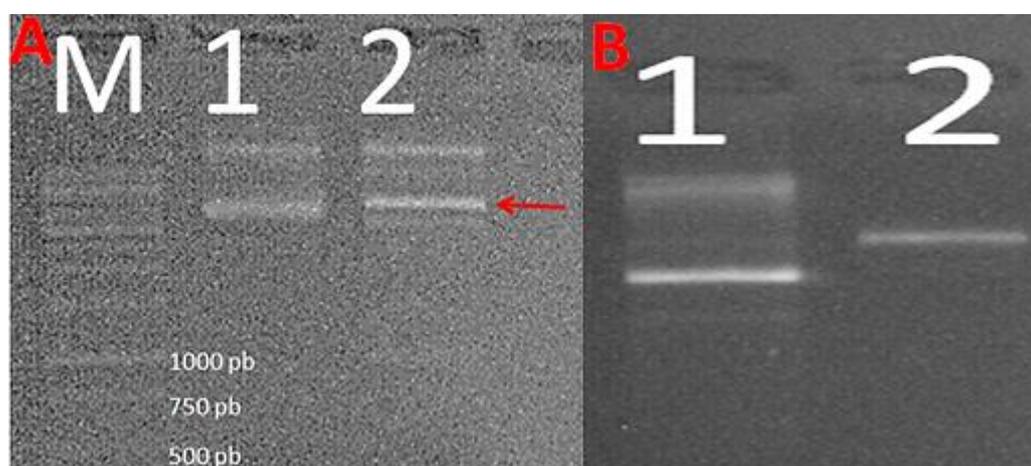
### 5.3.2 pCDM

Para se transformar o vetor pDM-GFP em um vetor “Caça Promotor” seria preciso inativar um dos dois sítios da endonuclease *SpeI*, o sítio situado na posição 1174. Desta forma seria possível retirar o promotor DM com dupla digestão com *SpeI* e *NdeI* sem a remoção do marcador genético de seleção, a sequência codificadora de GFP. Para se eliminar esse sítio utilizou-se a estratégia de digestão parcial do pDM-GFP com *SpeI* seguida de tratamento com T4 DNA polimerase seguido de religação. Esse novo vetor passou a ser chamado de pCDM.

Para sua construção, um clone recombinante de *E. coli* JM110 com o plasmídeo pDM-GFP foi escolhido para inóculo e seu DNA plasmidial extraído. Para a construção do novo vetor, o plasmídeo pDM-GFP foi digerido utilizando cinéticas enzimáticas analíticas

(diferentes concentrações da enzima e tempos de digestões diferentes) com a enzima de restrição *SpeI* para conhecer a melhor condição de digestão parcial.

Após testes de digestões analíticas do vetor pDM-GFP com a enzima de restrição *SpeI*, resolveu padronizar a enzima para 1U/ $\mu$ L, seguido de inativação enzimática com adição de tampão de amostra de eletroforese TBE azul e EDTA 200mM nos sistemas digeridos. Após a digestão com *SpeI*, o sistema digerido foi submetido a eletroforese e o fragmento correspondente ao vetor linearizado foi purificado. A fotografia 22-A a seguir mostra o perfil de restrição do vetor pDM-GFP com *SpeI*. A fotografia 22-B mostra o fragmento de interesse digerido parcialmente com a enzima *SpeI* e purificado.

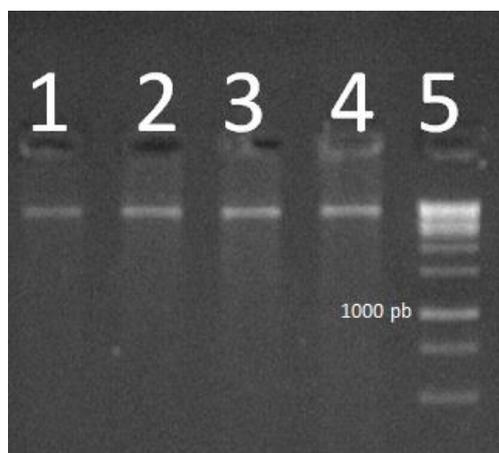


**Figura 22: Análise eletroforética do perfil de restrição do pDM-GFP com a endonuclease *SpeI* e purificação.** Na parte A, na canaleta 1 está o pDM-GFP sem tratamento com endonuclease e na canaleta 2 o plasmídeo pDM-GFP digerido parcialmente com *SpeI*. A seta em vermelho mostra o plasmídeo linearizado com *SpeI*. Na canaleta 1 da parte B, está identificado o pDM-GFP sem endonuclease e na canaleta 2 o vetor pDM-GFP digerido parcialmente com *SpeI* após purificação.

Após a análise eletroforética, o fragmento purificado como mostrado na canaleta B-2, foi submetido a tratamento com T4 DNA polimerase para destruição dos sítios de *SpeI*. O fragmento digerido parcialmente com *SpeI* foi submetido a reação de ligação.

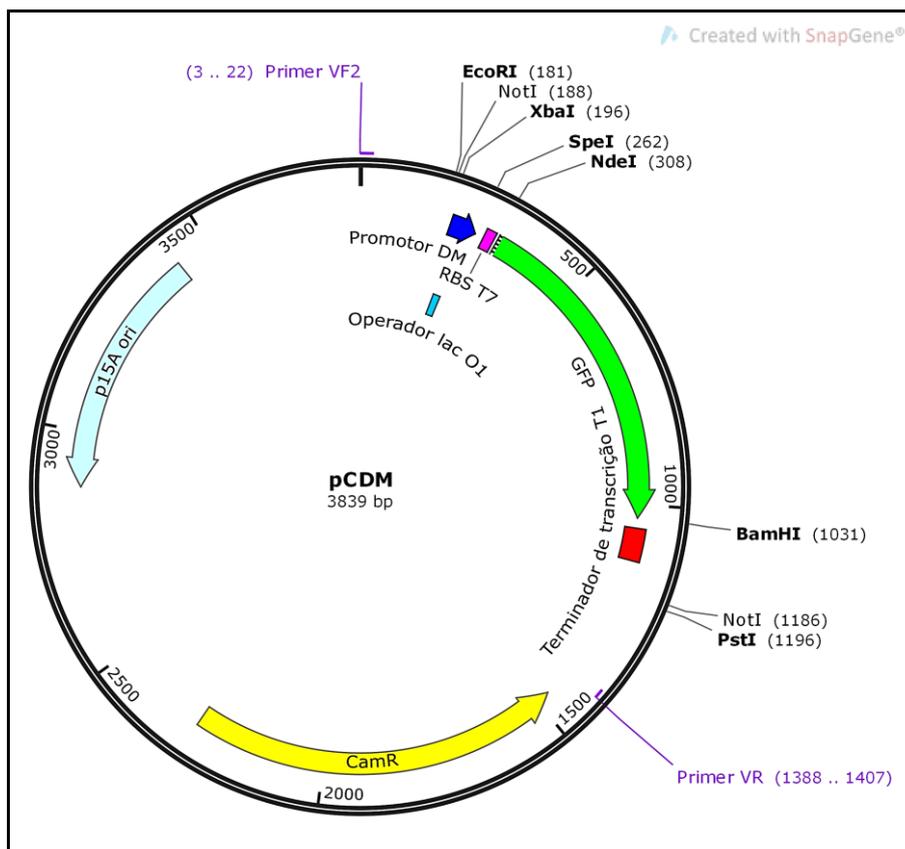
O sistema de ligação foi misturado com células *E. coli* DH5 $\alpha$  eletrocompetentes e transformadas geneticamente pela técnica de eletroporação. As placas foram incubadas durante a noite e clones foram selecionados no outro dia para pré-inóculos. Os pré-inóculos foram incubados durante a noite e em seguida seus DNAs plasmidiais extraídos.

Para confirmar que o sítio 1174 localizado nos DNAs plasmidiais dos clones foi destruído os DNAs plasmidiais extraídos de clones selecionados foram submetidos à dupla digestão com endonucleases *SpeI* e *EcoRI*. A fotografia 23 a seguir mostra os DNAs dos plasmídeos digeridos com as endonucleases de restrição *SpeI* e *EcoRI* após a eletroforese em gel de agarose e visualização sob luz UV.



**Figura 23: Análise eletroforética dos DNAs plasmidiais extraídos dos novos clones com as endonucleases *SpeI* e *EcoRI*.** As canaletas 1, 2, 3, e 4 estão os DNAs plasmidiais de novos clones tratados com *SpeI* e *EcoRI*.

Foi constatado na fotografia 23 (canaletas 1 a 4) um fragmento de DNA maior de 3174 pb, sendo que o fragmento menor esperado de 81 pb não pode ser visualizado no gel devido a seu pequeno tamanho. Este resultado mostrou que o sítio 1174 de *SpeI* foi destruído nestes DNAs plasmidiais. O novo vetor passou a se chamar de pCDM (baixo número de cópia). O seu mapa físico está representado na figura 24 a seguir.



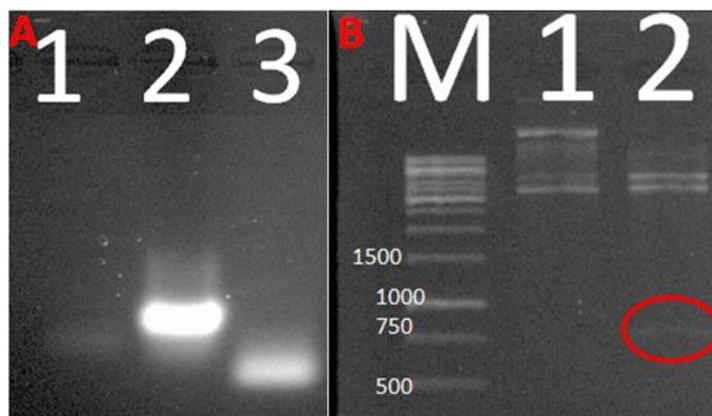
**Figura 24: O mapa físico do vetor pCDM.** A figura mostra os principais componentes do vetor como: os sítios das endonucleases, o promotor DM e o operador lac, o gene da GFP, o terminador de transcrição *rrnB*, o marcador de seleção para Canamicina e a origem de replicação. No vetor pCDM existe um único sítio para a enzima *SpeI* como estratégia de clonagem e seleção de novos promotores.

No vetor pCDM existe um único sítio para a enzima *SpeI* (posição 262) como estratégia de clonagem e seleção de novos promotores.

### 5.3.3 pDM02

Para a construção do novo vetor contendo o *polylinker* no lugar do gene estrutural da GFP, foram conferidos “*in silico*” os sítios das endonucleases do backbone do pDM-GFP, para se escolher enzimas que não cortassem no seu interior. O *polylinker* foi desenhado e sintetizado quimicamente na forma de dois oligonucleotídeos complementares contendo os sítios das endonucleases de restrição *NdeI*, *HpaI*, *AfIII*, *KpnI*, *BglIII*, *PmeI* e *BamHI* de forma que em sua forma duplex contivessem os sítios das endonucleases *NdeI* e *BamHI* nas suas extremidades já na forma colantes (“stick ends”) para facilitar a clonagem.

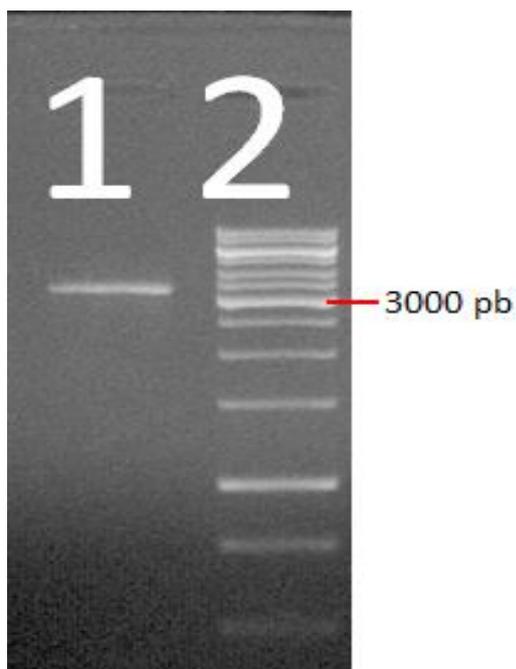
O pDM-GFP foi submetido a digestão com as endonucleases *NdeI* e *BamHI* para retirada do gene da GFP e clonagem do oligonucleotídeo dPLK. As sequências que compõem o *polylinker* após sua concepção e síntese foram redissolvidas e hibridizadas. A fotografia 25 a seguir, mostra os perfis eletroforéticos dos oligonucleotídeos hibridizados e do vetor pDM-GFP tratado com as endonucleases *NdeI* e *BamHI*.



**Figura 25: Fotografia do perfil eletroforético dos oligonucleotídeos pLK1 e pLK2 e do pDM-GFP com as endonucleases *NdeI* e *BamHI*.** Na canaleta 1, na parte A da fotografia, mostra somente o pLK1; na canaleta 2 mostram o pLK1 e pLK2 hibridizados e na canaleta 3 mostra somente o pLK2. Na parte B, a amostra 1 está o pDM-GFP sem digestão com endonucleases e a amostra 2 mostra o plasmídeo pDM-GFP tratado com *NdeI* e *BamHI*. O círculo em vermelho mostra o fragmento que corresponde ao gene da GFP.

Como foi possível observar no resultado da fotografia 25-A, na canaleta **2**, está o perfil eletroforético dos oligonucleotídeos pLK1 e pLK2 após tratamento térmico para anelamento. Na canaleta **2**, a banda de fluorescência apresentou mais intensidade e menor mobilidade eletroforética que as bandas das canaletas **1** e **3**. Por isso deduzido que o anelamento dos dois oligonucleotídeos ocorreu, pois os dois oligonucleotídeos possuem as sequências perfeitamente complementares e o DNA duplex absorve muito mais brometo de etídio que suas versões fitas simples. Na figura 25-B pode-se observar na canaleta **2** o plasmídeo parcialmente digerido com as referidas endonucleases.

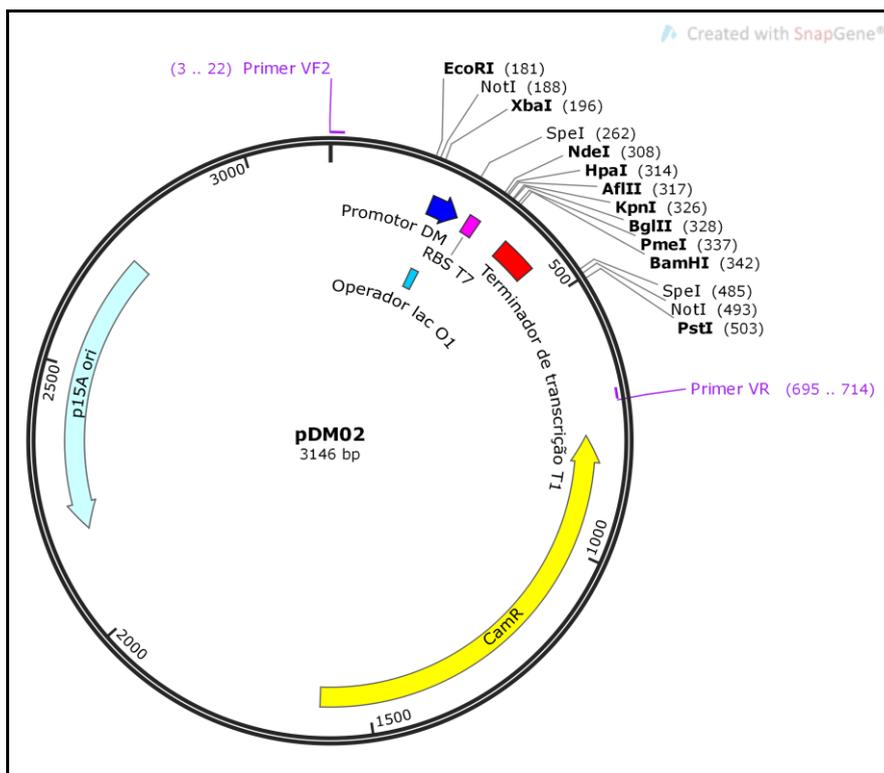
Foi realizada uma digestão do pDM-GFP com mais material e maior proporção de endonucleases e a banda desejada (vetor sem GFP) foi purificado e analisado por eletroforese, sendo que o resultado está mostrado na figura 26.



**Figura 26:** Análise eletroforética do vetor pDM-GFP purificado. Na canaleta 1 está identificado o pDM-GFP com *NdeI* e *BamHI* e após purificação.

Foi observado na fotografia 26 que houve a purificação do fragmento linearizado correspondente a 3212 pb que corresponde ao vetor pDM desprovido do gene da GFP. A partir disto, o vetor linearizado e o oligonucleotídeo dPLK foram submetidos a reação de ligação. Após a ligação, o sistema foi misturado com células de *E. coli* DH5 $\alpha$  eletrocompetentes e transformadas por eletroporação. Após a eletroporação, as células foram semeadas em alíquotas em placas de Petri contendo meio seletivo LB sólido com canamicina e glicose e incubadas durante a noite.

Foram selecionados clones de *E. coli* DH5 $\alpha$  contendo o novo vetor e cultivados em meio LB contendo canamicina e glicose e em seguida seus plasmídeos extraídos. O novo vetor passou a se chamar de pDM02 com 3146 pares de bases, contendo o *polylinker* em sua sequência com os sítios das endonucleases de restrição, o gene de resistência à canamicina e a Ori de baixo número de cópias do p15A (~10 a 12 cópias/célula). O seu mapa físico está representado na figura 27 a seguir.



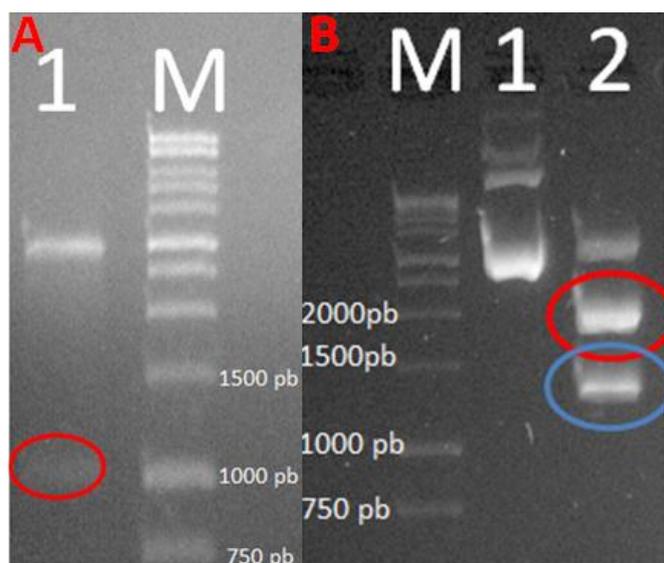
**Figura 27:** A figura apresenta o mapa físico do vetor de expressão pDM02. Pode-se identificar os principais componentes do vetor, como: o promotor sintético regulado (DM) e o operador lac; o RBS T7; a região de múltiplos sítios de clonagem composto pelos sítios das endonucleases (*NdeI*, *HpaI*, *AflII*, *KpnI*, *BglII*, *PmeI* e *BamHI*); o terminador de transcrição funcional T1, do gene de *rrnB* de *E. coli*; o marcador de seleção que confere resistência ao antibiótico canamicina (CamR); e a origem de replicação de baixo número de cópias – p15A (~10 a 12 cópias). Os sítios das endonucleases estão destacados em negrito.

### 5.3.4 pDMU02

Para a construção do vetor pDMU02 (alta cópia), o vetor pCDM foi tratado com as endonucleases *EcoRI* e *PstI* resultando em um fragmento de 1015 pb e este fragmento foi clonado no vetor pRIQT tratado com as mesmas endonucleases.

Células de *E. coli* JM110 e DH5 $\alpha$  foram transformadas geneticamente por eletroporação respectivamente com os plasmídeos pCDM e pRIQT. Clones transformantes foram cultivados em meio seletivo e seus plasmídeos extraídos.

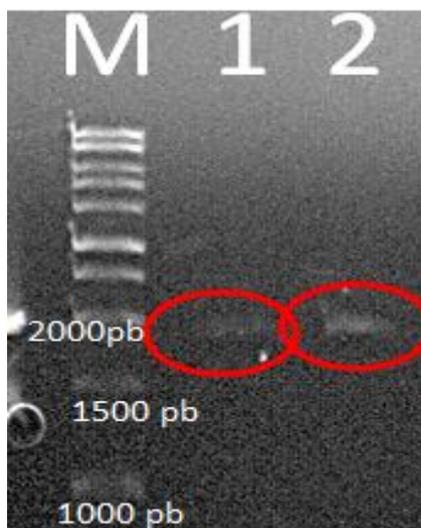
Os plasmídeos pCDM e pRIQT foram tratados com as endonucleases *EcoRI* e *PstI*. A fotografia 28 a seguir mostra os DNAs dos plasmídeos pCDM e pRIQT digeridos com as endonucleases *EcoRI* e *PstI* após a eletroforese em gel de agarose e visualização sob luz UV.



**Figura 28: Análise eletroforético dos vetores pCDM e pRIQT tratados com as endonucleases *EcoRI* e *PstI*.** Na fotografia A, a canaleta 1 está o pCDM digerido com as endonucleases *EcoRI* e *PstI*. O círculo em vermelho mostra o fragmento correspondente ao cassete de expressão. Na fotografia B, a amostra 1 está o pRIQT sem tratamento com endonucleases e a amostra 2 identifica o plasmídeo pRIQT digerido com as endonucleases *EcoRI* e *PstI*. O círculo em vermelho mostra o fragmento correspondente ao vetor linearizado e o círculo em azul corresponde ao fragmento do cassete de expressão que contém o gene *lacIqTS*.

Foi analisada a fotografia 28-A e observado na canaleta 1 o fragmento correspondente ao cassete de expressão (marcado círculo em vermelho) tratados com as endonucleases *EcoRI* e *PstI*. Na fotografia 28-B, na canaleta 2, mostra os fragmentos correspondente ao vetor pRIQT linearizado (marcado por círculo vermelho) e o cassete de expressão contendo o gene *lacIqTS* (marcado em círculo azul) tratados com as endonucleases *EcoRI* e *PstI*.

Após a análise do resultado com *EcoRI* e *PstI*, o sistema do pRIQT foi submetido a eletroforese em gel de agarose e o fragmento correspondente ao vetor linearizado foi purificado. O vetor pCDM desprovido do cassete de expressão não foi purificado. A fotografia 29 a seguir mostra o resultado da purificação do vetor pRIQT após a eletroforese em gel de agarose e visualizado no transluminador sob luz UV.



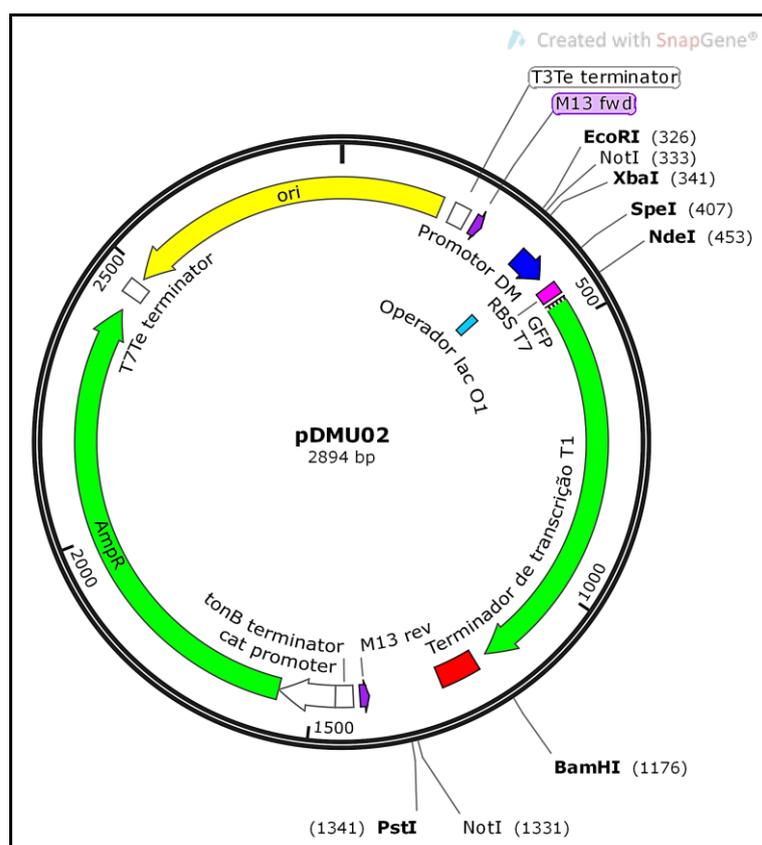
**Figura 29:** Análise eletroforética do vetor pRIQT digerido com *EcoRI* e *PstI* e purificado. Nas canaletas 1 e 2 está identificado em círculos vermelhos o pRIQT digerido com as endonucleases *EcoRI* e *PstI* e após tratamento de purificação.

Foi observado na fotografia, nas canaletas **1** e **2**, o fragmento correspondente ao marcador de seleção para ampicilina com a Ori de alto número de cópias (marcado em círculos vermelhos) com tamanho de 1879 pb que corresponde ao vetor pRIQT linearizado e purificado.

Após a obtenção do backbone de alto número de cópias do pRIQT purificado (desprovido do gene lacIqTS) e do cassete de expressão juntamente com o backbone do pCDM, os fragmentos foram submetidos a reação de ligação. Após a ligação uma alíquota foi misturada com células de *E. coli* da linhagem DH5 $\alpha$  eletrocompetentes e transformadas geneticamente por eletroporação.

A seguir, após a eletroporação, as células foram semeadas em alíquotas em placas de Petri contendo meio LB sólido com ampicilina e glicose e não com canamicina e glicose. Esta estratégia permitiu a seleção de clones possuindo o vetor com o cassete de expressão clonado somente nos backbones de marcador de seleção para ampicilina. Após o plaqueamento, as placas de Petri foram incubadas durante a noite. Foram selecionados clones transformantes de *E. coli* DH5 $\alpha$  das placas de Petri contendo o novo vetor após a incubação e seus plasmídeos foram extraídos.

O novo vetor denominado de pDMU02, possui um cassete de expressão composto pelo promotor sintético DM, o sítio de ligação ao ribossomo do fago T7, o gene repórter da GFP (*Green Fluorescent Protein*); um forte terminador de transcrição do gene *rrnB* de *E. coli*; a Ori de alto número de replicação no hospedeiro (pUC – 500~700 cópias) e é portador do marcador de seleção para ampicilina. No vetor pDMU02 existe um único sítio para a enzima *SpeI* (posição 407) que é utilizado na estratégia de clonagem e seleção de novos promotores, como mostrado na figura 30 a seguir.



**Figura 30: Mapa físico do vetor caça promotor pDMU02.** A figura mostra os principais componentes do vetor como: os sítios das endonucleases, o promotor DM, o gene da GFP, o terminador de transcrição *rrnB*, o marcador de seleção para ampicilina e a origem de replicação de alto número de cópias. No vetor pDMU02 existe um único sítio para a enzima *SpeI* como estratégia de clonagem e seleção de novos promotores.

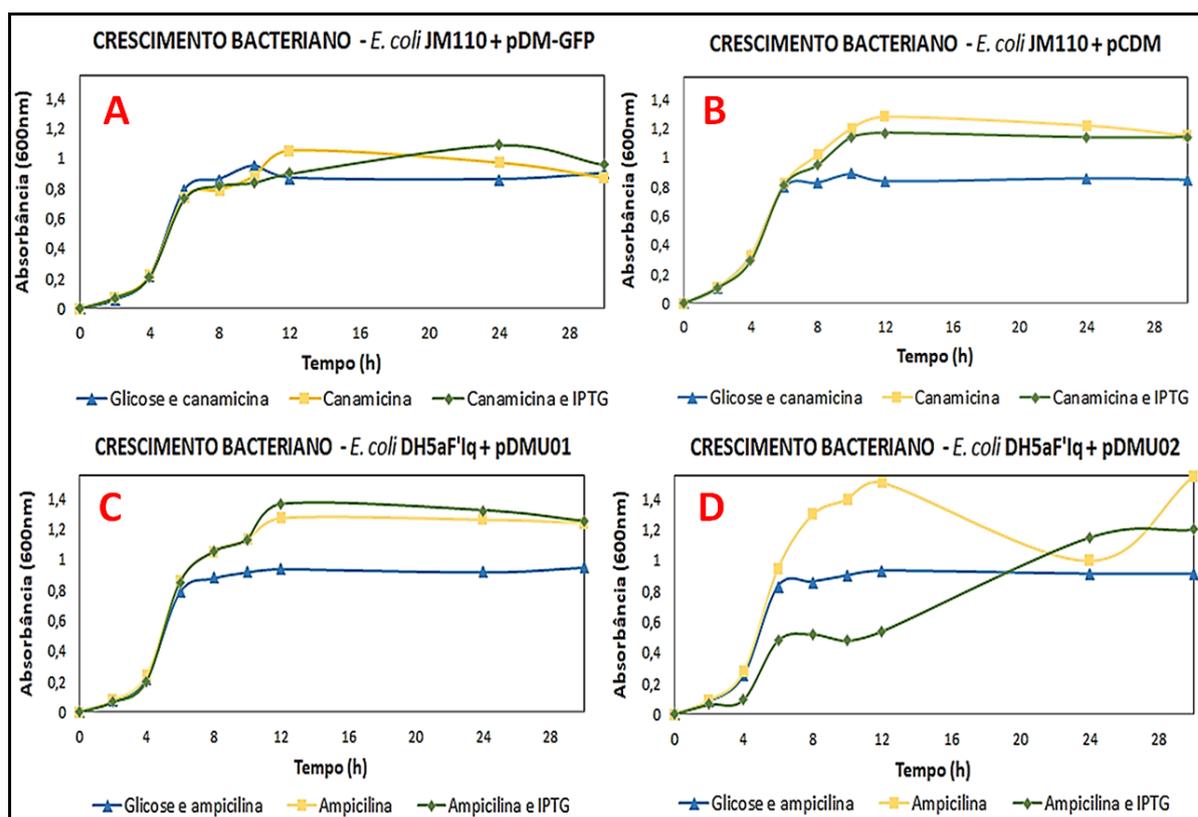
#### 5.4 ANÁLISE DA EXPRESSÃO DA GFP PELOS PLASMÍDEOS RECOMBINANTES CONSTRUÍDOS: pDM-GFP, pDMU01, pCDM e pDMU02

Os plasmídeos que compõem o novo conjunto de vetores construídos: pDM-GFP, pDMU01, pCDM e pDMU02 foram introduzidos em células *E. coli* por eletroporação. As

placas foram incubadas e um clone isolado dos sistemas: *E. coli* JM110+ pDM-GFP, *E. coli* JM110+pCDM, *E. coli* DH5 $\alpha$ F'IQ+pDMU01 e *E. coli* DH5 $\alpha$ F'IQ + pDMU02 foram inoculados em pré-inóculos em meio LB contendo glicose e o antibiótico seletivo.

A seguir foi inoculada uma alíquota de cada um dos quatro pré-inóculos em Erlenmeyers separadamente contendo meio LB líquido nas seguintes condições: um sistema somente com antibiótico seletivo, um sistema com antibiótico seletivo e suplementado com glicose e um sistema antibiótico seletivo e o indutor IPTG. As culturas foram incubadas e alíquotas retiradas para mensurar o crescimento celular e para a análise quantitativa da expressão e regulação do gene da GFP sob o controle promotor sintético DM. Os resultados foram analisados no *software IBM SPSS*, com significância de 0,05 e elaborados os gráficos para análise dos resultados do crescimento celular e da fluorescência emitida.

A figura 31 a seguir mostra o crescimento bacteriano dos sistemas do conjunto de vetores: pDM- GFP + *E. coli* JM110; pCDM + *E. coli* JM110; *E. coli* DH5 $\alpha$ F'IQ + pDMU01 e *E. coli* DH5 $\alpha$ F'IQ + pDMU02.



**Figura 31:** A figura mostra o resultado do crescimento celular dos sistemas do conjunto de vetores. Na parte A, B, C e D da figura, estão o crescimento celular dos sistemas *E. coli* JM110 + pDM-GFP; *E. coli* JM110 + pCDM; *E. coli* DH5 $\alpha$ F'IQ + pDMU01 e *E. coli* DH5 $\alpha$ F'IQ + pDMU02, respectivamente, em função do tempo de cultivo de um clone crescido com o seu antibiótico seletivo, suplementado ou não com glicose e IPTG. O processo de suplementação com glicose ou indução com o IPTG ocorreu a partir do início do cultivo.

Como pode ser observado na figura 31, o crescimento bacteriano do sistema **A**, JM110 + pDM-GFP, não foi afetado pela produção da GFP quando este foi tratado com o indutor IPTG ou com somente o antibiótico canamicina (linhas amarela e verde no gráfico). As culturas dos sistemas **B** e **C**, pCDM + *E. coli* JM110 e *E. coli* DH5 $\alpha$ F'IQ + pDMU01, apresentaram resultados semelhantes de crescimento celular até 12 horas de cultivo nas mesmas condições, com crescimento levemente maior (O.D. = 1,2 e 1,4) e depois entraram em fase estacionária (linhas amarela e verde no gráfico).

Pode ter acontecido o alongamento das cepas *E. coli*, e conseqüente maior dispersão de luz pelas células, nos cultivos com somente o antibiótico seletivo e com o antibiótico e o indutor IPTG (linhas amarela e verde no gráfico). Esse fato pode ocorrer devido à produção e acumulação da GFP no citosol das hospedeiras de *E. coli*, resultando em O.D. maior (JUSTICE et al., 2008).

O sistema **D**, *E. coli* DH5 $\alpha$ F'IQ + pDMU02, apresentou crescimento celular diferente no meio com o antibiótico seletivo e o indutor IPTG. A cultura entrou em fase estacionária (O.D. = 0,4-0,6) no período de 6 até 12 horas. Após 24 horas o crescimento celular dobrou (O.D. = 1,2) e manteve-se em fase estacionária (linha verde no gráfico). No meio com somente o antibiótico o crescimento celular, este mesmo cultivo foi capaz de crescer em até 12 horas com O. D. = 1,4-1,6 (linha amarela no gráfico).

Os sistemas **A**, **B**, **C** e **D** apresentaram resultados semelhantes de crescimento celular constante nos meios com glicose e com o antibiótico seletivo (O.D. = 0,8-1) entrando em fase estacionária após 8 horas. O número de cópias dos vetores (alto ou baixo número de cópias) sob o controle do promotor DM parece não ter influenciado na expressão nestas condições de cultivo.

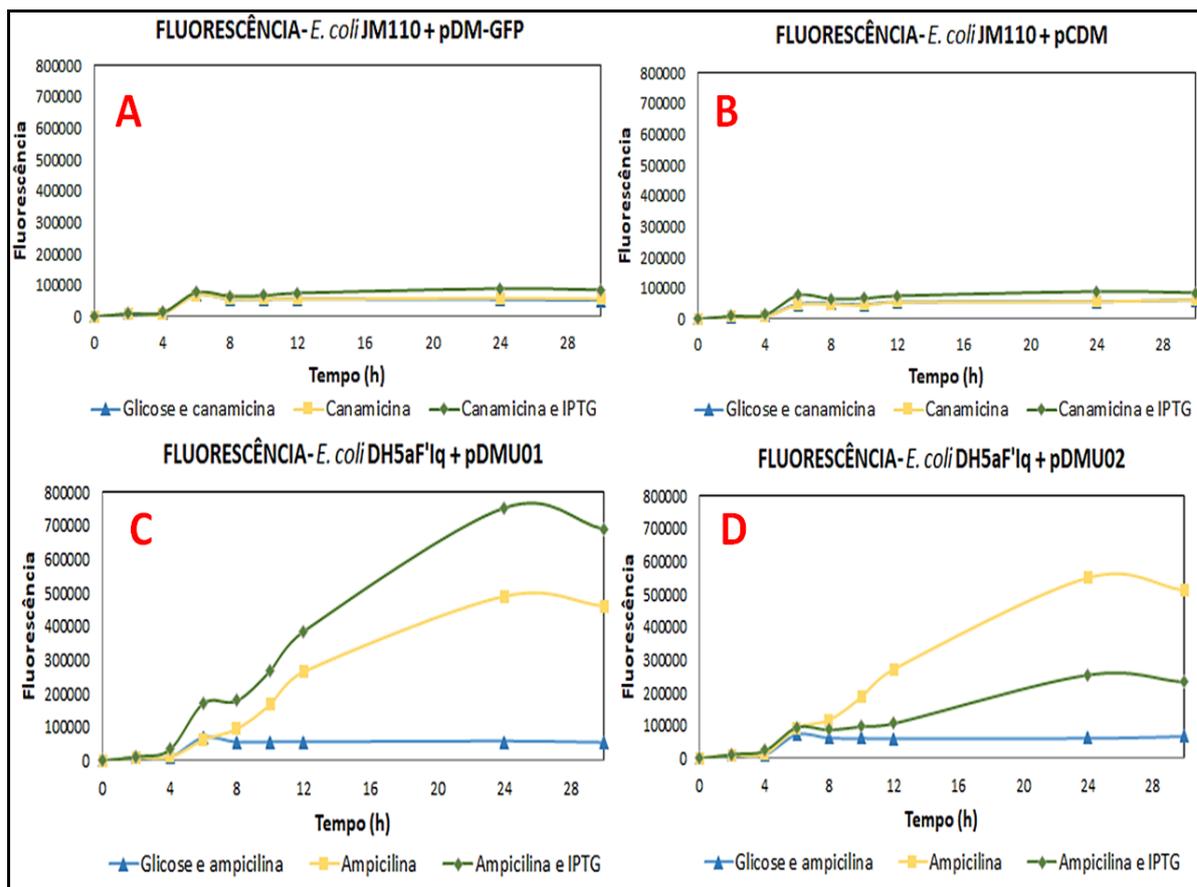
Os metabolismos de carbono e nitrogênio estão ligados, exigindo regulação compartilhada entre esses dois sistemas. *E. coli* pode utilizar muitas fontes diferentes de carbono e nitrogênio. Isso gera muitas combinações possíveis de nutrientes. A fonte de carbono preferida para *E. coli*, como para muitas outras bactérias, é a glicose, suportando uma taxa de crescimento mais rápida em comparação com outros açúcares. O exemplo mais conhecido de utilização preferencial de glicose vem do trabalho de Monod sobre o deslocamento diaúxico glicose-lactose: *E. coli* primeiro cresce rapidamente em glicose e,

quando a glicose acaba, muda para crescer mais lentamente em lactose ou outros açúcares (BREN et al., 2016).

A melhor fonte de nitrogênio para *E. coli* é a amônia. *E. coli* também pode crescer, embora mais lentamente, em uma variedade de outras fontes de nitrogênio, incluindo muitos aminoácidos. Assim, sob condições limitantes de nitrogênio que reduzem o fluxo para a biossíntese de aminoácidos, ocorre à repressão da captação de glicose e a inibição da síntese de cAMP pela adenilato ciclase, fornecendo circuitos que coordenam entre assimilação de carbono e nitrogênio. O baixo crescimento na presença da glicose parece surgir do desequilíbrio carbono/nitrogênio que resulta em cAMP insuficiente (BREN et al., 2016).

Estudo com espectrometria de massas mostrou que a glicose com algumas fontes pobres de nitrogênio, como arginina e outros aminoácidos, se torna uma das piores fontes de carbono, levando a níveis baixos de cAMP e resultando em um crescimento lento (BREN et al., 2016). O aumento experimental de cAMP por suprimento externo, meios genéticos ou inibição da absorção de glicose, leva novamente a um crescimento mais rápido sem efeitos negativos perceptíveis na sobrevivência ao estresse ou no rendimento final de biomassa (BREN et al., 2016).

Foi analisada a expressão do gene da GFP sob o controle do promotor DM nos diferentes sistemas. A figura 32 a seguir mostra a fluorescência emitida dos sistemas: pDM-GFP + *E. coli* JM110; pCDM + *E. coli* JM110; *E. coli* DH5 $\alpha$ F'IQ + pDMU01 e *E. coli* DH5 $\alpha$ F'IQ + pDMU02.



**Figura 32:** A figura mostra o perfil de expressão do gene da GFP dos sistemas do conjunto de vetores. Na parte A, B, C e D da figura, estão perfil de fluorescência dos sistemas *E. coli* JM110 + pDM-GFP; *E. coli* JM110 + pCDM; *E. coli* DH5αF'IQ + pDMU01 e *E. coli* DH5αF'IQ + pDMU02, respectivamente, em função do tempo de cultivo de um clone crescido com o seu antibiótico seletivo, suplementado ou não com glicose e IPTG. O processo de suplementação com glicose ou indução com o IPTG ocorreu a partir do início do cultivo.

As células de *E. coli* expressaram baixos níveis de GFP na presença do antibiótico seletivo e do indutor IPTG nos sistemas **A** e **B**, pDM-GFP e pCDM + *E. coli* JM110 (linha verde nos gráficos **A** e **B**). Os dois sistemas de baixo número de cópias, nestes dois gráficos **A** e **B** tiveram seus maiores ápices de expressão em 24 horas com 89.000 e 97.000 UR de fluorescência, respectivamente, e um menor ainda nível de fluorescência na presença de glicose e do antibiótico seletivo mostrando regulação nos vetores na ausência do indutor IPTG (linhas azuis nos gráficos **A** e **B**).

No sistema **C**, *E. coli* DH5αF'IQ + pDMU01, as células programadas com o vetor de alta cópia sob o controle do promotor DM expressaram altos níveis da GFP na presença do antibiótico seletivo e o indutor IPTG atingindo seu ápice no intervalo de 24 horas com 752.940 UR de fluorescência (linha verde no gráfico). Este sistema **C**, expressou o gene da GFP com somente o antibiótico em até 24 horas com 489.750 UR de fluorescência (linha amarela no gráfico) e o crescimento bacteriano não foi afetado pela produção da GFP

(Figura 31, linha amarela no gráfico **C**). No cultivo com glicose, este sistema expressou baixo nível (53.000 UR de fluorescência) mostrando a modulação por glicose do promotor DM em um vetor de alto número de cópias (linha azul no gráfico).

No sistema **D**, *E. coli* DH5 $\alpha$ F'IQ + pDMU02, a expressão foi baixa do gene da GFP na presença de glicose e do antibiótico seletivo (68.000 UR de fluorescência), mostrando a inibição da expressão da GFP neste vetor de alta cópia na presença de glicose e ausência do indutor IPTG. Houve detecção de maior nível de fluorescência da GFP no sistema somente com o antibiótico seletivo de forma similar aos resultados obtidos com o vetor pDMU01 (550.000 UR de fluorescência) linha amarela no gráfico **D**).

No sistema **D**, com o indutor IPTG e o antibiótico seletivo, o nível de fluorescência detectado foi menor que somente com o antibiótico. Isso pode ser interpretado como uma intensa expressão do gene da GFP deve ter acarretado síntese da GFP sem sua estrutura tridimensional correta resultando em corpúsculos de inclusão onde a GFP permanecia não fluorescente. Essa interpretação é compatível com o fato da cultura da bactéria com o indutor e antibiótico (sem glicose) ter crescido menos que nas outras condições. Isso provavelmente se deve a alta expressão da GFP ter sobrecarregado o metabolismo geral da *E. coli*.

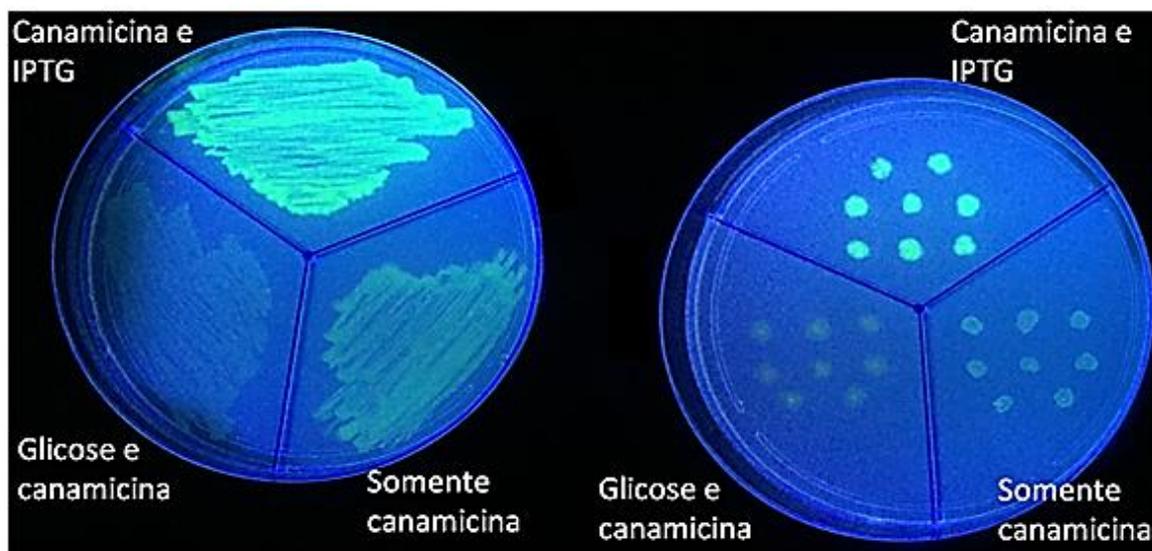
Nos sistemas **C** e **D**, nas culturas realizadas somente com o antibiótico, houve expressão de níveis razoáveis de GFP, da ordem de 500.000 UR de fluorescência, esse “vazamento” pode ter sido devido ao alto nível de cópias dos vetores com o número insuficiente de elementos regulatórios (Repressor e Proteína CAP).

## 5.5 ANÁLISE DA REGULAÇÃO E EXPRESSÃO DA GFP PELO VETOR pDM-GFP

Para análise qualitativa da força e regulação do promotor DM em um vetor de baixo número de cópias, um clone de *E. coli* JM110+ pDM-GFP foi submetido a réplica-plate em nas seguintes condições: somente com canamicina, com canamicina e glicose e com canamicina e o indutor IPTG. Um clone isolado foi selecionado e pela técnica de estria de colônia foi analisado nas mesmas condições.

Após a incubação, as placas foram analisadas por observação com iluminação UV longa (luz negra) e a figura 33 a seguir na placa de Petri da esquerda, mostra um clone isolado de *E. coli* JM110 transformado com o plasmídeo pDM-GFP cultivado em meio LB em diferentes condições de crescimento/indução e na placa de Petri da direita,

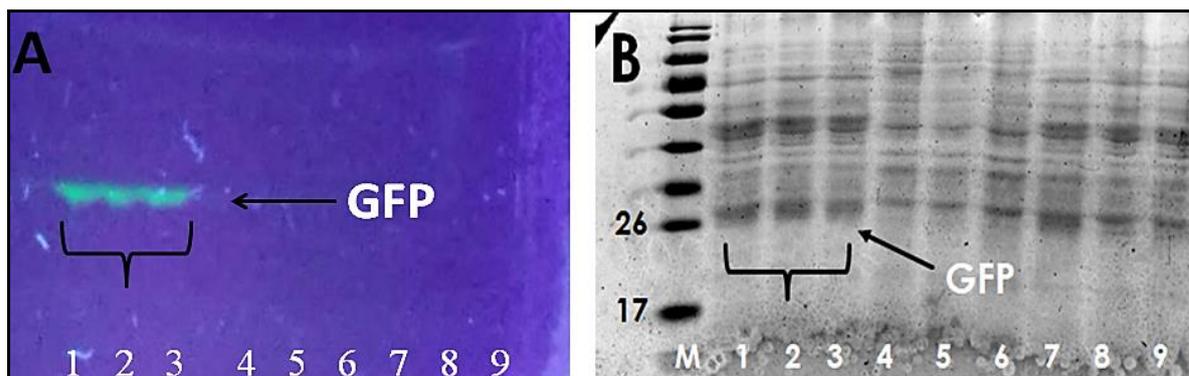
estão diferentes clones de *E. coli* JM110 programados com o pDM-GFP nas mesmas condições de cultivo e indução.



**Figura 33: Fotografia dos clones recombinantes programados com o vetor pDM-GFP.** A placa de Petri na esquerda contém um clone isolado e a placa da direita representa os clones crescidos dispostos nas mesmas posições em condições de cultivo. Os clones foram analisados em meio LB: com somente canamicina, canamicina e glicose, e canamicina e IPTG (indutor); e iluminados com luz negra (300 nm).

Após o resultado qualitativo das placas de Petri, foi observado que no meio contendo glicose e canamicina o gene da GFP não era expresso e a produção da proteína não ocorreu. No sistema somente com canamicina, o gene da GFP teve pouca expressão. No sistema com o antibiótico canamicina e o indutor IPTG, a produção da proteína ocorreu com mais intensidade, evidenciado o modelo de repressão/indução do promotor DM.

Prosseguiu-se para a análise quantitativa da expressão e regulação do gene da GFP. Um clone isolado foi selecionado para análise em meio líquido em triplicata nas três condições: em somente canamicina, com canamicina e glicose e com canamicina e o indutor IPTG. Acompanhou-se o crescimento e indução do clone de *E. coli* portador do pDM-GFP e quando a cultura atingiu 8 horas, uma alíquota foi coletada para análise das proteínas por eletroforese em gel desnaturante de poliácridamida 12 %. A figura 34 a seguir (sistemas **A** e **B**), mostra a análise eletroforética das proteínas produzidas por um clone de *E. coli* transformado com o plasmídeo recombinante pDM-GFP quando cultivados em meio LB líquido em diferentes condições de crescimento/indução.



**Figura 33: O resultado da análise por eletroforese em gel desnaturante de poliácridamida (SDS-PAGE) e fluorescência da produção de GFP por um clone de *E. coli* JM110 contendo o vetor pDM-GFP em diferentes condições de cultivo.** A fotografia A mostra o gel obtido em transluminador (300 nm) e a B corado com o azul de Coomassie. As amostras nas raias 1, 2 e 3 representam as proteínas de clones crescidos com canamicina e IPTG; as amostras nas raias 4, 5 e 6 com canamicina e glicose; e amostras nas raias 7, 8 e 9 somente com canamicina. A raia M mostra o marcador molecular em kDa. A proteína recombinante GFP localiza-se na posição de aproximadamente 25 kDa (seta).

No sistema **A**, nas amostras 1, 2 e 3, as proteínas fluorescentes são visualizadas no gel quando este foi exposto em transluminador (300 nm). Não foi visualizada fluorescência nas amostras crescidas em glicose e canamicina (amostras 4, 5 e 6) e e nas amostras crescidas somente com canamicina (7, 8 e 9). O sistema **B**, mostra o gel corado com azul de Coomassie, onde pode-se observar o perfil das bandas maior e atípico nas amostras 1, 2 e 3 quando estas são induzidas por IPTG não sendo visualizado este perfil quando são comparadas com as amostras 4, 5 e 6 (crescimento com glicose) e 7, 8 e 9 (crescimento com somente canamicina).

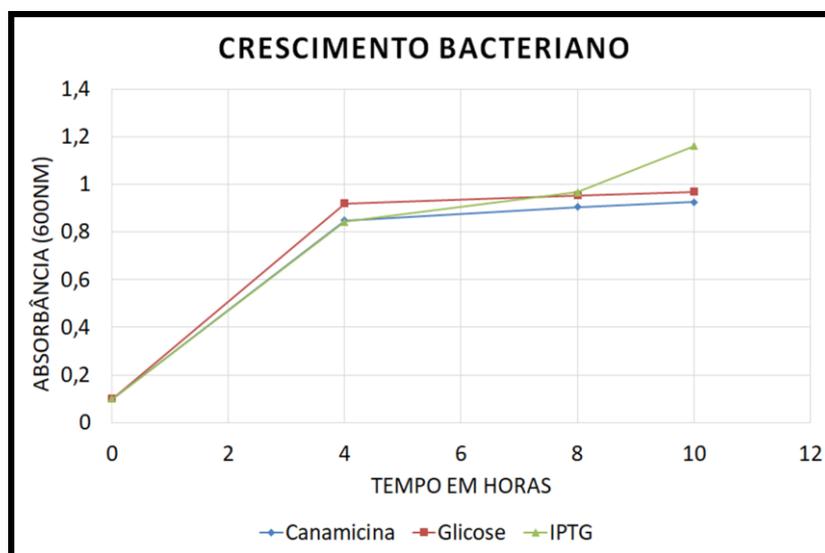
Os resultados de expressão do gene da GFP indicam que o promotor DM apresenta funcionalidade de força intermediária, mas com bom nível de regulação, ou seja, praticamente nenhuma proteína é expressa em *E. coli* na ausência do IPTG. Entretanto, quando o pDM-GFP é induzido com IPTG, a GFP (26,9 kDa) é produzida.

## 5.6 ANÁLISE DA EXPRESSÃO E REGULAÇÃO DO pDM-TEK

Um clone recombinante (*E. coli* JM110+pDM-TEK) foi selecionado para expressão e regulação do gene TEK em meio líquido em triplicatas nas condições: três somente com canamicina, três com canamicina e glicose e três com canamicina e o indutor IPTG.

A indução da expressão do gene da proteína híbrida TEK por *E. coli* JM110 programada com o plasmídeo pDM-TEK, foi realizada quando as culturas atingiram a absorbância igual a 0,1 e então foi adicionado IPTG até o tempo total de 10 horas. A figura

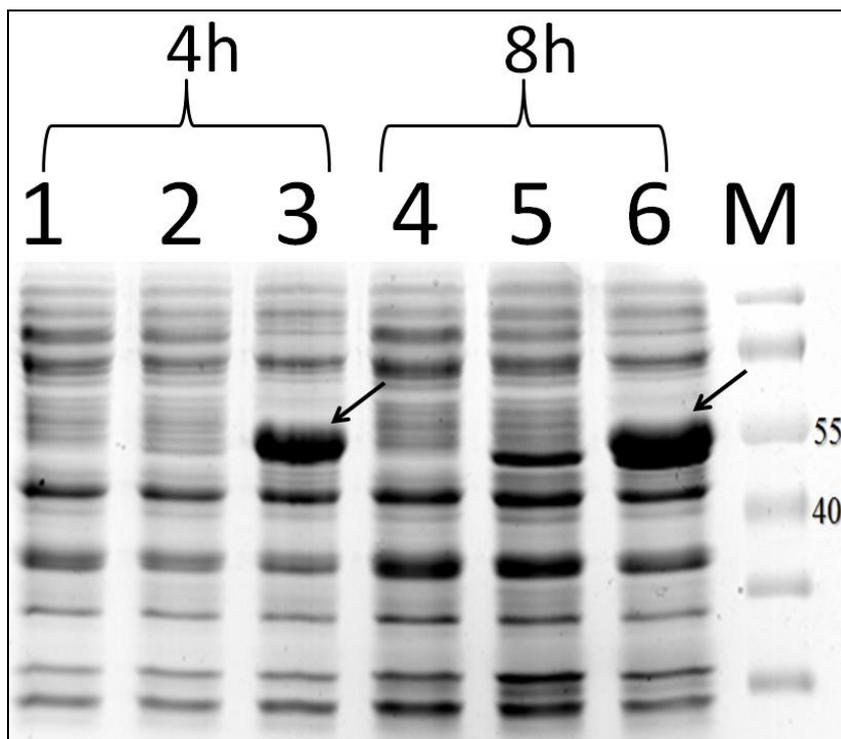
35 a seguir mostram gráfico do crescimento celular de um clone *E. coli* programado com o pDM-TEK.



**Figura 34: Análise do crescimento celular do pDM-TEK.** O perfil do crescimento celular (ABS. 600 nm x tempo) de *E. coli* JM110 programada com o plasmídeo pDM-TEK com canamicina e IPTG, com canamicina e glicose e somente com canamicina. O processo de indução com IPTG ocorreu a partir do início do cultivo.

Como pode ser observado na Figura 35, as células de *E. coli* JM110 nas 3 diferentes condições de crescimento apresentaram praticamente o mesmo perfil de crescimento, inclusive na presença do indutor IPTG, indicando que a proteína híbrida expressa não exibiu toxicidade para as células. Essa não toxicidade pode ser explicada das seguintes maneiras: a proteína híbrida é expressa inativa devido a sua própria estrutura artificial, ou por acumular-se no interior das células de *E. coli* na forma de corpúsculos de inclusão.

Nos intervalos de 4 e 8 horas de indução, alíquotas de 2 mLs dos sistemas foram coletadas para análise das proteínas por eletroforese em gel desnaturante de poliacrilamida 12%. A figura 36 a seguir mostra a análise eletroforética das proteínas produzidas por um clone de *E. coli* JM110 transformado com o plasmídeo recombinante pDM-TEK quando cultivados em meio LB líquido com canamicina e em diferentes condições de crescimento/indução.



**Figura 35: O perfil eletroforético em gel de poliacrilamida desnaturante de proteínas expressas por um clone de *E. coli* JM110 contendo o vetor pDM-TEK em diferentes condições de cultivo.** Nas raias 1 e 4, cultivo com glicose e canamicina; raias 2 e 5, com apenas canamicina; raias 3 e 6, cultivo com canamicina e o indutor IPTG. Nas raias 1, 2 e 3 estão as amostras com 4 horas de cultivo; nas raias 4, 5 e 6 as amostras com 8 horas de cultivo. A raia M, mostra o marcador molecular em kDa. A proteína recombinante TEK localiza-se na posição de 50 kDa (setas).

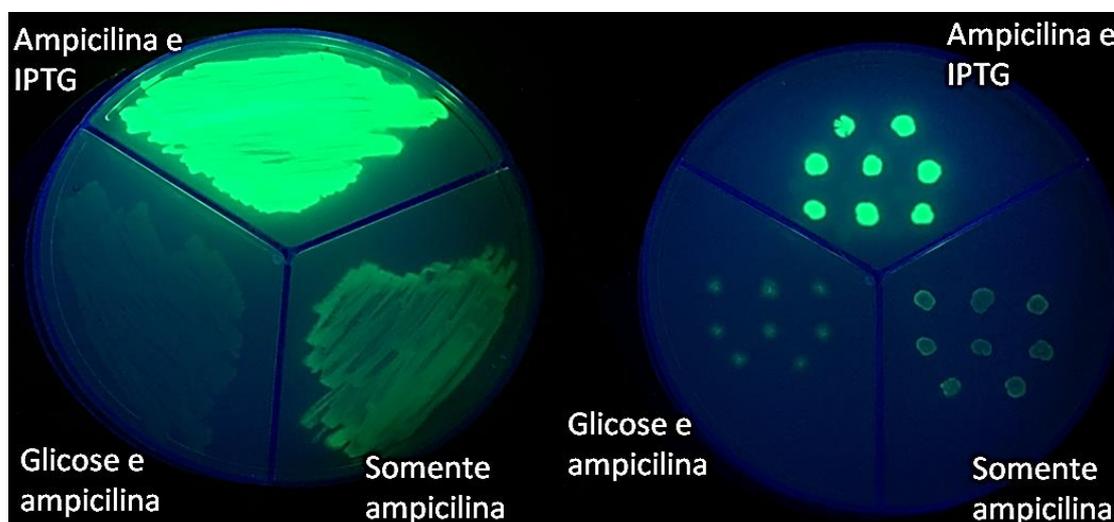
Pode-se observar altos níveis de expressão da proteína híbrida na presença do indutor tanto após 4 como 8 horas de indução (raias 3 e 6) e praticamente nenhuma indução da proteína híbrida na presença de glicose tanto após 4 como 8 horas de indução (raias 1 e 4), o que confirma que o promotor DM apresenta performance de alta expressão e boa regulação (estridente).

### 5.7 ANÁLISE DA REGULAÇÃO E EXPRESSÃO DA GFP PELO VETOR pDMU01

Para análise qualitativa da expressão do gene da GFP sob o controle do promotor DM em um vetor de alto número de cópias (pDMU01), clones foram submetidos à réplica-plate em placas de Petri nas seguintes condições: somente com ampicilina, com ampicilina e glicose e com ampicilina e o indutor IPTG.

Após a incubação, as placas foram analisadas por observação em lâmpada UV e a figura 37 a seguir mostra na placa de Petri da esquerda um clone isolado de *E. coli*

transformado com o plasmídeo pDMU01 cultivado em meio LB com ampicilina em diferentes condições de crescimento/indução e na placa de Petri da direita diferentes clones de *E. coli* programados com o pDMU01 nas mesmas condições de cultivo e indução.



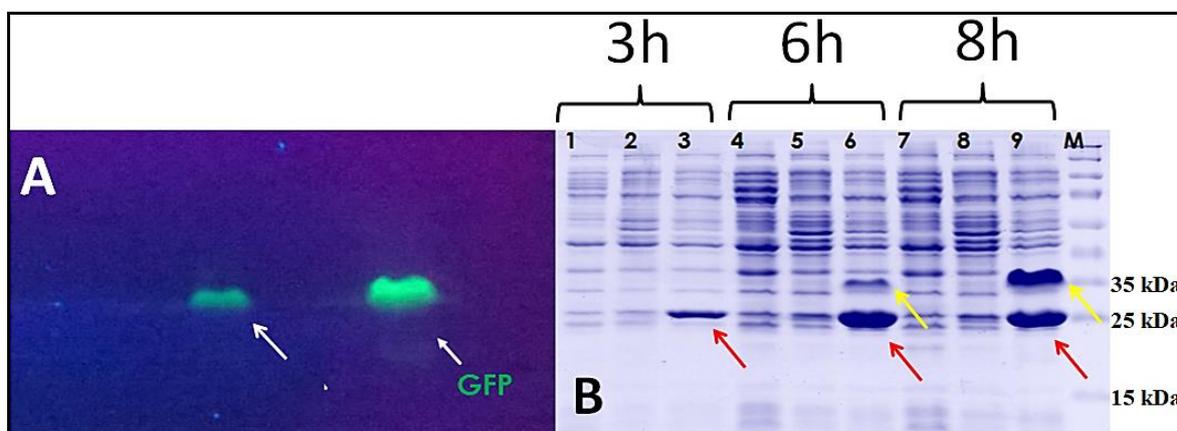
**Figura 36:** Fotografia dos clones recombinantes programados com o vetor pDMU01. A placa de Petri na esquerda conte um clone isolado e a placa da direita representa os clones crescidos dispostos nas mesmas posições em condições de cultivo. Os clones foram analisados em meio LB: com somente ampicilina, ampicilina e glicose, e ampicilina e IPTG (indutor); iluminados com luz negra (300 nm).

Os resultados visualmente dos clones dispostos na placa de Petri da direita e o clone isolado na placa de Petri da esquerda, cultivado em com glicose e ampicilina, mostraram que o gene da GFP não era expresso e a produção da proteína não ocorreu. No sistema somente com ampicilina, ocorreu a expressão do gene da GFP com nível de expressão basal, mostrando vazamento do sistema de regulação do promotor DM devido ao alto número de cópias do vetor, o que não ocorreu no pDM-GFP. No sistema com a ampicilina e o indutor IPTG a produção da proteína ocorreu com muito mais intensidade e em altos níveis de expressão, que pode ser explicado pelo alto número de cópias do vetor.

Prosseguiu-se para a análise quantitativa da expressão e regulação do gene da GFP em meio líquido. Acompanhou-se o crescimento e indução de um clone de *E. coli* portador do vetor pDMU01 e alíquotas foram coletados nos intervalos selecionados de 3, 6 e 8 horas de crescimento celular com 1, 3 e 5 horas de indução para análise das proteínas por eletroforese em gel desnaturante de poliacrilamida 12%. Como o fluoróforo da GFP é tridimensional, para ser possível sua visualização no gel de poliacrilamida por meio da fluorescência, após a solubilização das proteínas celulares no tampão de amostras para o

processo eletroforético, essa solução não foi fervida, para que a GFP não desnaturasse. O gel contendo as proteínas foi visualizado em UV antes de ser corado.

A figura 38 a seguir (sistemas **A** e **B**) mostra a análise eletroforética das proteínas produzidas de um clone de *E. coli* transformado com o plasmídeo recombinante pDMU01.



**Figura 37:** O resultado da análise por eletroforese em gel desnaturante de poliacrilamida e fluorescência da produção de GFP por um clone de *E. coli* JM110 contendo o vetor pDMU01 em diferentes condições de cultivo. A fotografia A mostra o gel obtido em transluminador (300 nm) e a B corado com o azul de Coomassie. As amostras nas raias 1, 4 e 7 representam as proteínas de clones crescidos com ampicilina e glicose; as amostras nas raias 2, 5 e 8 com somente ampicilina e amostras nas raias 3, 6 e 9 com ampicilina e IPTG. A raia M mostra o marcador molecular em kDa. A proteína recombinante GFP localiza-se na posição de aproximadamente 25 a 35 kDa (setas) nas raias 3, 6 e 9 na sua estrutura ativa (setas amarelas) e na sua estrutura inativa (setas vermelhas) nos intervalos selecionados de 3, 6 e 8 horas de crescimento celular com 1, 3 e 5 horas de indução, respectivamente.

Foram observados os resultados nos géis, sendo que no sistema **A**, as proteínas fluorescentes na sua conformação ativa foram visualizadas (indicado pelas setas brancas). Houve uma expressão praticamente nula das proteínas fluorescentes nas amostras em presença de glicose e com somente ampicilina nas canaletas **1, 2, 4, 5, 7 e 8**.

O sistema **B** mostra o gel corado com azul de Coomassie onde pode se observar o perfil das bandas de GFP na sua conformação fluorescente (indicado pelas setas amarelas) e na sua conformação não fluorescente (setas vermelhas). Com 3 horas de crescimento bacteriano e 1 hora de indução a proteína começou a ser produzida (canaleta 3) na conformação não fluorescente. Com 5 horas de crescimento e 3 horas de indução a proteína GFP continuou sendo produzida na sua conformação não fluorescente e na forma estruturalmente correta (canaleta 6) pela hospedeira *E. coli*. Em 8 horas de crescimento bacteriano e 5 horas de indução a proteína continuou sendo expressa de forma não enovelada corretamente, porém observou-se uma maior expressão da sua forma fluorescente e os níveis

de expressão foram os mais altos, considerando a expressão representada pela soma das duas bandas (seta vermelha e amarela na canaleta 9).

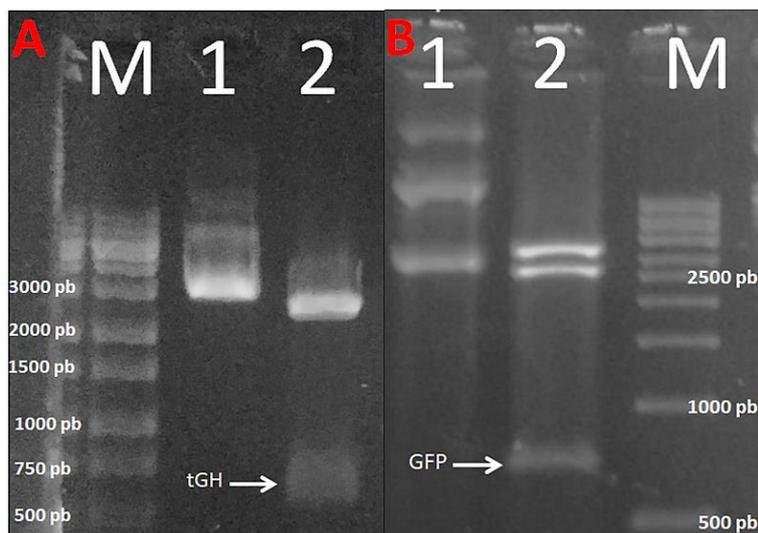
Estes resultados de expressão do gene da GFP (Figura 38, **A** e **B**) indicam que o promotor DM apresenta funcionalidade de força intermediária (entre média e forte), mas estridente nível de regulação, ou seja, praticamente nenhuma proteína foi expressa na ausência do indutor (IPTG) em *E. coli* quando o cassete de expressão foi trocado de um vetor de baixo número de cópias para alto número de cópias. Entretanto, quando o promotor é induzido com IPTG, há a produção da GFP fluorescente e não fluorescente (bandas separadas de peso molecular aparente de 25 e 35 kDa) em 8 horas com altíssima concentração (maior que 80% das proteínas totais), mostrando que esse vetor é capaz de programar em *E. coli* super-expressão da proteína recombinante.

Este sistema de expressão *E. coli* + pDMU01 mostrou que as proteínas fluorescentes começam a ser detectadas pelo equipamento com 6 horas de crescimento e 3 horas de indução (Figura 32-C na linha verde) e os resultados no gel indica a produção também da proteína na sua versão fluorescente (Figura 32-B no poço 6). Após 8 horas de crescimento e 5 horas de indução ocorre a produção da proteína fluorescente provavelmente devido as chaperonas moleculares e proteases que auxiliam respectivamente no enovelamento correto de proteínas e na destruição de proteínas não funcionais.

## 5.8 ANÁLISE DA REGULÇÃO E EXPRESSÃO DO tGH PELO VETOR pDMU01

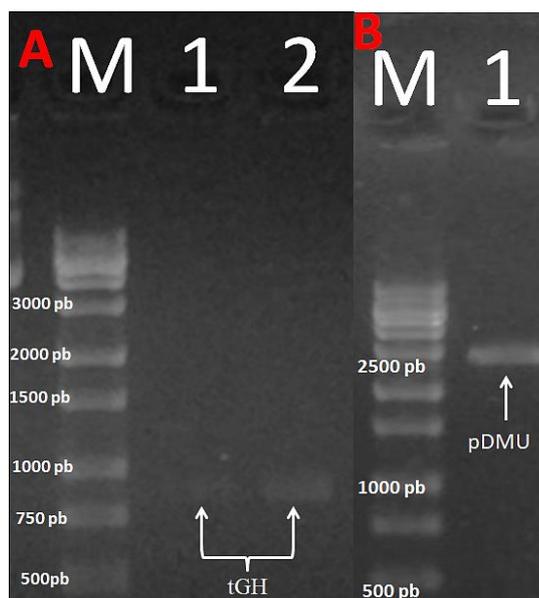
### 5.8.1 Construção do vetor de expressão pDMU-tGH

Para a construção do vetor pDMU-tGH, ou seja, a clonagem do tGH no vetor pDMU01, o fragmento entre os sítios das endonucleases *NdeI* e *BamHI* do pGSM-tGH contendo a sequência codificadora do tGH foi introduzido no vetor pDMU01 do qual o gene da GFP foi previamente tratado com as mesmas endonucleases. Inicialmente os plasmídeos foram extraídos de clones recombinantes de *E. coli* e então procedeu-se análise de restrição, os resultados estão mostrados na figura 39.



**Figura 38: Resultado da análise de restrição dos plasmídeos vetores pGSM-tGH e pDMU01 com as endonucleases *NdeI* e *BamHI*.** Na figura A, na canaleta 1 está o pGSM-tGH sem tratamento com endonucleases e na canaleta 2 o plasmídeo pGSM-tGH com *NdeI* e *BamHI*. Na figura B, na canaleta 1 está o pDMU01 sem tratamento enzimático e na canaleta 2 o pDMU01 com *NdeI* e *BamHI*. As setas em branco identificam os fragmentos do tGH e da GFP.

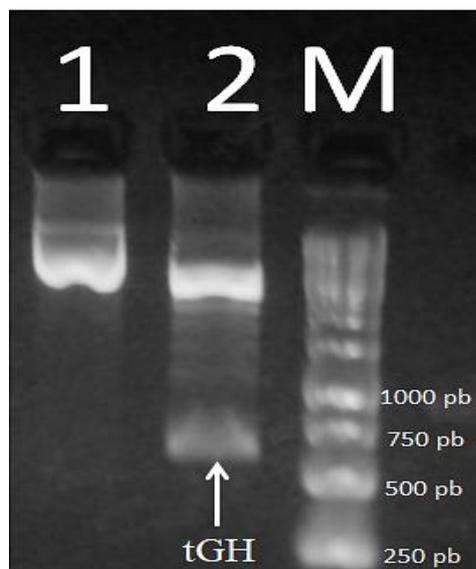
Na fotografia 39-A foi possível observar o fragmento gênico do tGH (cerca de 584 pb) na canaleta 2. Foi observado na parte B, na canaleta 2, o vetor pDMU01 com *NdeI* e *BamHI*. As setas brancas identificam os fragmentos correspondentes aos genes do tGH e da GFP (723 pb) após a tratamento com *NdeI* e *BamHI*. Após a análise de restrição, os fragmentos correspondentes ao vetor linearizado e o gene tGH foram purificados. A fotografia 40 a seguir o resultado da análise eletroforética dos produtos de purificação.



**Figura 39: Análise eletroforética dos fragmentos tGH e pDMU purificados.** Na parte A, nas canaletas 1 e 2 estão identificados pelas setas brancas o fragmento do tGH. Na parte B da figura, canaleta 1 está identificado o pDMU01.

Foi possível observar na fotografia 40, na parte **A**, os fragmentos correspondentes ao tGH nas canaletas 1 e 2, e na parte **B** da figura, o pDMU01 desprovido da GFP (canaleta 1). Após a purificação os fragmentos foram ligados e células de *E. coli* JM110 eletrocompetentes e transformadas geneticamente com o sistema de ligação.

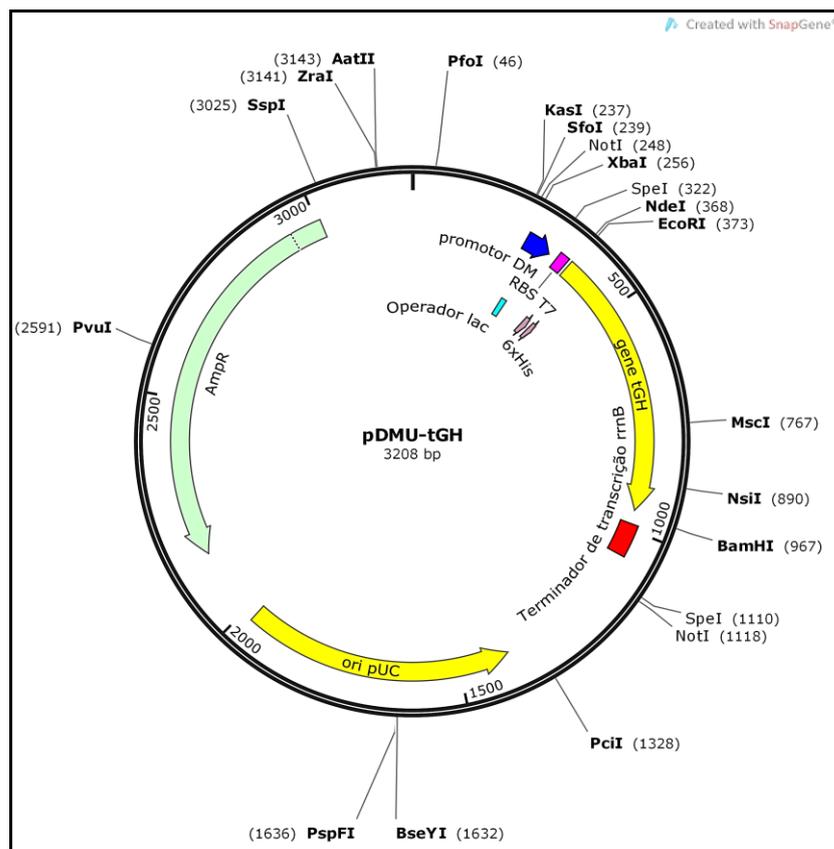
Diversos clones recombinantes foram obtidos e na Figura 41 está mostrada a análise de restrição do plasmídeo de um dos clones que foi denominado de pDMU-tGH.



**Figura 40: Análise eletroforética dos produtos de restrição do novo vetor com *NdeI* e *BamHI*.** Na canaleta 1, está o DNA plasmidial do vetor sem tratamento com endonucleases e na canaleta 2 está o DNA plasmidial do novo vetor digerido com *NdeI* e *BamHI*. A seta em branco mostra o fragmento correspondente ao gene do tGH.

Foi possível observar na fotografia 41 o fragmento do DNA do tGH (indicado pela seta branca) após tratamento com as endonucleases *NdeI* e *BamHI*. Este fato comprova que houve a clonagem do gene no novo vetor.

O novo plasmídeo recombinante construído foi denominado de pDMU-tGH (3208 pares de bases) possuindo o gene de resistência a ampicilina, a origem de replicação de alto número de cópias (pUC – 500~700 cópias) e com o gene tGH sob controle do promotor DM. O seu mapa físico está representado na figura 42 a seguir.



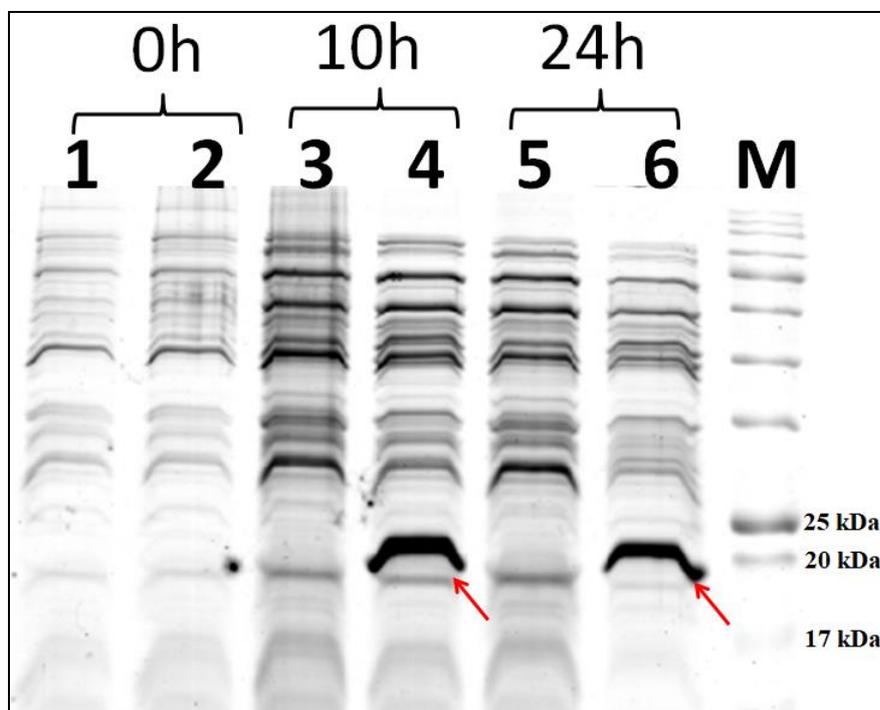
**Figura 41: Mapa físico do vetor pDMU-tGH.** A figura mostra os principais componentes do vetor como: os sítios das endonucleases, o promotor DM e o operador *lac*, o RBS do T7, o gene da recombinante do tGH, o terminador de transcrição *rrnB*, o marcador de seleção para ampicilina e a origem de replicação de alto número de cópias.

### 5.8.2 Expressão e regulação do gene tGH em *E. coli* JM110 e análise da expressão em SDS-PAGE

Para análise da regulação e expressão do gene tGH sob o controle do promotor DM, um clone de *E. coli* JM110 portador do vetor pDMU-tGH foi selecionado para pré-inóculo contendo glicose e ampicilina. No dia seguinte, foi inoculada uma alíquota do pré-inóculo em dois Erlenmeyers com somente ampicilina e as culturas incubadas em shaker com agitação.

O crescimento celular foi acompanhado. Quando a cultura bacteriana atingiu a O.D. igual a 0,4, retirou-se alíquotas para análise das proteínas, então em um dos frascos adicionou-se glicose e no outro IPTG. A partir deste momento, considerou-se como tempo inicial de indução (0 hora). Nos intervalos de 10 horas de crescimento bacteriano e 6 horas de indução, e 24 horas de crescimento bacteriano e 20 horas de indução, alíquotas

dos sistemas foram coletadas para mensurar o crescimento celular e análise das proteínas sintetizadas. Em seguida as amostras foram novamente normalizadas em relação às O.D.s e analisadas por eletroforese em gel desnaturante de poliacrilamida 12% (Fotografia 43).



**Figura 42: Análise da expressão do tGH por SDS-PAGE por um clone de *E. coli* JM110 contendo o vetor pDMU-tGH em diferentes condições de cultivo.** As amostras nas raias 1 e 2 representam a 0 hora com somente ampicilina. As raias 3 e 4 representam as proteínas de clones crescidos com com ampicilina e glicose e ampicilina e IPTG, respectivamente, após 10 horas de crescimento bacteriano e 6 horas de indução. As amostras nas raias 5 e 6 representam as proteínas com ampicilina e glicose e ampicilina e IPTG, respectivamente, em 24 horas de crescimento bacteriano e 20 horas de indução. A raia M mostra o marcador molecular em kDa. A proteína recombinante tGH localiza-se na posição de aproximadamente 24 kDa (setas vermelhas).

Pode-se observar claramente nos perfís eletroforéticos mostrados na figura 43 que o tGH foi expresso apenas nas condições com IPTG (canaletas 4 e 6) e em todas outras situações não foi possível detectar o tGH. Pode-se observar também que maior nível de expressão do tGH após 6 h de indução que após 20 h de indução. A diminuição do nível de tGH após 20h de indução provavelmente deve-se ao fato da cultura ter entrado em fase de degradação, com morte celular e consequente proteólise. Para determinar o melhor tempo de indução há necessidade de realizar experimentos com tempos de indução menos espaçados.

É surpreendente que praticamente não foi possível detectar expressão do tGH na presença da glicose mostrando que esse sistema pode ser modulado conforme a intenção do experimentador variando-se os níveis de glicose e IPTG.

## 5.9 MELHORAMENTO DO PROMOTOR DM POR SÍNTESE QUÍMICA DE DNA RANDÔMICA

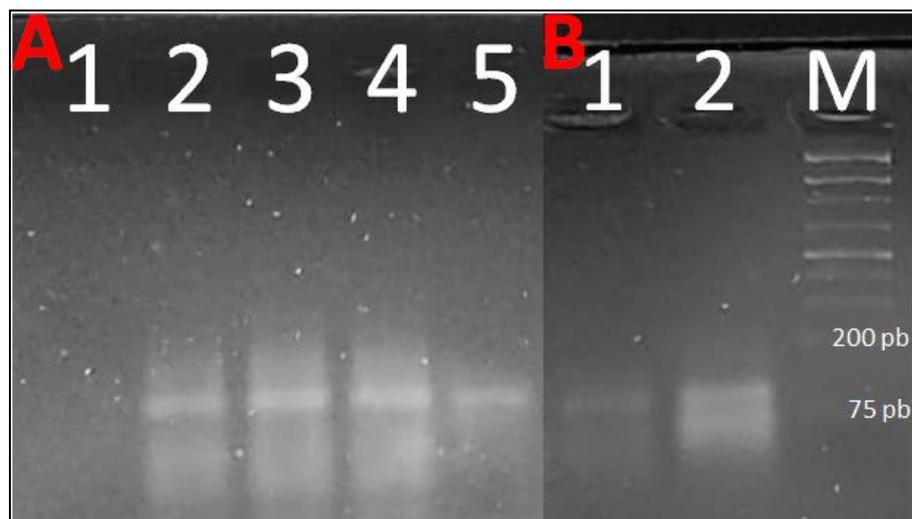
### 5.9.1 Construção da biblioteca dos novos promotores

5.9.1.1 Clonagem dos novos promotores no vetor pCDM e Transformação Genética em *E. coli* JM110

Para o melhoramento do promotor DM, foi realizada a construção de uma biblioteca de novos promotores tomando como base o promotor DM e como estratégia a síntese química com degeneração. Para obtenção dos novos promotores, foram sintetizados os oligonucleotídeos pDSM-1 e pDSM-2 (Quadro 2, figuras 12 e 13) que foram então hibridizados (figura 44-A).

Observando-se as canaletas 2, 3 e 4, devido ao aumento da fluorescência, pode-se inferir que ocorreu a hidridização (anelamento) como esperado, pois, os oligonucleotídeos possuem a sequência do operador lac e a região do elemento -10 estendido perfeitamente complementares entre si.

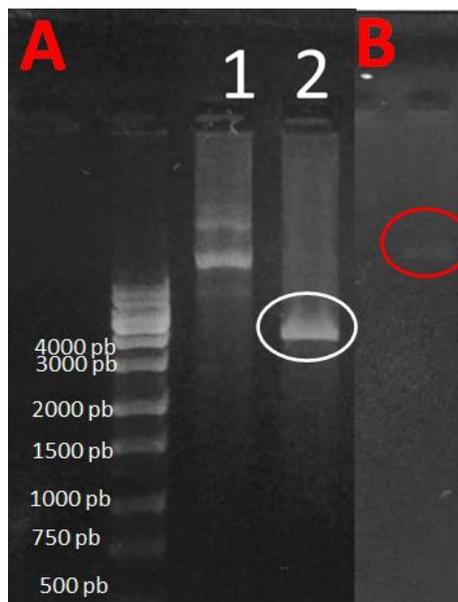
Após o anelamento, realizou-se o preenchimento das regiões fita-simples utilizando a enzima *Klenow*, a análise eletroforética do produto da reação está mostrada na figura 44-B . O aumento da fluorescência na canaleta 2 indica o funcionamento da *Klenow* pois DNA dupla fita absorve mais brometo de etídio e, portanto, fluoresce mais.



**Figura 43: Fotografia da análise do gel de eletroforese da hibridização dos oligonucleotídeos pDSM-1 e pDSM-2 e tratamento com *Klenow*.** Figura A - 1: somente oligonucleotídeo pDSM-1; 2, 3 e 4: oligonucleotídeos pDSM-1 e pDSM-2 anelados; e 5: somente oligonucleotídeo pDSM-2. Figura B - 1: oligonucleotídeos pDSM-1 e pDSM-2 somente anelados; 2: oligonucleotídeos pDSM-1 e pDSM-2 anelados e tratados com *Klenow*.

Os oligonucleotídeos dupla fita representando os diferentes promotores sintéticos foram purificados por eletroforese e por difusão do gel para o sobrenadante após sua maceração, isso foi necessário pois não foi possível purificar do gel usando kits comerciais não pois suas colunas não retinham oligos dupla fita menores que 100pb.

Para a clonagem dos promotores nos vetores, o vetor pCDM e os oligonucleotídeos foram digeridos com as endonucleases *EcoRI* e *SpeI*. O fragmento correspondente ao vetor pCDM linearizado (desprovido do promotor sintético DM) foi também purificado (45-A e B).



**Figura 44: Fotografia do perfil eletroforético do vetor pCDM com as endonucleases *SpeI* e *EcoRI* e purificado.** Na parte A, na canaleta 1 está o pCDM sem tratamento com endonuclease e na canaleta 2 o plasmídeo pCDM com as endonucleases *EcoRI* e *SpeI*. O círculo em branco mostra o vetor pCDM linearizado e na parte B o vetor pCDM após purificação (marcado em círculo vermelho).

Após as purificações o vetor foi ligado aos promotores e o produto utilizado para transformar geneticamente células de *E. coli* JM110 eletrocompetentes, obtendo-se 111 colônias transformantes em resistentes à canamicina na presença de glicose.

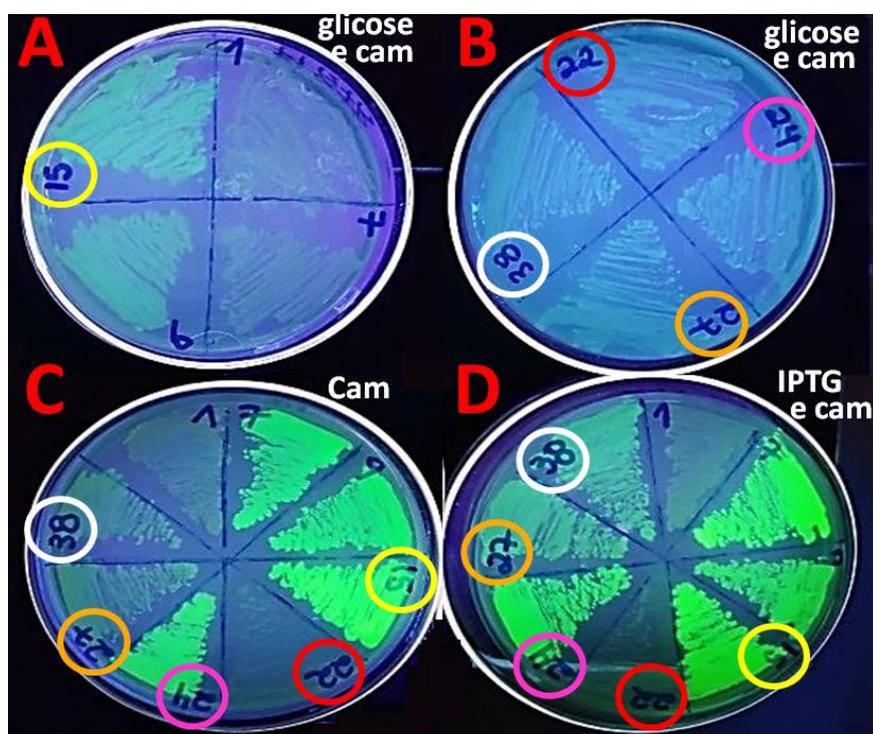
### **5.9.2 Análise da regulação e expressão do gene da GFP pelos diferentes clones recombinantes**

Para uma melhor caracterização da atividade e regulação do gene da GFP sob o controle dos novos promotores sintéticos, foram selecionados as 111 colônias recombinantes que cresceram em glicose e canamicina e reagrupadas em três placas de Petri (P1, P2 e P3). Estas colônias P1, P2 e P3 foram submetidas a uma nova análise por réplica-plate em diferentes condições com: 1) somente suplementado com o repressor glicose e o antibiótico canamicina; 2) somente o antibiótico canamicina; 3) e somente o antibiótico canamicina e o indutor IPTG. Um clone contendo o vetor pCDM foi adicionado para comparação com os novos clones recombinantes.

Foi possível observar clones exibindo diferentes níveis de regulação e expressão do gene da GFP em placas de Petri em diferentes condições de cultivo (Figura 46-A, B, C e D). Poucas colônias não apresentaram fluorescência, ou seja, não expressaram o gene da GFP em nenhuma das diferentes condições de cultivo, como exemplo, o clone 22.P1 (marcado em vermelho). Alguns clones não se mantiveram regulados por glicose e por isso expressaram um pouco o gene da GFP, indicando a perda de repressão por glicose nesta condição de cultivo. Nos sistemas com antibiótico e IPTG (Placas C e D) houve expressão da GFP de maneira mais expressiva, como exemplo, o clone 15.P1 (marcado em amarelo).

Alguns clones exibiram a regulação por glicose, ou seja, não expressaram o gene da GFP quando comparadas com o pCDM (clone 38.P1, marcado em círculo branco) nas mesmas condições de cultivo. Entretanto, alguns novos clones, exibiram uma maior atividade de expressão do gene da GFP no sistema de cultivo com somente com o antibiótico e com o sistema de cultivo com o indutor IPTG, como exemplo, o clone 24.P1 (marcado em círculo rosa).

Alguns clones mantiveram de maneira semelhante o mesmo perfil de regulação e expressão do gene da GFP em todas as condições de cultivo quando comparados com o pCDM, como exemplo o clone 27.P1 (marcado em círculo laranja).



**Figura 45: Fotografia de uma análise dos clones recombinantes em diferentes condições de cultivo e indução.** Na fotografia da placa A e B estão alguns diferentes clones em LB com glicose e o antibiótico. Na placa C e D, estão os mesmos clones dispostos em LB com somente o antibiótico canamicina e submetidos com canamicina e o indutor IPTG, respectivamente. Os círculos marcados com as diferentes cores mostram os clones com diferentes níveis de expressão e regulação do gene da GFP em diferentes condições de cultivo. O vetor pCDM está marcado em círculo branco.

Para confirmar a presença dos novos promotores sintéticos, alguns clones foram selecionados e seus plasmídeos extraídos. A clonagem dos novos promotores foram confirmadas digerindo-se os plasmídeos com a enzima *PstI* mostrando a liberação de um fragmento de DNA de aproximadamente 944 pb.

Após a confirmação da presença dos novos promotores nos plasmídeos dos clones escolhidos, 18 clones da placa 1 (P1), 16 clones da placa 2 (P2) e 17 clones da placa 3 (P3) com diferentes regulações e intensidades de GFP foram selecionados para análise das sequências de DNA dos promotores.

### 5.9.3 Sequenciamento e análise das novas sequências promotoras dos clones

De acordo com nosso modelo de randomização de promotores, a biblioteca continha possíveis 65.536 sequências diferentes. Para estudar a base molecular das diferentes forças dos promotores, foram escolhidos os 51 DNAs plasmidiais de clones recombinantes e submetidos à PCR para amplificação de um amplicom de 504 pb contendo as regiões promotoras utilizando o par de primers VF1 e RCP1. Após a PCR, os produtos amplificados foram visualizados em gel de agarose e sequenciados.

Após o sequenciamento foi avaliada a qualidade e o tamanho das sequências de cada fragmento pelo software Electropherogram quality analysis (<http://www.biomol.unb.br/phph/index.html>). Este software analisa as sequências geradas pelo sequenciador automático e mostra a qualidade usando o programa PHRED. A partir disso, as sequências foram alinhadas com o programa BioEdit para análise das regiões que compõem os promotores com a sequência do promotor DM adicionada em todas as análises para comparação.

Foram sequenciados os promotores dos clones das placas 1, 2 e 3, que perderam a regulação por glicose e com altos níveis de expressão foram alinhadas. A figura 47a seguir mostra os promotores após as análises de alinhamento (5' → 3').





Existem muitos promotores eficientes que utilizam o elemento UP, como o promotor pL do fago  $\lambda$  e o promotor P1 do gene *rrnB1* de *E. coli*. O quadro 4 a seguir mostra comparativamente os elementos UPs obtidos neste trabalho e com os da literatura (RUFF et al., 2015; LEWIS et al., 2016).

**Quadro 4: Sequências do UP elemento.**

Modelo	Sequência do elemento UP 5' → 3'	Tamanho
Promotor pL	AAAAAATAAATTCATATAAAAAA	23 pb
Promotor P1 do gene <i>rrnB</i>	AAAATTATTTTAAATTCCTC	21 pb
Promotor DM	TTAAAATTATTTTCAAACGG	20 pb
Randomizado neste trabalho	WWAAAWWTWTTTTWWAAWAK	20 pb
Consensual	NNNAAAWWTWTTTTNNNAAANN	23 pb

(W = A ou T; N = qualquer nucleotídeo)

Foram mantidas as sequências nucleotídicas conservadas da região espaçadora que compõem uma parte do operador lac O1 e do elemento -10 estendido. Na região -35 (TTGTGA) do promotor DM foi modificada a parte da sequência TGA. A sequência consenso da região -35 é TTGACA, com o -35 = T; -34 = T; e -33 = G sendo estas bases as mais altamente conservadas (HOOK-BARNARD & HINTON, 2007; ROSINSKI-CHUPIN et al., 2015; RUFF et al., 2015). Foi decidido manter o TTG intacto e o restante da sequência (TGA) foi randomizado para **WSW** (**W** = T ou A e **S** = G ou C) na tentativa de obter promotores com a sequência consensual ou restaurar a sequência completa do operador lac O1 para manter a regulação por IPTG.

A região do elemento -10 estendido localizada logo a montante da região -10, não foi randomizada por apresentar similaridade com a sequência consenso TRTGn, e com isso serviu para compor a região de anelamento como estratégia para produzir-se as degenerações em outras posições (HOOK-BARNARD & HINTON, 2007; RUFF et al., 2015).

Decidiu-se degenerar a região -10 em uma única base, trocando o nucleotídeo G para **K** (G ou T) na tentativa novamente de refazer a região -10 do promotor DM (GATAAT) ou tentar adicionar uma base nucleotídica T para ficar igual à sequência consenso da região -10

(TATAAT) (HOOK-BARNARD & HINTON, 2007; ROSINSKI-CHUPIN et al., 2015; RUFF et al., 2015). As novas sequências promotoras passaram a ter a região consensual TATAAT e em outros promotores a sequência GATAAT foi restaurada. Não foi possível determinar exatamente a contribuição de cada base, mas sabe-se que o G é muito utilizado em promotores de fagos e o T é o preferencial de sigma 70 (MOREIRA et al., 2018).

Na região discriminadora foi mantido os nucleotídeos GGG consensuais, nas posições -7, -6 e -5, após a região -10 e foram realizadas degenerações nas posições -4 -3 -2 e -1 do discriminador(GGGGACCA) para tentar diminuir o espaçamento de 7 para 6 bases nucleotídicas nesta região do promotor. A estratégia foi realizar três degenerações como sendo duas seguidas para **S** = C ou G e uma para **R** = G ou A na região discriminadora do promotor DM (GGGASSRA). Desta maneira poderíamos restaurar o espaçamento original do discriminador do promotor com uma nova combinação de sequência (GGGA G/C G/C G/A A) respeitando o conteúdo preferencial em G e C nesta região e a obtenção de sequências com G ou A na posição -1 (HOOK-BARNARD & HINTON, 2007; ROSINSKI-CHUPIN et al., 2015; RUFF et al., 2015).

Os resultados mostraram que a estratégia de modificação da sequência do discriminador em alguns promotores pode ter diminuído de 7 para 6 nucleotídeos nas sequências analisadas dos clones que mantiveram a regulação por glicose, como os clones 7.P1, 14.P1, 20.P2 e 29.P3 que tiveram um aumento de expressão provavelmente devido a redução do número de nucleotídeos nesta posição, já que a sequência TGA foi restaurada.

Os novos clones regulados por glicose e com maiores níveis de expressão da GFP exibiram o nucleotídeo A na posição +1, estes clones podem ter apresentado novos discriminadores com espaçamento de 6 nucleotídeos como o consensual (HOOK-BARNARD & HINTON, 2007; ROSINSKI-CHUPIN et al., 2015; RUFF et al., 2015).

No design do vetor pDM02 não foi planejada a adição do sítio de reconhecimento da CAP 5'-AATGTGATCTAGATCACATTT-3' na região de -61,5 ou em -41,5 como existentes em promotores dependentes da CAP, mas isso ocorreu conferindo essa característica a esse promotor o que faz com que seja também modulado por glicose.

As sequências dos clones que mantiveram a regulação em glicose, o motivo TGA na região -35 foi restaurado, como existente na sequência original do promotor DM. Os clones

que perderam a regulação em glicose apresentaram a região -35 diferente da sequência original.

Uma nova saga foi iniciada para descobrir a regulação pela CAP, pois em nossos resultados na presença de glicose, ocorre a repressão catabólica, ou seja, as proteínas não são expressas pelo promotor DM na presença de glicose. Nossas investigações começaram supondo que a sequência do operador lac sobreposta na região -35 pode ter mimetizado uma região de ancoramento para a CAP e que nesse caso estaria atuando como molécula repressora em promotores dependentes de CAP tipo II (LAWSON et al., 2004).

No banco de dados do RegulonDB (<http://regulondb.ccg.unam.mx/>) existem 430 promotores regulados pela CAP. A sequência do promotor DM (-58 até o +1) foi comparada com as do banco de dados e não mostrou identidade com nenhum promotor. Foi realizada a busca por outras regiões do promotor e por fim foi comparada a sequência do operador lacO1 de 21 pb (5'- AATTGTGAGCGGATAACAATT-3'). Nesse caso três sequências de promotores de operon lac foram encontradas, os promotores lacZp1, lacZp2 e lacZp3 (TIERRAFRÍA et al., 2022).

O promotor lacZp1 é o principal promotor do operon onde a CAP se liga nas posições +11,5 e -61,5. O promotor lacZp2 é regulado nas posições +33,5 e -39,5 e o promotor lacZp3 é regulado nas posições +26 e -46,5. As posições denotam os sítios de ancoramento para CAP em relação a posição +1 dos três promotores (lacZp1, lacZp2 e lacZp3). Estes resultados nos indicou que a CAP pode reconhecer como seu próprio sítio de ligação, a sequência do operador lac O1 em algumas posições próximo do núcleo do promotor (TIERRAFRÍA et al., 2022).

A seguinte análise foi inserir no campo de busca do Regulon, o termo “galp” que significa “promotor galP” do operon que codifica as enzimas para o metabolismo da galactose (galETKM) e o promotor mais bem caracterizado com o sítio de ancoramento para a CAP próximo da região -35. As sequências do operon de galactose foram investigadas e existem três promotores regulados pela CAP a ser: galEp, galEp1 e galEp2. O promotor galEp e galEp1 são regulados pela CAP na posição -41,5 e o promotor galEp2 é regulado pela CAP na posição -36,5. Os dois principais promotores regulados pela CRP estão situados de maneira opostas; um (galP1) é ativado e o outro (galP2) reprimido pela CRP na região -41,5 (WEICKERT & ADHYA, 1993).

Uma outra análise foi inserindo no campo de busca o termo “-35”, (a região -35) na tentativa de obter sequências reguladoras reconhecidas pela CAP nesta posição ou em adjacentes. Não foi encontrado algum promotor regulado pela CAP na posição -35, porém existe em adjacentes como no -34,5 e no -35,5, como os promotores proPp1 e o ptsHp2, respectivamente. O promotor ptsHp2 apresenta a sequência de ligação para a CAP como TTTATGATTTGGTTCAATTC não demonstrando tanta similaridade com a sequência consenso AATGTGATCTAGATCACATT, e o motivo TGA e TCA são fortemente conservados na sequência de ligação para CAP.

A CAP tem um sítio de reconhecimento assimétrico com dois meios-sítios distintos, um dos quais é semelhante ao sítio canônico de ligação, e o outro meio-sítio menos conservado (TIERRAFRÍA et al., 2022). CAP reconhece seu sítio de DNA por meio de uma combinação de “leitura direta” (reconhecimento de sequência de DNA mediado por ligação direta de hidrogênio ou interações van der Waals com pares de bases de DNA) e “leitura indireta” (reconhecimento de sequência de DNA mediado pela detecção de efeitos dependentes de sequência de DNA na posição de fosfato de DNA, solvatação de fosfato de DNA ou suscetibilidade à deformação do DNA) (BUSBY & EBRIGHT, 1999; LAWSON et al., 2004). No promotor DM o motivo TGA presente no promotor está pontuado especificamente nas posições -33, -32 e -31 e não está presente o outro motivo TCA. Mas diante de que o sítio da ligação da CAP poderia ser a própria região operadora O1, a CAP pode atuar reconhecendo como AATTGTGAGCGGATAACAATT, com um espaçamento de 5 nucleotídeos e não como na sua sequência de reconhecimento original AATGTGATCTAGATCACATT com 6 nucleotídeos.

A interação espontânea dependente de sistema CAP-cAMP desempenha um papel na instabilidade da expressão recombinante e não são projetados vetores de expressão com a regulação por fatores de transcrição, por exemplo, a adição de 1% de glicose a células de fase logarítmica praticamente eliminaria a expressão da proteína recombinante. Trabalhos com modificações nesta sequência ocasionaram a destruição da sequência de ligação reconhecida por CAP (GROSSMAN et al., 1998).

Foi desenvolvido um promotor mutante derivado do promotor lac denominado de placUV5, esse promotor difere do promotor lac selvagem em 2 nucleotídeos na região conservada -10, tornando-a idêntica à sequência consenso TATAAT (a selvagem é TATGTT) o que o torna 2,5 vezes mais forte que o promotor lac nativo. Além disso, o

PlacUV5 não contém o sítio de ligação da proteína CAP e, portanto, não é passível de regulação por cAMP/glicose. Esse promotor integra cassetes de expressão quando se necessita expressar um gene em baixo/médio nível de uma forma bem regulada (MOREIRA et al., 2018). O promotor lacUv5 possui mutações no sítio de ligação na CAP para torná-lo insensível a repressão catabólica nas posições -55 (C para T) e -66 (G para A), bem como mutações UV5 na região -10 (-9, -8 GT > AA) para aumentar a sua interação com o fator sigma 70 (FULLER, 1982; YU et al., 2021).

Os resultados encontrados no presente trabalho reforçam que o motivo TGA da região -35 modificado em alguns promotores fez com que a regulação por glicose mediada por via CAP fosse inativada. A definição do promotor DM visava adicionar na região espaçadora (entre as regiões -35 e -10) o operador lac O1, por isso na região -35 adicionou-se TTGTGA ao invés de TTGACA. De uma maneira surpreendente mesmo sem ter adicionado o sítio de ligação de CAP propositalmente, a expressão do gene da GFP estava sendo regulada por glicose.

Visando melhorar o nível de expressão desse promotor utilizou-se a estratégia de síntese química com nucleotídeos degenerados e obteve-se vários promotores diferentes. Alguns clones perderam a regulação por glicose e o sequenciamento das revelou que os promotores com mutações no motivo TGA não estavam mais tão bem reguladas por glicose, o que indica que a sequência sobreposta como operador na região -35 é um sítio de ligação para CAP. A literatura mostra vários promotores sendo regulados para o CAP (atuando como ativadora, repressora ou as duas ações), inclusive os promotores da família do lac e gal, que são bem caracterizados e possuem a montante, a jusante ou próximo núcleo o sítio de ligação para CAP na região -35, -10 e na região espaçadora (TIERRAFRÍA et al., 2022).

Além disso, pensa-se que o dímero de CAP funcional contém apenas uma única molécula de cAMP. É geralmente aceito que as proteínas reguladoras, como a CAP, são dímeros para aumentar sua especificidade de ligação ao DNA. No entanto, uma vez ligado ao DNA, um único domínio de ativação adequadamente posicionado pode ser suficiente para a ativação da transcrição (GASTON et al., 1990).

Os promotores lac e híbridos que utilizam a regulação do lac, são conhecidos por terem transcrição vazada (“leak”), ou seja, a transcrição ocorre um pouco mesmo quando o indutor está ausente. A maior expressão basal foi observada para os vetores contendo

promotores *trc* e *tac*. Este fato foi observado em culturas de *E. coli* transformadas com vetores *P<sub>trc</sub>* e *P<sub>tac</sub>* com *ori* de baixa cópia (*p15A*). Apesar de carregar a versão *LacI<sub>q</sub>* do gene do repressor o uso destes vetores pode ser uma desvantagem quando se deseja expressar na bactéria uma proteína tóxica (LOZANO TEROL et al., 2021).

No trabalho de Schuller e colaboradores, 2020, os promotores que continham apenas um sítio *lacO* exibiram uma força de promotor consideravelmente maior, mas também maior vazamento da expressão em promotores /operadores que continham dois sítios *lacO*. Quando os dois *lacO* estavam localizados a uma distância de 62 pb, eles exibiram uma afinidade de ligação muito forte com a molécula repressora, mas estas condições resultaram em uma parada completa na produtividade da produção das proteínas recombinantes em células parcialmente induzidas (SCHULLER et al., 2020).

Para contornar problemas de expressão e regulação de proteínas, um conjunto de promotores bacterianos foi construído por Hothersall e colaboradores, 2021, com uma variedade de diferentes promotores de força baixa, intermediária e alta. Os desenhos gênicos dos promotores foram com base nos promotores *lac* ou *tac* e com sequências de DNA do operador *lac O1* ou/e *O3* foram utilizadas para produção das proteínas-alvo: GFP, hormônio de crescimento humano (hGH) e fragmentos de anticorpo de região variável de cadeia simples (scFvs). Além disso, devido à sua regulação rigorosa, demonstraram o que este sistema também pode expressar uma proteína tóxica da membrana externa com sucesso em diferentes cepas *E. coli* (HOTHERSALL et al., 2021).

Os clones 16.P1, 25.P1, 35.P1 (placa 1); 4.P2, 6.P2, 8.P2, 12.P2, 22.P2, 31. P2, 36.P2 (placa 2) e 1.P3, 2.P3, 4.P3, 16.P3, 17.P3, 18.P3, 20.P3, 22.P3 (da placa 3) exibiram visualmente os maiores níveis de expressão e foram analisados com mais detalhes. Esses clones perderam a regulação por glicose e expressaram níveis maiores do gene da GFP nesta condição. Foi constatado que os clones 16.P1, 25.P1, 35.P1, 22.P2, 31. P2, 36.P2, 1.P3, 4.P3, 17.P3, possuíam na região -35 um ou dois nucleotídeos do consenso ACA. Os clones 6.P2 e 22.P3 possuem o ACA na região -35 e com isso restaurou-se a sequência consenso nesses promotores e os clones 12.P2, 2.P3, 16.P3, 18.P3 e 20.P3 apresentaram na região -35 a sequência TTGTGT e com diferenças nas regiões -10 entre a sequência GATAAT ou a consenso TATAAT. Esses resultados reforçam o conhecimento de que a região TTGTGA é um importante sítio de ligação da CAP pois as alterações na sensibilidade dos promotores à glicose ocorreram nos promotores com alterações na sequência TGA.

Os promotores e o conjunto de vetores construídos nesse trabalho permitem a abordagem de diferentes estratégias de clonagem e eficiente expressão de genes heterólogos, inclusive para a expressão de produtos tóxicos. A metodologia aqui desenvolvida pode ser aplicada a outros microrganismos, podendo gerar promotores de diversas forças e diferentes níveis de regulação para serem usados para a produção de proteínas recombinantes em larga escala e mesmo para se realizar engenharia metabólica para diferentes fins.

## 6 CONSIDERAÇÕES E PERSPECTIVAS

Neste trabalho idealizou-se, construiu-se e testou-se um promotor sintético DM, que mostrou excelente regulação e nível médio de força. O promotor, motivo de patente de invenção (anexa), foi inserido em um cassete de expressão e todo o conjunto foi clonado em um vetor de baixo número de cópias (pDM-GFP) e testado sua funcionalidade em *E. coli*.

Para estudos posteriores da regulação e análise da funcionalidade do cassete de expressão, foi construído um conjunto de vetores baseado no promotor DM. O promotor DM foi inserido por diferentes estratégias em diferentes chassis plasmidiais para clonagem e expressão de genes heterólogos. Os plasmídeos recombinantes resultantes foram denominados de: pDM02, pCDM, pDMU01 e pDMU02. Os vetores construídos podem ser utilizados para clonagem de promotores e clonagem e expressão de genes homólogos e heterólogos com o uso de diferentes marcadores de seleção e origens de replicação. Os genes clonados podem ser expressos via o promotor DM por indução através de um único indutor (IPTG).

Os estudos realizados nesse trabalho indicaram que o promotor DM embora tenha bom nível de regulação, sua força poderia ainda ser aumentada. Nesse contexto visando aumentar a força desse promotor foi construída uma biblioteca de promotores sintéticos para serem utilizados em eficientes vetores de expressão de genes heterólogos na *E. coli*. Para confirmação de que o promotor é regulado por CAP, um experimento com uma hospedeira *E. coli* deficiente de CAP ou cAMP deveria ser realizado (GROSSMAN et al., 1998).

Em outra vertente, espera-se determinar a força do promotor DM pela técnica de RT-qPCR utilizando o vetor pDM-GFP comparando com a expressão de genes constitutivos (*housekeeping*) do genoma de *E. coli* JM110 descritos na literatura.

Considera-se que os novos promotores podem ser utilizados na construção de um maior conjunto de vetores de expressão com diferentes características para serem utilizados em biologia sintética e engenharia metabólica.

## 7 REFERENCIAL TEÓRICO

ABDELAAL, A. S.; YAZDANI, S. S. A genetic toolkit for co-expression of multiple proteins of diverse physiological implication. **Biotechnology Reports**, 2021.

ADRIO, J. L.; DEMAIN, A. L. Recombinant organisms for production of industrial products. **Bioengineered Bugs**, 2010.

AMORES, G. R.; GUAZZARONI, M. E.; SILVA-ROCHA, R. Engineering Synthetic cis-Regulatory Elements for Simultaneous Recognition of Three Transcriptional Factors in Bacteria. **ACS Synthetic Biology**, 2015.

ANDREWS, B. et al. A tightly regulated high level expression vector that utilizes a thermosensitive lac repressor: Production of the human T cell receptor V $\beta$ 5.3 in *Escherichia coli*. **Gene**, 1996.

ANDRIANANTOANDRO, E. et al. Synthetic biology: New engineering rules for an emerging discipline. **Molecular Systems Biology**, 2006.

ARPINO, J. A. J. et al. Tuning the dials of synthetic biology. **Microbiology (United Kingdom)**, 2013.

BAESHEN, M. N. et al. Production of biopharmaceuticals in *E. Coli*: Current scenario and future perspectives. **Journal of Microbiology and Biotechnology**, 2015.

BAESHEN, N. A. et al. Cell factories for insulin production. **Microbial Cell Factories**, v. 13, n. 1, 2014.

BARINOVA, N. et al. Structural modules of RNA polymerase required for transcription from promoters containing downstream basal promoter element GGGA. **Journal of Biological Chemistry**, v. 283, n. 33, p. 22482–22489, 2008.

BARRICK, D. et al. Quantitative analysis of ribosome binding sites in *E.coli*. **Nucleic Acids Research**, 1994.

BEREZHNOY, A. Y.; SHCKORBATOV, Y. G. Dependence of the *E. coli* promoter strength and physical parameters upon the nucleotide sequence. **Journal of Zhejiang University. Science. B**, v. 6, n. 11, p. 1063–8, 2005.

BERLEC, A.; ŠTRUKELJ, B. Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells. **Journal of Industrial Microbiology and Biotechnology**, 2013.

BERMAN, M. L.; LANDY, A. Promoter mutations in the transfer RNA gene *tyrT* of *Escherichia coli*. **Proceedings of the National Academy of Sciences of the United States of America**, 1979.

BICH, L. et al. Biological regulation: controlling the system from within. **Biology and Philosophy**, v. 31, n. 2, p. 237–265, 2016.

BORUKHOV, S.; NUDLER, E. RNA polymerase holoenzyme: Structure, function and biological implications. **Current Opinion in Microbiology**, 2003.

- BOSDRIESZ, E. et al. How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. **FEBS Journal**, 2015.
- BREN, A. et al. Glucose becomes one of the worst carbon sources for E.coli on poor nitrogen sources due to suboptimal levels of cAMP. **Scientific Reports**, 2016.
- BROWNING, D. F. et al. Exploitation of the Escherichia coli lac operon promoter for controlled recombinant protein production. **Biochemical Society Transactions**, v. 47, n. 2, p. 755–763, 2019.
- BROWNING, D. F.; BUSBY, S. J. W. Local and global regulation of transcription initiation in bacteria. **Nature Reviews Microbiology**, v. 14, n. 10, p. 638–650, 2016.
- BROWNING, D. F. D. F.; BUSBY, S. J. W. S. The regulation of bacterial transcription initiation. **Nature reviews. Microbiology**, v. 2, n. 1, p. 57–65, 2004.
- BURDEN, S.; LIN, Y. X.; ZHANG, R. Improving promoter prediction for the NNPP2.2 algorithm: A case study using Escherichia coli DNA sequences. **Bioinformatics**, v. 21, n. 5, p. 601–607, 2005.
- BURGESS, R. R. Separation and Characterization of the Subunits of Ribonucleic Acid Polymerase\*. **THE JOURNAL OF BIOLOGICAL CHEMISTRY CHEMISTRY**, v. 244, n. 22, p. 6168–176, 1969.
- BUSBY, S. et al. The selection and characterization of two novel mutations in the overlapping promoters of the Escherichia coli galactose operon. **Gene**, 1984.
- BUSBY, S.; EBRIGHT, R. H. **Transcription activation by catabolite activator protein (CAP)** **Journal of Molecular Biology**, 1999.
- CAMBRAY, G. et al. Measurement and modeling of intrinsic transcription terminators. **Nucleic Acids Research**, v. 41, n. 9, p. 5139–5148, 2013.
- CAMPBELL, E. A.; WESTBLADE, L. F.; DARST, S. A. Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. **Current opinion in microbiology**, v. 11, n. 2, p. 121–7, 2008.
- CASES, I.; DE LORENZO, V. Promoters in the environment: Transcriptional regulation in its natural context. **Nature Reviews Microbiology**, 2005.
- CHALFIE, M. et al. Green fluorescent protein as a marker for gene expression. **Science**, v. 263, n. 5148, p. 802–805, 1994.
- CHANDRANGSU, P.; HELMANN, J. D. Sigma Factors in Gene Expression. In: **eLS**. [s.l: s.n.].
- CHANG, A. C. Y.; COHEN, S. N. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. **Journal of Bacteriology**, 1978.
- CHEN, H. et al. Determination of the optimal aligned spacing between the shine - dalgarno sequence and the translation initiation codon of escherichia coli m RNAs. **Nucleic Acids**

**Research**, 1994.

CHEN, R. Bacterial expression systems for recombinant protein production: E. coli and beyond. **Biotechnology Advances**, v. 30, n. 5, p. 1102–1107, 2012.

CHEN, X. et al. DCEO Biotechnology: Tools to Design, Construct, Evaluate, and Optimize the Metabolic Pathway for Biosynthesis of Chemicals. **Chemical Reviews**, 2018.

CHOI, J. H.; KEUM, K. C.; LEE, S. Y. Production of recombinant proteins by high cell density culture of Escherichia coli. **Chemical Engineering Science**, 2006.

COSTA, L. S. et al. Daily rhythms of the expression of genes from the somatotrophic axis: The influence on tilapia (*Oreochromis niloticus*) of feeding and growth hormone administration at different times. **Comparative Biochemistry and Physiology Part - C: Toxicology and Pharmacology**, 2016.

COZZARELLI, N. R.; KELLY, R. B.; KORNBERG, A. A minute circular DNA from Escherichia coli 15. **Proceedings of the National Academy of Sciences of the United States of America**, v. 60, n. 3, p. 992–999, 1968.

CROOKS, G. E. et al. WebLogo: A sequence logo generator. **Genome Research**, 2004.

CUMMINGS, D. E.; MERRIAM, G. R. Growth Hormone Therapy in Adults. **Annual Review of Medicine**, 2003.

CZARNIECKI, D.; NOEL, R. J.; REZNIKOFF, W. S. The -45 region of the Escherichia coli lac promoter: CAP-dependent and CAP-independent transcription. **Journal of Bacteriology**, 1997.

DADE-ROBERTSON, M. et al. Design and modelling of an engineered bacteria-based, pressure-sensitive soil. **Bioinspiration and Biomimetics**, 2018.

DEMAIN, A. L.; VAISHNAV, P. Production of recombinant proteins by microbes and higher organisms. **Biotechnology Advances**, 2009.

DEUTCH, C. E. Transformation of Escherichia coli with the pGLO Plasmid: Going beyond the Kit . **The American Biology Teacher**, 2018.

DRENNAN, A. et al. Key roles of the downstream mobile jaw of escherichia coli RNA polymerase in transcription initiation. **Biochemistry**, 2012.

DÜRWARD, H.; HOFFMANN-BERLING, H. Endonuclease I-deficient and ribonuclease I-deficient Escherichia coli mutants. **Journal of Molecular Biology**, 1968.

EBRIGHT, R. H.; EBRIGHT, Y. W.; GUNASEKERA, A. Consensus DNA site for the Escherichia coli catabolite gene activator protein (CAP): CAP exhibits a 450-fold higher affinity for the consensus DNA site than for the E.coli lac DNA site. **Nucleic Acids Research**, v. 17, n. 24, p. 10295–10305, 1989.

ENGSTROM, M. D.; PFLEGER, B. F. Transcription control engineering and applications in synthetic biology. **Synthetic and Systems Biotechnology**, 2017.

- ESTREM, S. T. et al. Identification of an UP element consensus sequence for bacterial promoters. **Proceedings of the National Academy of Sciences**, v. 95, n. 17, p. 9761–9766, 1998.
- ESTREM, S. T. et al. Bacterial promoter architecture: Subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase  $\alpha$  subunit. **Genes and Development**, 1999.
- FEKLISTOV, A. et al. A Basal Promoter Element Recognized by Free RNA Polymerase  $\sigma$  Subunit Determines Promoter Recognition by RNA Polymerase Holoenzyme. **Molecular Cell**, v. 23, n. 1, p. 97–107, 2006.
- FEKLISTOV, A. RNA polymerase: In search of promoters. **Annals of the New York Academy of Sciences**, v. 1293, n. 1, p. 25–32, 2013.
- FEKLÍSTOV, A. et al. Bacterial Sigma Factors: A Historical, Structural, and Genomic Perspective. **Annual Review of Microbiology**, v. 68, n. 1, p. 357–376, 2014.
- FERNÁNDEZ-CASTANÉ, A. et al. Evidencing the role of lactose permease in IPTG uptake by *Escherichia coli* in fed-batch high cell density cultures. **Journal of Biotechnology**, 2012.
- FERRER-MIRALLES, N. et al. Microbial factories for recombinant pharmaceuticals. **Microbial Cell Factories**, 2009.
- FULLER, F. A family of cloning vectors containing the lacUV5 promoter. **Gene**, v. 19, n. 1, p. 43–54, 1982.
- GAAL, T. et al. Localization of the intrinsically bent DNA region upstream of the *E. coli* rrnB P1 promoter. **Nucleic Acids Research**, 1994.
- GALE, G. A. R.; WANG, B.; MCCORMICK, A. J. Evaluation and Comparison of the Efficiency of Transcription Terminators in Different Cyanobacterial Species. **Frontiers in Microbiology**, v. 11, 2021.
- GASTON, K. et al. Stringent spacing requirements for transcription activation by CRP. **Cell**, 1990.
- GEZVAIN, K. ; LANDICK, R. The structure of bacterial RNA polymerase. **The bacterial chromosome**, n. 40, p. 283–296, 2004.
- GHAVIM, M. et al. High level expression of recombinant human growth hormone in *Escherichia coli*: Crucial role of translation initiation region. **Research in Pharmaceutical Sciences**, 2017.
- GHOSH, T.; BOSE, D.; ZHANG, X. Mechanisms for activating bacterial RNA polymerase. **FEMS Microbiology Reviews**, 2010.
- GILBERT, W.; MAXAM, A. The Nucleotide Sequence of the lac Operator. **Proceedings of the National Academy of Sciences**, 1973.
- GOPAL, G. J.; KUMAR, A. Strategies for the Production of Recombinant Protein in *Escherichia coli*. **The Protein Journal**, v. 32, n. 6, p. 419–425, 2013.

- GOURSE, R. L.; ROSS, W.; GAAL, T. UPs and downs in bacterial transcription initiation: The role of the alpha subunit of RNA polymerase in promoter recognition. **Molecular Microbiology**, v. 37, n. 4, p. 687–695, 2000.
- GRANA, D.; YOUNDERIAN, P.; SUSSKIND, M. M. Mutations that improve the ANT promoter of salmonella phage P22. **Genetics**, 1985.
- GRANT, S. G. N. et al. Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. **Proceedings of the National Academy of Sciences of the United States of America**, 1990.
- GRINDLEY, N. D.; JOYCE, C. M. Genetic and DNA sequence analysis of the kanamycin resistance transposon Tn903. **Proceedings of the National Academy of Sciences**, 1980.
- GROSSMAN, T. H. et al. Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. **Gene**, 1998.
- GRUBER, T. M.; GROSS, C. A. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. **Annual Review of Microbiology**, v. 57, n. 1, p. 441–466, 2003.
- GUALERZI, C. O.; PON, C. L. **Initiation of mRNA translation in bacteria: Structural and dynamic aspects** *Cellular and Molecular Life Sciences*, 2015.
- GUAZZARONI, M. E.; SILVA-ROCHA, R. Expanding the Logic of Bacterial Promoters Using Engineered Overlapping Operators for Global Regulators. **ACS Synthetic Biology**, 2014.
- HANAHAN, D. Studies on transformation of *Escherichia coli* with plasmids. **Journal of Molecular Biology**, v. 166, n. 4, p. 557–580, 1983.
- HARLEY, C. B.; REYNOLDS, R. P. Analysis of *E. coli* promoter sequences. **Nucleic Acids Research**, v. 15, n. 5, p. 2343–2361, 1987.
- HAUGEN, S. P. et al. rRNA Promoter Regulation by Nonoptimal Binding of  $\sigma$  Region 1.2: An Additional Recognition Element for RNA Polymerase. **Cell**, v. 125, n. 6, p. 1069–1082, 2006.
- HAUGEN, S. P. et al. Fine structure of the promoter- $\sigma$  region 1.2 interaction. **Proceedings of the National Academy of Sciences of the United States of America**, 2008.
- HAUGEN, S. P.; ROSS, W.; GOURSE, R. L. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. **Nature Reviews Microbiology**, 2008.
- HAWLEY, D. K.; MCCLURE, W. R. Compilation and analysis of *Escherichia coli* promoter DNA sequences. **Nucleic acids research**, v. 11, n. 8, p. 2237–55, 1983.
- HEISS, S. et al. Evaluation of novel inducible promoter/repressor systems for recombinant protein expression in *Lactobacillus plantarum*. **Microbial Cell Factories**, v. 15, n. 1, 2016.
- HELLMAN, L.; FRIED, M. Electrophoretic Mobility Shift Assay (EMSA) for Detecting Protein- Nucleic Acid Interactions. **Nature protocols**, v. 2, n. 8, p. 1849–1861, 2007.

HENDERSON, K. L. et al. Mechanism of transcription initiation and promoter escape by E. coli RNA polymerase. **Proceedings of the National Academy of Sciences of the United States of America**, 2017.

HENKIN, T. M. Control of transcription termination in prokaryotes. **Annual Review of Genetics**, 1996.

HENKIN, T. M. Transcription termination control in bacteria. **Current Opinion in Microbiology**, 2000.

HEYDUK, E.; HEYDUK, T. DNA template sequence control of bacterial RNA polymerase escape from the promoter. **Nucleic Acids Research**, 2018.

HIGUCHI, R.; KRUMMEL, B.; SAIKI, R. K. A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. **Nucleic acids research**, v. 16, n. 15, p. 7351–67, 1988.

HOCHKOEPLER, A. Expanding the landscape of recombinant protein production in Escherichia coli. **Biotechnology Letters**, 2013.

HOOK-BARNARD, I. G.; HINTON, D. M. Transcription Initiation by Mix and Match Elements: Flexibility for Polymerase Binding to Bacterial Promoters. **Gene Regulation and Systems Biology**, v. 1, p. 117762500700100, 2007.

HOTHERSALL, J. et al. The PAR promoter expression system: Modified lac promoters for controlled recombinant protein production in Escherichia coli. **New Biotechnology**, 2021.

HÖYBYE, C.; CHRISTIANSEN, J. S. Growth hormone replacement in adults -current standards and new perspectives. **Best Practice and Research: Clinical Endocrinology and Metabolism**, 2015.

HUANG, C.-J. J.; LIN, H.; YANG, X. M. Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. **Journal of Industrial Microbiology & Biotechnology**, v. 39, n. 3, p. 383–399, 2012.

IRLA, M. et al. Genome-based genetic tool development for Bacillus methanolicus: Theta- and rolling circle-replicating plasmids for inducible gene expression and application to methanol-based cadaverine production. **Frontiers in Microbiology**, 2016.

ISHIHAMA, A. Functional modulation of Escherichia coli RNA polymerase. **Annual review of microbiology**, v. 54, p. 499–518, 2000.

ISHIHAMA, A. Prokaryotic genome regulation: a revolutionary paradigm. **Proc.Jpn.Acad.Ser.B Phys.Biol.Sci.**, 2012. Disponível em:  
<[https://www.jstage.jst.go.jp/article/pjab/88/9/88\\_PJA8809B-02/\\_pdf](https://www.jstage.jst.go.jp/article/pjab/88/9/88_PJA8809B-02/_pdf)>

JACOB, F.; MONOD, J. Genetic regulatory mechanisms in the synthesis of proteins. **Journal of Molecular Biology**, 1961.

JANA, S.; DEB, J. K. Strategies for efficient production of heterologous proteins in Escherichia coli. **Applied Microbiology and Biotechnology**, 2005.

- JENSEN, P. R.; HAMMER, K. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. **Applied and Environmental Microbiology**, v. 64, n. 1, p. 82–87, 1998.
- JESSE, J. AND BLODGETT, K. (1988). Max efficiency DH5aF'IQ: A new M13 and phagemid host. *Focus* 10, 69.
- JIN, H. et al. Influences on gene expression in vivo by a Shine-Dalgarno sequence. **Molecular Microbiology**, 2006.
- JONASSON, P. et al. Genetic design for facilitated production and recovery of recombinant proteins in *Escherichia coli*. **Biotechnol Appl Biochem**, v. 35, n. Pt 2, p. 91- 105., 2002.
- JUSTICE, S. S. et al. Morphological plasticity as a bacterial survival strategy. **Nature Reviews Microbiology**, 2008.
- KACZANOWSKA, M.; RYDÉN-AULIN, M. Ribosome Biogenesis and the Translation Process in *Escherichia coli* . **Microbiology and Molecular Biology Reviews**, 2007.
- KANG, M. K. et al. Synthetic biology platform of CoryneBrick vectors for gene expression in *Corynebacterium glutamicum* and its application to xylose utilization. **Applied Microbiology and Biotechnology**, 2014.
- KINGSFORD, C. L.; AYANBULE, K.; SALZBERG, S. L. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. **Genome Biology**, 2007.
- KIRYU, H.; OSHIMA, T.; ASAI, K. Extracting relations between promoter sequences and their strengths from microarray data. **Bioinformatics**, 2005.
- KRÜGER, A. et al. A host - vector toolbox for improved secretory protein overproduction in *Bacillus subtilis*. **Applied Microbiology and Biotechnology**, v. 9, p. 5137–5151, 2022.
- LAWSON, C. L. et al. Catabolite activator protein: DNA binding and transcription activation. **Current Opinion in Structural Biology**, 2004.
- LEE, D. J.; MINCHIN, S. D.; BUSBY, S. J. W. Activating Transcription in Bacteria. **Annual Review of Microbiology**, v. 66, n. 1, p. 125–152, 2012.
- LEE, J.; BORUKHOV, S. Bacterial RNA Polymerase-DNA Interaction—The Driving Force of Gene Expression and the Target for Drug Action. **Frontiers in Molecular Biosciences**, v. 3, 2016.
- LEWIN, A. et al. Viral promoters can initiate expression of toxin genes introduced into *Escherichia coli*. **BMC Biotechnology**, v. 5, 2005.
- LEWIS, D. E. A.; GUSSIN, G. N.; ADHYA, S. New Insights into the Phage Genetic Switch: Effects of Bacteriophage Lambda Operator Mutations on DNA Looping and Regulation of PR, PL, and PRM. **Journal of Molecular Biology**, v. 428, n. 22, p. 4438–4456, 2016.
- LEWIS, M. et al. Crystal structure of the lactose operon repressor and its complexes with

DNA and inducer. **Science**, 1996.

LEWIS, M. Allosterity and the lac operon. **Journal of Molecular Biology**, 2013.

LI, R. et al. Effects of cooperation between translating ribosome and RNA polymerase on termination efficiency of the Rho-independent terminator. **Nucleic Acids Research**, 2015.

LI, Y. et al. Expression of common carp growth hormone in the yeast *Pichia pastoris* and growth stimulation of juvenile tilapia (*Oreochromis niloticus*). **Aquaculture**, 2003.

LOZANO TEROL, G. et al. Impact of the Expression System on Recombinant Protein Production in *Escherichia coli* BL21. **Frontiers in Microbiology**, 2021.

MA, J.; CAMPBELL, A.; KARLIN, S. Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. **Journal of Bacteriology**, 2002.

MAKRIDES, S. C. Strategies for achieving high-level expression of genes in *Escherichia coli*. **Microbiological reviews**, v. 60, n. 3, p. 512–538, 1996.

MANIATIS, T.; FRITSCH, E. F.; SAMBROOK, J. **Molecular Cloning: A Laboratory Manual**. Cold Spring Harbor laboratory press. [s.l.: s.n.].

MARSCHALL, L.; SAGMEISTER, P.; HERWIG, C. Tunable recombinant protein expression in *E. coli*: enabler for continuous processing? **Applied Microbiology and Biotechnology**, 2016.

MAZUMDER, A.; KAPANIDIS, A. N. Recent Advances in Understanding  $\sigma$ 70-Dependent Transcription Initiation Mechanisms. **Journal of Molecular Biology**, 2019.

MCCARTY, N. S.; LEDESMA-AMARO, R. Synthetic Biology Tools to Engineer Microbial Communities for Biotechnology. **Trends in Biotechnology**, 2019.

MELANÇON, P. et al. The Anti-Shine-Dalgarno Region in *Escherichia coli* 16S Ribosomal RNA Is Not Essential for the Correct Selection of Translational Starts. **Biochemistry**, 1990.

MERRICK, M. J. In a class of its own--the RNA polymerase sigma factor sigma 54 (sigma N). **Molecular microbiology**, v. 10, n. 5, p. 903–909, 1993.

MITCHELL, J. E. et al. Identification and analysis of “extended -10” promoters in *Escherichia coli*. **Nucleic Acids Research**, v. 31, n. 16, p. 4689–4695, 2003.

MOREIRA, S. et al. Uma breve história de promotores bacterianos : da sequência e mecanismo. **Amazonia, Scientia**, p. 17–37, 2018.

MÜHLMANN, M. et al. Optimizing recombinant protein expression via automated induction profiling in microtiter plates at different temperatures. **Microbial Cell Factories**, 2017.

MURAKAMI, K. S. et al. Structural basis of transcription initiation: An RNA polymerase holoenzyme-DNA complex. **Science**, v. 296, n. 5571, p. 1285–1290, 2002.

MURAKAMI, K. S.; DARST, S. A. Bacterial RNA polymerases: The whole story. **Current Opinion in Structural Biology**, 2003.

MUTALIK, V. K. et al. Precise and reliable gene expression via standard transcription and translation initiation elements. **Nature Methods**, 2013.

OEHLER, S. et al. The three operators of the lac operon cooperate in repression. **The EMBO journal**, v. 9, n. 4, p. 973–979, 1990.

OEHLER, S.; ALBERTI, S.; MÜLLER-HILL, B. Induction of the lac promoter in the absence of DNA loops and the stoichiometry of induction. **Nucleic Acids Research**, 2006.

OLINS, P. O. et al. The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in Escherichia coli. **Gene**, v. 73, n. 1, p. 227–235, 1988.

OLINS, P. O.; RANGWALA, S. H. A novel sequence element derived from bacteriophage T7 mRNA acts as an enhancer of translation of the lacZ gene in Escherichia coli. **Journal of Biological Chemistry**, 1989.

OLIVEIRA, C.; DOMINGUES, L. Guidelines to reach high-quality purified recombinant proteins. **Applied Microbiology and Biotechnology**, 2018.

OMOTAJO, D. et al. Distribution and diversity of ribosome binding sites in prokaryotic genomes. **BMC Genomics**, 2015.

OROSZ, A.; BOROS, I.; VENETIANER, P. Analysis of the complex transcription termination region of the Escherichia coli rrnB gene. **European journal of biochemistry / FEBS**, 1991.

OVERTON, T. W. Recombinant protein production in bacterial hosts. **Drug Discovery Today**, 2014.

PACHECO, B. et al. A screening strategy for heterologous protein expression in Escherichia coli with the highest return of investment. **Protein Expression and Purification**, 2012.

PASOTTI, L.; ZUCCA, S. Advances and computational tools towards predictable design in biological engineering. **Computational and Mathematical Methods in Medicine**, 2014.

PEMBERTON, I. K. et al. The G + C-rich discriminator region of the tyrT promoter antagonises the formation of stable preinitiation complexes. **Journal of Molecular Biology**, v. 299, n. 4, p. 859–864, 2000.

PENUMETCHA, P. et al. Improving the Lac System for Synthetic Biology. **BIOS**, 2010.

PETI, W.; PAGE, R. Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost. **Protein Expression and Purification**, 2007.

PHILLIPS, K. N. et al. Diversity in lac operon regulation among diverse escherichia coli isolates depends on the broader genetic background but is not explained by genetic relatedness. **mBio**, 2019.

PINTO., K. R. EXPRESSÃO HETERÓLOGA DA ENZIMA ENTEROQUINASE EM *Escherichia coli* *Escherichia coli*. [s.l.] UFAM, 2017.

PONNAMBALAM, S. et al. Transcription initiation at the *Escherichia coli* galactose operon promoters in the absence of the normal -35 region sequences. **Journal of Biological Chemistry**, 1986.

PRIBNOW, D. Bacteriophage T7 early promoters: Nucleotide sequences of two RNA polymerase binding sites. **Journal of Molecular Biology**, v. 99, n. 3, 1975.

RANGEL-CHAVEZ, C.; GALAN-VASQUEZ, E.; MARTINEZ-ANTONIO, A. Consensus architecture of promoters and transcription units in *Escherichia coli*: design principles for synthetic biology. **Mol. BioSyst.**, v. 13, n. 4, p. 665–676, 2017.

RAO, L. et al. Factor independent activation of *rrnB* p1. An “Extended” promoter with an upstream element that dramatically increases promoter strength. **Journal of Molecular Biology**, v. 235, n. 5, p. 1421–1435, 1994.

RAVISHANKAR, S. et al. An IPTG inducible conditional expression system for mycobacteria. **PLoS ONE**, v. 10, n. 8, 2015.

RETNONINGRUM, D. S. et al. Construction and Characterization of a Medium Copy Number Expression Vector Carrying Auto-Inducible *dps* Promoter to Overproduce a Bacterial Superoxide Dismutase in *Escherichia coli*. **Molecular Biotechnology**, v. 61, n. 4, p. 231–240, 2019.

RHODIUS, V. A.; MUTALIK, V. K.; GROSS, C. A. Predicting the strength of UP-elements and full-length *E. coli*  $\sigma$  e promoters. **Nucleic Acids Research**, v. 40, n. 7, p. 2907–2924, 2012.

RINGQUIST, S. et al. Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. **Molecular Microbiology**, 1992.

ROBERTS, J. W.; SHANKAR, S.; FILTER, J. J. RNA Polymerase Elongation Factors. **Annual Review of Microbiology**, v. 62, n. 1, p. 211–233, 2008.

RODRIGUE, S. et al. The  $\sigma$  factors of *Mycobacterium tuberculosis*. **FEMS Microbiology Reviews**, v. 30, n. 6, p. 926–941, 2006.

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

ROSANO, G. L.; MORALES, E. S.; CECCARELLI, E. A. New tools for recombinant protein production in *Escherichia coli*: A 5-year update. **Protein Science**, 2019.

ROSINSKI-CHUPIN, I. et al. Single nucleotide resolution RNA-seq uncovers new regulatory mechanisms in the opportunistic pathogen *Streptococcus agalactiae*. **BMC Genomics**, v. 16, n. 1, 2015.

ROSS, W. et al. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. **Science**, v. 262, n. 5138, p. 1407–1413, 1993.

ROSS, W.; GOURSE, R. L. Sequence-independent upstream DNA- CTD interactions strongly stimulate Escherichia coli RNA polymerase-lacUV5 promoter association. **Proceedings of the National Academy of Sciences**, v. 102, n. 2, p. 291–296, 2005.

RUDGE, T. J. et al. Characterization of Intrinsic Properties of Promoters. **ACS Synthetic Biology**, 2016.

RUFF, E. F.; THOMAS RECORD, M.; ARTSIMOVITCH, I. Initial events in bacterial transcription initiation. **Biomolecules**, 2015.

SACHDEVA, P. et al. The sigma factors of Mycobacterium tuberculosis: Regulation of the regulators. **FEBS Journal**, 2010.

SAECKER, R. M.; RECORD, M. T.; DEHASETH, P. L. Mechanism of bacterial transcription initiation: RNA polymerase - Promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. **Journal of Molecular Biology**, 2011.

SAHDEV, S.; KHATTAR, S. K.; SAINI, K. S. Production of active eukaryotic proteins through bacterial expression systems: A review of the existing biotechnology strategies. **Molecular and Cellular Biochemistry**, 2008.

SAIDA, F. et al. Expression of Highly Toxic Genes in E. coli: Special Strategies and Genetic Tools. **Current Protein and Peptide Science**, v. 7, n. 1, p. 47–56, 2006.

SAÏDA, F. Overview on the Expression of Toxic Gene Products in Escherichia coli . **Current Protocols in Protein Science**, 2007.

SALIS, H. M. The ribosome binding site calculator. In: **Methods in Enzymology**. [s.l: s.n.].

SANCHEZ-GARCIA, L. et al. Recombinant pharmaceuticals from microbial cells: A 2015 update. **Microbial Cell Factories**, 2016.

SANTOS, C. N. S.; STEPHANOPOULOS, G. Combinatorial engineering of microbes for optimizing cellular phenotype. **Current Opinion in Chemical Biology**, 2008.

SCHNEIDER, T. D.; STEPHENS, R. M. Sequence logos: A new way to display consensus sequences. **Nucleic Acids Research**, 1990.

SCHULLER, A. et al. Escherichia coli  $\sigma$ 70 promoters allow expression rate control at the cellular level in genome-integrated expression systems. **Microbial Cell Factories**, 2020.

SCHUMANN, W.; FERREIRA, L. C. S. Production of recombinant proteins in Escherichia coli. **Genetics and Molecular Biology**, 2004.

SELAS CASTIÑEIRAS, T. et al. E. coli strain engineering for the production of advanced biopharmaceutical products. **FEMS Microbiology Letters**, 2018.

SHIMADA, T. et al. The whole set of constitutive promoters recognized by RNA polymerase RpoD holoenzyme of Escherichia coli. **PLoS ONE**, v. 9, n. 3, 2014.

SHINE, J.; DALGARNO, L. The 3' terminal sequence of Escherichia coli 16S ribosomal

RNA: complementarity to nonsense triplets and ribosome binding sites. **Proceedings of the National Academy of Sciences of the United States of America**, 1974.

SHINE, J.; DALGARNO, L. Terminal-Sequence Analysis of Bacterial Ribosomal RNA. **European journal of biochemistry**, 1975.

SHOJAOSADATI, S. A. et al. Recent Advances in High Cell Density Cultivation for Production of Recombinant Protein. **Iran J Biotech**, v. 6, n. 2, p. 63–84, 2008.

SIEBENLIST, U.; GILBERT, W. Contacts between Escherichia coli RNA polymerase and an early promoter of phage T7. **Proceedings of the National Academy of Sciences of the United States of America**, 1980.

SINGH, A. et al. Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. **Microbial Cell Factories**, 2015.

SIVASHANMUGAM, A. et al. Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. **Protein Science**, v. 18, n. 5, p. 936–948, 2009.

SMILLIE, C. et al. Mobility of Plasmids. **Microbiology and Molecular Biology Reviews**, 2010.

SØRENSEN, H. P. et al. Advanced genetic strategies for recombinant protein expression in Escherichia coli. **Journal of Biotechnology**, v. 115, n. 2, p. 113–128, 2005.

STUDIER, F. W. Protein production by auto-induction in high density shaking cultures. **Protein expression and purification**, 2005.

THIEL, K. et al. Translation efficiency of heterologous proteins is significantly affected by the genetic context of RBS sequences in engineered cyanobacterium Synechocystis sp. PCC 6803. **Microbial Cell Factories**, 2018.

TIERRAFRÍA, V. H. et al. RegulonDB 11.0: Comprehensive high-throughput datasets on transcriptional regulation in Escherichia coli K-12. **Microbial Genomics**, v. 8, n. 5, p. 1–17, 2022.

TRIPATHI, L.; ZHANG, Y.; LIN, Z. Bacterial Sigma Factors as Targets for Engineered or Synthetic Transcriptional Control. **Frontiers in Bioengineering and Biotechnology**, v. 2, 2014.

U K LAEMMLI. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. **nature**, 1970.

ULLMANN, A. Escherichia coli Lactose Operon. **Encyclopedia of Life Sciences**, 2009.

VAN HIJUM, S. A. F. T.; MEDEMA, M. H.; KUIPERS, O. P. Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. **Microbiology and molecular biology reviews : MMBR**, v. 73, n. 3, p. 481–509, Table of Contents, 2009.

VAN HOEK, M.; HOGEWEG, P. The effect of stochasticity on the Lac operon: An evolutionary perspective. **PLoS Computational Biology**, 2007.

VENUGOPAL, T. et al. Molecular cloning of growth hormone-encoding cDNA of an Indian major carp, *Labeo rohita*, and its expression in *Escherichia coli* and zebrafish. **General and Comparative Endocrinology**, 2002.

VIMBERG, V. et al. Translation initiation region sequence preferences in *Escherichia coli*. **BMC Molecular Biology**, 2007.

WEICKERT, M. J.; ADHYA, S. The galactose regulon of *Escherichia coli*. **Molecular Microbiology**, 1993.

WIGNESHWERARAJ, S. et al. Modus operandi of the bacterial RNA polymerase containing the sigma54 promoter-specificity factor. **Molecular microbiology**, v. 68, n. 3, p. 538–46, 2008.

YABUTA, M.; ONAI-MIURA, S.; OHSUYE, K. Thermo-inducible expression of a recombinant fusion protein by *Escherichia coli* lac repressor mutants. **Journal of Biotechnology**, 1995.

YANG, J. et al. Screening through the PLICable promoter toolbox enhances protein production in *Escherichia coli*. **Biotechnology Journal**, 2016.

YANISCH-PERRON, C.; VIEIRA, J.; MESSING, J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. **Gene**, v. 33, n. 1, p. 103–119, 1985.

YOUSEFIAN, M.; SHIRZAD, E. The review of the effect of growth hormone on immune system, metabolism and osmoregulation of fish. **Australian Journal of Basic and Applied Sciences**, 2011.

YU, T. C. et al. Multiplexed characterization of rationally designed promoter architectures deconstructs combinatorial logic for IPTG-inducible systems. **Nature Communications**, 2021.

## 8 APÊNDICES

Pré-inóculo - Selecionar uma colônia isolada em placa de Petri contendo antibiótico seletivo com palito limpo e autoclavado. Inocular a colônia em 5 mL de meio LB líquido em tubo de ensaio com rosca e incubar a 37°C sob agitação durante a noite.

Sequência palindrômica - Uma sequência palindrômica é uma sequência de ácidos nucleicos em uma molécula de DNA ou RNA de fita dupla que apresenta simetria dupla, isto é, em que a leitura em certa direção (por exemplo, 5' → 3') em uma fita é idêntica à sequência na mesma direção (por exemplo, 5' → 3') na fita complementar.

## 9 ANEXOS

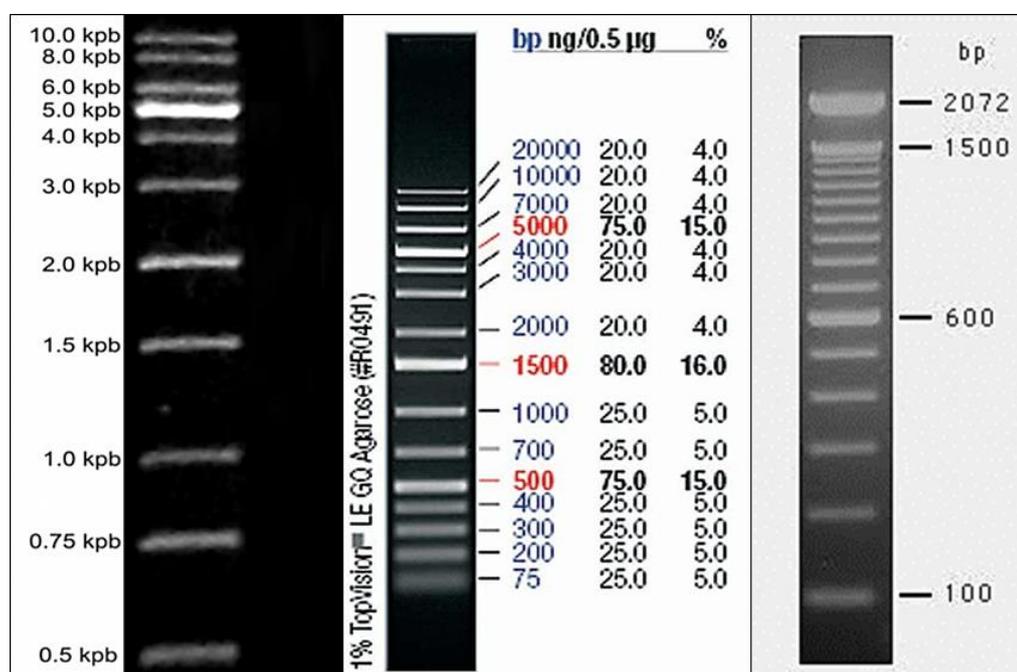
### ANEXO 1

<b>SOLUÇÕES, TAMPÕES E REAGENTES</b>	<b>COMPOSIÇÃO</b>
<b>Tampão TE</b>	Tris-HCl - 20 mM; EDTA – 0,1 mM; pH 7,5
<b>Tampão de corrida TEB 10x</b>	Tris base - 0,89 M; Ácido Bórico - 0,89M; EDTA - 0,08M; pH 8.0- 8.4
<b>Tampão de amostra TEB 5x</b>	Glicerol 30% (p/v); azul de bromofenol - 0,01% (p/v)
<b>Tampão de lavagem</b>	Tris-HCl (pH 7,5) 20 mM; EDTA 2 mM; NaCl 200 mM; Etanol absoluto 96 mL
<b>Tampão A</b>	Tris-HCl pH 8 20mM; NaCl 500mM
<b>Tampão TN</b>	Tris-HCl 10mM e NaCl 0,15M pH 8,0
<b>Glicose 100%</b>	Solução estoque: 100% Solução de trabalho: 1%
<b>Ampicilina</b>	Solução estoque: 1g/mL Solução de trabalho: 100 µg/mL
<b>Glicogênio</b>	Solução estoque: 2 µg/µL
<b>Kanamicina</b>	Solucao estoque: 34 mg/mL Solução de trabalho: 50 µg/mL
<b>IPTG</b>	Solucao estoque: 1M Solução de trabalho: 1mM
<b>Brometo de etídio</b>	Solução estoque: 10 mg/mL Solução trabalho: 1 µg/mL
<b>Etanol</b>	70% (V/V)
<b>MgCl<sub>2</sub></b>	25 mM
<b>dNTP</b>	2,5 mM
<b>Triton</b>	Solução estoque: 100X Solução trabalho: 100X
<b>PMSF (fluoreto de fenilmetilsulfonil)</b>	Solução estoque: 100mM Solução trabalho: 100mM
<b>Marcador Ladder DNA GeneRuler™ 1 kb (Fermentas);</b>	250-10,000 pb; 75-20000 pb; 100-10000 pb.

<b>Plus; Mix</b>	
<b>Soluções do Kit GE HealthCare</b>	<p><b>Solução I:</b> Tris-HCl (pH 8) 25 mM EDTA 10 mM Glicose 10 mM</p> <p><b>Solução II:</b> NaOH 0,2 M SDS 1% (p/v)</p> <p><b>Solução III:</b> KAC 3M Ácido acético 2M pH 4,8-5,0</p>

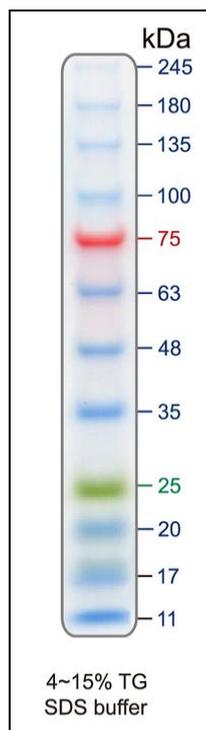
## ANEXO 2

### Marcadores Moleculares de DNA



O ladder 1Kb da Quatro G (500 – 10000 pb) possui um fragmento de 750 pb entre a banda de 500 e 1000 pb. Foi utilizado também o Marcador GeneRuler™ 100 pb Mix (100 – 2072 pb).

### Marcador Molecular de proteínas



### ANEXO 3

#### Preparo de células eletrocompetentes e eletroporação

Foi realizado um pré-inóculo de células competentes *E. coli* com o antibiótico canamicina ou ampicilina e incubado em agitador de bancada (tipo shaker) de 160 rpm a 37°C durante a noite. Dessa cultura, 1 mL foi inoculado em 200 mL de meio LB isento de antibiótico e após foi incubada sob agitação de 150 rpm a 37°C.

Foram retiradas alíquotas de 1 mL, em intervalos de 1 hora até atingir O.D.600 = 0,6. Foi utilizado um branco contendo o mesmo meio para cultivo na ausência de crescimento bacteriano. Assim que a O.D. foi alcançada, foi pego tubos falcons pré-resfriados no gelo e distribuído 50 mL de cultura em cada um.

As células foram centrifugadas a 4.000 rpm por 10 minutos a 4°C. O sobrenadante foi descartado de imediato e ressuspendido gentilmente com 10 mL de glicerol 10% estéril gelado até soltar o pellet. As células foram novamente centrifugadas a 4.000 rpm por 10 minutos a 4°C e o sobrenadante descartado. Esta operação foi repetida seguidamente por 3 vezes e as células foram reunidas em um único tubo falcon no final.

Logo após as lavagens e centrifugações, o pellet foi ressuspendido gentilmente em 1 mL de glicerol 10% gelado e estéril. As células eletrocompetentes foram separadas em alíquotas de 50 ou 100  $\mu$ L distribuídas em microtubos, congeladas imediatamente em nitrogênio líquido e estocadas a  $-80^{\circ}\text{C}$ .

Uma alíquota do sistema de ligação foi misturado em 50 ou 100  $\mu$ L de células *E. coli* eletrocompetentes que estavam a  $-80^{\circ}\text{C}$  e transferidos para uma cuveta de eletroporação (0,2 cm). Foi aplicado um pulso elétrico de 1900 mV no equipamento eletroporador (“*Eletroporator 2510 Eppendorff*”), e imediatamente após dois pulsos, adicionou-se 500  $\mu$ L de meio LB na cuveta revigorando as células transformadas. As amostras foram incubadas a  $37^{\circ}\text{C}$  por 1 hora com agitação (160 rpm) para estabilização e em seguida, alíquotas do sistema (vetor + fragmento + célula) foram plaqueados em placas de Petri com meio LB ágar sólido contendo o antibiótico seletivo canamicina ou ampicilina. Ao final do plaqueamento as placas foram incubadas invertidas a  $37^{\circ}\text{C}$  durante a noite.

#### **ANEXO 4**

Preparo de células quimiocompetentes pelo método de Cloreto de cálcio e choque térmico

Uma colônia isolada foi retirada com palito estéril potiadado da placa de Petri contendo antibiótico seletivo canamicina ou ampicilina e inoculada em pré-inóculo. O pré inóculo foi incubado a  $37^{\circ}\text{C}$  sob agitação durante a noite.

No outro dia foi inoculado 500  $\mu$ L do pré-inóculo de bactéria em 50 mL de meio LB e incubado a  $37^{\circ}\text{C}$  sob agitação até a cultura atingir uma densidade óptica a 600 nm de 0,2 a 0,4. Foi utilizado um branco contendo o mesmo meio para cultivo na ausência de crescimento bacteriano. Após a aferição da absorbância foi transferido 40 mL da cultura bacteriana para um tubo Falcon (de 50 mL) e centrifugado a 3.000 g a  $4^{\circ}\text{C}$  por 15 minutos. Após a centrifugação o sobrenadante foi descartado.

O pellet (massa celular) foi ressuspendido em 20 mL de solução de Tris-Ca gelada ( $4^{\circ}\text{C}$ ) através de suave agitação e posteriormente centrifugada a a 2.000 g a  $4^{\circ}\text{C}$  por 10 minutos. Após a centrifugação o sobrenadante foi descartado.

O pellet foi novamente ressuspensionado cuidadosamente em 1 mL de solução Tris-Ca gelada (4°C) e após isso as células foram mantidas no gelo por pelo menos 45 minutos.

Após o tempo mínimo de 45 minutos, a partir do que, as células passam a ser consideradas “competentes”, podendo ser usadas imediatamente ou durante 24 horas depois, foi transferido 100 µL de células competentes do tubo falcon para um microtubo de 1,5 mL com 10 a 50 ng de DNA plasmidial ou 5-10 µL do vetor. O sistema (células + vetor) foi colocado em banho de água-gelo por pelo menos 15-30 minutos.

Em seguida, o sistema foi submetido a um choque térmico incubando 37°C por 5 min utilizando banho-maria e foi adicionado 1 mL de meio LB aos tubos submetidos a choque térmico. O sistema foi incubado por 60 min. a 37°C, e após plaqueado em alíquotas que variavam de 50 até 300 µL das suspensões de células transformadas em placas contendo meio LB-sólido com o antibiótico adequado. Ao final as placas eram incubadas invertidas a 37°C durante a noite (12-15h).

#### Controles da transformação:

- Negativo – Um sistema sem DNA – plaquear 50 ou 100 µL da célula hospedeira em placa de Petri com LB sem antibiótico.
- Positivo – fazer um sistema de transformação com 25 ng de um plasmídeo conhecido – plaquear 50 ou 100 µL em placas com e sem antibiótico.

#### Solução:

**Tris-Ca: Tris-HCl 10mM pH 7,5 e CaCl<sub>2</sub> 80 mM** – esterilizado por fervura em forno de micro-ondas.

## ANEXO 5

### EXTRAÇÃO PLASMIDIAL POR LISE ALCALINA

*illustra™ plasmidPrep Mini Spin kit*

Inocular uma colônia isolada no dia anterior e transferir para um pré-inóculo contendo antibiótico seletivo apropriado e incubar a 37°C por 16 horas com agitação (150 rpm). Transferir 1.5 mL da cultura para um microtubo de 1,5 mL e centrifugar a 12.000 g por 1 minuto.

Ressuspender o pellet em 175  $\mu$ L da solução **Lysis buffer type 7**. Adicionar 175  $\mu$ L da solução **Lysis buffer type 8** e gentilmente inverter o tubo algumas vezes. Adicionar 175  $\mu$ L da solução **Lysis buffer type 9** e gentilmente inverter o tubo algumas vezes até precipitar bem. Centrifugar a 16.000 g por 4 minutos.

Após, coletar o sobrenadante e transferir para um tubo novo contendo uma coluna dentro e centrifugar por 30 segundos a 16.000 g. Descartar o centrifugado do tubo e adicionar na coluna 400  $\mu$ L da solução **Wash buffer type 1**, centrifugar por 1 minuto a 16.000 g.

Após a centrifugação, descartar o centrifugado e o tubo, e transferir a coluna para um novo microtubo 1,5 mL novo e estéril. Adicionar 50-100  $\mu$ L da solução **Elution buffer type 4** no centro da coluna e centrifugar por 30 segundos a 16.000 g. Repetir este procedimento mais 2 vezes sempre usando o centrifugado e passando na coluna novamente. Analisar o DNA plasmidial extraído em gel de agarose ou armazenar em  $-20^{\circ}\text{C}$ .

*Cellco Fast-n-easy Plasmid Mini-Prep kit (Column based isolation of plasmid DNA)*

## **1 Recuperação das células e lise**

Recupere as células a partir de uma cultura celular (1-3 mL) por centrifugação ressuspender o pellet bacteriano em 300  $\mu$ L de **Lysis Buffer** por pipetagem ou vortex por 1 minuto.

## **2 Neutralização**

Adicione 300  $\mu$ L do **Neutralization Buffer** (contendo RNase) na amostra e misture gentilmente por inversão do tubo 4-6 vezes (não usar vortex). Centrifugue a 10.000 g por 5 min. A cor da mistura de ligação deverá mudar para amarelo brilhante, indicando um pH de 7.5 (requerido para a ligação ótimo do DNA na coluna). A coloração laranja ou violeta indica que o pH > 7.5 e implicará numa adsorção do DNA ineficiente. Neste caso, é recomendado o ajuste do pH pela adição de pequeno volume de acetato de sódio 3M (pH 5.0) antes de prosseguir com as demais etapas.

## **3 Ativação da coluna**

Coloque a coluna (*spin column*) em um tubo coletor de 2 mL e adicionar 100  $\mu$ L de **Activation Buffer** na coluna. Centrifugar a 10.000 g por 30 segundos.

#### 4 Carregamento da coluna

Aplique o sobrenadante da etapa 2 na coluna ativada por decantação ou pipetagem. Centrifugar a 10.000 g por 30 segundos. Descartar o que passou pela coluna.

#### 5 Lavagem da coluna

Adicionar 500  $\mu$ L de **Washing Buffer** na coluna. Centrifugar a 10.000 g por 30 s. Descartar o que passou pela coluna.

Lavagem Secundária (opcional): Recomendada apenas para fragmentos de DNA menores que 200 bp, se DNA altamente puro (para sequenciamento, transfecção, etc.) for necessário.

Adicione 700  $\mu$ L de **Washing Buffer** na coluna. Centrifugar a 10.000 g por 30 s e descartar o que passou pela coluna. Centrifugar a 10.000 g por 2 minutos para remover qualquer resíduo de **Washing Buffer**.

#### 6 Eluição

Colocar a coluna em um microtubo de 1.5 mL novo e adicionar 30-50  $\mu$ L de **Elution Buffer** ou água no centro da coluna. Incubar por 1 minuto à temperatura ambiente e após centrifugar a 10.000 g por 1 minuto para eluir o DNA.

### ANEXO 6

Purificação de DNA a partir de géis de agarose TBE e TAE (*Illustra GFX PCR DNA and Gel Band Purification Kit*)

#### 1. Captura da Amostra

Pesar um microtubo de 1,5 mL livre de contaminações e registrar o seu peso. Usar um bisturi limpo, comprimento de onda longo (365 nm), luz ultravioleta e tempo de exposição mínimo, cortar a banda de agarose contendo a amostra de interesse. Colocar a

banda em gel de agarose no microtubo de 1,5 mL e registrar seu peso. Repetir este procedimento se houver mais amostras.

Pesar o microtubo com a banda de agarose e calcula o peso da banda de agarose. Adicionar 10  $\mu$ L de **Capture buffer type 3** para cada 10 mg de agarose. Misturar por inversão e incubar a 60 °C durante 15-30 minutos até a agarose estar totalmente dissolvida.

## **2. Ligação da Amostra na coluna**

Para cada purificação que está a ser executada, colocar uma coluna GFX MicroSpin em um microtubo. Centrifugar a amostra rapidamente. Transferir até 800  $\mu$ L de amostra no tubo de coluna, incubar à temperatura ambiente durante 1 minuto e centrifugar a amostra no tubo de coleta com a coluna a 16.000 g por 30 segundos. Eliminar o centrifugado do microtubo e colocar a coluna GFX MicroSpin de volta para dentro do microtubo.

## **3. Lavagem das amostras**

Adicionar 500  $\mu$ L de **Wash buffer type 1** dentro da coluna GFX MicroSpin e centrifugar a 16.000 g por 30 segundos. Descartar o centrifugado e o tubo de coleta e transferir a coluna GFX MicroSpin a um microtubo de 1,5 mL livre de contaminações.

## **4. Eluição**

Adicionar 10-50  $\mu$ L de **Elution buffer type 4 or type 6** para o centro da membrana na coluna GFX MicroSpin e incubar à temperatura ambiente durante 1 minuto. Centrifugar a amostra 16.000 g por 16 000  $\times$  g durante 1 minuto para recuperar o DNA purificado. Guardar o DNA purificado a -20 ° C.

## **ANEXO 7**

### **Protocolo de purificação de DNA por precipitação**

Adicionar no microtubo da ligação 2  $\mu$ L de glicogênio (2  $\mu$ g/ $\mu$ L), 0.1V de NaCl 3M) e 2.5V de Álcool absoluto. Deixar o sistema durante a noite por -20°C. Após transpassado a noite, centrifugar por 15 minutos a 4°C por 12000rpm ou 16000g. Com auxílio de uma pipeta, retirar o sobrenadante com cuidado. Adicionar 1 mL de álcool 70% gelado pelas bordas, em movimento circulares, gentilmente, para lavar o eppendorf. Incubar por 15

minutos no gelo. Centrifugar por 10 minutos a 4°C. Retirar o sobrenadante gentilmente. Tentar deixar o tubo o mais sem líquido possível e deixar o tubo aberto secando no fluxo. Redissolver o DNA purificado e concentrado em 2-10 µL H<sub>2</sub>O ultrapura tipo 1.



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

(21) BR 102019000470-3 A2



(22) Data do Depósito: 10/01/2019

(43) Data da Publicação Nacional: 28/07/2020

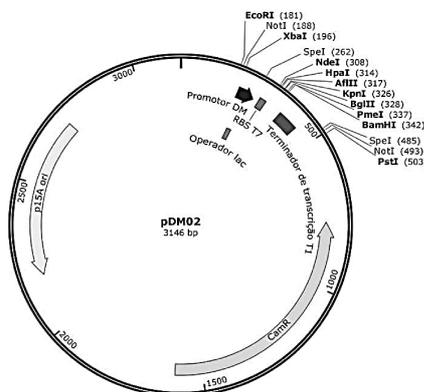
(54) **Título:** UM PROMOTOR SINTÉTICO PARA ALTA EXPRESSÃO DE GENES HETERÓLOGOS EM ESCHERICHIA COLI EM PERFORMANCE BEM REGULADA

(51) **Int. Cl.:** C12N 15/11; C12N 15/70; C12R 1/19.

(71) **Depositante(es):** FUNDAÇÃO UNIVERSIDADE DO AMAZONAS.

(72) **Inventor(es):** DIEGO DA SILVA MOREIRA; SPARTACO ASTOLFI FILHO; MARIA CLARA TAVARES ASTOLFI; KEROLLEN RUNA PINTO.

(57) **Resumo:** A presente invenção refere-se à engenharia do novo promotor sintético DM, funcional no organismo modelo *Escherichia coli*, com performance de forte expressão e alto nível de regulação, construído com inspiração no sistema do promotor *lac*. Visando aprimorar sua regulação para aplicações industriais, o design foi realizado a partir da sobreposição da sequência operadora *lac* à região -35 de reconhecimento da subunidade  $\sigma^{70}$ , precisamente impedindo o início da transcrição pela RNA polímerase holoenzima. Esta estratégia possibilitou alto nível de expressão e regulação gênica, demonstrada pela análise quantitativa da síntese da Green Fluorescent Protein (GFP) e da proteína híbrida dationredoxina humana fusionada ao domínio catalítico da enteroquinase bovina sob controle do promotor sintético DM, nos vetores pDM-GFP e pDM-TEK. O presente promotor, clonado no versátil vetor de expressão pDM02, apresenta vasta aplicação biotecnológica, expandindo as ferramentas para construção de inovadores circuitos genéticos. Principalmente, a presente invenção abre caminho para expressão de genes de proteínas recombinantes de interesse industrial, inclusive as tóxicas, cuja elevada produção e minuciosa regulação tem sido um desafio.



## UM PROMOTOR SINTÉTICO PARA ALTA EXPRESSÃO DE GENES HETERÓLOGOSEM *Escherichia coli* EM PERFORMANCE BEM REGULADA

### **Campo da Invenção, Finalidade e Aplicação**

[001]Dentre as ferramentas utilizadas pela engenharia genética, os vetores de expressão estão entre as mais importantes, pois programam a expressão da síntese de proteínas heterólogas nas células hospedeiras. Desta forma, são utilizados para a expressão em altos níveis de genes de interesse industrial, como os codificadores de: insulina, hormônio de crescimento, anticorpos monoclonais, enzimas, dentre outras. Esses vetores também são utilizados para realizar edição genômica e engenharia metabólica em células para os mais variados propósitos, inclusive para finalidades de terapia genética e para produção industrial de substâncias de baixo peso molecular como substâncias antimaláricas, antineoplásicas e antibióticas (COHEN, 1973; DEMAÏN & VAISHNAV, 2009; KEASLING, 2012; VITORINO & BESSA, 2017).

[002] Um dos principais componentes de um vetor de expressão é o “cassete de expressão”, composto por um conjunto seletivo de sequências de DNA que garantem a expressão de genes codificantes de proteínas de interesse industrial na célula hospedeira. Um “cassete de expressão” contém principalmente: 1) um promotor de transcrição capaz de recrutar a RNA polimerase holoenzima ( $\alpha 2$ ,  $\beta$ ,  $\beta'$ ,  $\omega$  e  $\sigma$ ) e iniciar a síntese do RNA mensageiro (mRNA); 2) uma região de múltiplos sítios de enzimas de restrição (“*polylinker*”) para permitir a introdução das informações genéticas a serem expressas; 3) e um terminador de transcrição – para interromper a síntese do mRNA pela RNA polimerase (HALL et al, 1995; BROWNING & BUSBY, 2004; SINGHA et al, 2017).

[003] Um dos componentes mais importantes de um vetor de expressão é o promotor, uma sequência de DNA que é reconhecida pela subunidade sigma da RNA polimerase para dar início à transcrição, ou seja, a biossíntese de RNA. Os promotores podem ser passíveis de regulação (regulados) ou não (constitutivos) (GOLDSTEIN & DOI, 1995; HAUGEN et al., 2008).

[004] Visando garantir a alta produção de proteínas em células hospedeiras durante o processo fermentativo em escala industrial, há a necessidade de que a expressão gênica seja precisamente regulada. Desta forma, há a capacidade de ativar/induzir a síntese de

genes somente após o crescimento celular atingir o nível adequado. A expressão em altos níveis no início da fase de crescimento das células nos processos fermentativos pode acarretar sobrecarga do maquinário celular e a morte das mesmas, inviabilizando o processo e consequentemente comprometendo a produção da proteína recombinante de interesse industrial. Para a expressão de proteínas heterólogas tóxicas em níveis médios e altos no interior das células hospedeiras é imprescindível que os seus genes sejam bem regulados. Por isso a descoberta ou invenção de novos promotores bem regulados é importante e fundamental para o desenvolvimento da moderna biotecnologia industrial (SAIDA, 2006; OVERTON, 2014).

[005] Atualmente, com o desenvolvimento da engenharia metabólica e a biologia sintética, muitas vezes há a necessidade de se expressar simultaneamente diferentes proteínas heterólogas no interior de uma mesma célula. Para isso, seus genes devem ser expressos por meio de uma gama de cassetes de expressão, com sequências suficientemente diferentes entre si para evitar que ocorra recombinação entre elas e consequentemente instabilidade do sistema genético e inviabilização do processo como um todo (GLICK, 1995; TERPE, 2006; KEASLING, 2012; BLAZECK & ALPER, 2013; ENGSTROM & PFLEGER, 2017).

[006] A presente sequência promotora, invenção motivo dessa solicitação de patente, pode ser aplicada nas duas circunstâncias acima descritas: 1) na produção industrial em altos níveis de proteínas recombinantes; 2) e na engenharia metabólica do maquinário celular envolvendo a expressão simultânea de diferentes genes heterólogos; pois apresenta expressão forte, bem regulada e sua sequência difere dos outros promotores conhecidos atualmente.

### **Breve descrição das Figuras**

[007] A Figura 1 apresenta o mapa físico do vetor de expressão pDM02. Pode-se identificar os principais componentes do vetor, como: o promotor sintético regulado (DM) e o operador *lac*; o sítio de ligação ao ribossomo (RBS) derivado do gene da proteína 10 do fago T7; a região de múltiplos sítios de clonagem composto pelos sítios das enzimas (*NdeI*, *HpaI*, *AflIII*, *KpnI*, *BglIII*, *PmeI* e *BamHI*); o terminador de transcrição funcional T1, do gene de *rrnB* de *E. coli*; o marcador de seleção que confere resistência ao antibiótico canamicina (CamR); e a origem de replicação de baixo

número de cópias – p15A (~10 a 12 cópias). Os sítios das enzimas de restrição destacados em negrito são únicos no vetor.

[008] A Figura 2 mostra o mapa físico do vetor de expressão pDM-GFP. A sequência codificante do gene da proteína verde fluorescente (GFP) foi clonada no vetor pDM02 entre os sítios de *NdeI* e *BamHI*.

[009] A Figura 3 apresenta o mapa físico do plasmídeo recombinante pDM-TEK. O gene híbrido, no qual possui uma porção do gene da tiorredoxina humana e a parte catalítica da enzima enteroquinase bovina, foi clonado no vetor pDM02 entre os sítios de *NdeI* e *BamHI*.

[010] A Figura 4 apresenta o perfil de expressão do gene da GFP, indicado pela fluorescência emitida pela proteína GFP em função do tempo de cultivo de *E. coli* JM110, programada pelo promotor sintético DM, clonado no plasmídeo pDM-GFP. O processo de indução com IPTG ocorreu a partir do início do cultivo.

[011] A Figura 5 mostra os resultados da análise por eletroforese em gel desnaturante de poliacrilamida 12% (SDS-PAGE) e fluorescência da produção de GFP por um clone de *E. coli* JM110 em diferentes condições de cultivo. Nas raias 1, 2 e 3, cultivo com canamicina e o indutor IPTG; raias 4, 5 e 6, com apenas canamicina; raias 7, 8 e 9 cultivo com glicose e canamicina. A raia M mostra o marcador molecular em kDa. A proteína recombinante GFP localiza-se na posição de aproximadamente 25 kDa (seta).

[012] A Figura 6 apresenta perfil do crescimento celular (ABS. 600nm x tempo) de *E. coli* JM110 programada com o plasmídeo pDM-TEK. O processo de indução com IPTG ocorreu a partir do início do cultivo.

[013] A Figura 7 apresenta o perfil eletroforético em gel de poliacrilamida desnaturante [12%] de proteínas expressas por um clone de *E. coli* JM110 em diferentes condições de cultivo. Nas raias 1 e 4, cultivo com glicose e canamicina; raias 2 e 5, com apenas canamicina; raias 3 e 6, cultivo com canamicina e o indutor IPTG. Nas raias 1, 2 e 3 amostras com 4 horas de cultivo; nas raias 4, 5 e 6 amostras com 8 horas de cultivo. A raia M, mostra o marcador molecular em kDa. A proteína recombinante TEK localiza-se na posição de 50 kDa (seta).

### Breve descrição das sequências

[014] A Sequência 1 compreende o novo promotor sintético regulado - DM. O promotor apresenta a sequência do *UP element* (TTAAAATTATTTTCAAACGG), a região -35 (TTGTGA) sobreposta ao operador *lac* (AATTGTGAGCGGATAACAATT), a região -10 (GATAAT), o elemento -10 estendido (TGTGTG), o discriminador (GGGGACC) e o sítio de início da transcrição ou +1 composto por uma base nucleotídica adenina (A).

[015] A Sequência 2 consiste na sequência do cassete de expressão contido no vetor pDM02. A porção sublinhada mostra o promotor sintético regulado descrito na Sequência 1; a região em itálico identifica o sítio de ligação ao ribossomo; marcado em negrito estão sítios de enzimas de restrição (“*polilinker*”) e o terminador de transcrição sublinhado duas vezes. E nas extremidades estão os sítios de reconhecimento das enzimas de restrição *EcoRI*, *NotI*, *XbaI* e *SpeI*, *NotI* e *PstI*.

[016] A Sequência 3 refere-se à sequência codificante do gene da proteína verde fluorescente (GFP). Nas extremidades estão os sítios de reconhecimento das enzimas de restrição *NdeI* e *BamHI*.

[017] A Sequência 4 consiste da sequência de parte do gene da tiorredoxina humana fusionado com a porção catalítica do gene da enteroquinase bovina, referida no texto como sequência TEK. A sequência que está sublinhada corresponde a sequência de parte do gene da tiorredoxina. As sequências que estão delimitadas em negrito e com identificação por escrito (Sítio de reconhecimento ENTK) correspondem a dois sítios de reconhecimento da enzima enteroquinase. A sequência que está pontilhada corresponde ao domínio *His-tag* utilizado para purificação. A sequência que não está sublinhada corresponde à cadeia leve (catalítica) da enzima enteroquinase. E nas extremidades estão os sítios de reconhecimento das enzimas de restrição *NdeI* e *BamHI*.

[018] Sequência 5 refere-se à sequência produzida por síntese química, contendo múltiplos sítios de clonagem “*polylinker*”, com sítios das seguintes enzimas de restrição: 5’ *NdeI*, *HpaI*, *AflIII*, *KpnI*, *BglIII*, *PmeI* e *BamHI* 3’.

[019] A Sequência 6 mostra a sequência completa do vetor de expressão pDM02.

**Promotores e vetores de expressão derivados do operon *lac* – Estado da técnica.**

[020] Os vetores de expressão funcionais em *Escherichia coli* para a produção de proteínas recombinantes necessitam ser bem regulados já que a alta e constante expressão de um gene heterólogo, com o acúmulo da respectiva proteína, pode ser prejudicial ou mesmo letal para as células, especialmente quando são tóxicas (SCHUMANN & FERREIRA, 2004; SAIDA, 2007; MARSCHALL et al, 2017).

[021] A regulação da expressão gênica em *E. coli*, como em todos os seres procariontes, ocorre principalmente ao nível da transcrição, tendo como principal elemento gênico funcional, o promotor. Existem dois grandes tipos de promotores presentes nas células, os ditos constitutivos ou não reguláveis e os reguláveis, ou seja, os que são suscetíveis a regulação (ANTHONY et al, 2004; HEISS et al, 2016).

[022] Um promotor é uma sequência de DNA curta (40 ~ 80 pb) capaz de recrutar a RNA polimerase holoenzima ( $\alpha$ ,  $\beta$ ,  $\beta'$ ,  $\omega$  e  $\sigma$ ) e iniciar a síntese de RNA. Em *E. coli*, os promotores mais comuns empregados em circuitos genéticos modificados são aqueles que interagem e são reconhecidos pelo fator  $\sigma$ 70. Os promotores reconhecidos por  $\sigma$ 70 em *E. coli* possuem elementos conservados denominados de região -35 (TTGACA) e região -10 (TATAAT) relativamente ao ponto de início da transcrição (+1). Estes hexâmeros são separados pela distância de 17 nucleotídeos, podendo ser um pouco maior ou menor, mas normalmente 17 nucleotídeos é a distância existente na maioria dos promotores fortes, pois corresponde ao espaçamento ideal para acoplamento da subunidade sigma (BROWNING & BUSBY, 2004; RANGEL-CHAVEZ et al., 2017).

[023] O complexo gênico regulável mais estudado e utilizado em engenharia genética é o operon responsável pelo metabolismo da lactose (*lac*) de *Escherichia coli*. Esse operon regulável é capaz de ser ativado/não ativado e reprimido/induzido. O operon *lac* é constituído por três genes estruturais designados por *lacZ*, *lacY* e *lacA*, que, respectivamente, codificam as proteínas  $\beta$ -galactosidase, permease e transacetilase sob o controle do promotor *lac*, que atua juntamente com outros elementos regulatórios para captação e uso da lactose, para utilizá-la como substrato energético, quando há pouca oferta de glicose para a célula (JACOB & MONOD, 1961; COOPER & MAGASANIK, 1974; LEWIS, 2005).

[024] O operon *lac* apresenta duas formas de regulação que atuam de forma complementar. O sistema de repressão/indução que envolve diretamente a região promotora/operadora onde se liga a proteína repressora *LacI* e o sistema de ativação que envolve uma sequência específica de DNA próxima ao promotor onde se liga a proteína *CRP* (*Cyclic-AMP Receptor Protein*), também denominada de CAP (Proteína Ativadora de Catabolismo) (KUHLMAN et al, 2007; ULLMAN, 2009).

[025] O sistema de repressão/indução funciona da seguinte forma: a *E. coli* tem um gene constitutivo (não regulado) denominado de *lacI*, que sintetiza a proteína repressora *LacI*, constituída por 360 aminoácidos, um produto difusível ativo que reconhece e se liga com afinidade ao locus operador (*lacO*) na ausência de lactose. Tal ligação impede que a RNA polimerase se associe ao promotor *lac* (*lacP*), bloqueando a transcrição deste operon. Na presença de lactose, essa é isomerizada no interior da célula, originando a alolactose, que se liga à proteína repressora *lacI* e altera sua conformação, fazendo com que essa perca a afinidade pelo operador *lac*, liberando-o. Dessa forma, a RNA polimerase pode reconhecer o promotor (*lacP*) e iniciar o processo de transcrição dos genes estruturais do operon *lac*, ou seja, dos genes *lacZ*, *lacY* e *lacA*. Esse processo de liberação da transcrição se chama de indução, onde a alolactose atua como molécula indutora. Normalmente, em laboratório, para induzir-se o operon *lac* utiliza-se um indutor análogo, não metabolizável e mais eficiente denominado de isopropil- $\beta$ -D-1-tiogalactopiranosídeo (IPTG) (JACOB & MONOD, 1961; ULLMAN, 2009; GARCIA et al., 2012; FAUST et al., 2015; BROWNING & BUSBY, 2016).

[026] O operador *lac* (*lacO*), que tem uma afinidade alta ao repressor *LacI*, é uma sequência palindrômica imperfeita de 21 pares de nucleotídeos (AATTGTGAGCGGATAACAATT) que na versão natural do promotor do operon *lac* está localizado da posição +1 no sítio de iniciação da transcrição, até a posição +21 (GILBERT & MÜLLER-HILL, 1966; LEWIS, 2005). O operon *Lac* natural contém também dois outros operadores *lac* situados nas posições -82 e +412 e postula-se que atuam sinergisticamente (OEHLER et al., 1990).

[027] O sistema de ativação, sensível a glicose, envolve *cAMP*, a proteína *CRP* (ou CAP) e o locus de ligação do complexo *CRP-cAMP*. Na ausência de glicose, aumenta o nível de *cAMP* nas células de *E. coli*, e este liga-se a proteína *CRP*. O complexo *cAMP-CRP* liga-se em uma sequência específica de DNA que se situa a montante do

promotor/operador. O complexo dímero *cAMP-CRP* interage com a RNA polimerase de forma que resulta na ativação do processo de transcrição dos genes estruturais do operon *lac* (BECKWITH, 1987; BUSBY & EBRIGHT, 1999; LEWIS, 2005).

[028] Pelo acima exposto, o maior nível de transcrição do operon *lac* ocorre então na ausência de glicose (sistema ativado) e na presença de alolactose ou IPTG (sistema induzido). Por outro lado, o menor nível de transcrição ocorre na presença de glicose e ausência de um indutor. O sistema dual de regulação ocorre simultaneamente e de forma complementar, satisfazendo as necessidades do metabolismo celular. Entretanto, esse sistema de regulação gênica para expressão de proteínas heterólogas em altos níveis em *E. coli*, apresenta limitações para aplicação industrial por dois principais motivos: 1) O promotor *lac* (*lacP*) é um promotor de força fraca e por si só não programa a produção em altos níveis de proteína heterólogas na células; 2) Embora o sistema de regulação seja dual, ainda assim a regulação não é boa (estridente), apresentando expressão basal. Portanto, mesmo nas melhores condições de repressão e não ativação, ainda ocorre um baixo nível de transcrição, cerca de 2% do total, e esse fato inviabiliza a expressão de proteínas tóxicas para as células bacterianas (BECKWITH et al., 1972; CAMSUND, 2014; ROSANO & CECCARELLI, 2014).

[029] Esses dois motivos, além da necessidade de existirem promotores disponíveis com sequências diferentes para programação de células simultaneamente com múltiplos genes heterólogos, foi o que motivou os autores a inventarem o novo promotor DM e a desenvolver o vetor de expressão pDM02, que é capaz de programar a expressão de proteínas heterólogas em *E. coli* em altos níveis e de forma bem regulada.

[030] Com o desenvolvimento da engenharia genética/biologia sintética, diversos promotores foram clonados ou construídos por síntese química, constituindo as estruturas de diferentes cassetes de expressão. A seguir serão apresentados vetores de expressão contendo promotores de diferentes origens:

-Promotor *lac* natural, utilizado conjuntamente com parte da sequência codificante do gene *lacZ* na composição dos vetores de clonagem e expressão de proteínas de fusão (híbridas), denominados de pUC8, pUC9, pUC18 e pUC19. Esses vetores são amplamente utilizados para clonagem molecular de genes, por conta do elevado número de cópias de suas origens de replicação. Porém, só

expressam proteínas na forma de fusão com  $\beta$ -galactosidase, com nível de expressão e regulação baixo (RINGS, 1982; YANISCH-PERRON et al., 1985; STRIEDNER et al., 2003; MARSCHALL et al., 2017);

- Promotor *tac*, um promotor híbrido formado por parte dos promotores *lace trp* (do operon triptofano). O promotor *tac* contém a região consensual -35 (TTGACA) derivada do promotor *trp* e a região consensual -10 (TATAAT) derivada do promotor *lac* e entra na composição dos vetores da série pTTQs ou pTRCs, que expressam médios e altos níveis de proteínas heterólogas com médio nível de regulação (de BOER et al., 1983; STARK, 1987; WONG & SUTHERLAND, 1990; 1997);
- Promotor  $P_{BAD}$  do operon da arabinose é regulado por L-arabinose e glicose, entra na composição dos vetores da série pBADs que expressam genes heterólogos com força média e relativamente bem regulados em bactérias (GUZMAN et al., 1995; GRUBER & HUANG, 2011; MARSCHALL et al., 2017);
- O promotor *T7* do bacteriófago T7, só é transcrito pela própria polimerase do vírus T7. Entra na formação dos vetores da série pETs, que expressam proteínas heterólogas em altos níveis, porém com baixo nível de regulação (STUDIER et al., 1997; PIOLI et al., 1999; OVERTON, 2014);
- Promotores *pR* e *pL*, derivados do fago  $\lambda$ , regulados pelo repressor termossensível  $CI^{857}$ , são induzidos pelo aumento da temperatura de 30°C para 40°C. Os vetores de expressão baseados nesses promotores expressam, respectivamente, médios e altos níveis de proteínas heterólogas de forma bem regulada. Entretanto, para manter-se não induzido termicamente, a célula hospedeira *E. coli* necessita crescer abaixo de sua temperatura ótima, a 30°C, sendo uma limitação para o processo fermentativo industrial. Além disso, durante a indução a 40°C, o aumento da temperatura pode desencadear na ativação do sistema de estresse celular térmico, influenciando no metabolismo celular (KELLY et al., 2005; LIMA et al., 2010; VALDEZ-CRUZ et al., 2010).

### **Descrição detalhada do promotor e do vetor de expressão objetos da invenção**

[031] O promotor DMinventado pelos presentes autores, cuja sequência completa está apresentada na Sequência 1, foi construído por síntese química simultaneamente com

todo o cassete de expressão (Sequência 2). Além do promotor, o cassete de expressão contém também a sequência codificadora da região 5' UTR do mRNA do gene da proteína 10 do bacteriófago T7 que contém um eficiente sítio de ligação do ribossomo (OLINS, 1988), uma região de múltiplos sítios de clonagem, e um terminador de transcrição, o T1 do gene de *rrnB* de *E. coli* (OROSZ et al., 1991).

[032] Como pode ser observado na Sequência 1, o operador *lac* (*lacO*) do sistema do operon *lac* foi sobreposto à região conservada -35 e à região espaçadora situada entre a região -35 e a -10 (TATA ou *Pribnowbox*) do promotor. Portanto, a região -35 no presente design apresenta a sequência TTGTGA, ao invés da sequência consensual TTGACA. Dessa forma, diferentemente do promotor *lac* natural, no promotor sintético DM a região de reconhecimento -35 da subunidade sigma 70 está sobreposta à região operadora, onde se acopla a proteína repressora *LacI*, aprimorando conseqüentemente a eficiência de bloqueio da transcrição deste complexo promotor-operador-regulador. Essa estratégia fez com que o promotor operasse com bom nível de regulação por meio do gene *lacI<sup>q</sup>* que codifica a proteína repressora *LacI*. Para completar os 17 nucleotídeos preferenciais da região espaçadora entre as regiões -35 e -10, optou-se por colocar o elemento TGTG antes da região -10, considerado uma extensão da mesma, capaz de torná-la mais eficiente e potencializar a força da expressão (BURR et al., 2000).

[033] O cassete de expressão contendo a região estrutural do gene da GFP (*Green Fluorescent Protein*) (Sequência 3) construído por síntese química (Genone Biotecnologies Ltda.) foi clonado em um plasmídeo denominado pSB3K3-TH2, o plasmídeo resultante foi denominado pDM-GFP (Figura 2). Além do cassete de expressão, esse plasmídeo possui origem de replicação de baixo número de cópias (10-12 cópias) proveniente do plasmídeo p15A (CHANG & COHEN, 1978) e o gene de resistência a canamicina originário do transposon Tn903 (GRINDLEY & JOYCE, 1980).

[034] Para estudar a expressão e regulação de genes heterólogos sob o controle do promotor DM, uma outra sequência gênica, também produzida por síntese química, codificante da proteína híbrida entre tiorredoxina humana e enteroquinase bovina (TEK) (Sequência 4) foi subclonada entre os sítios das enzimas de restrição *NdeI* e *BamHI* do plasmídeo recombinante pDM-GFP, em substituição a sequência codificante do gene da GFP, originando o plasmídeo recombinante pDM-TEK (Figura 3).

[035] Estudos de expressão do gene da GFP e da proteína híbrida entre tiorredoxina humana e enteroquinase bovina (TEK), contidos respectivamente nos plasmídeos recombinantes pDM-GFP e pDM-TEK, demonstram que o promotor DM é capaz de expressar altos níveis de proteínas heterólogas em *E. coli* performance muito bem regulada (estringentemente) (Figuras 4, 5, 6 e 7).

[036] Para criar um vetor de expressão com maior flexibilidade para clonagem e expressão de outras sequências, o gene de GFP do plasmídeo pDM-GFP foi substituído por uma sequência com múltiplos sítios de restrição (*polylinker*) (Sequência 5), originando o vetor pDM02 (Figura 1) (Sequência 6).

### **Exemplos**

#### **Exemplo 1 – Síntese química do promotor DM, do cassete de expressão, construção do plasmídeo pDM-GFP e expressão do gene da GFP em *Escherichia coli*.**

[037] Conforme já descrito nesse documento, o promotor DM (Sequência 1) foi engenheirado para ser muito bem regulado (estringentemente), posicionado no cassete de expressão a montante de sequências robustas de *UP element* e a jusante da sequência codificadora do sítio de ligação do ribossomo (*ShineDalgarno*) e do terminador de transcrição (Sequência 2). Após sua concepção teórica e o posicionamento adequado da região estrutural do gene da GFP (Sequência 3) no interior do cassete de expressão, a síntese química de todo o conjunto foi encomendada à empresa GENONE BIOTECNOLOGIES LTDA. Após a síntese a sequência foi clonada no plasmídeo pSB3K3-TH2, originando o plasmídeo recombinante pDM-GFP (Figura 2).

[038] O plasmídeo pDM-GFP que foi recebido liofilizado (4µg) e foi dissolvido em 20 µL de água ultra pura estéril. A seguir, para obtenção de massa plasmidial e análise da expressão do gene da GFP, o plasmídeo foi introduzido em células de *E. coli* JM110 (YANISCH-PERRON et al., 1985) da seguinte forma: 0,2 µL da solução de DNA plasmidial foram misturados com 50 µL de células *E. coli* JM110 eletrocompetentes, e submetidas ao procedimento de eletroporação. Após esse procedimento, as células foram misturadas com 0,5 mL de meio líquido LB (LuriaBertani) e colocadas para recuperação a 37°C por 1 hora com leve agitação. Em seguida, alíquotas dos sistemas foram semeadas em placas de Petri com meio LB ágar (Peptona, 10g/l. Extrato de

levedura, 5 g/l. NaCl, 10 g/l. Ágar, 15g/l. pH final 7,0) suplementado com canamicina (50 µg/ µL). As placas foram incubadas invertidas a 37°C durante 16 horas.

[039] Colônias transformantes foram selecionadas e incubadas em 5 mL de meio LB líquido contendo 50 µg/mL de canamicina em frasco fechado e estéril. Os pré-inóculos foram incubados sob agitação constante (140 rpm) por um período de 14 horas a 37°C. Os plasmídeos de clones recombinantes de *E. coli* portadores do vetor pDM-GFP selecionados foram extraídos com o Kit *IllustraplasmidPrep Mini Spin Kit (GE HealthcareBioscience)* de acordo com as especificações do fabricante. Após, alíquotas foram analisadas por eletroforese em gel de agarose 0,8% e visualizado/fotografado sob luz ultravioleta (UV). As técnicas de digestão de DNA com enzimas de restrição e de ligação foram realizadas de acordo com protocolos descritos por SAMBROOK et al., 2001.

[040] Para avaliar a expressão do gene da GFP em meio líquido, um clone recombinante portador do vetor pDM-GFP foi inoculado em um tubo de ensaio contendo 5 mL de meio LB líquido contendo o antibiótico seletivo canamicina 50 µg/mL e 1% de glicose. A cultura foi incubada durante a noite a 37°C sob agitação de 150 rpm. No dia seguinte, foram inoculados 500 µL do pré-inóculo em triplicatas em Erlenmeyers de 250 mL estéreis e fechados contendo 50 mL de meio LB líquido suplementados com: ou somente canamicina 50µg/mL, ou canamicina 50µg/mL e glicose (1%) ou canamicina 50µg/mL e o indutor IPTG (1mM). As culturas foram então incubadas por 8 horas à 37°C com agitação de 150 rpm. Em intervalos de 1 hora alíquotas de 2 mL foram coletadas para mensurar o crescimento celular a partir da absorbância a 600nm em espectrofotômetro e para análise das proteínas por eletroforese em gel desnaturante de poliacrilamida 12% (SDS-PAGE), conforme método descrito por Laemmli, 1970. Como o fluoróforo da GFP é tridimensional, para ser possível sua visualização no gel de poliacrilamida por meio da fluorescência, após a solubilização das proteínas celulares no tampão de amostras para o processo eletroforético, essa solução não foi fervida, para que a GFP não desnaturasse.

[041] Para a análise quantitativa da expressão e regulação do gene da GFP com a produção da proteína verde fluorescente (GFP), sob o controle promotor sintético regulado DM, em células de *E. coli* JM110 programadas com o plasmídeo recombinante pDM-GFP, clones foram submetidos a diferentes condições de cultivo e indução. Para

quantificara fluorescência da GFP, alíquotas de 1 mL de cada cultura eram coletadas e centrifugadas a 12.000g por 5 minutos. Os sedimentos foram ressuspensos em 1 mL de tampão TN (Tris-HCl 10mM e NaCl 0,15M pH 8,0) e dessa suspensão celular, 100µL foram transferidos para placas de culturas de 96 poços de cor negra. A fluorescência emitida foi medida em espectro fluorímetro *HidexChameleon* (comprimento de onda de excitação de 340 nm e de emissão de 535nm). Durante o desenvolver das culturas a produção de GFP era acompanhada também visualizando-se as células bacterianas por meio de microscópio de fluorescência (dados não mostrados).

[042] Como pode ser observado na Figura 4, as células de *E. coli* expressam altos níveis de GFP (13.000 UR de fluorescência) na presença de IPTG 1mM e nível basal de fluorescência na presença apenas do antibiótico (canamicina). É notável que praticamente nenhuma fluorescência seja detectada em células crescidas na presença de glicose 1%, o que demonstra ser possível modular o referido promotor entre condições de expressão praticamente nula até níveis de alta expressão de GFP.

[043] A análise da expressão da GFP por eletroforese está mostrada na Figura 5 onde pode-se observar alta expressão da GFP pela banda fluorescente de cerca de 25 kDa, indicada pela seta nas raias 1, 2 e 3, referentes as proteínas das células cultivadas na presença do indutor IPTG 1mM. Este resultado se contrasta com a expressão praticamente nula nas outras duas condições: apenas com canamicina (raias 4,5 e 6) e com canamicina e glicose 1% (raias 7, 8 e 9). Esses resultados estão de acordo com os de análise quantitativa de fluorescência (Figura 4) e demonstram que o promotor DM apresenta performance de alta expressão e boa regulação.

### **Exemplo 2 – Construção do plasmídeo pDM-TEK e expressão do gene híbrido da tiorredoxina humana e enteroquinase bovina em *E. coli*.**

[044] Visando analisar o funcionamento do promotor DM na expressão de outra sequência codificante, selecionou-se o gene híbrido da tiorredoxina humana com o gene da enteroquinase bovina (TEK), uma vez que a enteroquinase, além de ser tóxica por ser uma protease, tem diversas aplicações industriais e alto valor de mercado.

[045] Para a construção do plasmídeo pDM-TEK (Figura 3), o plasmídeo pDM-GFP foi digerido com as enzimas de restrição *NdeI* e *BamHI* retirando a sequência codificante do gene da GFP. Por meio do processo de clivagem do plasmídeo pENTK com as

enzimas de restrição *NdeI* e *BamHI* e posterior purificação por eletroforese em gel de agarose 0,8%, foi possível obter a sequência codificante do gene híbrido da proteína tiorredoxina humana e enteroquinase bovina (TEK) (Sequência 4) de aproximadamente 1254 pares de bases. A seguir, a sequência gênica TEK foi ligada utilizando T4 DNA ligase ao vetor de expressão pDM, também digerido com as mesmas enzimas, para originar o plasmídeo recombinante pDM-TEK (Figura 3). As técnicas de digestão de DNA com enzimas de restrição e de ligação foram realizadas de acordo com protocolos descritos por SAMBROOK et al., 2001.

[046] Para avaliar a expressão e regulação do gene híbrido da tiorredoxina humana e enteroquinase bovina (TEK), clones recombinantes de *E. coli* JM110 portadores dos vetores de expressão pDM-TEK foram selecionados e inoculados em 5 mL de meio LB líquido em tubos fechados e estéreis suplementado com o antibiótico seletivo canamicina 50µg/mL e 1% de glicose. Os pré-inóculos foram incubados sob agitação constante (150 rpm) por um período de 14 horas a 37°C. No dia seguinte, foram inoculados 500 µL dos pré-inóculos em triplicatas em Erlenmeyers de 250 mL contendo cada 50 mL de meio LB líquido suplementados separadamente com: somente canamicina 50µg/mL, ou canamicina 50µg/mL e glicose (1%), ou canamicina 50µg/mL e o indutor IPTG (1 mM).

[047] A indução da expressão do gene da proteína híbrida TEK por *E. coli* JM110 programada com o plasmídeo pDM-TEK, com adição de IPTG 1mM, foi realizada quando as culturas atingiram a absorbância em 600 nm igual a 0,1. A partir deste momento, alíquotas de 2 mLs foram retiradas a cada 2 horas até o tempo total de 8 horas para as leituras de absorbância (elaboração das curvas de crescimento) e também para análise das proteínas por eletroforese em gel de poliacrilamida 12% (SDS-PAGE). As proteínas recombinantes produzida por clones de *E. coli* portadores de plasmídeos pDM-TEK foram analisada por eletroforese em gel desnaturante de poliacrilamida 12% (SDS-PAGE) segundo Laemmli (1970).

[048] Como pode ser observada na Figura 6, as células de *E. coli* JM110 nas 3 diferentes condições de crescimento apresentaram praticamente o mesmo perfil de crescimento, inclusive na presença do indutor IPTG, indicando que a proteína híbrida expressa não apresentou toxicidade para as células, não alterando o perfil de crescimento. Essa não toxicidade pode ser explicada das seguintes maneiras: a proteína híbrida é expressa

inativa devido a sua própria estrutura artificial, ou por acumular-se no interior das células de *E. coli* na forma de corpúsculos de inclusão.

[049] A análise da expressão da TEK em *E. coli* JM110 por eletroforese em gel desnaturante de poliacrilamida está mostrada na Figura 7. Pode-se observar altos níveis de expressão da proteína híbrida na presença do indutor tanto após 4 como 8 horas de indução (raias 3 e 6) e praticamente nenhuma indução da proteína híbrida na presença de glicose tanto após 4 como 8 horas de indução (raias 1 e 4), o que confirma que o promotor DM apresenta performance de alta expressão e boa regulação (estridente).

### **Exemplo 3 – Construção do vetor de expressão pDM02**

[050] Visando maior versatilidade no uso do promotor DM, contido no cassete de expressão do plasmídeo pDM-GFP, o gene da GFP foi substituído por um fragmento de DNA sintetizado quimicamente portador de múltiplos sítios de clonagem (*polylinker*), contendo os seguintes sítios de enzimas de restrição: *NdeI* (CATATG) e *BamHI* (GGATTC) nas extremidades em forma de *stick-ends*.

[051] O *polylinker* foi desenhado para ser construído a partir de 2 oligonucleotídeos, a saber o oligo PLK01 e o oligo PLK02, cujas sequências estão apresentadas na forma de DNA dupla-fita na Sequência 5. Na forma dupla fita a molécula exibe sítios de *NdeI* e *BamHI* em suas extremidades. Ambos os nucleotídeos foram sintetizados pela empresa EXXTEND BIOTECNOLOGIA LTDA.

[052] Após a síntese, os oligonucleotídeos PLK01 e PLK02 foram dissolvidos em água ultrapura estéril para ficarem na concentração de 100 pmoles/ $\mu$ L. Para a obtenção do oligonucleotídeo dupla-fita (dPLK), alíquotas de cada um dos oligos contendo 1  $\mu$ g foram misturadas, adicionando-se Tris-HCl 1M, NaCl 3M e água, para que ficassem com volume final de 100  $\mu$ L e respectivas concentrações finais de 10mM de Tris-HCl e 50mM de NaCl. Para anelamento, a solução contendo os dois oligonucleotídeos foi fervida em banho-maria por 2 minutos e deixada esfriar gradativamente a temperatura ambiente por 30 minutos. Após resfriamento, a solução foi armazenada a 4°C até o uso.

[053] Para a construção do plasmídeo pDM02, o plasmídeo pDM-GFP foi extraído de clone de *E. coli* JM110 geneticamente transformado. A seguir, foi digerido com as enzimas de restrição *NdeI* e *BamHI* para retirada da sequência codificante do gene da GFP. O plasmídeo pDM desprovido do gene foi purificado por eletroforese em gel de

agarose 0,8% e extraído do gel de agarose por meio do kit *illustra™ GFX™ PCR DNA and gel Band Purification Kit* da GE Healthcare (USA). A seguir, 50ng do oligonucleotídeo sintético (dPLK) foi ligado com a enzima T4 DNA ligasea 100ng do vetor pDM linearizado, originando o vetor de clonagem e expressão pDM02. As técnicas de digestão com enzimas de restrição, de ligação e eletroforese foram realizadas de acordo com o descrito por SAMBROOK et al., 2001; ou de acordo com os fabricantes das enzimas e kits.

[054] O mapa físico do vetor de expressão resultante pDM02 está apresentada na Figura 1, onde pode-se identificar o cassete de expressão com o *polylinker*, o gene de resistência a canamicina e a origem de replicação do plasmídeo. Dessa forma, com a adição do *polylinker*, o vetor pode ser utilizado de forma mais versátil e ampla, com mais opções de sítios de enzimas de restrição, na clonagem de sequências de DNA a serem expressas.

### Referências Bibliográficas

ANTHONY, L.C.; SUZUKI, H.; FILUTOWICZ, M. (2004). Tightly regulated vectors for the cloning and expression of toxic genes. *J. Microbiol. Methods.* **2**:243-250 <https://doi.org/10.1016/j.mimet.2004.04.003>.

BECKWITH, J. (1987). The Lactose Operon. In *Escherichia coli* and *Salmonella typhimurium. Cellular and Molecular Biology.* (edited by NEIDHARDT, F. C.; INGRAHAM, J. L.; LOW, K. B.; MAGASANIK, M.; SCHAECHTER, M.; UMBARGER, H. E. American Society for Microbiology, Washington, D.C. pp. 1444–1452).

BECKWITH, J.; GRODZICKER, T.; ARDITTI, R. (1972). Evidence for two sites in the lac promoter region. *J. Mol. Biol.* **69**:155-160. [https://doi.org/10.1016/0022-2836\(72\)90031-9](https://doi.org/10.1016/0022-2836(72)90031-9).

BLAZECK J. & ALPER H. S. (2013). Promoter engineering recent advances in controlling transcription at the most fundamental level. *Biotechnol. J.* **8**:46-58. doi: 10.1002/biot.201200120.

BROWNING, D. F. & BUSBY, S. J. W. (2004). The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* **2**:57–65 doi: 10.1038/nrmicro787

BROWNING, D. F. & BUSBY, S. J. (2016). Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* **14**:638–650. doi:10.1038/nrmicro.2016.103.

BURR T.; MITCHELL J.; KOLB A.; MINCHIN S.; BUSBY S. (2000). DNA sequence elements located immediately upstream of the -10 hexamer in *Escherichia coli* promoters: a systematic study. *Nucleic. Acids. Res.* **28**:1864–1870. doi: 10.1093/nar/28.9.1864.

BUSBY, S. & EBRIGHT, R.H. (1999). Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol.*, **293**:199-213. <https://doi.org/10.1006/jmbi.1999.3161>

CAMSUND, D.; HEIDORN, T.; LINDBLAD, P. (2014). Design and analysis of LacI-repressed promoters and DNA-looping in a *cyanobacterium*. *J. Biol Eng.* **8**:4. doi: 10.1186/1754-1611-8-4.

CHANG A. C. & COHEN S. N. (1978). Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* **134**:1141–1156.

COOPER, T. G. & MAGASANIK, B. (1974) Transcription of the lac Operon of *Escherichia coli*. *The Journal of Biological Chemistry.* **249**:6556- 6561, 1974.

COHEN, S. N.; CHANG, A. C.; BOYER, H. W.; HELLING, R. B. (1973). Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **70**: 3240–3244. doi: 10.1073/pnas.70.11.3240

de BOER, H. A.; COMSTOCK, L. J.; VASSER M. (1983). The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proc. Natl. Acad. Sci. U. S. A.* **80**:21-25.

DEMAIN, A. L. & VAISHNAV, P. (2009). Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* **27**:297–306. doi: 10.1016/j.biotechadv.2009.01.008.

ENGSTROM, M. D. & PFLEGER, B. F. (2017) Transcription control engineering and applications in synthetic biology. *Synthetic and Systems Biotechnology.* **3**:176-191. <https://doi.org/10.1016/j.synbio.2017.09.003>.

- FAUST, G.; STAND, A.; WEUSTER-BOTZ, D. (2015). IPTG can replace lactose in auto-induction media to enhance protein expression in batch-cultured *Escherichia coli*. *Eng. Life Sci.***15**:824-829. doi: 10.1002/elsc.201500011.
- GARCIA, H.G.; SANCHEZ, A.; BOEDICKER, J.Q.; OSBORNE, M.; GELLES, J.; KONDEV, J.; PHILLIPS, R. (2012). Operator sequence alters gene expression independently of transcription factor occupancy in bacteria. *Cell Rep.* **2**(1):150–61. <http://dx.doi.org/10.1016/j.celrep.2012.06.004>.
- GILBERT W & MÜLLER-HILL B. (1966). Isolation of the lac repressor. *Proc. Natl. Acad. Sci. USA.* **8**:1891–1898. doi: 10.1073/pnas.56.6.1891.
- GLICK, B. R. (1995). Metabolic load and heterologous gene expression. *Biotech. Adv.***2**:247-261. [https://doi.org/10.1016/0734-9750\(95\)00004-A](https://doi.org/10.1016/0734-9750(95)00004-A).
- GOLDSTEIN M. A. & DOI R. H. (1995). Prokaryotic promoters in biotechnology. *Biotechnology Annual Review*, **1**:105-128. [https://doi.org/10.1016/S1387-2656\(08\)70049-8](https://doi.org/10.1016/S1387-2656(08)70049-8)
- GRINDLEY N. D. & JOYCE C. M. (1980). Genetic and DNA sequence analysis of the kanamycin resistance transposon Tn903. *Proc. Natl. Acad. Sci. U S A.***12**:7176–7180.
- GRUBER, T. M. & HUANG, L. L. Mutant arabinose promoter for inducible gene expression. 2011. Número de registro: US7998702 B2. Data de depósito: 17/09/2008 Data de Concessão: 16/08/2011.
- GUZMAN, L. M.; BELIN, D.; CARSON, M. J.; BECKWITH, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose pBAD promoter. *J. Bacteriol.* **177**, 4121–4130. doi: 10.1128/jb.177.14.4121-4130.1995
- HALL, R. M. & COLLIS, C. M. (1995) Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Mol. Microbiol.***4**:593–600. doi: 10.1111/j.1365-2958.1995.tb02368.x.
- HAUGEN, S. P.; ROSS, W.; GOURSE, R. L. (2008). Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat. Rev. Microbiol.***7**:507-519. doi: 10.1038/nrmicro1912

HEISS, S.; HORMANN, A.; TAUER, C.; SONNLEITNER, M.; EGGER, E.; GRABHERR, R.; HEINL, S. (2016). Evaluation of novel inducible promoter/repressor systems for recombinant protein expression in *Lactobacillus plantarum*. *Microb. Cell. Fact.* **15**:50. <https://doi.org/10.1186/s12934-016-0448-0>.

KEASLING J. D. (2012). Synthetic biology and the development of tools for metabolic engineering. *Metab. Eng.* **14**:189-195. <https://doi.org/10.1016/j.ymben.2012.01.004>.

KELLY, A.; FILIPE, S.; GNADT, N.; BARBAS, A.; PISSARRA, P. Expression vectors and promoters for heterologous gene expression. (2005). Número de registro: US20050119462 A1. Data de depósito: 23/12/2002 Data de Concessão: 02/06/2005.

KUHLMAN, T.; ZHANG, Z.; SAIER, M. H.; HWA, T. (2007). Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proceedings of the National Academy of Sciences.* **14**:6043-6048; doi: 10.1073/pnas.0606717104.

LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680–685.

LEWIS, M. (2005). The lac repressor. *C. R. Biol.* **328**(6):521-548. <https://doi.org/10.1016/j.crvi.2005.04.004>

LIMA, B. D., SOUZA, H. R. T., THIEMANN, J. F., VILELA, L; ASTOLFI-FILHO, S. Vetor para expressão de proteína heteróloga e métodos para extrair proteína recombinante e purificar insulina recombinante isolada. 2010. Número de registro: PI9810650-3. Data de depósito: 02/07/1998 Data de Concessão: 14/02/2010

MARSCHALL, L.; SAGMEISTER, P.; HERWIG, C. (2017). Tunable recombinant protein expression in *E. coli*: promoter systems and genetic constraints. *ApplMicrobiolBiotechnol.* **2**:501-512. <http://doi.org/10.1007/s00253-016-7550-4>.

OEHLER, S.; EISMANN E.R.; KRAMER, H.; MÜLLER-HILL, B. (1990). The three operators of the *lac* operon cooperate in repression. *EMBO J.* **9**:973-9

OLINS P. O.; DEVINE C. S.; RANGWALA S. H.; KAVKA K. S. (1988). The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in *Escherichia coli*. *Gene.* **73**(1):227-235.

OROSZ, A.; BOROS, I.; VENETIANER, P. (1991). Analysis of the complex transcription termination region of the *Escherichia coli* *rrnB* gene. *Eur. J. Biochem.* **201**:653-659.

OVERTON, T. W. (2014). Recombinant protein production in bacterial hosts. *Drug Discovery Today*. <https://doi.org/10.1016/j.drudis.2013.11.008>.

PIOLI, D.; HOCKNEY, R. C.; KARA, B. V.; BUNDELL, K. R. T7 promoter-based expression system. 1999. Número de registro: WO1999005297 A1. Data de depósito: 21/07/1998 Data de Concessão: 4/02/1999.

RANGEL-CHAVEZ, C.; GALAN-VASQUEZ, E.; MARTINEZ-ANTONIO, A. (2017). Consensus architecture of promoters and transcription units in *Escherichia coli*: design principles for synthetic biology†. *Mol. BioSyst.* **13**:665. doi: 10.1039/c6mb00789a

RINGS A. D. Method for microbial polypeptide expression. 1982. Número de registro: US4366246 A. Data de depósito: 05/11/1979 Data de Concessão: 28/12/1982.

ROSANO G. L. & CECCARELLI E. A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* **5**:172. doi:10.3389/fmicb.2014.00172.

SAIDA, F.; UZAN, M.; ODAERT, B.; BONTEMS, F. (2006) Expression of highly toxic genes in *E. coli*: special strategies and genetic tools. *Curr. Protein Pept. Sci.* **7**:47–56.

SAIDA, F. (2007). Overview on the Expression of Toxic Gene Products in *Escherichia coli*. *Current Protocols in Protein Science.* **5**:5.19. doi: 10.1002/0471140864.ps0519s50

SAMBROOK, J. & RUSSELL, D. (2001). *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

SCHUMANN, W. & FERREIRA, L.C.S. (2004). Production of recombinant proteins in *Escherichia coli*. *Genet. Mol. Biol.* **3**:442-453.

SINGHA, T. K.; GULATI, P.; MOHANTY, A.; KHASA, Y. P.; KAPOOR, R. K.; KUMAR, S. (2017) Efficient genetic approaches for improvement of plasmid based

expression of recombinant protein in *Escherichia coli*: A review. *Process Biochemistry* **55**:17–31. <https://doi.org/10.1016/j.procbio.2017.01.026>

STARK, M. J. R. (1987). Multicopy expression vectors carrying the lac repressor gene for regulated high-level expression of genes in *Escherichia coli*. *Gene*. **51**:255–267.

STRIEDNER, G.; CSERJAN-PUSCHMANN, M.; POTSCHACHER, F.; BAYER, K. (2003). Tuning the transcription rate of recombinant protein in strong *Escherichia coli* expression systems through repressor titration. *Biotechnol. Prog.* **19**(5):1427–1432. doi:10.1021/bp034050u.

STUDIER, F. W.; DAVANLOO, P.; ROSENBERG, A. H.; MOFFATT, B. A.; DUNN, J. J. (1997). Cloning and expression of the gene for bacteriophage T7 RNA polymerase. Número de registro: US5693489 A. Data de depósito: 14/06/1994 Data de Concessão: 02/12/1997.

TERPE, K. (2006) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biot.* **5**:523–533. doi: 10.1007/s00253-002-1158-6.

ULLMAN, A. (2009). *Escherichia coli* Lactose Operon. Encyclopedia of Life Sciences. doi: 10.1002/9780470015902.a0000849.pub2.

VALDEZ-CRUZ, N. A.; CASPETA, L.; PÉREZ, N. O.; RAMÍREZ, O. T.; TRUJILLO-ROLDÁN, M. A. (2010). Production of recombinant proteins in *E. coli* by the heat inducible expression system based on the phage lambda pL and/or pR promoters. *Microbial CellFactories*. **9**: 18. doi: 10.1186/1475-2859-9-18.

VITORINO, L. C. & BESSA, L. A. (2017). Technological Microbiology: Development and Applications. *Front. Microbiol.* **8**:827. <https://doi.org/10.3389/fmicb.2017.00827>.

WONG, R. W. K. & SUTHERLAND, M. L. Excretion of heterologous proteins from *E. coli*. (1990). Número de registro: EP0357391 A2. Data de depósito: 30/08/1989 Data de Concessão: 07/03/1990.

WONG, R. W. K. & SUTHERLAND, M. L. Excretion of heterologous proteins from *E. coli*. (1997). Número de registro: US5646015 A. Data de depósito: 11/05/1995 Data de Concessão: 08/07/1997.

YANISCH-PERRON, C.; VIEIRA, J.; MESSING, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*.**33**:103-119. [https://doi.org/10.1016/0378-1119\(85\)90120-9](https://doi.org/10.1016/0378-1119(85)90120-9).

## REINVINDICAÇÕES

### UM PROMOTOR SINTÉTICO PARA ALTA EXPRESSÃO DE GENES HETERÓLOGOS EM *Escherichia coli* EM PERFORMANCE BEM REGULADA

1. UM PROMOTOR SINTÉTICO PARA ALTA EXPRESSÃO DE GENES HETERÓLOGOS EM *Escherichia coli* EM PERFORMANCE BEM REGULADA **caracterizado** pela sequência completa (Sequência 1), que possui a região operadora *lac* sobreposta a região -35, possibilitando com essa estratégia alto nível de regulação em *E. coli* e em outras espécies bacterianas;
2. UM PROMOTOR SINTÉTICO PARA ALTA EXPRESSÃO DE GENES HETERÓLOGOS EM *Escherichia coli* EM PERFORMANCE BEM REGULADA , de acordo com a reivindicação 1, caracterizado por quaisquer derivados produzidos por mutagênese ou síntese química;
3. UM PROMOTOR SINTÉTICO PARA ALTA EXPRESSÃO DE GENES HETERÓLOGOS EM *Escherichia coli* EM PERFORMANCE BEM REGULADA, de acordo com a reivindicação 1, caracterizado pelo cassete de expressão de sequência completa (Sequência 2), denominado de DM juntamente com a sequência codificadora da região SD do gene da proteína 10 do fago T7, *polylinker* e o terminador de transcrição T1 e quaisquer dos seus derivados;
4. UM PROMOTOR SINTÉTICO PARA ALTA EXPRESSÃO DE GENES HETERÓLOGOS EM *Escherichia coli* EM PERFORMANCE BEM REGULADA, de acordo com as reivindicações 1, 2 e 3, caracterizado pelo vetor de expressão pDM02, e seu mapa físico (Figura 1), capaz de expressar altos níveis de proteínas heterólogas e quaisquer de seus derivados;
5. UM PROMOTOR SINTÉTICO PARA ALTA EXPRESSÃO DE GENES HETERÓLOGOS EM *Escherichia coli* EM PERFORMANCE BEM REGULADA, acordo com as reivindicações 1,2, 3, 4, caracterizado pela expressão de GFP de forma estritamente regulada pelo plasmídeo pDM-GFP e seu mapa físico (Figura 2);
6. UM PROMOTOR SINTÉTICO PARA ALTA EXPRESSÃO DE GENES HETERÓLOGOS EM *Escherichia coli* EM PERFORMANCE BEM REGULADA, de acordo com as reivindicações 1, 2, 3, 4, caracterizado pela

expressão de enteroquinase de forma estritamente regulada pelo plasmídeo recombinante pDM-TEK.

## LISTA DE FIGURAS

Figura 1

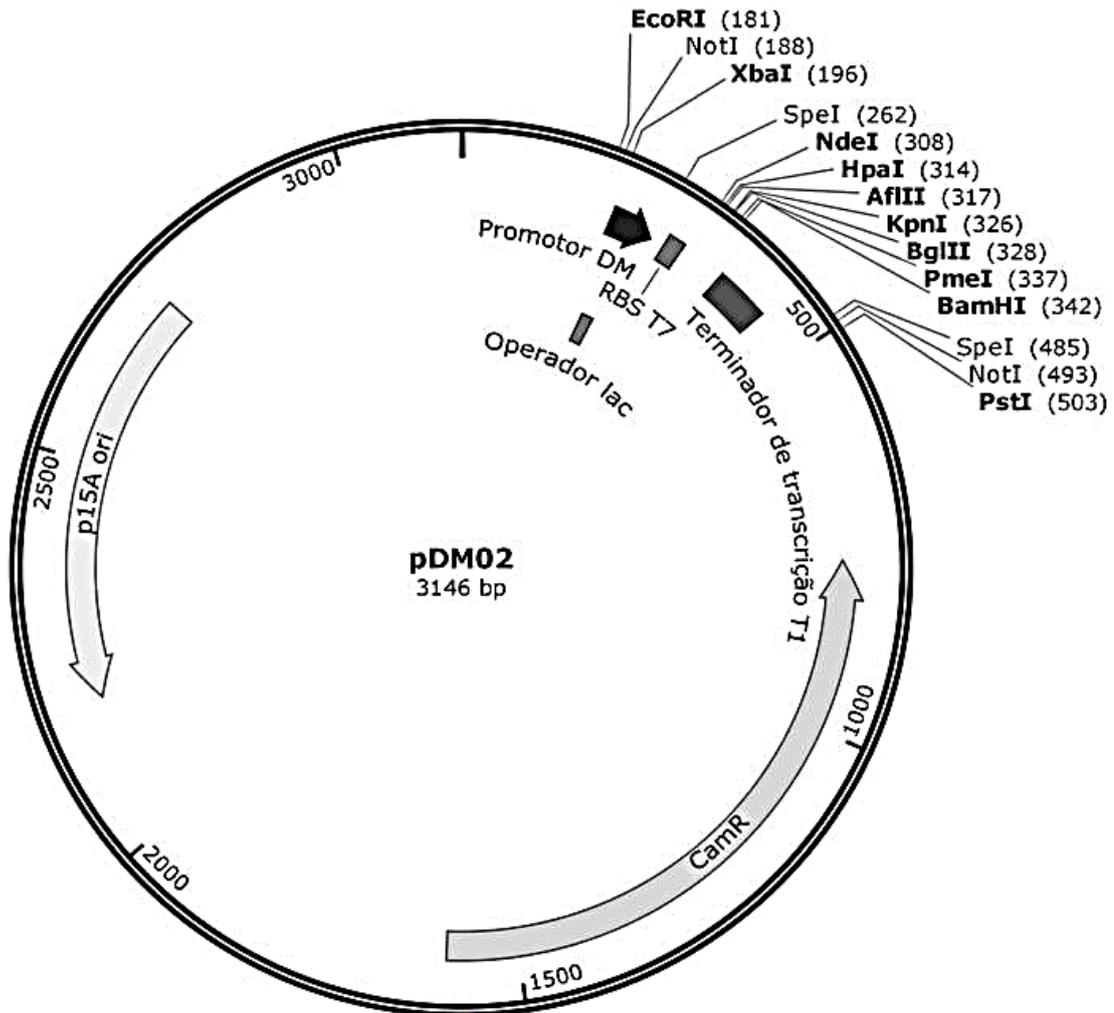


Figura 1: O mapa físico do vetor de expressão pDM02 que contém o *polylinker* situado entre os sítios das enzimas de restrição *NdeI* (308) e *BamHI* (342). O vetor de expressão pDM02 apresenta o promotor regulado DM e o operador *lac*; o sítio de ligação ao ribossomo (RBS) derivado do gene da proteína 10 do fago T7; o terminador de transcrição funcional T1, do gene de *rrnB* de *E. coli*; o marcador de seleção que confere resistência ao antibiótico canamicina (CamR); e a origem de replicação de baixo número de cópias – p15A (~10 a 12 cópias). O vetor pDM02 apresenta 3146 pares de bases.

Figura 2

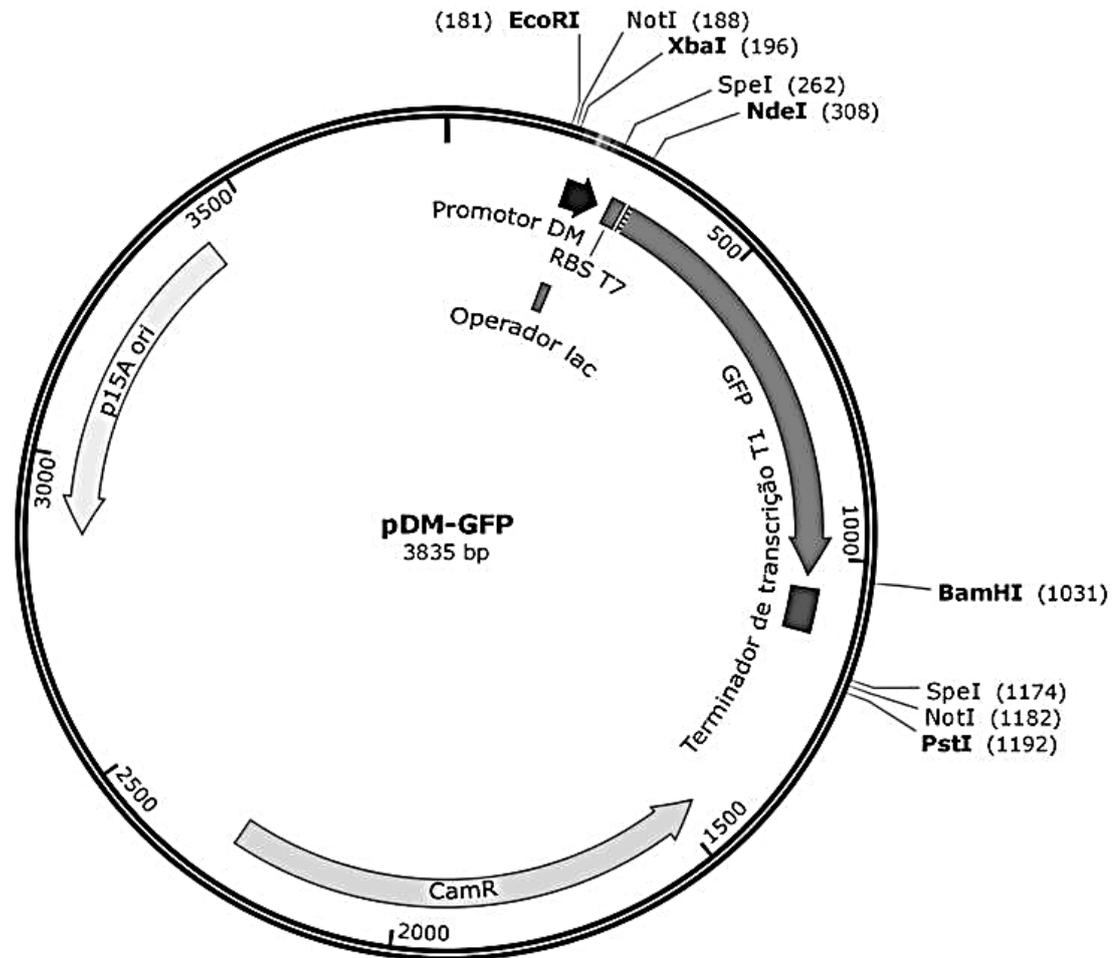


Figura 2: O mapa físico do vetor de expressão pDM-GFP. A sequência codificante do gene da proteína verde fluorescente (GFP) foi clonada no vetor pDM02 entre os sítios de *NdeI* (308) e *BamHI* (1031). O vetor pDM-GFP apresenta 3835 pares de bases.

Figura 3

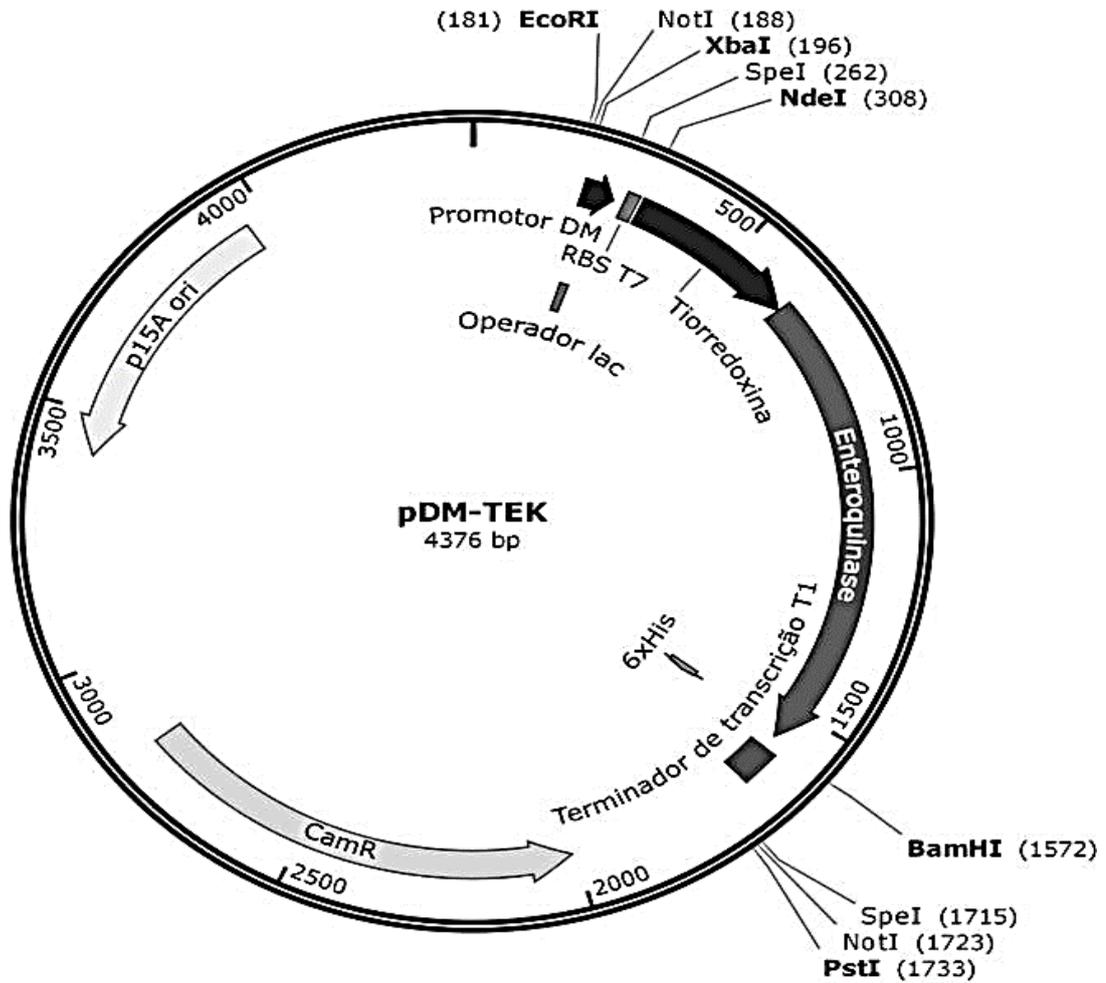


Figura 3: O mapa físico do plasmídeo recombinante pDM-TEK. O gene híbrido da tiorredoxina humana e a parte catalítica da enzima enteroquinase bovina, foi clonado no vetor pDM02 entre os sítios de *NdeI* (308) e *BamHI* (1572). O vetor pDM-TEK apresenta 4376 pares de bases.

Figura 4

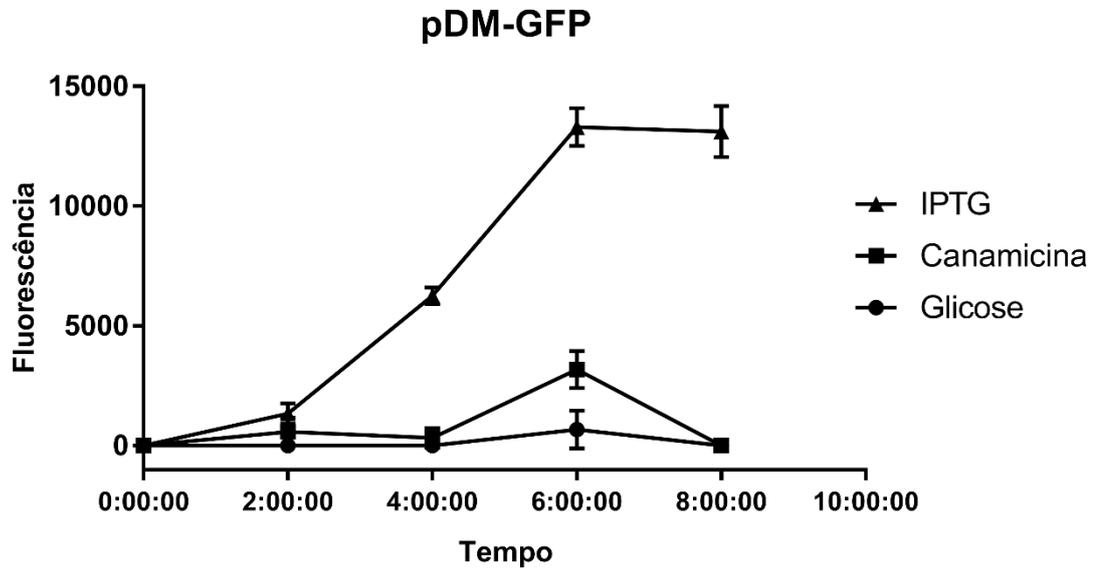


Figura 4: O perfil de expressão do gene da GFP, indicado pela fluorescência emitida pela proteína GFP em função do tempo de cultivo de *E. coli* JM110, programada pelo promotor sintético DMclonado no plasmídeo pDM-GFP. O processo de indução com IPTG ocorreu a partir do início do cultivo.

Figura 5



Figura 5: O resultado da análise por eletroforese em gel desnaturante de poliacrilamida 12% (SDS-PAGE) e fluorescência da produção de GFP por um clone de *E. coli* JM110 contendo o vetor pDM-GFP em diferentes condições de cultivo. Nas raias 1, 2 e 3, cultivo com canamicina e o indutor IPTG; raias 4, 5 e 6, com apenas canamicina; raias 7, 8 e 9 cultivo com glicose e canamicina. A raia M mostra o marcador molecular em kDa. A proteína recombinante GFP localiza-se na posição de aproximadamente 25 kDa (seta).

Figura 6

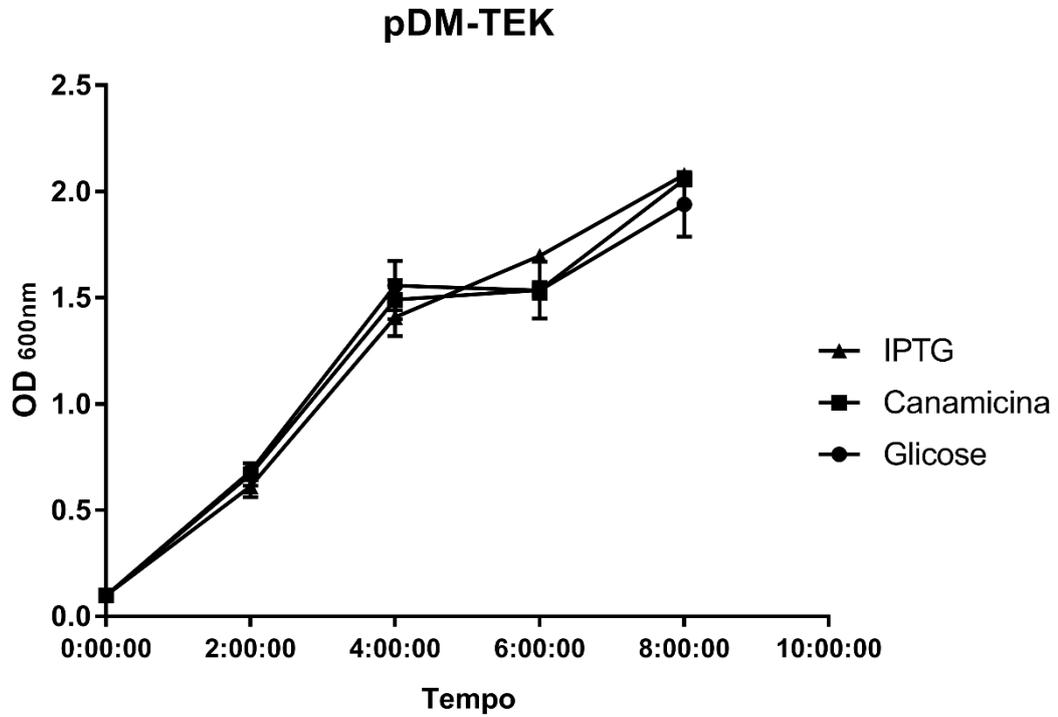


Figura 6: O perfil do crescimento celular (ABS. 600nm x tempo) de *E. coli* JM110 programada com o plasmídeo pDM-TEK. O processo de indução com IPTG ocorreu a partir do início do cultivo.

Figura 7

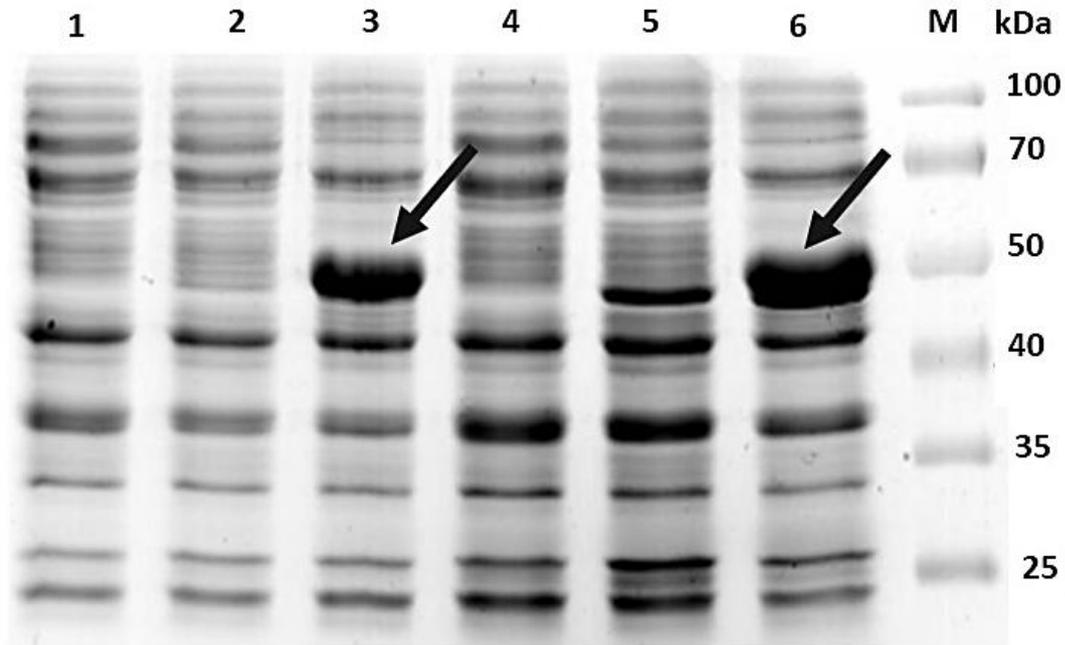


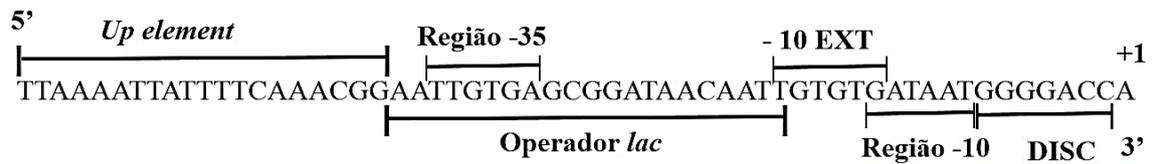
Figura 7: O perfil eletroforético em gel de poliacrilamida desnaturante [12%] de proteínas expressas por um clone de *E. coli* JM110 contendo o vetor pDM-TEK em diferentes condições de cultivo. Nas raias 1 e 4, cultivo com glicose e canamicina; raias 2 e 5, com apenas canamicina; raias 3 e 6, cultivo com canamicina e o indutor IPTG. Nas raias 1, 2 e 3 amostras com 4 horas de cultivo; nas raias 4, 5 e 6 amostras com 8 horas de cultivo. A raia M, mostra o marcador molecular em kDa. A proteína recombinante TEK localiza-se na posição de 50 kDa (seta).

**RESUMO****UM PROMOTOR SINTÉTICO PARA ALTA EXPRESSÃO DE GENES HETERÓLOGOSEM *Escherichia coli* EM PERFORMANCE BEM REGULADA**

A presente invenção refere-se à engenharia do novo promotor sintético DM, funcional no organismo modelo *Escherichia coli*, com performance de forte expressão e alto nível de regulação, construído com inspiração no sistema do promotor *lac*. Visando aprimorar sua regulação para aplicações industriais, o design foi realizado a partir da sobreposição da sequência operadora *laço* à região -35 de reconhecimento da subunidade  $\sigma 70$ , precisamente impedindo o início da transcrição pela RNA polímerase holoenzima. Esta estratégia possibilitou alto nível de expressão e regulação gênica, demonstrada pela análise quantitativa da síntese da *Green Fluorescent Protein* (GFP) e da proteína híbrida dationredoxina humana fusionada ao domínio catalítico da enteroquinase bovina sob controle do promotor sintético DM, nos vetores pDM-GFP e pDM-TEK. O presente promotor, clonado no versátil vetor de expressão pDM02, apresenta vasta aplicação biotecnológica, expandindo as ferramentas para construção de inovadores circuitos genéticos. Principalmente, a presente invenção abre caminho para expressão de genes de proteínas recombinantes de interesse industrial, inclusive as tóxicas, cuja elevada produção e minuciosa regulação tem sido um desafio.

## LISTAGEM DE SEQUÊNCIAS

## Sequência 1



Sequência 1: O novo promotor sintético regulado - DM. O promotor apresenta a sequência do *UP element* (TTAAAATTATTTCAAACGG), a região -35 (TTGTGA) sobreposta ao operador *lac* (AATTGTGAGCGGATAACAATT), a região -10 (GATAAT), o elemento -10 estendido (TGTGTG), o discriminador (GGGGACC) e o sítio de início da transcrição ou +1 composto por uma base nucleotídica adenina (A).

## Sequência2

<sup>EcoRI</sup>  
CGAATTCGCGGCCGCTTCTAGATTAAAATTATTTCAAACGGAATTGTGAGCGG  
<sub>NotI</sub>  
ATAACAATTGTGTGATAATGGGGACCATACTAGTACTAGAAATAATTTTGTTTAA  
<sup>HpaI</sup> <sup>KpnI</sup> <sup>PmeI</sup>  
CTTTAAGAAGGAGATATACATATGTTAACTTAAGGTACCAGATCTGTTTAAAC  
<sub>NdeI</sub> <sub>AflII</sub> <sub>BglII</sub>  
**GGATCC**ACTAGAGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACT  
<sub>BamHI</sub>  
GGGCCTTTCGTTTTATCTGTTGTTTGTTCGGTGAACGCTCTCTACTAGAGTCACA  
CTGGCTCACCTTCGGGTGGGCCTTCTGCGTTTATATACTAGTAGCGGCCGCTG  
<sub>SpeI</sub> <sup>NotI</sup> <sub>PstI</sub>  
CAGC  
<sub>PstI</sub>

Sequência 2: A sequência do cassete de expressão contido no vetor pDM02. A porção sublinhada mostra o promotor sintético regulado descrito na Sequência 1; a região em itálico identifica o sítio de ligação ao ribossomo; marcado em negrito estão sítios de enzimas de restrição (“*polilinker*”) e o terminador de transcrição sublinhado duas vezes. E nas extremidades estão os sítios de reconhecimento das enzimas de restrição *EcoRI*, *NotI*, *XbaI* e *SpeI*, *NotI* e *PstI*.

## Sequência3

<sup>NdeI</sup>  
 5' CATATGGCTAGCAAAGGAGAAGAAGCTTTTCACTGGAGTTGTCCAATTCTTGTTGAATTA  
 GATGGTGATGTTAATGGGCACAAATTTTCTGTTCAGTGGAGAGGGTGAAGGTGATGCTACA  
 TACGGAAAACCTTACCCTTAAATTTATTTGCACTACTGGAAAACCTACCTGTTCCATGGCCAA  
 CACTTGTCACTACTTTCTCTTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCACATGAA  
 ACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATC  
 TTTCAAAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCC  
 TTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTCGGACA  
 CAAACTGGAGTACAACCTATAACTCACACAATGTATACATCACGGCAGACAAACAAAAGAA  
 TGGAATCAAAGCTAACTTCAAATTCGCCACAACATTGAAGATGGCTCCGTTCAACTAGC  
 AGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCAT  
 TACCTGTCTACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTC  
 CTTCTTGAGTTTGTAAGTCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAATAA  
 GGATCC 3'  
<sup>BamHI</sup>

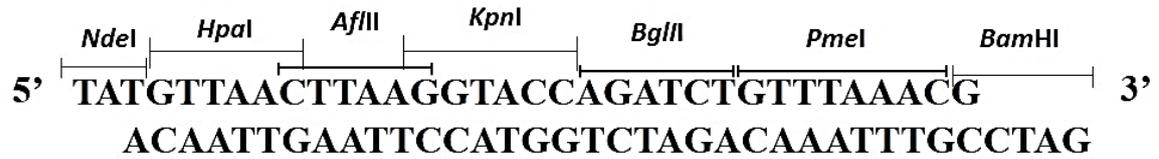
Sequência 3: A sequência codificante do gene da proteína verde fluorescente (GFP). Nas extremidades estão os sítios de reconhecimento das enzimas de restrição *NdeI* e *BamHI*. Esta sequência está inserida no vetor pDM-GFP.

## Sequência 4

<sup>*NdeI*</sup>  
CATATGAGCGATAAGATTATTCACCTGACTGACGACAGCTTTGACACGGATGTACTCAAAGCGGACGGCGGATCCTCGTCGATTTCTGGGCA  
GAGTGGTGCGGTCCGTGCAAGATGATCGCGCCGATTCTGGATGAAATCGTGACGAATATCAGGGCAAACTGACCGTTGCCGAAACTGAACATC  
GATCAGAACCCTGGCACTGCGCCGAAATATGGCATCCGTGGTATCCCGACTCTGCTGTGTTCAAGAACGGTGAAGTGGCGGCAACCAAAGT  
GGGTGCACTGTCTAAAGGTCAGTTGAAAGAGTTCCTCGACGCTAACCTGGCCGGTACCGGTAGTGGCCACATGGCGTCAGCACATGCCAGCA  
GCTCGGGCCTGGTGCCGCGGTTCAGGCATGAAAGAAACCCCGCGCAAAGTTGAGCGTCAGCACATGGATTCTCCTGATTAGGCAGC  
<sup>Sítio de reconhecimento ENTK</sup> **GACGATGACGATAAA**<sup>Sítio de reconhecimento ENTK</sup> **GACGATGACGATAAA**ATTGTGGCGGTAGCGATATCGCGAAGGCGCGTGGCCGT  
GGTGGTTGCGCTGTATTTGACGATCAGCAAGTGTGCGGTGCCAGCCTGGTATCGCGGATTGGTTAGTTTCAGCGGCCATTGCGTGTATGG  
CCGCAACATGGAACCGAGCAAATGGAAGGCAGTTCTGGGATTACACATGGCGTCTAACTTGACCAGCCCGCAGATTGAGACCCGCTGATCG  
ATCAGATCGTGATTAATCCGCACTACAACAAACGTCGCAAGAATAACGACATTGCCATGATGCACCTTCAAATGAAAGTGAACATAACAGATTA  
CATTAGCCGATATGTCTACCAGAGGAAAACAGGTGTTCCGCCTGGCCGTATTGACGATCGCTGGTGGGAGCGCTGATTATCAGGGC  
TCGACCGCCGACGTGTGCAAGAGGCAGATGTACCGCTCCTGAGCAATGAGAAATGTCAGCAACAGATGCCGGAGTATAACATTACGGAAAA  
CATGGTCTGCCGCGGCTACGAAGCCGGTGGGGTTGATAGTTGCCAAGGCGACTCAGGCGGTCCACTTATGTGTCAGGAAAATAACCGATGGTT  
GCTGGCTGGCGTAACTTCCTTCGGTTACCAATGTGCGCTCCCTAATCGTCCGGCGTCTATGCGCGTGTGCCGCGTTTACGGAATGGATTCAG  
TCTTTCTGCATGCGACCAGCGGCTCGGGTCATCACCATCACCATCACTAAATGAATAGGGATCC  
<sup>*His-tag*</sup> <sup>*BamHI*</sup>

Sequência 4: A sequência de parte do gene da tiorredoxina humana fusionado com a porção catalítica do gene da enteroquinase bovina (TEK). A sequência que está sublinhada corresponde a sequência de parte do gene da tiorredoxina. As sequências que estão delimitadas em negrito e com identificação por escrito (Sítio de reconhecimento ENTK) correspondem a dois sítios de reconhecimento da enzima enteroquinase. A sequência que está pontilhada corresponde ao domínio *His-tag* utilizado para purificação. A sequência que não está sublinhada corresponde à cadeia leve (catalítica) da enzima enteroquinase. E nas extremidades estão os sítios de reconhecimento das enzimas de restrição *NdeI* e *BamHI*. Esta sequência está inserida no vetor pDM-TEK.

## Sequência 5



Sequência 5: A sequência contendo múltiplos sítios de clonagem (*polylinker*) inserida no vetor pDM02. O *polylinker* contém sítios das seguintes enzimas de restrição: 5' *Nde*I, *Hpa*I, *Afl*III, *Kpn*I, *Bgl*III, *Pme*I e *Bam*HI 3'.

## Sequência 6

GGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAAATAGGCGTATCACGAGGCAGAATTCAGATAAA  
AAAAATCCTTAGCTTTTCGCTAAGGATGATTTCTGCGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAC  
CCGGGATCGAATTCGCGCCGCTTCTAGATTAATAATTTTCAAACGGAATTGTGAGCGGATAACAATTGTGTGATAATGGGGA  
CCATACTAGTACTAGAAATAATTTTGTAACTTTAAGAAGGAGATATACATATGTTAACTTAAGGTACCAGATCTGTTAAACGGA  
TCCACTAGAGCCAGGCATCAAATAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTGTGCGGTGAAC  
GCTCTCTACTAGAGTCACTGGCTCACCTTCGGGTGGGCCTTTCGCGTTTATATACTAGTAGCGGCCGCTGCAGCATCCTCGAG  
GTTCCCTTAGTGAGGGTAAATTGCGAGCTTGGCGTAATCATGGTCATCCGGCAAAAAAGGGCAAGGTGTACCACCCTGCCCTT  
TTCTTTAAAACCGAAAAGATTACTTCGCGTTATGCAGGCTTCCTCGCTACTGACTCGCTGCGCTCGGCTCGTTCGGCTGCGCG  
AGCGGTATCAGCTCACTCAAAGGCGGTAATCTCGAGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTACAACCAATTAACCA  
ATTCTGATTAGAAAACTCATCGAGCATCAAATGAACTGCAATTTATCATATCAGGATTATCAATACCATATTTTTGAAAAAGCC  
GTTTCTGTAATGAAGGAGAAAACCTACCGAGGCAGTTCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATCCGACTCGTCC  
AACATCAATAACCTATTAATTTCCCTCGTCAAAAATAAGGTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGT  
GAGAATGGCAAAAGCTTATGCATTTCTTCCAGACTTGTCAACAGGCCAGCCATTACGCTCGTCATAAAATCACTCGCATCAA  
CCAAACCGTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACCGCATCGCTGTAAAAGGACAATTACAAACAGGAATCGA  
ATGCAACCGCGCAGGAACACTGCCAGCGCATCAACAATATTTTACCTGAATCAGGATATTCTTAATACCTGGAATGCTGTTT  
TCCCGGGATCGCAGTGGTGAAGTACCATCATCAGGAGTACGGATAAAATGCTTGATGGTGGAAAGAGGCATAAATCCGT  
CAGCCAGTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAAACAACCTGCGCATCGG  
GCTTCCCATACAATCGATAGATTGTGCGACCTGATTGCCCGACATTATCGCGAGCCATTTATACCATATAAATCAGCATCCATGTT  
GGAATTAATCGCGCCCTCGAGCAAGACGTTTCCCGTTGAATATGGTCTATAACACCCCTGTATTACTGTTTATGTAAGCAGACA  
GTTTTATTGTTGATGATGATATAATTTTATCTTGTGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTGTTGAATAAATCG  
AACTTTGCTGAGTTGAAGGATCAGATCAGCATCTTCCCGACAACGCAGACCGTTCCGTGGCAAAGCAAAGTTCAAAATCAC  
CAACTGGTCCACCTACAACAAGCTCTCATCAACCGTGGTCCCTCACTTTCTGGCTGGATGATGGGGCGAATTCAGGCCTGGTAT  
GAGTCAGCAACCTTCTTACGAGGCAGACTCAGCGTAGCGAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGT  
CAGTGAAGTGCTTCATGTGGCAGGAGAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTGATACAGGATATATCCGCTTCT  
CGCTCACTGACTCGCTACGCTCGGTCGTTGACTGCGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTCTGGAAGATG  
CCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCTGACAAGCATCACGA  
AATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATAACCAGGCGTTTCCCTGGCGGCTCCCTCGTGGC  
TCTCTGTTCCTGCTTTCGGTTTACCGGTGTCATTCGCTGTATGGCCGCTTTGTCTCATTCCACGCTGACACTCAGTTCCG  
GGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGAACCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCT  
TGAGTCAACCCGAAAGACATGCAAAAGCACCCTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCAT  
GCGCCGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTACTGCGCTCCTCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAG  
CTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTTTAGAGCAAGAGATTACGCGCAGACCAAAACGATCTC  
AAGAAGATCATCTTATTAAGGGGTCTGACGCTCAGTGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAG  
GATCTTACCTAGATCCTTTAAATTAATAAATGAAGTTTTAAATCAATCTAAAAGTATATGAGTAAACTTGGTCTGACAGTTACCA  
ATGCTTAATCAGTGGACCTATCTCAGCGATCTGTCTATTTCGTTTCCATAGTTGCTGACTCCCCGCTGTTAGATAACTAC  
GATACGGGAGGGCTTACCATCTGGCCCCAGTGTGCAATGATACCGGAGACCCAGCTCACCAGGCTCCAGATTATCAGCAATA  
AACCAGCCAGCCGAAAGGGCCGAGCGCAGAAAGTGGTCTGCAACTTTATCCGCTCCATCCAGTCTATTCCAT

Sequência 6: A sequência completa do vetor de expressão pDM02.