



**UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA**

CATARINA ALFAYA MOREIRA

**DESENVOLVIMENTO DE PROTEÍNA SINTÉTICA QUIMÉRICA
BASEADA EM TALE (TRANSCRIPTION ACTIVATOR-LIKE
EFFECTOR) PARA APLICAÇÃO EM SISTEMA BIOSSENSOR DE
DNA**

Salvador
2017

CATARINA ALFAYA MOREIRA

**DESENVOLVIMENTO DE PROTEÍNA SINTÉTICA QUIMÉRICA
BASEADA EM TALE (TRANSCRIPTION ACTIVATOR-LIKE
EFFECTOR) PARA APLICAÇÃO EM SISTEMA BIOSENSOR DE
DNA**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Biotecnologia da Universidade Federal da Bahia como requisito parcial para obtenção do título de Mestre em Biotecnologia.

Orientador: Prof. Dr. Luis Gustavo Carvalho Pacheco

Salvador
2017

Alfaya Moreira, Catarina
Desenvolvimento de proteína sintética quimérica baseada em tale (transcription activator-like effector) para aplicação em sistema biossensor de DNA / Catarina Alfaya Moreira. -- Salvador, 2017.

72 f. : il

Orientador: Luis Gustavo Carvalho Pacheco.
Dissertação (Mestrado - Programa de Pós-Graduação em Biotecnologia) -- Universidade Federal da Bahia, Instituto de Ciências da Saúde, 2017.

1. Biologia Sintética. 2. Transcription Activator-Like Effector. 3. Biossensor de DNA. 4. Cromoproteína. I. Carvalho Pacheco, Luis Gustavo. II. Título.

AGRADECIMENTOS

Agradeço primeiramente aos meus pais, por todo amor e carinho, por sempre acreditarem e investirem no meu potencial.

Ao Prof. Dr. Luis Gustavo Carvalho Pacheco, pelo valioso tempo de orientação. Gostaria de agradecer imensamente por todo incentivo e conhecimento que me foi repassado ao longo desse tempo.

Hiago Soares, por todo amor, carinho, apoio e incentivo, sempre me lembrando que sou capaz de ir mais longe. Obrigada por estar sempre ali para me dar sempre uns empurrõezinhos e puxões de orelha!

A Eliezer Santana, colega que se tornou amigo para a vida. Obrigada pelas várias conversas sobre futuro.

Aos professores e colegas do LAA, por todo suporte, e ao grupo de pesquisa ao qual tive o prazer de fazer parte (*GenoFun*), obrigada por tudo. Especialmente ao Filipe Sampaio, que me ajudou bastante desde a graduação. Conte comigo!

A todos os colegas de graduação, pelas experiências compartilhadas. Nunca esquecerei!

Aos amigos do Ciências sem Fronteiras, que foram minha família durante o período que passei no Canadá, saudades!

*“Suba o primeiro degrau com fé. Não é necessário
que você veja toda a escada. Apenas dê o primeiro
passo.”*

Martin Luther King

RESUMO

Moreira, Catarina Alfaya. DESENVOLVIMENTO DE PROTEÍNA SINTÉTICA QUIMÉRICA BASEADA EM TALE (TRANSCRIPTION ACTIVATOR-LIKE EFFECTOR) PARA APLICAÇÃO EM SISTEMA BIOSSENSOR DE DNA. 72f 2017. Dissertação de Mestrado – Instituto de Ciências da Saúde, Universidade Federal da Bahia, Salvador, 2017.

As abordagens de Biologia Sintética possuem potencial de aumentar a produção de ferramentas diagnósticas adequadas para utilização na prática, como os biossensores de DNA. O controle dos sistemas biológicos permite o aprimoramento de tecnologias em uso e surgimento de novas tecnologias, como o diagnóstico molecular através da identificação de sequências específicas de DNA. A maioria dos métodos moleculares de diagnóstico de DNA se baseiam em hibridização. Contudo, atualmente, as proteínas ligadoras de DNA tem tido destaque, devido às suas numerosas vantagens. As principais vantagens são a identificação de sequências em DNA dupla-fita, eliminando o passo de desnaturação, e a alta especificidade pelo alvo. Neste contexto, o objetivo deste trabalho foi desenvolver uma proteína sintética quimérica para aplicação em sistema biossensor de DNA que permitirá rápida identificação de uma sequência específica de DNA de *Corynebacterium striatum*, um importante patógeno emergente. Foi construída, neste estudo, uma proteína quimérica biossensora composta por: (I) um domínio de ligação ao DNA derivado de proteínas TALEs (proteínas efetoras do tipo ativadores transcricionais) o qual reconhece especificamente uma sequência de 18 pb de um gene único de *C. striatum* codificador de uma permease para transporte de ferro; combinado a (II) uma cromoproteína tsPurple (Tinsel Purple) para permitir detecção visual da sequência alvo. Para construção do gene codificador da proteína TALE específica, o gene específico de *C. striatum* foi analisado com a ferramenta Tal Effector Nucleotide Targeter (Cornell University) e uma sequência-alvo de 18pb foi definida. A construção de domínios TAL foi realizada utilizando a ferramenta GeneArt™ Perfect Match TAL (Thermo Scientific); o gene sintético encontra-se clonado no plasmídeo pTALE. Para construir as sequências gênicas quiméricas, foram desenhados primers de PCR para isolamento da sequência da tsPurple com a adição de sítios artificiais de restrição para EcoRI e BamHI; os produtos de PCR digeridos foram clonados no plasmídeo pTALE, downstream à sequência da TALE específica. Essas sequências fusionadas foram então sub-clonadas em um vetor de expressão pDEST™17, através de recombinação utilizando tecnologia Gateway para expressão das proteínas quiméricas. A expressão da proteína quimérica foi induzida pela adição de 1mM de IPTG. A expressão da proteína quimérica foi confirmada através de análises de SDS PAGE 10% , que demonstraram a superexpressão da proteína esperada de 151kDa; além disso, foi observada a forte coloração roxa no precipitado bacteriano, em razão da funcionalidade da cromoproteína TsPurple.

Palavras-chave: Biologia Sintética; Biossensor; Proteína Efetora do tipo Ativador Transcricional; Cromoproteína; Corinebactérias.

ABSTRACT

Moreira, Catarina Alfaya. DEVELOPMENT OF A SYNTHETIC CHIMERIC PROTEIN TALE-BASED (TRANSCRIPTION ACTIVATOR-LIKE EFFECTOR) FOR APPLICATION IN BIOSENSOR SYSTEM TO IDENTIFY DNA SEQUENCES. 72f 2017. Master Dissertation – Instituto de Ciências da Saúde, Universidade Federal da Bahia, Salvador, 2017.

Synthetic Biology approaches have the potential to increase the production of diagnostic tools suitable for use in practice, such as DNA biosensors. The control of biological systems allows the improvement of technologies in use and the advent of new technologies, such as molecular diagnosis through the identification of specific DNA sequences. Most molecular methods of DNA diagnosis rely on hybridization. However, currently, as DNA binding proteins has been studied due to its advantages. Some significant advantages are: the identification of double-stranded DNA sequences, avoiding the denaturation step, and high target specificity. In this context, the aim of this work was to develop a DNA biosensor system that allows rapid identification of a specific DNA sequence of *Corynebacterium striatum*. In this study was designed a chimeric biosensor protein composed of: (I) a DNA binding domain derived from TALE proteins (transcriptional activator-like effector proteins) which specifically recognizes an 18 bp sequence of a single gene from *C. Striatum* encoder of a permeability for iron transport; Combined with (II) a chromoprotein tsPurple (Tinsel Purple) to allow visual detection of the target sequence. For the construction of the gene encoding the specific TALE protein, the specific gene of *C. striatum* was analyzed with a Tal Effector Nucleotide Targeter tool (Cornell University) and a target sequence of 18bp was defined. The construction of TAL domains was performed using a GeneArt™ Perfect Match TAL (Thermo Scientific) tool; The synthetic gene was cloned in the pTALE plasmid. To construct the chimeric gene sequences, PCR primers were designed for isolation of the tsPurple sequence with an addition of artificial restriction sites to EcoRI and BamHI; The digested PCR products were cloned on pTALE, downstream to the specific TALE sequence. These fused sequences were then subcloned into a pDEST™17 expression vector, using Gateway technology for expression of the chimeric proteins. Expression of the chimeric protein was induced by the addition of 1mM IPTG. The chimeric protein expression was confirmed by analyzes of 10% SDS PAGE, which demonstrated the overexpression of an expected protein of 151kDa; In addition, the strong purple color was observed in the bacterial precipitate, due to the functionality of the chromoprotein TsPurple.

Keywords: Synthetic Biology; Biosensor; Transcriptional Activator-like Protein; Chromoprotein; Corinebacteria

LISTA DE FIGURAS

Figura 1 – Hierarquia da Biologia Sintética inspirada na Engenharia Computacional.

Figura 2 – Diversas aplicações da Biologia Sintética.

Figura 3 – Visão geral dos métodos diagnósticos de DNA.

Figura 4 – Representação esquemática do funcionamento de biosensores.

Figura 5 – Metodologias possíveis de serem aplicadas para criação de um biossensor.

Figura 6 – Interação dos domínios de Zinc Fingers com sequências de DNA.

Figura 7 – Representação dos domínios presentes numa proteína TALE de origem natural.

Figura 8 – Apresentação das RVDs padrão e os nucleotídeos com os quais elas especificamente interagem.

Figura 9 – Estrutura tridimensional obtida por cristalografia de raio-x da TALE PthXo1 do patógeno de arroz *Xanthomonas oryzae*.

Figura 10 – Imagem de uma placa com colônias de *Escherichia coli* transformadas com plasmídeo contendo a cromoproteína Tinsel Purple.

Figura 11 – Fluxograma da estratégia experimental adotada.

Figura 12 – Plasmídeo pTALE desenhado através da ferramenta DNA ATLAS do site ATUM.

Figura 13 – Representação esquemática da digestão do pTALE e consequente abertura do sítio múltiplo de clonagem.

Figura 14 – Mapa do plasmídeo BBa_K1357008.

Figura 15 – Representação da amplificação por PCR da sequência gênica *tsPurple* e adição dos sítios de restrição EcoRI (primer forward) e BamHI (primer reverse).

Figura 16 – Representação esquemática da digestão dupla do fragmento de PCR contendo o gene *tsPurple*.

Figura 17 – Representação esquemática da ligação dos produtos das digestões do pTALE e *tsPurple* com EcoRI e BamHI através das suas extremidades coesivas.

Figura 18 – Representação esquemática da reação de recombinação entre os plasmídeos pTSP1 e o pDESTTM17.

Figura 19 – Mapas dos plasmídeos pTALE, pDESTTM17TM, pTSP1 e pTSP2.

Figura 20 – Resultado obtido da ferramenta TAL Effector Targeter da Cornell University.

Figura 21 – Plasmídeo pTALE obtido a partir do kit de extração de DNA plasmidial *NucleoSpin® Plasmid* (Macherey-Nagel).

Figura 22 – Resultado das digestões do pTALE com as enzimas EcoRI e BamHI.

Figura 23 – Gel de agarose com amostras da segunda digestão simples do plasmídeo pTALE.

Figura 24 – Extração de DNA plasmidial do plasmídeo BBa_K1357008, contendo a sequência gênica da cromoproteína Tinsel Purple

Figura 25 – Resultado da amplificação por PCR do fragmento contendo o gene *tsPurple* a partir do plasmídeo BBa_K1357008.

Figura 26 – Resultado do gel de agarose 1% com o resultado do PCR convencional das amostras de ligação dos fragmentos digeridos com os primers *tsPurple* forward e *tsPurple* reverse.

Figura 27 – Resultado do gel de agarose 1% com o resultado do PCR convencional das amostras de ligação dos fragmentos digeridos com os primers T7 forward e *tsPurple* reverse.

Figura 28 – Resultado do gel de agarose 1% com o resultado do PCR convencional das amostras de ligação dos fragmentos digeridos e das primeiras recombinações.

Figura 29 – Fotos das placas de LB-ágar com ampicilina (à esquerda) e kanamicina (à direita) para seleção dos clones positivos para as recombinações.

Figura 30 – Reação de PCR convencional utilizando os primers T7 forward e *tsPurple* reverse para confirmação das recombinações.

Figura 31 – Fotos dos *pellets* das culturas durante indução da expressão da proteína quimérica biossensora.

Figura 32 – Sequência de aminoácidos da proteína quimérica biossensora, ponto isoelétrico e peso molecular estimado.

Figura 33 – SDS-PAGE 10% das amostras de indução dos clones 10 (à esquerda) e 19 (à direita).

Figura 34 – Predição da estrutura tridimensional da proteína TALE (azul) e da proteína resultante da fusão da TALE com a Tinsel Purple (vermelho).

Figura 35 – Teste de solubilidade das amostras de indução dos clones 10 e 19, obtidos após recombinação.

LISTA DE TABELAS

Tabela 1 – Genes repórter mais comumente utilizados.

Tabela 2 – Linhagens de *E. coli* selecionadas para o estudo e seus respectivos genótipos.

Tabela 3 – Plasmídeos utilizados no presente estudo, suas respectivas origens de replicação e marcadores de resistência.

Tabela 4 – Reação de digestão para o pTALE e fragmento amplificado por PCR do gene *tsPurple*.

Tabela 5 – Reação de PCR convencional realizado com o kit Pht para isolamento da sequência codificadora da cromoproteína Tinsel Purple e adição dos sítios de restrição

Tabela 6 – Reação de ligação realizada utilizando a enzima T4 DNA ligase da *New England Biolabs*.

Tabela 7 – Reação de PCR para confirmação das ligações e recombinações.

LISTA DE ABREVIATURAS

Amp^r – Resistência a Ampicilina

CAT – Cloranfenicol acetiltransferase

Cm^r – Resistência a Cloranfenicol

CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats

DBP – DNA Binding Protein

DNA – Deoxyribonucleic acid (Ácido Desoxirribonucléico)

dsDNA – Ácido Desoxirribonucléico dupla fita

FISH – Fluorescent in situ Hybridization

GFP – Green Fluorescent Protein

gRNA – Ácido Ribonucléico guia

IGEM – International Genetically Engineered Machine

IPTG – Isopropyl β -D-1-thiogalactopyranoside

IUPAC – International Union of Pure and Applied Chemistry

Kan^r – Resistência a Kanamicina

LB – meio de cultura Lúria Bertani

LFNAB – Lateral Flow Nucleic Acid Binding

MCS – Multiple Cloning Site (Sítio Múltiplo de Clonagem)

ml – mililitro

ng - nanograma

NTR – N-terminal Region (Região N-terminal)

PAM – Motivo Adjacente ao Protoespaçador

pb – pares de bases

PCR – Reação em Cadeia da Polimerase

PNA – Peptidic Nucleic Acid (Ácido nucléico peptídico)

RBS – Ribosome Binding Site (Sítio de Ligação Ribossomal)

RNA – Ácido Ribonucléico

RVD – Repeat-Variable Diresidue

SDS PAGE – Sodium Dodecyl Sulfate - PolyAcrylamide Gel Electrophoresis

ssDNA –Ácido Desoxirribonucléico simples fita

ssRNA – Ácido Ribonucléico simples fita

TALE – Transcription Activator-like Effector

TB – Tuberculose

X-GAL

XSMA – *C. xerosis*, *C. striatum*, *C. minutissimum*, *C. amycolatum*

ZFP – Zinc Finger Protein

ZnF – Zinc Finger

µg – micrograma

µl – microlitro

SUMÁRIO

INTRODUÇÃO	15
1 FUNDAMENTAÇÃO TEÓRICA	18
1.1 Biologia Sintética	18
1.2 Métodos diagnósticos para identificação de sequências de DNA.....	21
1.1 Biossensores para diagnóstico pela identificação de sequências de DNA	22
1.4 Proteínas ligadoras de DNA.....	25
1.4.1 Zinc Fingers	25
1.4.2 Sistema CRISPR-Cas	27
1.4.3 Proteínas Efetoras do tipo Ativadores Transcricionais (Transcripton Activator-Like Effectors – TALEs).....	28
1.6 Genes reporter: Cromoproteínas	32
1.7 Corinebactérias emergentes e seu diagnóstico.....	34
2 OBJETIVOS	35
2.1 Objetivo geral.....	35
2.2 Objetivos específicos	35
3 METODOLOGIA	36
3.1 Estratégia experimental	36
3.2 Linhagens de Escherichia coli presentes no estudo.....	37
3.3 Plasmídeos utilizados para a estratégia de clonagem.....	37
3.4 Estratégia de clonagem para obtenção da proteína quimérica biossensora.....	39
3.4.1 Seleção da sequência alvo para construção da Proteína TALE específica	39
3.4.2 Clonagem e digestão do plasmídeo contendo proteína TALE	39
3.4.3 Obtenção do plasmídeo contendo sequência gênica da cromoproteína Tinsel Purple	42
3.4.4 Isolamento da sequência gênica tsPurple por PCR convencional e digestão dupla	43
3.4.5 Fusão das sequências TALE e tsPurple para obtenção da proteína quimérica biossensora	45
3.4.6 Recombinação dos plasmídeos pTSP1 e pDEST™ 17 pelo sistema de clonagem Gateway	47
3.4.7 Expressão da proteína quimérica biossensora.....	50
4 RESULTADOS E DISCUSSÃO	51
4.1 Desenho e construção do plasmídeo com sequência gênica da proteína TALE específica	51

4.2 Obtenção da sequência gênica da cromoproteína Tinsel Purple.....	53
4.3 Isolamento da sequência gênica tsPurple por PCR convencional e digestão dupla .	54
4.4 Ligação do plasmídeo pTALE e fragmento contendo o gene tsPurple	55
4.5 Recombinação dos plasmídeos pTSP1 e pDEST™17 pelo sistema de clonagem Gateway	56
4.6 Expressão da proteína quimérica biossensora.....	59
5 CONSIDERAÇÕES FINAIS.....	65
REFERÊNCIAS	66
APÊNDICE I – SEQUÊNCIA DO GENE <i>efeU</i> DE <i>Corynebacterium striatum</i>	70
APÊNDICE II – SEQUÊNCIA GÊNICA DA CROMOPROTEÍNA TINSEL PURPLE	71
APÊNDICE III – SEQUÊNCIA DOS PRIMERS UTILIZADOS NO ESTUDO	72

INTRODUÇÃO

A Biologia Sintética, aliada a outras áreas médicas e científicas tem enfrentado desafios, como a produção de métodos diagnósticos de baixo custo e de aplicação rápida e simples. As abordagens de Biologia Sintética possuem potencial de impulsionar, significativamente, o aumento de tecnologias promissoras em ferramentas diagnósticas adequadas para utilização na prática, como os biossensores de DNA. Dispositivos baseados em papel estão prontos para desempenhar um papel proeminente em aplicações *in vitro*, pois se encaixam à necessidade de ferramentas de diagnóstico de baixo custo, práticas e simples para uso fora do laboratório. Outro exemplo são os sensores *in vivo* existentes, que já começaram a transição de seus atuais modelos de pesquisa para aplicativos para monitoramento em tempo real e diagnóstico precoce e personalizado de pacientes (SLOMOVIC; PARDEE; COLLINS, 2015).

Um diagnóstico precoce é capaz de reduzir a mortalidade e economizar nos gastos com tratamentos (WEI; CHENG, 2016). Segundo Shorr e colaboradores (2007), a aplicação de métodos para diagnósticos precoce no ano de 2007, poderia ter evitado cerca de 92 mil mortes por sepse e ter promovido uma economia de pelo menos 1,5 bilhões de dólares dos custos em saúde nos Estados Unidos. Estudos mais recentes apontam que só em 2011, o custo total de tratamentos de sepse, nos Estados Unidos, foi de 20,3 bilhões (KHWANNIMIT; BHURAYANONTACHAI, 2015).

A identificação de um agente infeccioso por métodos moleculares, por exemplo, através de sua sequência de DNA é um método mais acurado e até muitas vezes mais rápido que os métodos convencionais de cultura, análise bioquímica, dentre outros. Assim como qualquer tecnologia de diagnóstico, os diagnósticos de DNA também requerem um método de detecção e um transdutor de sinal. Grande parte dos métodos de detecção atuais de reconhecimento de uma sequência específica de DNA se baseia no DNA de cadeia simples (ssDNA), para pareamento das bases com alta especificidade com uma molécula complementar. Essa outra molécula pode ser outro ssDNA, ssRNA, ácido nucléico peptídico (PNA), ou um análogo molecular de pareamento de bases. O reconhecimento específico ou hibridização é a base de tecnologias comuns como a amplificação por PCR, *Southern blot*, *Northern blot*, *microarray* de DNA e hibridização fluorescente *in situ* (FISH). Porém, existem outras formas de identificar sequências alvo sem a necessidade do ssDNA. Um exemplo são as poliamidas, pequenos compostos químicos que podem ser produzidos para se ligarem com alta especificidade a sequência alvo no sulco menor do DNA dupla fita (dsDNA) (GHOSH et al., 2006).

Nesse contexto, as proteínas ligadoras de DNA (DBPs) surgiram nos últimos anos como uma alternativa mais rápida, simples e, potencialmente, com melhor custo-benefício na área diagnóstica (STRAUS; LAHAYE, 2013). As principais proteínas ligadoras de DNA são as Zinc Fingers, CRISPR (Repetições Palindrômicas Curtas Agrupadas e Regularmente Interespaçadas) e TALE (Proteínas efetoras do tipo ativador transcricional). As primeiras proteínas sintéticas produzidas para detecção de novos alvos de DNA foram baseadas nos domínios de zinc finger do tipo Cys2His2 (KABADI; GERSBACH, 2014). Em 2012, foi desenvolvida, pela primeira vez, uma versão do sistema CRISPR/Cas9 para direcionamento de proteínas para sequências alvo de DNA por meio de interações RNA-DNA, ao invés de interações proteína-DNA. Foi descrito pela primeira vez em 2009 como os domínios das TALEs são capazes de reconhecer alvos de DNA. A partir de então, ficou conhecido um novo domínio modular mais facilmente reprogramável do que domínios Zinc Fingers, capaz de se ligar a um DNA alvo. Devido ao seu código de reconhecimento mais simples, muitos estudos tem investido em TALEs sintéticas para modular expressão gênica e têm obtido muito sucesso (KABADI; GERSBACH, 2014).

As DBPs são capazes de identificar sequências em DNA dupla-fita. A relevância disso é que raramente o DNA é encontrado na forma de fita simples, e a necessidade de equipamentos como termocicladores encarecem a proposta de diagnóstico simples e barato para locais de difícil acesso à tecnologia (GHOSH et al., 2006; KIM et al., 2011). A produção de proteínas efetoras do tipo ativadores transcricionais (TALEs) sintéticas é uma área de pesquisa que tem crescido bastante nos últimos anos. As TALEs possuem grande potencial, uma vez que são capazes de se ligar a dupla-fita de DNA, permitindo que se liguem diretamente aos cromossomos dentro do núcleo das células alvo, permitindo o uso para diagnóstico *in vivo* (POLITZ; COPELAND; PFLEGER, 2013). Sabe-se muita coisa a respeito dessas proteínas, porém muitos detalhes ainda precisam ser estudados para melhor esclarecer a natureza biofísica da interação específica das TALEs com os nucleotídeos da sequência alvo. De forma geral, são inúmeras as atuais e possíveis aplicações futuras para essa classe de proteína ligadora de DNA (BOGDANOVE; VOYTAS, 2011).

Não há na literatura, até o momento, nenhum estudo utilizando proteínas TALE para fins diagnósticos e identificação de patógenos. Por isso, essa proteína foi escolhida para ser estudada a sua aplicação em sistema biossensor de DNA de agentes patogênicos. Como prova de conceito, selecionamos a detecção da bactéria *Corynebacterium striatum*, que é considerada atualmente como um patógeno emergente multiresistente a drogas de alta relevância em ambientes hospitalares (HAHN et al., 2017), cujo gene-alvo para identificação específica, o gene *efeU*, foi recentemente descrito pelo nosso grupo de pesquisa na Universidade Federal da Bahia (SANTOS, 2016). Como método de detecção visual da

resposta do biossensor, foi escolhida a cromoproteína Tinsel Purple, que dispensa o uso de equipamentos sofisticados para leitura dos resultados.

1 FUNDAMENTAÇÃO TEÓRICA

1.1 BIOLOGIA SINTÉTICA

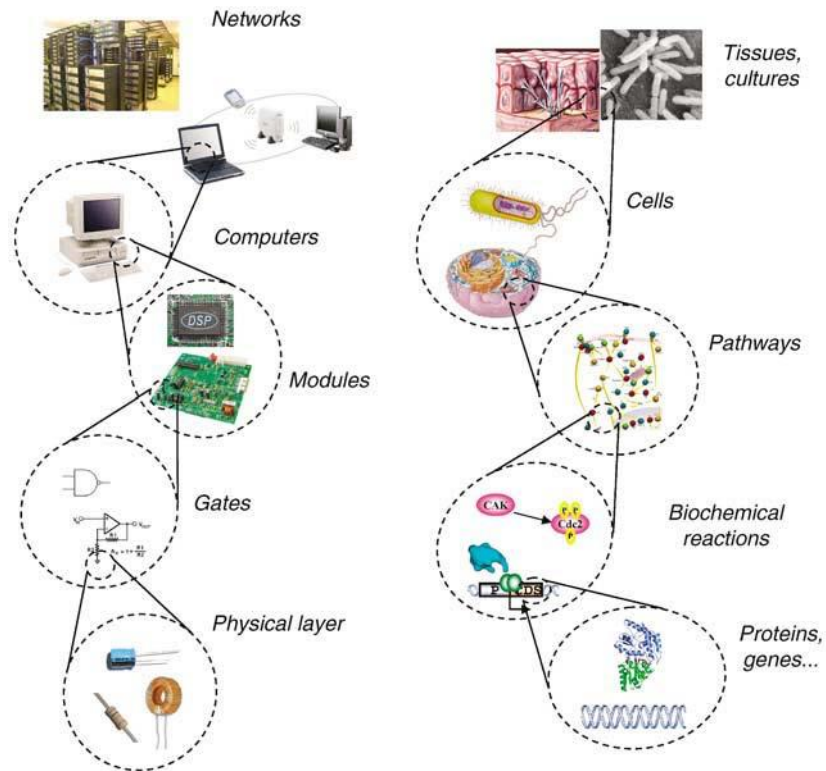
Com o advento da tecnologia do DNA recombinante, o uso de organismos vivos tornou-se parte dos processos de produção de uma diversidade de biomateriais. Há anos, a indústria biotecnológica vem explorando essa tecnologia com grande sucesso. Porém, na maioria dos casos, essa tecnologia se baseia em produção por “tentativa e erro” (COOLING et al., 2010).

A Biologia Sintética é uma área que tem crescido consideravelmente, gerando resultados importantes e grandes expectativas no intervalo de uma década, o que fez com que fosse de fato reconhecida como mais um ramo de pesquisa no grande ramo da Biologia (CAMERON; BASHOR; COLLINS, 2014). A criação dessa área se deu pela necessidade de simplificar e compreender sistemas biológicos (vírus, bactérias, eucariotos). Com essa finalidade, pesquisadores constroem/reconstroem sistemas bioquímicos e genéticos, *in vivo* e *in silico*, e criam modelos computacionais (utilizando a Bioinformática) para mimetizar o comportamento dessas vias de forma que seja possível chegar a uma aplicação real (PADILLA-VACA; ANAYA-VELÁZQUEZ; FRANCO, 2015). De forma geral, o principal objetivo da Biologia Sintética é tornar os sistemas biológicos mais previsíveis, eficientes, sistemáticos e robustos (WANG; WEI; SMOLKE, 2013). Para tanto, pesquisadores trabalham na tentativa de gerar organismos ou sistemas biológicos recodificados geneticamente a fim de que se tornem mais previsíveis e sistemáticos, produzam resultados mais robustos, sejam capazes de serem aplicados em diferentes escalas de trabalho ou produção e sejam mais eficientes que os sistemas ou organismos selvagens (PADILLA-VACA; ANAYA-VELÁZQUEZ; FRANCO, 2015).

Muitos pesquisadores fazem uma analogia da Biologia Sintética com a Engenharia Computacional para conceituar a sua hierarquia e metodologia (Figura 1). De fato, cada componente ou parte individual constitui um bloco ou dispositivo mais complexo. O design de uma nova característica ou comportamento desejado é visualizado ou idealizado do topo (*top-bottom*), ou seja, do(s) módulo(s) com todas as partes unidas. Porém, para sintetizar o(s) módulo(s) é necessário começar de baixo para cima (*bottom-up*), parte por parte. Fazem parte da base dessa hierarquia o DNA, RNA, proteínas e metabólitos que seriam análogos aos resistores, capacitores e camadas físicas dos transistores na Engenharia Computacional. A próxima camada na hierarquia seria o dispositivo, que compreende a regulação do fluxo de informação (reações bioquímicas), que na Computação seria o equivalente às portas lógicas engunheiradas. Já na camada do módulo, os dispositivos são agrupados de forma lógica para

formar uma via complexa ou circuitos integrados. Da mesma forma nos computadores, com os dispositivos agrupados formando módulos, esses módulos podem ser conectados para dar origem a circuitos integrados ou redes (ANDRIANANTOANDRO et al., 2006).

Figura 1 – Hierarquia da Biologia Sintética inspirada na Engenharia Computacional.



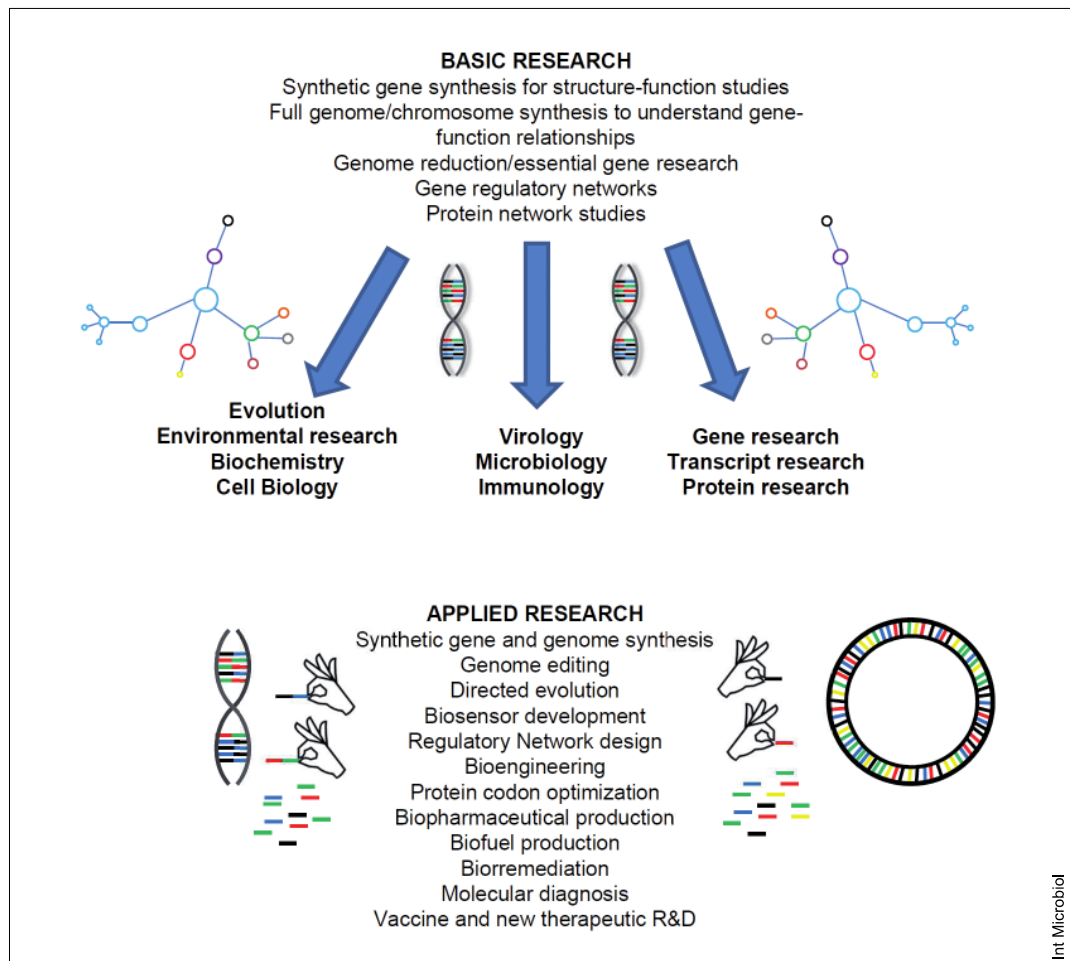
Fonte: (ANDRIANANTOANDRO et al., 2006)

A Biologia Sintética possui inúmeras aplicações, como pode ser observado na Figura 2. Atualmente existe uma grande quantidade de informação que pode ser obtida através da genômica, transcriptômica, proteômica, e metabolômica, porém é necessário tempo, esforço e *know-how* bem maior para compreensão dos dados. Já a busca direta da informação desejada torna o trabalho mais simples. Por exemplo, se é desejado avaliar e modular o comportamento de uma via metabólica, com a Biologia Sintética é possível mimetizar a via metabólica através da construção de circuitos genéticos, *in silico* e *in vivo* (PADILLA-VACA; ANAYA-VELÁZQUEZ; FRANCO, 2015).

Há vários anos, microorganismos vêm sendo utilizados para uma infinidade de aplicações biotecnológicas, como por exemplo, produção de fármacos. Contudo, com a capacidade de “engenheirar” um sistema biológico é possível aprimorar as aplicações conhecidas e torná-las mais eficientes, bem como criar novas possibilidades e novas

aplicações (ANDRIANANTOANDRO et al., 2006). KOBAYASHI e colaboradores, em 2004, produziram um sistema que acopla genes regulatórios com módulos biosensores, para que seja possível detectar dano ao DNA, com formação de um biofilme como resposta. A exposição transiente das células engenheiradas à luz ultravioleta causava dano ao DNA e induzia a formação dos biofilmes.

Figura 2 – Diversas aplicações da Biologia Sintética.

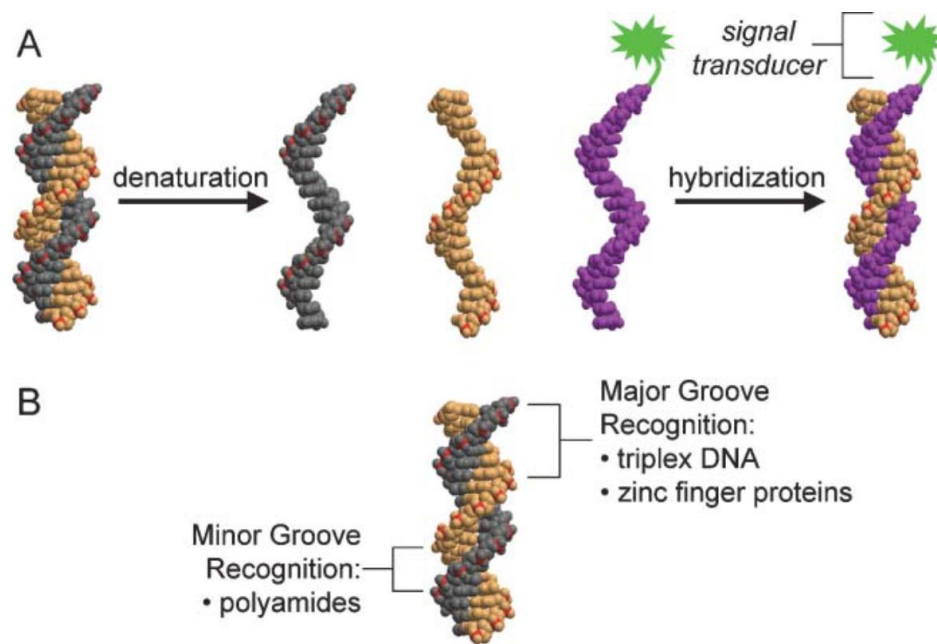


Fonte: (PADILLA-VACA; ANAYA-VELÁZQUEZ; FRANCO, 2015)

1.2 MÉTODOS DIAGNÓSTICOS PARA IDENTIFICAÇÃO DE SEQUÊNCIAS DE DNA

Diagnóstico molecular baseado na identificação de sequências de DNA contribuem continuamente para os avanços da pesquisa e medicina. A identificação de um agente infeccioso, por exemplo, através de sua sequência de DNA é um método menos subjetivo, mais acurado e até mesmo mais rápido que métodos convencionais de cultura. Assim como qualquer outro método diagnóstico, o diagnóstico baseado no reconhecimento de sequências de DNA necessita de um meio de detecção e um meio de transmissão de sinais. A maior parte dos meios de detecção de sequências específicas de DNA utilizam uma fita simples DNA (ssDNA), ou dupla-fita de DNA desnaturada, para parear com alta especificidade à uma molécula complementar, seja ela DNA, RNA ou proteína. O pareamento específico, ou hibridização, é a base das metodologias comuns como a amplificação por PCR (Reação em Cadeia da Polimerase), *Southern blot*, *Northern blot*, microarranjo de DNA (*microarray*) e *Hibridização Fluorescente in situ*, ou FISH (GHOSH et al., 2006). O método de PCR tem sido considerado muito mais rápido e sensível para diagnóstico de patógenos quando comparado com métodos padrão-ouro (por exemplo, cultura bacteriana). Entretanto, por muitas vezes é necessário um diagnóstico de múltiplos patógenos simultaneamente. O PCR multiplex é capaz de detectar diversos genes numa reação única, porém a técnica de microarranjos consegue detectar um número muito maior de sequências de uma só vez, apesar de necessitar de uma sonda para cada novo alvo e ser mais demorado. Além disso, todos os métodos citados acima necessitam do passo de desnaturação do DNA dupla-fita, implicando na necessidade de um termociclador. Nesse sentido, as proteínas ligadoras de DNA (DBPs) surgem como uma alternativa mais rápida, simples e, potencialmente, com maior custo-benefício, uma vez que podem identificar sequências em DNA dupla-fita. Isso é importante, pois raramente o DNA é encontrado na forma de fita simples, seja naturalmente ou após amplificação por PCR (GHOSH et al., 2006; KIM et al., 2011).

Figura 3 – Visão geral dos métodos diagnósticos de DNA.

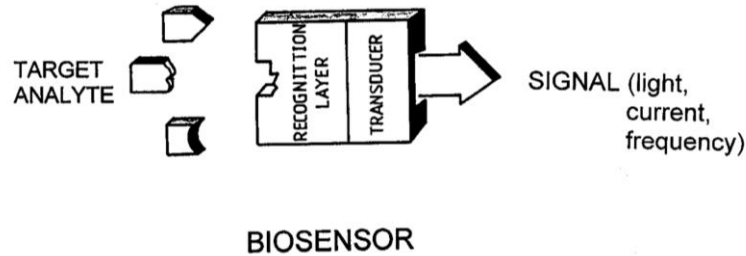


Fonte: (GHOSH et al., 2006)

1.3 BIOSSENSORES PARA DIAGNÓSTICO PELA IDENTIFICAÇÃO DE SEQUÊNCIAS DE DNA

Há um grande interesse na produção de biossensores, tanto na área de pesquisa quanto no âmbito comercial (TELES; FONSECA, 2008). De acordo com a nomenclatura IUPAC, biossensores são definidos como "dispositivos que utilizam reações bioquímicas específicas mediadas por enzimas isoladas, sistema imune, tecidos, organelas ou células inteiras para detectar compostos químicos" (NIC et al., 2005). Biossensores são pequenos dispositivos capazes de detectar, por meio de reações biológicas, um determinado analito alvo. De modo geral, biossensores compreendem dois a três módulos: um sensor, um computador (opcional) e um repórter (Figura 4).

Figura 4 – Representação esquemática do funcionamento de biosensores. O dispositivo biosensor é capaz de detectar o alvo e transmitir o sinal que pode ser emissão de luz, por exemplo.



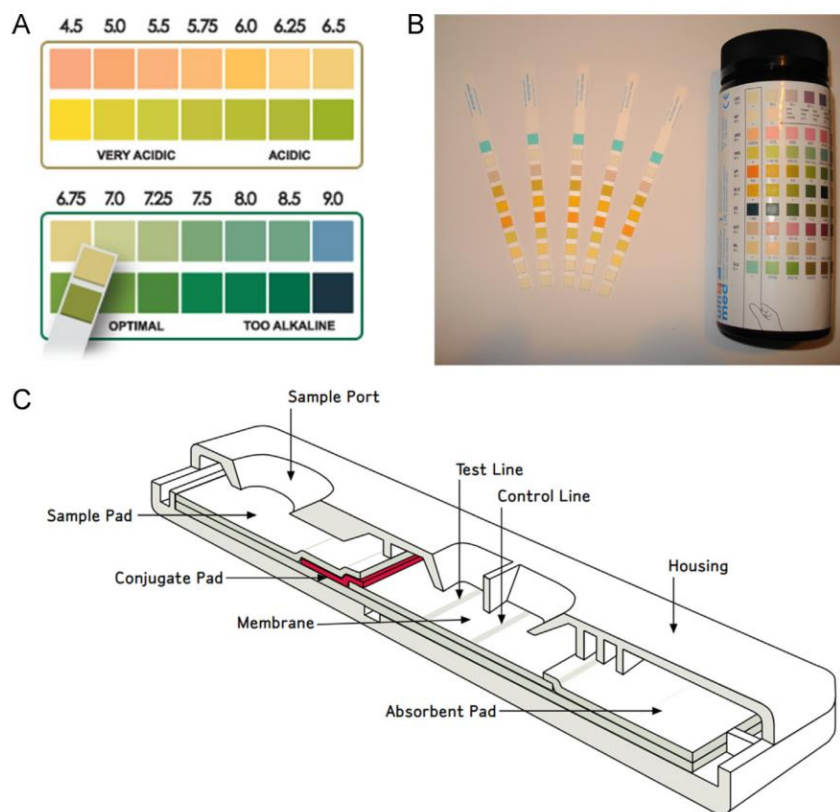
Fonte: (WANG, 2000)

Os biossensores de células inteiras, por exemplo, incorporam uma proteína sensora que busca o ligante de interesse e é acoplado com um sistema regulador genético para produzir um sinal capaz de ser quantificado, tal como a fluorescência (DONALDSON & DATTELBAUM, 2014). Além de células inteiras, moléculas, anticorpos e enzimas também podem ser usados em sistemas biossensores (RAUT et al., 2012; COURBET et al., 2015). Quando combinados com novas tecnologias, como à base de papel, em chip, microfabricação e/ou biomateriais dinâmicos, os sistemas biossensores possuem grande potencial como ferramentas portáteis de diagnóstico em pontos de atendimento (COURBET, 2015)

Métodos moleculares para identificação de sequências de DNA têm desempenhado papel importante para o diagnóstico médico, controle de qualidade alimentar e até mesmo monitoramento ambiental. A grande questão é o alto custo dos equipamentos e profissionais experientes que essas técnicas moleculares requerem. Na tentativa de oferecer uma alternativa para países em desenvolvimento, dispositivos portáteis (biossensores) têm sido cada vez mais estudados. Esses dispositivos podem ser a chave para um diagnóstico ou identificação precisa, eficiente e de baixo custo (ELLERBEE et al., 2009; YAGER et al., 2006; YING et al., 2017). As abordagens de Biologia Sintética possuem potencial de impulsionar tecnologias promissoras em ferramentas diagnósticas adequadas para utilização na prática, como os biossensores de DNA. Dispositivos baseados em papel estão prontos para desempenhar um papel proeminente em aplicações *in vitro*, pois se encaixam à necessidade de ferramentas de diagnóstico de baixo custo, práticas e simples para uso fora do laboratório. Outro exemplo são os sensores *in vivo* existentes, que já começaram a transição de seus atuais modelos de pesquisa para aplicativos para monitoramento em tempo real e diagnóstico precoce e personalizado de pacientes (SLOMOVIC; PARDEE; COLLINS, 2015).

Um bom exemplo de biossensor muito utilizado em análises biológicas e diagnóstico médico é o biossensor de fluxo lateral (lateral flow nucleic acid biosensor - LFNAB). Esse é um método promissor para pontos de atendimento rápido (GAO et al., 2014). A Figura 5 mostra alguns exemplos de metodologias para aplicação de um biossensor. Eles podem ser, por exemplo, de fluxo lateral, dispositivo analítico microfluidico baseado em membrana de papel e fitas-teste de pH.

Figura 5 – Metodologias possíveis de serem aplicadas para criação de um biossensor. Exemplos: fitas-teste de pH (A e B) e fluxo lateral (C).



Fonte: HU et al., 2014

Pode-se dizer que é possível que, num futuro próximo, as tecnologias dos biossensores estejam amplamente inseridas na área de diagnóstico, principalmente molecular, por se tratar de um diagnóstico mais sensível. Dispositivos biossensores têm importância na saúde, pois podem ser uma solução simples e eficiente para diagnóstico em locais com poucos recursos, como postos de saúde, distantes dos laboratórios de referência, ou em casos de situações emergenciais (TELES; FONSECA, 2008).

Um simples diagnóstico precoce pode reduzir a mortalidade e economizar nos gastos com saúde (WEI; CHENG, 2016). Segundo Shorr e colaboradores (2007), a aplicação de métodos para diagnósticos precoce no ano de 2007, poderia ter evitado aproximadamente 92 mil mortes por sepse e ter promovido uma economia de, no mínimo, 1,5 bilhões de dólares dos custos em saúde nos Estados Unidos. Estudos mais recentes apontam que, só em 2011 nos Estados Unidos, o custo total de tratamentos de sepse foi de 20,3 bilhões (KHWANNIMIT; BHURAYANONTACHAI, 2015). Um diagnóstico precoce pode otimizar o uso de antibióticos, por exemplo, e ajudar a aumentar as taxas de sobrevivência dos pacientes (Dellinger et al., 2013; Singer et al., 2016). Por exemplo, a tuberculose latente (LTBI) é uma condição pela qual os pacientes não apresentam TB ativa, mas podem desenvolver TB quando expostos a fatores de risco que enfraquecem seu sistema imunológico. Seu diagnóstico precoce poderia reduzir o risco de desenvolvimento da TB ativa, por meio da prevenção e métodos profiláticos. Organização Mundial de Saúde, 2015a, 2015b).

1.4 PROTEÍNAS LIGADORAS DE DNA

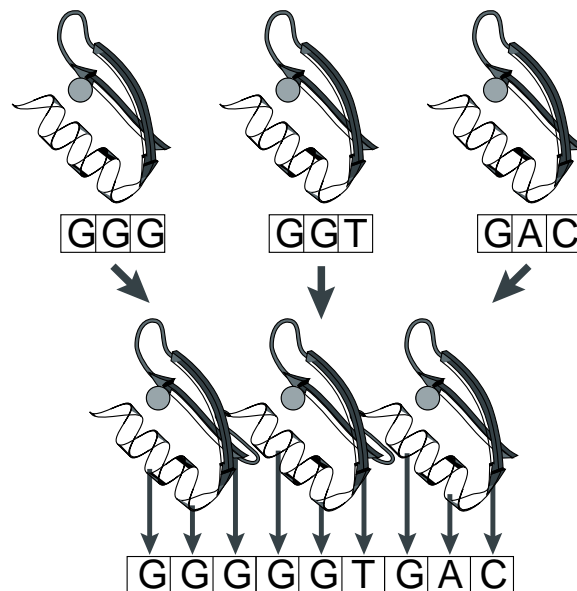
Um dos grandes diferenciais de proteínas ligadoras de DNA é que elas são capazes de se ligar a sequências específicas na dupla-fita de DNA, não sendo necessário a desnaturação e renaturação da fita para reconhecimento da sequência alvo. Dessa forma, podem ser aplicadas para alvos no genoma de organismos vivos (KIM et al., 2011). Proteínas ligadoras de DNA, produzidas sinteticamente, com a finalidade de controlar a expressão gênica têm permitido grandes avanços em pesquisas biomédicas. Fatores de transcrição sintéticos tem sido aplicados em testes pré-clínicos para o tratamento de doenças, e alguns também estão participando da fase de desenvolvimento clínico (KABADI; GERSBACH, 2014). Dentre as proteínas utilizadas para estudos envolvendo ligação à sequências específicas de DNA, as mais discutidas e utilizadas atualmente, tanto em pesquisa (*in vitro*) como em testes pré-clínicos, são as proteínas com domínios de Zinc Fingers (ZFP), CRISPR-Cas9 (*Clustered Regularly Interspaced Short Palindromic Repeats*) e as Proteínas Efetoras do tipo Ativador Transcricional (TALEs).

1.4.1 Zinc Fingers

Zinc Fingers (ZnF) são pequenos domínios protéicos, não reativos, estabilizados por um íon zinco, de onde surge então sua denominação. Cada domínio de zinc finger é capaz de se

ligar a três nucleotídeos específicos numa fita dupla de DNA, como pode ser observado na Figura 6. Os complexos formados por zinc fingers podem ser compostos de um ou mais domínios, podendo ser multidomínios (multifinger) quando arranjados em tandem. Dessa maneira, formam estruturas que são capazes de reconhecer seqüências de DNA mais extensas (URNOV et al., 2010).

Figura 6 – Interação dos domínios de Zinc Fingers com seqüências de DNA. Cada domínio ZnF tem a capacidade de interagir com três nucleotídeos de uma seqüência de DNA. Sendo assim, os três domínios ZnF apresentados como exemplo, quando em tandem, podem interagir com um total de nove nucleotídeos.



Fonte: (URNOV et al., 2010)

Desde a determinação de estruturas tridimensionais de Zinc Fingers ligadas a seqüências de DNA, as proteínas ligadoras de DNA tem sido consideradas uma ferramenta de grande importância para a Biotecnologia, uma vez que possuem vasta aplicabilidade (PAVLETICH; PABO, 1991). Essas proteínas podem ser programadas, de forma que possam localizar qualquer alvo específico num dado genoma. A depender da aplicação, as proteínas ligadoras de DNA podem ser acopladas a diferentes domínios, a saber: 1) domínios ativadores ou repressores transcricionais; 2) domínios para clivagem do DNA; ou 3) domínios para modificação ou remoção de marcas epigenéticas (STRAUS; LAHAYE, 2013).

A importância do zinco na estabilidade de estruturas protéicas foi primeiramente descrito em 1983, com o estudo da molécula do fator de transcrição IIIA (TFIIIA) (BERG; SHI, 1996).

Domínios ZnF estão presentes em inúmeras proteínas envolvidas em diversos processos celulares, a saber: replicação e reparo, transcrição e tradução, metabolismo e sinalização, proliferação celular e apoptose (KRISHNA; MAJUMDAR; GRISHIN, 2003; PAVLETICH; PABO, 1991; RICE; TURPIN, 1996).

As primeiras proteínas engenheiradas para detectar sequências específicas de DNA foram baseadas nas Zinc Fingers com domínios do tipo Cis_2His_2 , muito comuns no proteoma humano. Esse tipo de domínio forma uma estrutura bastante versátil para construção de proteínas ligadoras de DNA personalizadas para reconhecimento de qualquer sequência de DNA desejada. Cada domínio Cis_2His_2 é composto por cerca de trinta aminoácidos e forma uma estrutura do tipo $\beta\beta\alpha$, e essa estrutura é estabilizada por interações hidrofóbicas e pela quelação de um íon zinco a dois resíduos de Cisteína e dois resíduos de Histidina. Nessa estrutura, a região que determina a especificidade é a α -hélice, onde três ou quatro resíduos interagem com o DNA (GHOSH et al., 2006; KABADI; GERSBACH, 2014; KIM et al., 2011).

Embora as ZFP possuam, na literatura, resultados apresentando alta eficiência e especificidade, elas ainda são um desafio para os pesquisadores. Isso se deve à sua dificuldade na predição da afinidade e especificidade. O desenvolvimento de uma ZFP requer a construção e o *screening* utilizando um design racional, ou através de uma seleção de alta eficiência com base nas proteínas encontradas na vasta biblioteca de ZFPs previamente determinadas. Por esse motivo, muitos pesquisadores atualmente tem dado preferência às CRISPR-Cas9 e às TALEs, pois essas possuem um sistema de predição da especificidade e afinidade mais simples de compreender e engenheirar. Contudo, os conhecimentos adquiridos através de estudos com as ZFPs foram e são muito úteis para compreensão dessas outras proteínas (KABADI; GERSBACH, 2014).

1.4.2 Sistema CRISPR-Cas

O sistema CRISPR-Cas funciona como um mecanismo de defesa das bactérias e arqueobactérias contra infecções virais. Esse sistema é capaz de detectar e silenciar ácidos nucléicos não-próprios através de um RNA guia (gRNA). No sistema original, bactérias e arqueobactérias integram fragmentos pequenos de sequências de DNAs exógenos, chamados protoespaçadores, no *locus* genômico CRISPR. Os protoespaçadores, então, geram uma “memória” dos invasores anteriores. O *locus* CRISPR é transcrito e processado em RNAs pequenos, chamados crRNAs. Esses crRNAs então possuem complementariedade com sequências de DNA exógenas. No sistema CRISPR-Cas tipo II, os crRNAs são associados com crRNAs transativadores (tracrRNA) e com a endonuclease Cas9. O crRNA e

o tracrRNA recrutam então uma enzima RNaseIII para processamento e maturação do crRNA, para que este sirva como guia para a Cas9 atingir a sequência alvo (MAHFOUZ; PIATEK; STEWART JR, 2014; MALI et al., 2013; WIEDENHEFT; STERNBERG; DOUDNA, 2012).

Sendo assim, por complementariedade, o complexo CRISPR-Cas9 é capaz de induzir uma clivagem na fita de DNA exógena. Através da caracterização do sítio de clivagem da Cas9, hoje sabe-se que é necessário uma pequena sequência à jusante do protoespaçador para que a clivagem ocorra. Essa região é chamada Motivo Adjacente ao Protoespaçador (PAM) e possui alta variação na sua sequência nas diferentes espécies de bactérias. Sabendo que o gRNA é a molécula responsável pela especificidade ao alvo, o processo de construção de moléculas para diferentes sítios alvo torna-se muito mais fácil, já que somente é necessário modificação do gRNA. Porém, há relatos de um aumento na taxa de ligação fora do alvo, ou não-específico, mesmo em sequências antes consideradas conservadas. Outra característica negativa do sistema CRISPR-Cas é o tamanho da molécula Cas9, que dificulta a administração à células (JINEK et al., 2012; KABADI; GERSBACH, 2014; MAHFOUZ; PIATEK; STEWART JR, 2014; MALI et al., 2013; WIEDENHEFT; STERNBERG; DOUDNA, 2012).

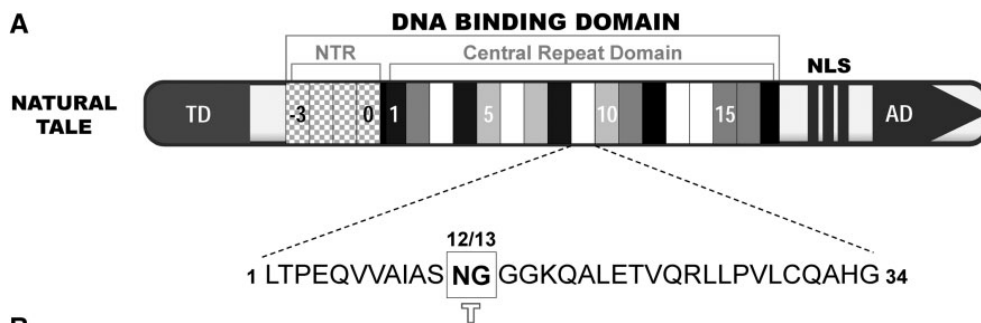
1.4.3 Proteínas Efetoras do tipo Ativadores Transcricionais (*Transcription Activator-Like Effectors* - TALEs)

Proteínas efetoras do tipo ativadores transcricionais (TALEs) são fatores de transcrição secretados por bactérias fitopatogênicas do gênero *Xanthomonas* e *Ralstonia* (BOCH et al., 2009; MOSCOU; BOGDANOVE, 2009; MUÑOZ BODNAR et al., 2013; NGA-SZE MAK et al., 2013). Esses organismos naturalmente produzem TALEs com o objetivo de modular a expressão gênica do hospedeiro, como um mecanismo de invasão, aumentando, dessa forma, a susceptibilidade da planta à infecção (CERMAK et al., 2011; KABADI; GERSBACH, 2014; POLITZ; COPELAND; PFLEGER, 2013). A primeira TALE natural identificada foi a *AvrBs3* de *Xanthomonas campestris* pv. *vesicatoria*, um patógeno de espécies de pimentas. A *AvrBs3* desencadeia uma resposta imune na planta que possui o gene *Bs3*, gene de resistência à doença (BOCH; BONAS, 2010; BOGDANOVE; SCHORNACK; LAHAYE, 2010).

As TALEs de origem natural são compostas por três partes essenciais: 1) um domínio central de repetições em tandem (*DNA Binding Domain* – *DBD*), responsável pela ligação da proteína a uma sequência de DNA específica; 2) porção N-terminal, com sinais de secreção e translocação; e 3) região C-terminal que possui sinal de localização nuclear e um domínio

ou proteína fusionada (SCHOLZE; BOCH, 2011). TALEs, originalmente, são compostas de domínios centrais contendo 7 a 34 repetições, cada uma contendo de 33-35 (normalmente 34) aminoácidos (Figura 7). A última repetição truncada na posição C-terminal, chamada *half-repeat* (meia-repetição), contém 20 aminoácidos (BOCH; BONAS, 2010).

Figura 7 – Representação dos domínios presentes numa proteína TALE de origem natural. Destaque para o Domínio Central de ligação ao DNA e a RVD (resíduos 12 e 13) de cada repetição.

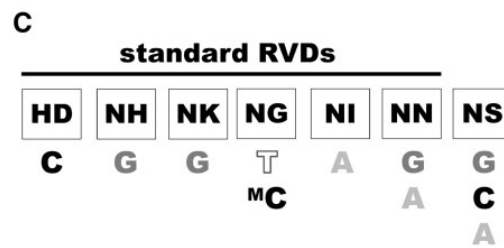


Fonte: (JANKELE; SVOBODA, 2014a)

Cada repetição do domínio central possui dois aminoácidos hipervariáveis (*Repeat-Variable Di-residue – RVD*), sendo eles os resíduos 12 e 13. O resíduo 12 não faz contato com o DNA, contudo, também é importante pois forma ligações de hidrogênio que estabilizam a estrutura da TALE. O resíduo 13, por sua vez, é responsável pela preferência de ligação do módulo de repetição à um nucleotídeo específico na sequência de DNA alvo. Ele se liga ao nucleotídeo na fita senso (sentido 5'-3'), no sulco maior (*major groove*) da dupla-hélice por meio de ligações de hidrogênio e interação de Van der Waals (DENG et al., 2012; JANKELE; SVOBODA, 2014a; KABADI; GERSBACH, 2014; NGA-SZE MAK et al., 2013).

A especificidade dos RVDs foram decifradas experimentalmente e através de ferramentas de Bioinformática. Atualmente, os RVDs bem caracterizados na literatura são (Figura 8): NI (Asparagina e Isoleucina) para Adenina (A), HD (Histidina e Aspartato) para Citosina (C), NG (Asparagina e Glicina) para Timina (T) e Citosina metilada (mC), NH (Asparagina e Histidina) e NK (Asparagina e Lisina) para Guanina (G), NN (duas Asparaginas) para G ou A, e NS (Asparagina e Serina) para todas as quatro bases. Os diresíduos NN e NS possuem, portanto, baixa especificidade (BOCH et al., 2009; MOSCOU; BOGDANOVE, 2009).

Figura 8 – Apresentação das RVDs padrão e os nucleotídeos com os quais elas especificamente interagem.



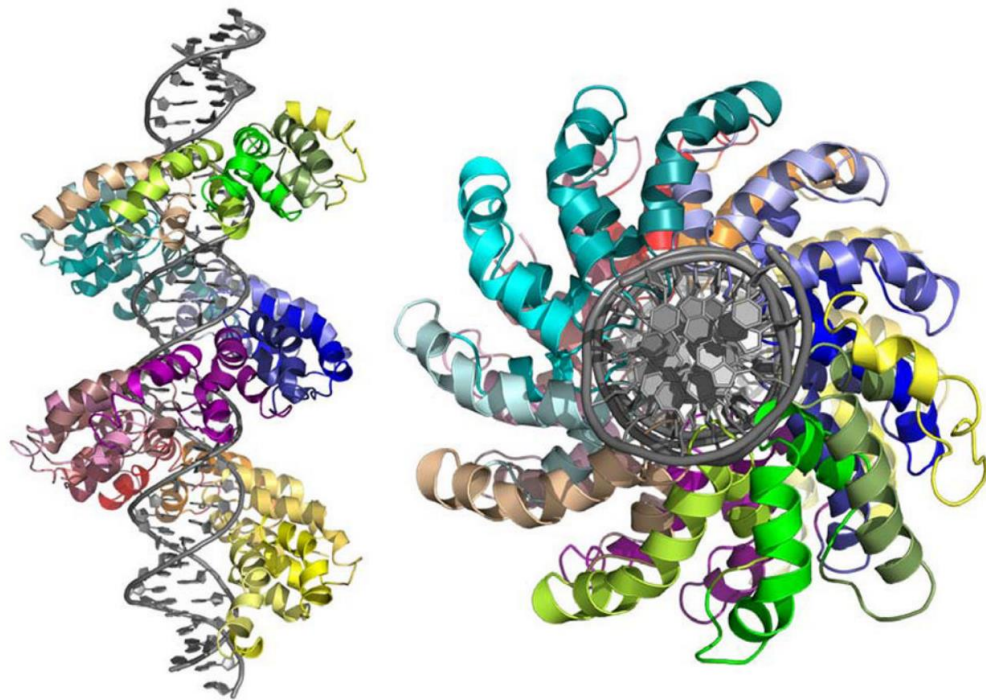
Fonte: (JANKELE; SVOBODA, 2014a)

A região N-terminal (NTR) é composta de quatro repetições antecedentes ao domínio central que contribuem diretamente para a carga total básica da TALE. Essa região é importante para a ligação da proteína ao DNA e faz interação com uma Timina conservada (posição 0 da sequência alvo). Nas TALEs naturais, produzidas por fitobactérias, a NTR possui ainda sinalização para secreção e translocação (CASPER-LINDLEY et al., 2002; JANKELE; SVOBODA, 2014a).

A Timina conservada na posição 0 (5'-T⁰) é observada na maioria das sequências alvo das TALEs de ocorrência natural. Elas parecem ser necessárias para a atividade das proteínas fusionadas à região C-terminal das TALEs (GAO et al., 2012; LAMB; MERCER; BARBAS, 2013; MAHFOUZ et al., 2011; MOSCOU; BOGDANOVE, 2009). Entretanto, já foram observadas proteínas fusionadas à TALEs que funcionavam perfeitamente mesmo na ausência de uma 5'-T⁰. Aparentemente, a 5'-T⁰ da sequência de DNA alvo parece não ter a mesma significância das TALEs de ocorrência natural para as TALEs desenhadas artificialmente com RVDs padrão. Isso sugere que as TALEs sintéticas se liguem com maior afinidade ao alvo, de forma que não é necessário a 5'-T⁰ (MECKLER et al., 2013; MILLER et al., 2011; SUN et al., 2012).

De forma independente, cada repetição do domínio central forma uma estrutura secundária de “hélice-loop-hélice”. Nessa estrutura as α -hélices são ligadas pelo loop que é formado pela RVD (Figura 9). A primeira α -hélice (α -helix S), a menor, se estende do resíduo 3 ao 11, enquanto que a segunda α -hélice (α -helix L), a maior, se estende do resíduo 15 ao 33 (JANKELE; SVOBODA, 2014a). O domínio central forma uma estrutura de super-hélice para a direita (LANGE; BINDER; LAHAYE, 2014).

Figura 9 – Estrutura tridimensional obtida por cristalografia de raio-x da TALE PthXo1 do patógeno de arroz *Xanthomonas oryzae*.



Fonte: (MAK et al., 2012)

A produção de TALEs sintéticas para regulação de genes é uma área de pesquisa que tem crescido bastante nos últimos anos. Os pesquisadores consideram duas regras importantes no momento de selecionar TALEs em potencial. A primeira é a inserção da 5'-T⁰, ou repetição 0 no desenho da proteína (para garantir a estabilidade da ligação ao alvo); e a segunda, é a meia-repetição (*half-repeat*) truncada na posição 3'. Essa última é importante para a ligação correta à sequência de DNA alvo (BOCH et al., 2009; MOSCOU; BOGDANOVE, 2009).

As TALEs têm surgido como ferramentas de destaque para fins de manipulação gênica. Elas possuem grande potencial, uma vez que são capazes de se ligar a dupla-fita de DNA, permitindo que se liguem diretamente aos cromossomos dentro do núcleo das células alvo e possam então promover edição gênica *in loci* (POLITZ; COPELAND; PFLEGER, 2013). Apesar da quantidade de informação conhecida a respeito dessas proteínas, muito ainda precisa ser estudado a fim de melhor elucidar a natureza biofísica da interação específica das RVDs com os nucleotídeos. Além disso, é preciso esclarecer a relação da extensão e composição da proteína com a afinidade e especificidade. Mais ainda, um estudo aprofundado da estrutura tridimensional seria capaz de definir de que forma exatamente a proteína interage com o DNA para que a arquitetura dos domínios ou das proteínas fusionadas na porção C-

terminal das TALEs possa ser otimizada. São inúmeras possíveis aplicações para essa classe de proteína ligadora de DNA (BOGDANOVE; VOYTAS, 2011).

1.5 GENES REPÓRTER: CROMOPROTEÍNAS

Convencionalmente, todos os processos de transformação de organismos devem conter genes marcadores que sejam capazes de identificar e isolar os transformados dos demais (SHIH et al., 2015). Os genes marcadores mais utilizados em sistemas bacterianos são chamados genes repórter. O termo gene repórter se refere a um gene cujo fenótipo resultante possa ser quantificado (exemplo: cor, fluorescência) e que seja facilmente distinguido das proteínas endógenas do organismo. Genes repórteres, como as proteínas fluorescentes e bioluminescentes, não exercem pressão seletiva nas células transformadas. Eles são amplamente utilizados para observar eventos celulares, a exemplo de transdução de sinais e expressão gênica (NAYLOR, 1999; SHIH et al., 2015). Alguns exemplos muito utilizados estão demonstrados na Tabela 1.

Tabela 1 – Genes repórter mais comumente utilizados.

Genes repórter	Fenótipos
Cloranfenicol acetiltransferase (CAT)	Usa a acetil-CoA para transferir dois grupos acetil para os grupos hidróxido do cloranfenicol, impedindo-o de se ligar ao ribossomo, neutralizando a sua ação antibiótica.
Luciferase	Enzimas que catalisam reações biológicas transformando energia química em energia luminosa.
Proteína fluorescente verde (GFP)	Proteína produzida pelo cnidário <i>Aequorea victoria</i> que emite fluorescência na zona verde do espectro visível.

LacZ (Beta-galactosidade)

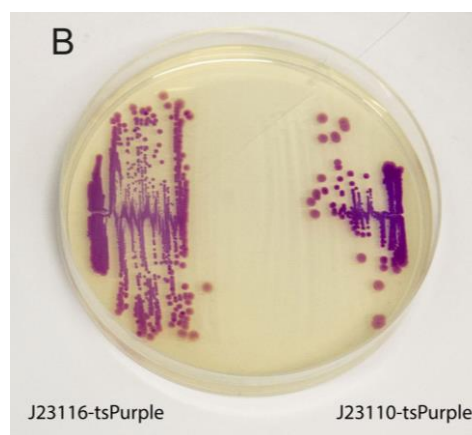
Coloração azul na presença do
substrato X-GAL

Fonte: Adaptado de NAYLOR, 1999

Alguns estudos, como o de avaliação de expressão gênica por exemplo, não necessitam de marcadores que exerçam pressão seletiva. Para tais estudos, pode-se utilizar alternativas mais rápidas e ecológicas, como o uso de genes repórter codificadores de cromoproteínas. Uma vez que proteínas fluorescente necessitam de equipamentos especiais para visualização e/ou quantificação da fluorescência emitida, as cromoproteínas surgem como uma alternativa mais simples. Cromoproteínas permitem a identificação e isolamento de organismos transformados por meio de simples detecção visual (MERCANTE; RUMJANEK; FRANCO, 2000; NAYLOR, 1999; SHIH et al., 2015).

A Tinsel Purple (tsPurple) é uma cromoproteína distribuída comercialmente pela empresa DNA 2.0 (ProteinPaintbox®). A proteína apresenta coloração violeta (como observado na Figura 10), funcionando portanto como um bom gene repórter. Sua sequência gênica foi adicionada ao Registro de Partes Biológicas (iGEM) pelo grupo iGEM11_Uppsala-Sweden no ano de 2011.

Figura 10 – Imagem de uma placa com colônias de *Escherichia coli* transformadas com plasmídeo contendo a cromoproteína Tinsel Purple.



Fonte: iGEM (http://parts.igem.org/wiki/images/9/92/AmilCP_and_tsPurple.jpg)

1.6 Corinebactérias emergentes e seu diagnóstico

A difteria é uma doença infecciosa que merece atenção devido aos constantes surtos e ondas epidêmicas com alta incidência e taxas de mortalidade cada vez mais elevadas (GALAZKA, 2000). *Corynebacterium diphtheriae* é o principal agente etiológico causador da doença, e por consequência, o organismo mais estudado do gênero *Corynebacterium spp.* O gênero é composto por diversas espécies de interesse médico, algumas dessas corinebactérias são capazes de produzir a toxina diftérica, enquanto existem outras corinebactérias não diftéricas (não produtoras de toxina), também chamadas difteróides (KONRAD; HÖRMANSDORFER; SING, 2015). Espécies de corinebactérias não-diftéricas são naturalmente encontradas na mucosa e microbiota normal da pele de seres humanos e animais. Atualmente, essas espécies têm sido frequentemente encontradas em amostras clínicas, tornando difícil determinar com eficiência se é uma simples contaminação ou se é clinicamente relevante (BERNARD, 2012).

Essas corinebactérias não-diftéricas têm sido relatadas na literatura como agentes oportunistas, capazes de causar infecções nosocomiais em humanos. As infecções associadas a esses agentes acabam por desencadear quadros de septicemias, endocardite, infecções de feridas cirúrgicas, de próteses e infecções relacionadas a cateteres (VILA et al., 2012). Em 2015, pesquisadores nos Estados Unidos apontaram as bactérias do gênero *Corynebacterium spp.* como segunda causa mais comum de infecção em ferimentos crônicos. Dentre as três espécies mais frequentes estão: *C. striatum* e *C. xerosis* (WOLCOTT et al., 2015). As corinebactérias não-diftéricas mais estudadas são as que compõem o chamado complexo XSMA, que compreende: *Corynebacterium xerosis*, *Corynebacterium striatum*, *Corynebacterium minutissimum* e *Corynebacterium amycolatum*. Isso se deve não só ao potencial patogênico dessas espécies, mas às dificuldades na identificação e diagnóstico preciso, pois essas elas possuem altas taxas de variabilidade bioquímica entre linhagens da mesma espécie.

Atualmente, o diagnóstico preciso das corinebactérias ainda é um desafio para a medicina. O teste mais utilizado para diferenciação destas bactérias é o API Coryne™ (bioMérieux). Esse teste é um sistema padronizado para a identificação de bactérias corineformes que é composto por um kit com 20 testes bioquímicos (substratos) para avaliar atividades enzimáticas e fermentação de carboidratos que geram resultados após 24 horas. Os resultados dos testes são lançados em uma base de dados online da empresa que emite uma identificação da espécie (API Coryne™, Guia do Usuário. BioMérieux).

Atualmente, o método mais confiável para a identificação das corinebactérias é o

sequenciamento do genoma total. Esse método é capaz de fornecer informações de genes únicos das espécies, perfil de resistência a antibióticos ou mesmo informar a ausência de genes já conhecidos e previamente relacionados com as espécies. Contudo, é uma metodologia que requer equipamentos de alto custo, bem como profissionais capacitados para executar a técnica e interpretar os resultados. Por esses motivos, esse tipo de metodologia ainda é restrita a laboratórios de referência (KOSER et al., 2012). Os genomas das quatro espécies do complexo XSMA já se encontram depositados em banco de dados online, o que torna mais fácil o desenvolvimento de novas técnicas moleculares para identificação correta das espécies de corinebactérias (MATTOS-GUARALDI et al., 2015; PACHECO et al., 2015).

2 OBJETIVOS

2.1 OBJETIVO GERAL

Desenvolver uma proteína ligadora de DNA do tipo TALE (*Transcription Activator Like Effector*) fusionada com a cromoproteína Tinsel Purple, com potencial de detectar uma sequência específica de DNA e emitir sinal colorimétrico para detecção visual.

2.2 OBJETIVOS ESPECÍFICOS

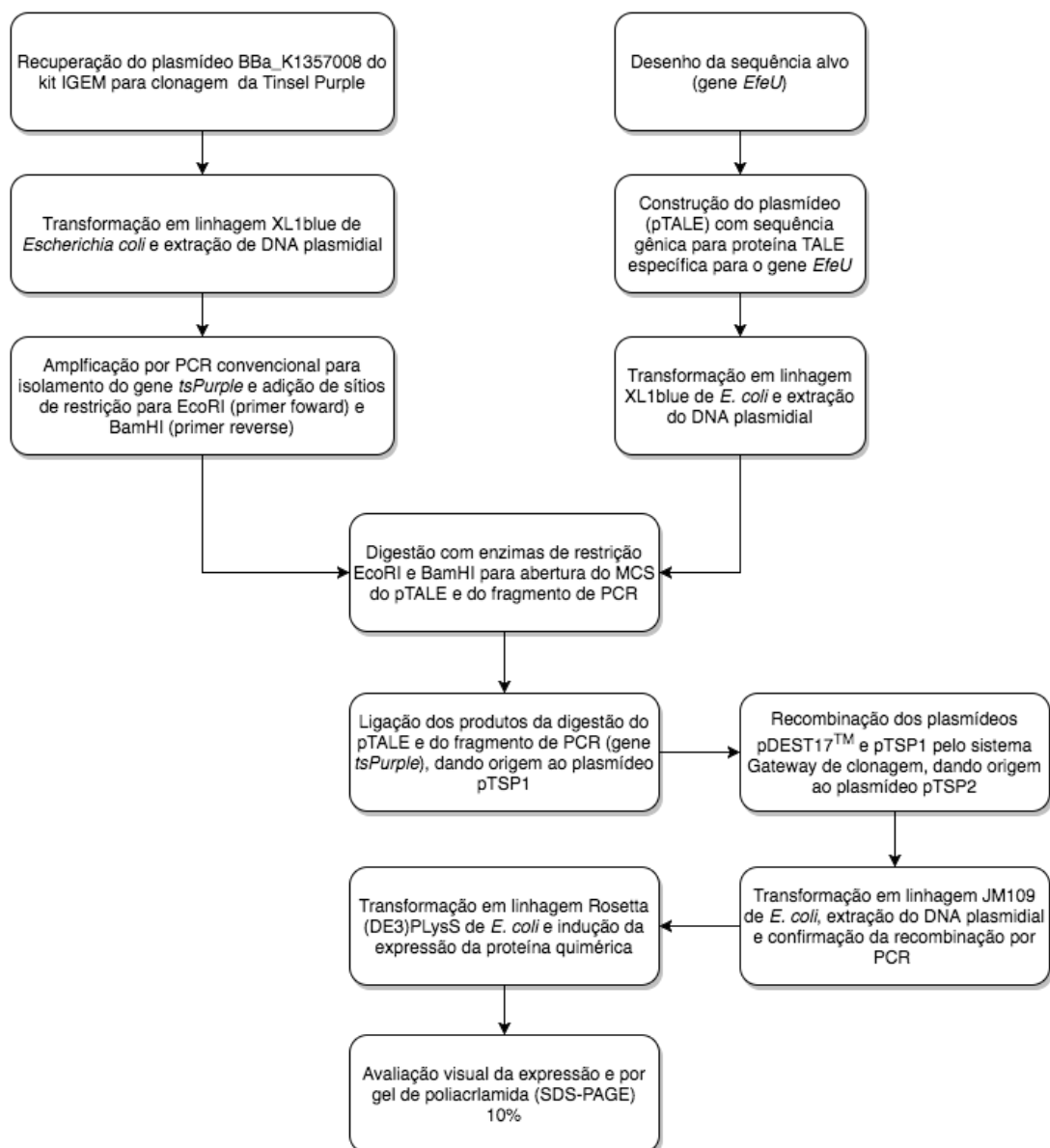
- ✓ Desenhar e construir a sequência codificadora para proteína TALE ligadora de DNA específica para *Corynebacterium striatum*.
- ✓ Obter a sequência gênica da cromoproteína Tinsel Purple (tsPurple).
- ✓ Fusionar a sequência gênica da TALE com a sequência gênica da cromoproteína.
- ✓ Clonar sequência fusionada num plasmídeo de expressão pelo sistema Gateway de clonagem.
- ✓ Induzir a expressão da proteína quimérica e avaliar a emissão de cor para validar sua viabilidade em diagnóstico.

3 METODOLOGIA

3.1 ESTRATÉGIA EXPERIMENTAL

O presente trabalho segue a metodologia de acordo com o apresentado no fluxograma abaixo.

Figura 11 – Fluxograma da estratégia experimental adotada.



3.2 LINHAGENS DE *ESCHERICHIA COLI* UTILIZADAS NO ESTUDO

As linhagens de *Escherichia coli* utilizadas no presente estudo, bem como seus respectivos genótipos, são apresentados na Tabela 2. As linhagens XL1blue e JM109 foram selecionadas como linhagens de clonagem, pois o laboratório possuía um estoque dessas células competentes. A linhagem Rosetta™(DE3)pLysS foi selecionada, pois a sequência codificadora da proteína TALE específica possuía muitos códons cuja frequência de uso era baixa em *Escherichia coli*, e essa linhagem é otimizada para expressão de genes contendo códons raros.

Tabela 2 – Linhagens de *E. coli* selecionadas para o estudo e seus respectivos genótipos.

Linhagens	Genótipo
JM109	<i>endA1 glnV44 thi-1 relA1 gyrA96 recA1 mcrB+ Δ(lac-proAB) e14- [F' traD36 proAB+ lacIq lacZΔM15] hsdR17(rK-mK+)</i>
XL1blue	<i>endA1 gyrA96(nalR) thi-1 recA1 relA1 lac glnV44 F[::Tn10 proAB+ lacIq Δ(lacZ)M15 Amy CmR] hsdR17(rK- mK+)</i>
Rosetta™(DE3)pLysS	<i>E. coli str. B F- ompT gal dcm lon? hsdSB(rB-mB-) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(ΔS) pLysSRARE[T7p20 ileX argU thrU tyrU glyT thrT argW metT leu W proL orip15A](CmR)</i>

Fonte: http://openwetware.org/wiki/E._coli_genotypes e Thermo Scientific.

3.3 PLASMÍDEOS UTILIZADOS PARA A ESTRATÉGIA DE CLONAGEM

Os plasmídeos utilizados no presente estudo estão dispostos na Tabela 3. O plasmídeo pSB1C3 foi obtido do kit IGEM distribuição 2015, e nele estava presente a sequência gênica codificadora da proteína Tinsel Purple. O plasmídeo pTALE foi o plasmídeo obtido da Thermo Scientific com a sequência gênica codificadora da proteína TALE específica. Esse plasmídeo possui resistência a kanamicina e possui um sítio múltiplo de clonagem na posição C-terminal da sequência gênica para proteína TALE. O pDEST™17 foi selecionado como vetor de expressão, pois ele possui sequência codificadora para uma cauda de Histidina a ser

traduzida na posição N-terminal do inserto, resistência a ampicilina e sítios de recombinação *att*. Os plasmídeos pTSP1 e pTSP2 foram resultantes das clonagens realizadas durante o estudo e possuem resistência a ampicilina, bem como o pDEST™17.

Tabela 3 – Plasmídeos utilizados no presente estudo, suas respectivas origens de replicação e marcadores de resistência.

Plasmídeos	Origem de Replicação / Marcador de Resistência
pSB1C3	pMB1 / Cm ^r
GeneArt™ PerfectMatch TAL vector (pTALE)	pUC / Kan ^r
pDEST™17	pMB1 / Amp ^r
pTSP1	pUC / Kan ^r
pTSP2	pMB1 / Amp ^r

Fonte: IGEM (<http://parts.igem.org/wiki/index.php/Part:pSB1C3>) e Thermo Scientific (<http://thermofischer.com/>)

3.4 ESTRATÉGIA DE CLONAGEM PARA OBTENÇÃO DA PROTEÍNA QUIMÉRICA BIOSSENSORA

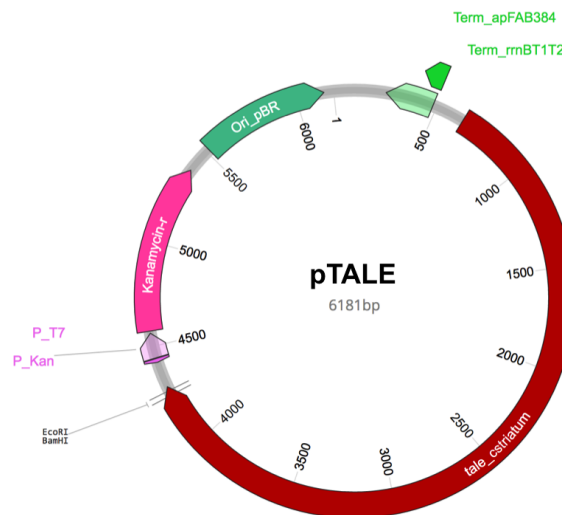
3.4.1 Seleção da sequência alvo para construção da Proteína TALE específica

A sequência alvo específica de *Corynebacterium striatum* foi determinada com base em um estudo prévio do nosso grupo de pesquisa na Universidade Federal da Bahia (SANTOS, 2016), o qual gerou um depósito de patente recente (SANTOS et al., 2017). Através das anotações de sequenciamento genômico de linhagens de *C. striatum*, foram realizadas análises bioinformáticas para seleção de genes únicos de *C. striatum* dentre as corinebactérias do

complexo XSMA. Foram eliminados genes hipotéticos, e então 68 genes restantes foram avaliados por BLASTn e BLASTx. Foi então selecionado o gene *efeU* que codifica uma permease do transporte de ferro (APÊNDICE I).

Com base no gene alvo (*efeU*), foi realizado o desenho da sequência codificadora da proteína TALE com auxílio da ferramenta TAL Effector Targeter (<https://tale-nt.cac.cornell.edu/>) disponível no website da Universidade de Cornell. A sequência foi fornecida à empresa *Thermo Scientific* para construção do plasmídeo com a sequência codificadora para a proteína TALE específica. O plasmídeo fornecido está apresentado na Figura 12 e sua sequência pode ser encontrada no APÊNDICE II.

Figura 12 – Plasmídeo pTALE desenhado através da ferramenta DNA ATLAS do site ATUM. Contem sequência codificadora para proteína TALE específica para o gene *efeU* da *C. striatum* e um sítio múltiplo de clonagem (MCS) com sítios de restrição para EcoRI e BamHI.



3.4.2 Clonagem e digestão do plasmídeo contendo proteína TALE

O pTALE foi transformado em *E. coli* da linhagem XL1blue por choque térmico utilizando-se aproximadamente 100ng do plasmídeo. Foram selecionadas as colônias resistentes a kanamicina numa concentração de 50ug/ml. As colônias selecionadas foram inoculadas em meio Luria-Bertani (LB). Após 24h de cultivo, foi realizada a extração do DNA plasmidial utilizando o kit *NucleoSpin® Plasmid* (Macherey-Nagel). Em seguida, o DNA obtido

foi dosado por espectrofotometria (μ drop, ThermoScientific) e sua integridade analisada por gel de agarose a 1%.

A partir daí, o primeiro passo das etapas de clonagem foi a digestão do sítio múltiplo de clonagem (MCS) do plasmídeo fornecido contendo a proteína TALE específica com as enzimas EcoRI e BamHI (Figura 13a), deixando extremidades coesivas. Cada reação de digestão (Tabela 4) foi realizada separadamente, durante um período de 16h, e entre cada reação, os fragmentos foram purificados do gel de agarose com uso do kit *Illustra GFX PCR and Gel Band Purification Kit* (GE Healthcare and Life Sciences). Conseqüentemente, foi obtido então o plasmídeo pTALE linearizado, com a abertura do MCS (Figura 13b).

Figura 13 – Representação esquemática da digestão do pTALE e conseqüente abertura do sítio múltiplo de clonagem.

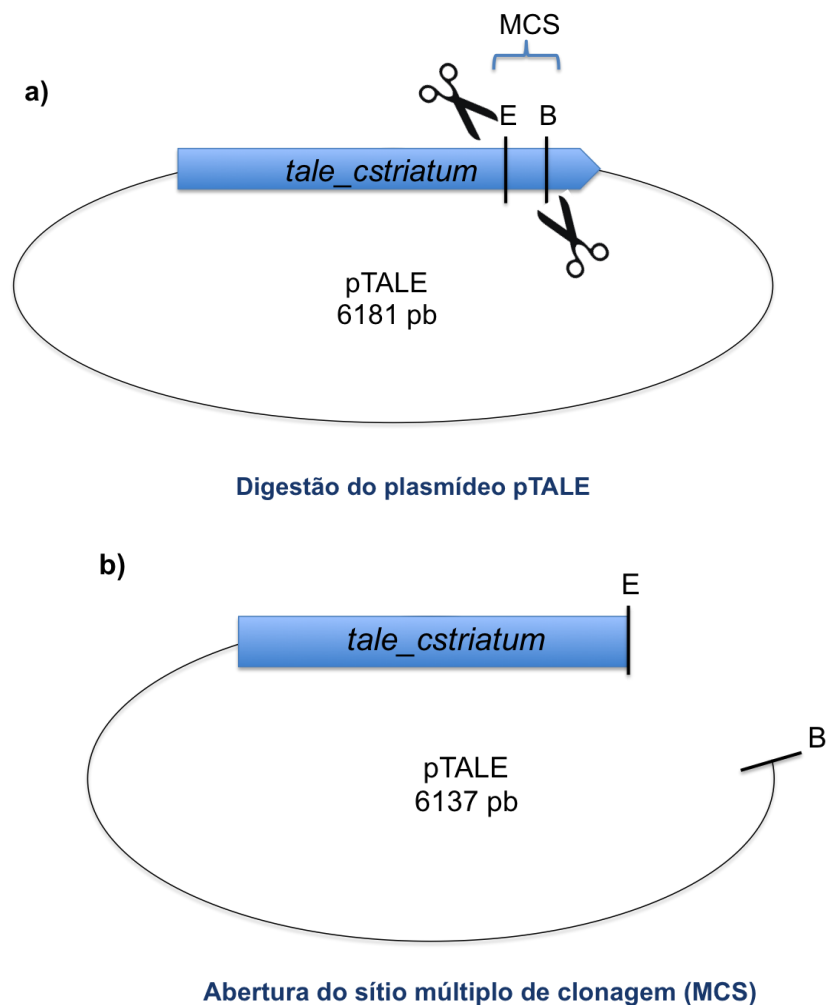


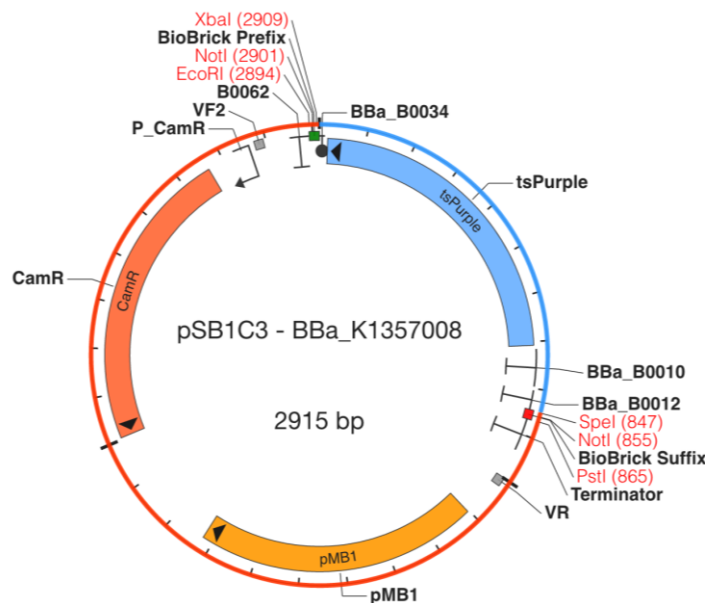
Tabela 4 – Reação de digestão para o pTALE e fragmento amplificado por PCR do gene *tsPurple*. Foram realizadas reações para cada enzima separadamente, com duração de 16h.

Componente	Volume para digestão pTALE	Volume para digestão tsPurple
Tampão 10x	2ul	2ul
BSA (2ug/ul)	1ul	1ul
DNA (500ng)	5ul (pTALE)	2ul (tsPurple)
EcoRI/BamHI (5U)	1,5ul	1,5ul
H2O miliQ	13,5ul	16,5ul
Volume final	20ul	20ul

3.4.3 Obtenção do plasmídeo contendo sequência gênica da cromoproteína Tinsel Purple

Após a digestão do pTALE, foi obtido o plasmídeo contendo a sequência gênica para a cromoproteína Tinsel Purple (APÊNDICE II). O plasmídeo (BBa_K1357008) foi recuperado do kit IGEM (distribuição 2015). No kit, os plasmídeos carregando os elementos genéticos padronizados e registrados são disponibilizados liofilizados. O plasmídeo do kit foi ressuspendido seguindo protocolo do IGEM. Primeiro foi localizado o poço correspondente à parte desejada. Em seguida, foi feito um pequeno furo no papel de proteção do poço com auxílio de uma ponteira (sem remover a cobertura). Foram pipetados 10 µl de água miliQ autoclavada e a homogeneização foi feita por pipetagem. Foram aguardados 5 minutos para garantir máxima ressuspensão e então a amostra foi armazenada em microtubos em freezer -20°C. O plasmídeo recuperado do kit continha uma sequência de RBS (ribosome binding site), o gene codificador da Tinsel Purple (*tsPurple*) e Terminadores transcrpcionais (Figura 14).

Figura 14 – Mapa do plasmídeo BBa_K1357008. Contem RBS (BBa_B0034), gene codificador da cromoproteína Tinsel Purple (tsPurple) e Terminadores (BBa_B0010 e BBa_B0012).



Fonte: Registry of Biological Parts (IGEM).

Seguiu-se então o protocolo de transformação da linhagem XL1blue de *Escherichia coli* por choque térmico. Para a transformação foram utilizados 5µl do plasmídeo contendo o gene codificador da cromoproteína Tinsel Purple. Foram selecionadas as colônias resistentes a cloranfenicol numa concentração de 34 µg/ml. As colônias selecionadas foram inoculadas em meio LB. Após 24h, foi realizada a extração de DNA plasmidial para recuperação da partes utilizando o kit *NucleoSpin® Plasmid* (Macherey-Nagel). Foi realizada a extração do DNA plasmidial, em seguida o DNA foi dosado por espectrofotometria (µdrop) e sua integridade analisada por gel de agarose a 1%.

3.4.4 Isolamento da sequência gênica *tsPurple* por PCR convencional e digestão dupla

Primers para isolamento da sequência codificadora da cromoproteína Tinsel Purple foram desenhados, desconsiderando-se o RBS e terminadores transcricionais, e foram adicionados sítios de restrição nas suas extremidades, EcoRI (primer *foward*) e BamHI (primer *reverse*). Todos os primers utilizados no estudo estão presentes no APÊNDICE III. Foi realizada uma reação de PCR convencional com o kit de Taq polimerase recombinante Pht (Phoneutria Biotecnologia) (Figura 15). O protocolo da reação realizada para amplificação da sequência consta na Tabela 5.

Figura 15 – Representação da amplificação por PCR da sequência gênica *tsPurple* e adição dos sítios de restrição EcoRI (primer forward) e BamHI (primer reverse).

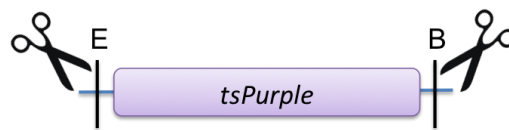


Tabela 5 – Reação de PCR convencional realizado com o kit Pht para isolamento da sequência codificadora da cromoproteína Tinsel Purple e adição dos sítios de restrição

Reagentes	Conc. Inicial	Conc. Final	Volume
Tampão	10x	1x	2,5ul
DNA polymerase Pht	5U/ul	1U	0,2ul
MgCl ₂	50mM	1,5mM	0,75ul
dNTP mix	10mM	200µM	0,5ul
Primer forward	100uM	2µM	0,5ul
Primer reverse	100uM	2µM	0,5ul
DNA template		10ng/µl > 100ng/µl	2ul
H2O miliQ	q.s.p.	q.s.p.	16,55ul
Volume final			25ul

Em seguida, o produto do PCR foi purificado e foi submetido à uma digestão dupla com ambos EcoRI e BamHI (Figura 16), deixando extremidades coesivas. As reações de digestão foram feitas durante um período de 16h, e o produto purificado da digestão com uso do kit *Illustra GFX PCR and Gel Band Purification Kit* (GE Healthcare and Life Sciences) foi utilizado para a ligação do fragmento de PCR (*tsPurple*) com o plasmídeo pTALE.

Figura 16 – Representação esquemática da digestão dupla do fragmento de PCR contendo o gene *tsPurple*.



Digestão do fragmento de PCR *tsPurple*

3.4.5 Fusão das sequências TALE e *tsPurple* para obtenção da proteína quimérica biossensora

Uma vez que, tanto o pTALE como o gene *tsPurple* se encontravam digeridos por EcoRI e BamHI, foi feita a ligação com a enzima T4 ligase da *Promega* através das suas extremidades coesivas complementares (Figura 17). A reação de ligação (Tabela 6) foi feita utilizando uma proporção de 1:1 de cada fragmento linearizado e purificado após digestão, durante um período de 16h, resultando no plasmídeo pTSP1. A sequência obtida após a reação de ligação dos fragmentos foi simulada para avaliação dos resultados. Foi obtido então o plasmídeo pTSP1, com a fusão das sequências gênicas, no *backbone* do plasmídeo pTALE.

Figura 17 – Representação esquemática da ligação dos produtos das digestões do pTALE e tsPurple com EcoRI e BamHI através das suas extremidades coesivas.

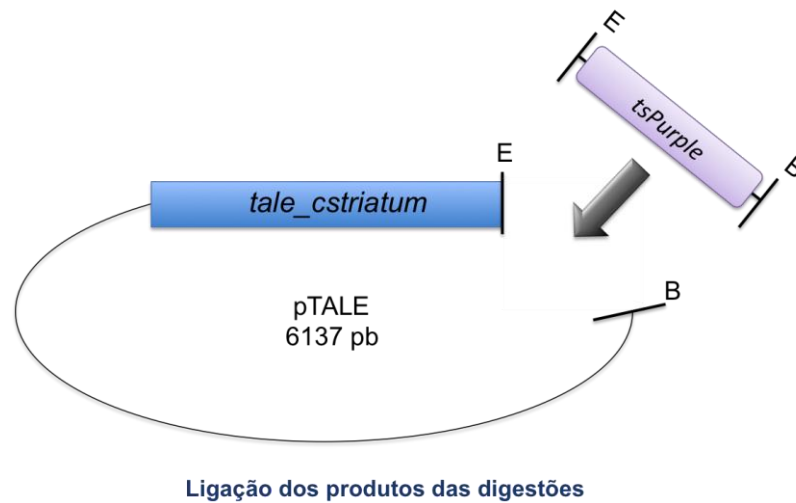


Tabela 6 – Reação de ligação realizada utilizando a enzima T4 DNA ligase da *New England Biolabs*.

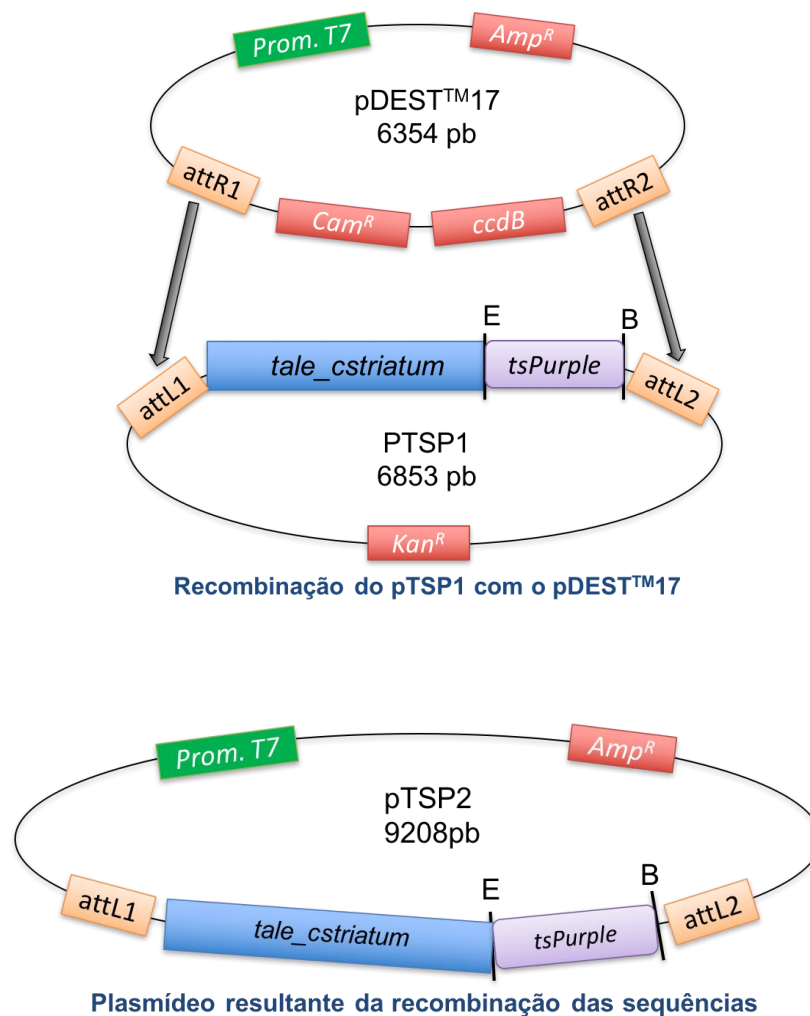
Componente	Volume
Tampão 10x	2ul
DNA vetor	8,5ul
DNA inserto	8,5ul
T4 DNA ligase	1ul
Volume final	20ul

A linhagem JM109 de *Escherichia coli* foi transformada com o pTSP1 (produto da reação de ligação) por choque térmico utilizando-se 3µl do plasmídeo. Foram selecionadas as colônias resistentes a kanamicina numa concentração de 50ug/ml. As colônias selecionadas foram inoculadas em meio LB. Após 24h, foi realizada a extração do DNA plasmidial utilizando o kit *NucleoSpin® Plasmid* (Macherey-Nagel). Em seguida, o DNA obtido foi dosado por espectrofotometria (µdrop) e sua integridade analisada por gel de agarose a 1%.

3.4.6 Recombinação dos plasmídeos pTSP1 e pDESTTM17 pelo sistema de clonagem Gateway

Levando-se em consideração de que o pTALE possui sítios de recombinação por recominase lambda (sítios *att*), foi realizada uma reação de recombinação do plasmídeo pTALE com o pDESTTM17 (vetor de expressão) utilizando a enzima *LR clonase II* da empresa *Invitrogen* (Figura 18).

Figura 18 – Representação esquemática da reação de recombinação entre os plasmídeos pTSP1 e o pDESTTM17 e o plasmídeo resultante, pTSP2 (9208pb).



O pDESTTM17 contém um promotor T7, que pode ser induzido com a adição de IPTG ao meio de cultura. Além disso, possui resistência a ampicilina no seu backbone e resistência a

cloranfenicol e o gene *ccdB* entre os sítios de recombinação. O gene citotoxigênico *ccdB* é um marcador de contra-seleção, pois somente os clones que possuem os plasmídeos recombinados, ou seja, sem o gene *ccdB*, podem crescer. A resistência a cloranfenicol e o gene *ccdB* foram então perdidos após a recombinação. A reação de recombinação ocorreu num período de 18h e em seguida a reação foi interrompida com a adição da *Proteinase K* por 15 minutos. O produto da recombinação foi utilizado para transformação da linhagem JM109 de *Escherichia coli* por choque térmico utilizando-se 4µl do DNA plasmidial. Foram selecionadas as colônias resistentes a ampicilina numa concentração de 100ug/ml e descartadas as colônias resistentes a kanamicina numa concentração de 50ug/ml, através de *replica plate*. As colônias selecionadas foram inoculadas em meio LB. Após 24h, foi realizada a extração do DNA plasmidial utilizando o kit *NucleoSpin® Plasmid* (NoLid) e em seguida, o DNA foi dosado por espectrofotometria (µdrop) e sua integridade analisada por gel de agarose a 1%.

A confirmação da recombinação foi feita por meio das análises das placas de cultura seletivas com os antibióticos Ampicilina e Kanamicina. Além disso foi realizado PCR para confirmar a ligação e recombinação (Tabela 7).

Tabela 7 – Reação de PCR para confirmação das ligações e recombinações.

Reagentes	Conc. Inicial	Conc. Final	Volume
Tampão	10x	1x	2,5ul
Taq DNA polimerase <i>Sinapse</i>	5U/ul	1U	0,2ul
dNTP mix	10mM	200µM	0,5ul
Primer forward	100uM	2µM	0,5ul
Primer reverse	100uM	2µM	0,5ul
DNA template		10ng/µl > 100ng/µl	1ul
H2O miliQ	q.s.p.	q.s.p.	19,8ul
		Volume final	25ul

Foram utilizados os primers M13 foward (-20) e M13 reverse, para os plasmídeos pTALE e pTSP1 (resultante da reação de ligação). Primers T7 foward e tsPurple reverse foram utilizados para amplificar o plasmídeo pTSP1. Já o plasmídeo pTSP2 (produto da recombinação) foi amplificado com os primers T7 foward e tsPurple reverse. Os backbones dos plasmídeos constam na Figura 19. Os plasmídeos pTALE e pTSP1 compartilham o mesmo backbone, assim como os plasmídeos pDEST™ 17™ e pTSP2.

Figura 19 – Mapas dos plasmídeos pTALE, pDEST™ 17™, pTSP1 e pTSP2. Criação dos mapas feita através do site *Addgene*.



3.4.7 Expressão da proteína quimérica biossensora

Considerando-se que a sequência do gene codificador da proteína TALE específica possuía muitos códons raros, optou-se por utilizar a linhagem Rosetta™(DE3) pLysS de *Escherichia coli*. Essa linhagem foi transformada por choque térmico, utilizando-se aproximadamente 100ng do plasmídeo pTSP2. Foram selecionadas as colônias resistentes a ampicilina numa concentração de 100ug/ml e descartadas as colônias resistentes a kanamicina numa concentração de 50ug/ml. As colônias selecionadas foram inoculadas em 5ml de meio LB contendo ampicilina a 100ug/ml para o pré-inóculo. Após 16h, foi feito o inóculo de 100µl do pré-inóculo em 10ml de meio LB contendo ampicilina a 100µg/ml. A densidade ótica foi medida com filtro de 600nm e quando as alíquotas atingiram 1,0 e 1,5 foi feita a indução da expressão da proteína quimérica com a adição de IPTG a 1mM na temperatura de 37°C a 200 RPM. Foram coletadas amostras dos tempos 0 (pré-IPTG), 4 e 24 horas. Os resultados da expressão obtidos foram avaliados visualmente e por eletroforese de poliacrilamida (SDS PAGE) 10%.

4 RESULTADOS E DISCUSSÃO

4.1 DESENHO E CONSTRUÇÃO DO PLASMÍDEO COM SEQUÊNCIA GÊNICA DA PROTEÍNA TALE ESPECÍFICA

Ao utilizar a sequência do gene *efeU* na ferramenta TAL Effector Targeter, foram desenhadas as melhores sequências alvo com base no tamanho, posição e localização, e a sequência selecionada foi a TCCGTCTTCGTCCTGGCAT (Figura 20), com 18 pb. Foi obtido o plasmídeo contendo a sequência gênica para a TALE específica para a sequência fornecida para a empresa *Thermo Scientific*.

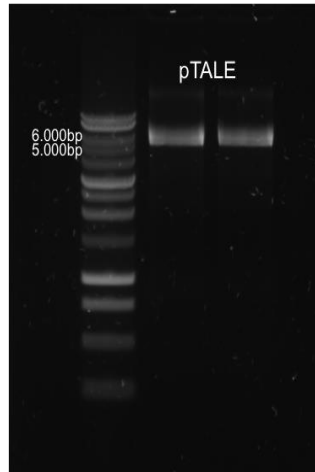
Figura 20 – Resultado obtido da ferramenta TAL Effector Targeter da Cornell University. Em vermelho foi selecionado o resultado utilizado para construção da proteína TALE.

Results:

Sequence Name	TAL start	TAL length	RVD sequence	Strand	Target sequence
Cs_TAL	66	19	HD HD NH NG HD NG NG HD NH NG HD HD NG NH NH HD NI NG NG	Plus	T CCGTCTTCGTCCTGGCA TT
Cs_TAL <i>efeU</i>	66	18	HD HD NH NG HD NG NG HD NH NG HD HD NG NH NH HD NI NG	Plus	T CCGTCTTCGTCCTGGCA T

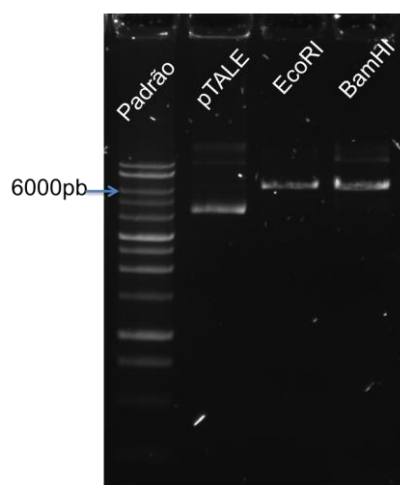
O plasmídeo obtido da empresa *Thermo Scientific* foi utilizado para transformação da linhagem XL1blue. Foi obtida uma concentração de 90,95ng/ul de DNA plasmidial após a extração de DNA de 5ml da cultura. Através da análise do gel de agarose 1%, foi visto que o plasmídeo de 6181 pb obtido da extração apresentou boa concentração e integridade (Figura 21).

Figura 21 – Plasmídeo pTALE obtido a partir do kit de extração de DNA plasmidial *NucleoSpin® Plasmid* (Macherey-Nagel). Canaleta 1: Padrão de peso molecular 1kB; Canaletas 2 e 3: pTALE.



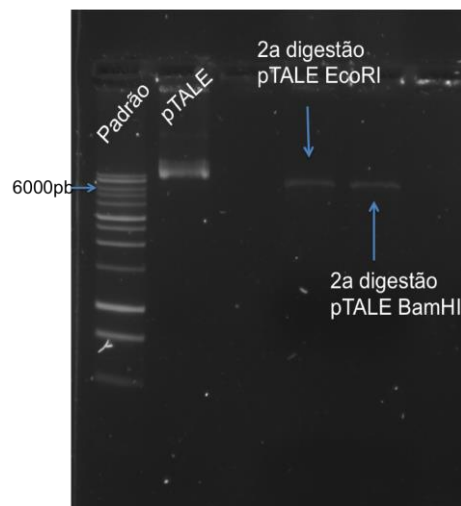
O plasmídeo pTALE foi então digerido com as enzimas EcoRI e BamHI, cada enzima por vez (digestões simples). A reação durou 16h e como pode ser visto no gel de agarose 1% na Figura 22, foram obtidas as bandas no tamanho esperado (6181 pb) correspondente ao plasmídeo linearizado. O plasmídeo não digerido pode ser visto na segunda canaleta, para comparação do resultado.

Figura 22 – Resultado das digestões do pTALE com as enzimas EcoRI e BamHI. O tamanho do plasmídeo pTALE é de 6181 pb. Canaleta 1: Padrão de peso molecular 1kB; Canaleta 2: pTALE; Canaleta 3: pTALE digerido com EcoRI; Canaleta 4: pTALE digerido com BamHI.



Na segunda reação de digestão, feita após a purificação da banda do gel de agarose, a amostra que havia sido digerida com EcoRI foi digerida com BamHI e a amostra digerida previamente com BamHI foi digerida com EcoRI. No gel de agarose (Figura 23) foi visualizada uma banda no tamanho esperado (6137 pb).

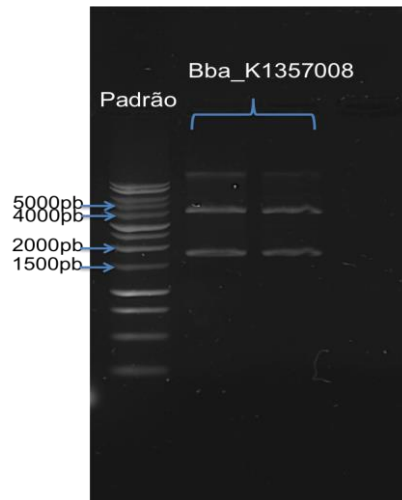
Figura 23 – Gel de agarose com amostras da segunda digestão simples do plasmídeo pTALE. Canaleta 1: Padrão de peso molecular de 1kB; Canaleta 2: pTALE; Canaleta 4: pTALE linearizado previamente com BamHI, digerido com EcoRI; Canaleta 4: pTALE linearizado previamente com EcoRI, digerido com BamHI.



4.2 OBTENÇÃO DA SEQUÊNCIA GÊNICA DA CROMOPROTEÍNA TINSEL PURPLE

A transformação do plasmídeo contendo a tsPurple na linhagem JM109 não apresentou crescimento de colônias. Por isso, foi feita a transformação em XL1blue. A extração do DNA plasmidial foi realizada e foi obtida uma dosagem média de 29ng/μl (Figura 24).

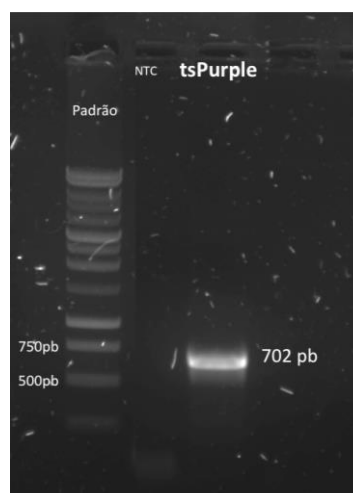
Figura 24 – Extração de DNA plasmidial do plasmídeo Bba_K1357008, contendo a sequência gênica da cromoproteína Tinsel Purple (duplicata - canaletas 2 e 3).



4.3 ISOLAMENTO DA SEQUÊNCIA GÊNICA *TSPURPLE* POR PCR CONVENCIONAL E DIGESTÃO DUPLA

Após a reação de PCR convencional da sequência gênica *tsPurple*, foi observado no gel de agarose 1% a banda esperada de 702 pb (Figura 25), correspondente ao gene e aos sítios de restrição *EcoRI* e *BamHI*, inseridos nas extremidades durante a amplificação.

Figura 25 – Resultado da amplificação por PCR do fragmento contendo o gene *tsPurple* a partir do plasmídeo Bba_K1357008. Como resultado, obteve-se uma banda esperada correspondente a 702 pb.



O produto do PCR foi purificado e foi submetido à uma digestão dupla com ambos EcoRI e BamHI. O produto purificado da digestão foi utilizado para a ligação do fragmento de PCR (*tsPurple*) com o plasmídeo pTALE.

4.4 LIGAÇÃO DO PLASMÍDEO pTALE E FRAGMENTO CONTENDO O GENE *TSPURPLE*

Após as digestões do plasmídeo pTALE e do produto de PCR contendo o gene *tsPurple*, foi feita a ligação. Para confirmar a ligação foram feitas reação de PCR convencional com os seguintes pares de primers: *tsPurple* forward e *tsPurple* reverse; e T7 forward e *tsPurple* reverse. A partir da análise dos resultados em gel de agarose (Figuras 26 e 27) foi possível confirmar a ligação através da visualização das bandas esperadas (702 pb para a reação com primers *tsPurple* forward e *tsPurple* reverse; e 907 pb para a reação com os primers T7 forward e *tsPurple* reverse) para os plasmídeos pTSP1 (produto da ligação).

Figura 26 – Resultado do gel de agarose 1% com o resultado do PCR convencional das amostras de ligação dos fragmentos obtidos por PCR com os primers *tsPurple* forward e *tsPurple* reverse. Canaleta 1: padrão de peso molecular 1kB; Canaleta 2: controle sem DNA molde; Canaleta 3: Ligação 1; Canaleta 4: Ligação 2; Canaleta 5: Ligação 3.

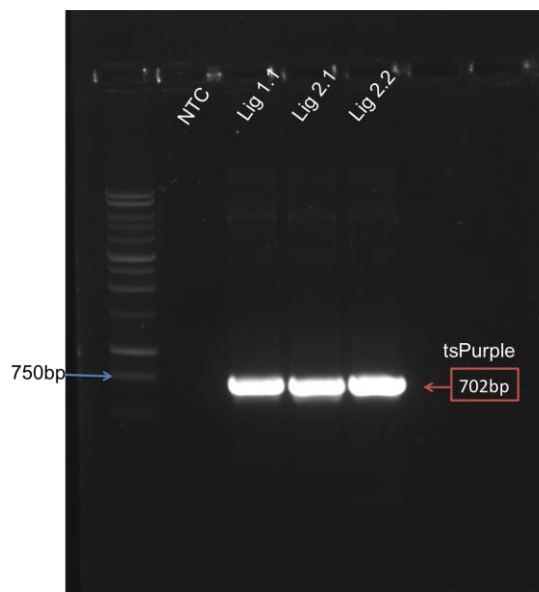
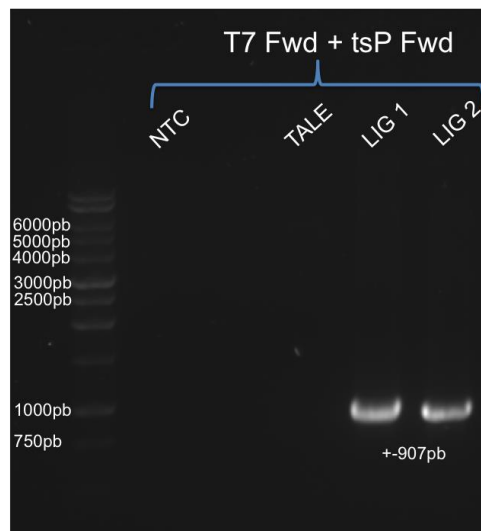


Figura 27 – Resultado do gel de agarose 1% com o resultado do PCR convencional das amostras de ligação dos fragmentos digeridos com os primers T7 forward e tsPurple reverse. Canaleta 1: padrão de peso molecular 1kB; Canaleta 2: controle negativo; Canaleta 3: produto da reação de PCR com pTALE; Canaleta 4: produto da reação de PCR com plasmídeo da primeira ligação; Canaleta 5: produto da reação de PCR com plasmídeo da segunda ligação.



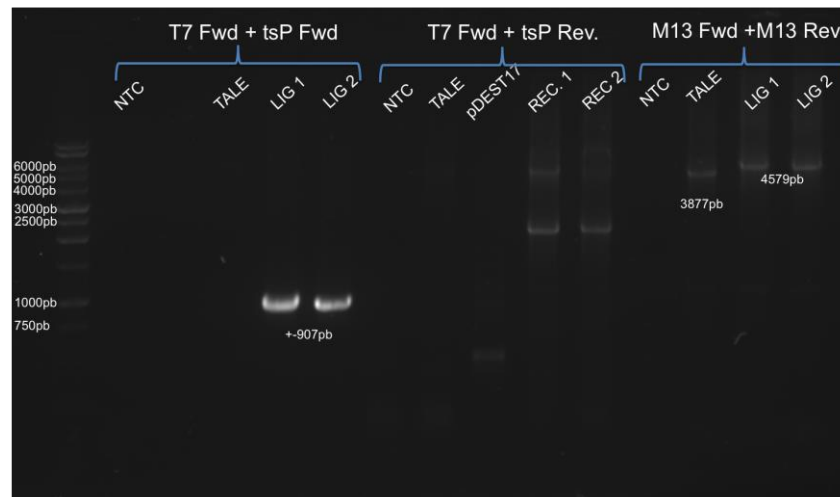
4.5 RECOMBINAÇÃO DOS PLASMÍDEOS PTSP1 E PDEST™ 17 PELO SISTEMA DE CLONAGEM GATEWAY

A primeira reação de recombinação realizada durou 1h com a *LR clonase II* e 15 minutos com a *Proteinase K*. Contudo, cresceram colônias dos clones tanto nas placas com ampicilina, como nas placas com kanamicina. Dessa forma, foi feito então a reação de PCR convencional para confirmação da recombinação e das ligações.

A partir da análise dos *backbones* dos plasmídeos utilizados no estudo, foi visto que o plasmídeo pTALE e o plasmídeo da ligação (pTSP1) que possui mesmo *backbone* pode ser amplificado com os primers M13 forward (-20) e M13 reverse, resultando em fragmentos de 3877 pb para o plasmídeo pTALE e 4579 pb para o plasmídeo pTSP1 (resultante da reação de ligação). Primers T7 forward e tsPurple reverse podem ser utilizados para amplificar o plasmídeo pTSP1, resultando em um fragmento de 907 pb. Já o plasmídeo pTSP2 (produto da recombinação) pode ser amplificado com os primers T7 forward e tsPurple reverse, resultando em um fragmento de 4574 pb. Com base nesses conhecimentos, os primers foram utilizados para a confirmação das ligações e recombinações. Dessa forma, como pode ser

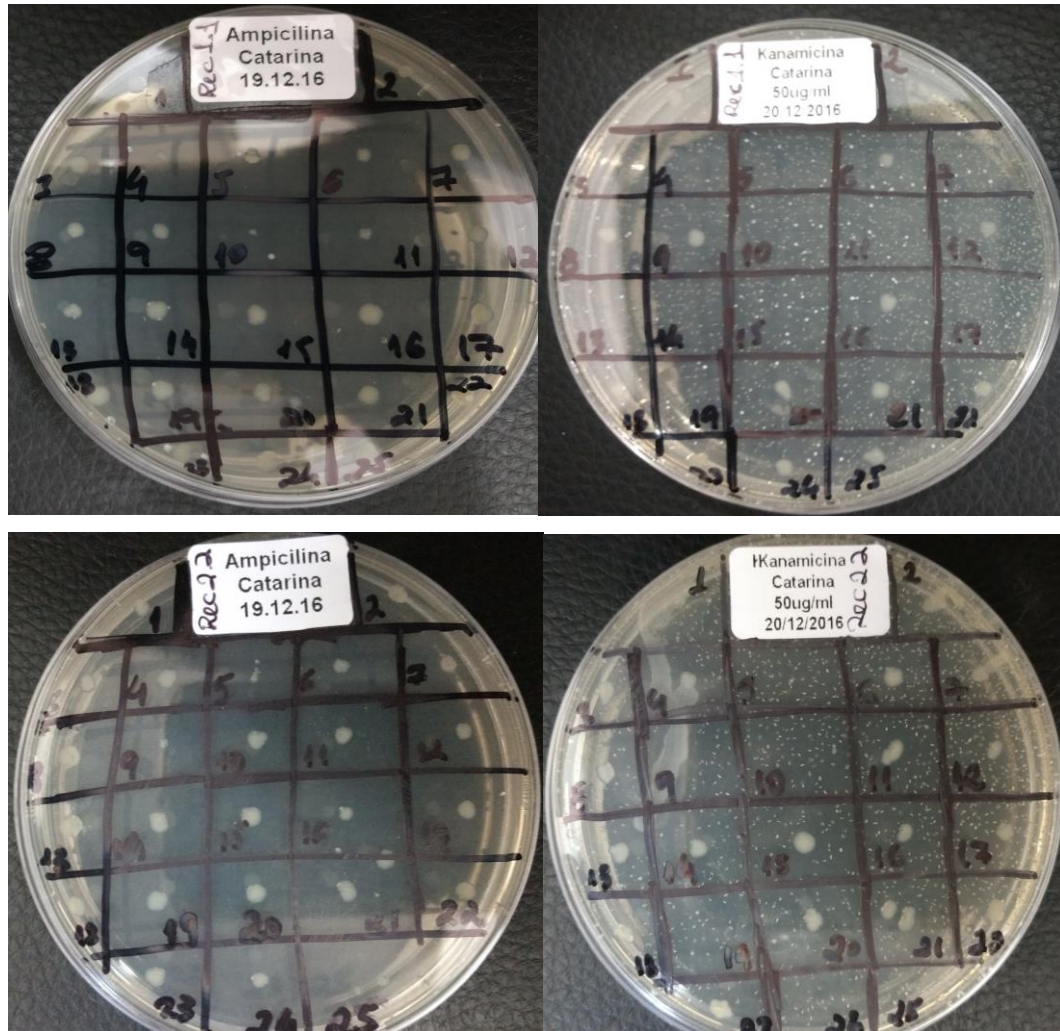
observado no gel de agarose da Figura 28, ficou confirmado então que de fato, foram obtidos os plasmídeos pTSP1. Isso pode ser confirmado com o aparecimento das bandas de 907 pb na reação 1 (primers T7 foward e tsPurple foward) e de 4579 pb na reação 2 (M13 foward e reverse). Entretanto, novos experimentos foram necessários para confirmação da recombinação.

Figura 28 – Resultado do gel de agarose 1% com o resultado do PCR convencional das amostras de ligação dos fragmentos digeridos e das primeiras recombinações. Foram utilizados os pares de primers: T7 foward e tsPurple foward (canaletas 2-5); T7 foward e tsPurple reverse (canaletas 6-10); e M13 foward e reverse (canaletas 11-14).



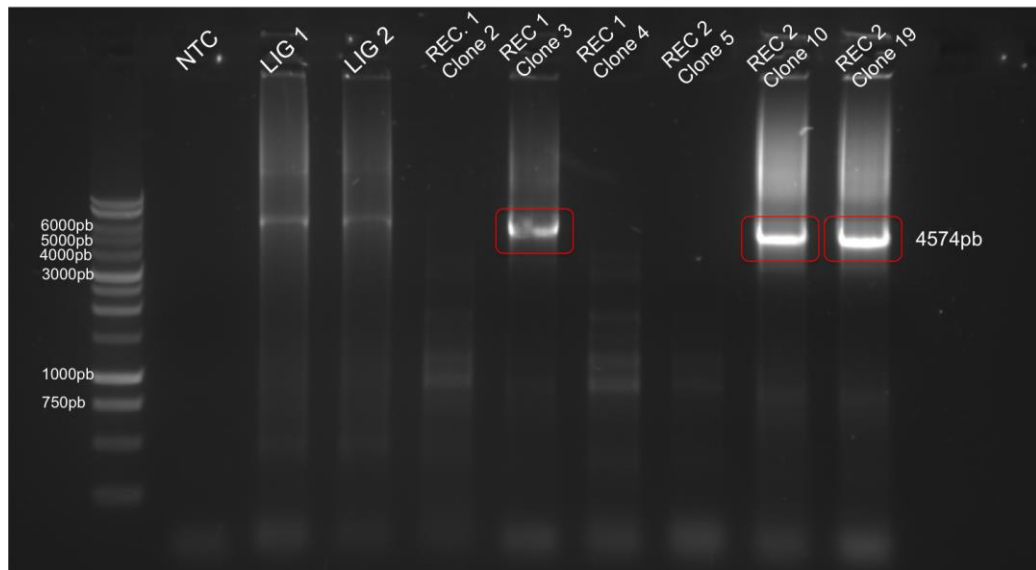
Com base nesses resultados, foram feitas então duas novas reações de recombinação de 18h (*overnight*) com a *LR clonase II* e 15min. com a *Proteinase K*. A reação 1 com a ligação 1 e a reação 2 com a ligação 2. Dessa vez, cresceram colônias em placas tanto com ampicilina como kanamicina. Foram selecionadas 25 colônias que cresceram na placa com ampicilina e foram inoculadas novamente em ampicilina e kanamicina, como mostrado na Figura 29.

Figura 29 – Fotos das placas de LB-ágar com ampicilina (à esquerda) e kanamicina (à direita) para seleção dos clones positivos para as recombinações.



Foram selecionados os clones 2, 3 e 4 da reação de recombinação 1, e os clones 5, 10 e 19 da reação de recombinação 2. Esses clones foram utilizados numa nova reação de PCR para confirmação da recombinação (Figura 30).

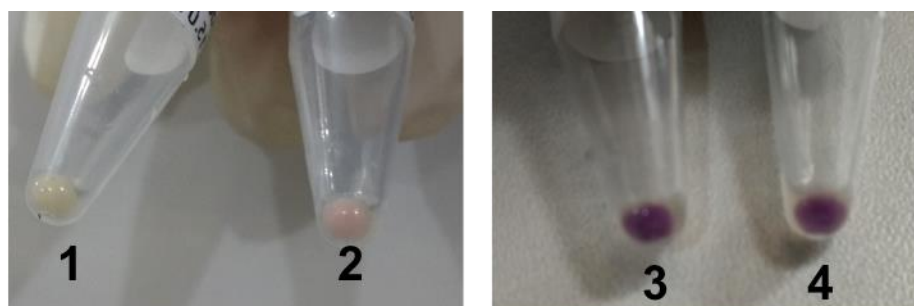
Figura 30 – Reação de PCR convencional utilizando os primers T7 forward e tsPurple reverse para confirmação das recombinações.



4.6 EXPRESSÃO DA PROTEÍNA QUIMÉRICA BIOSENSORA

A partir do resultado do PCR, foram considerados positivos os clones 3, 10 e 19. Esses clones foram então utilizados para transformação da linhagem Rosetta (DE3) PLysS de *E. coli*. Foi feita então a indução da expressão da proteína quimérica biossensora com 1mM de IPTG nas densidades óticas de 1,0 e 1,5. Observou-se então a emissão da coloração rosada no pellet da cultura com 4h de indução, e com 24h de indução, o precipitado da cultura bacteriana já apresentava coloração totalmente roxa (Figura 31).

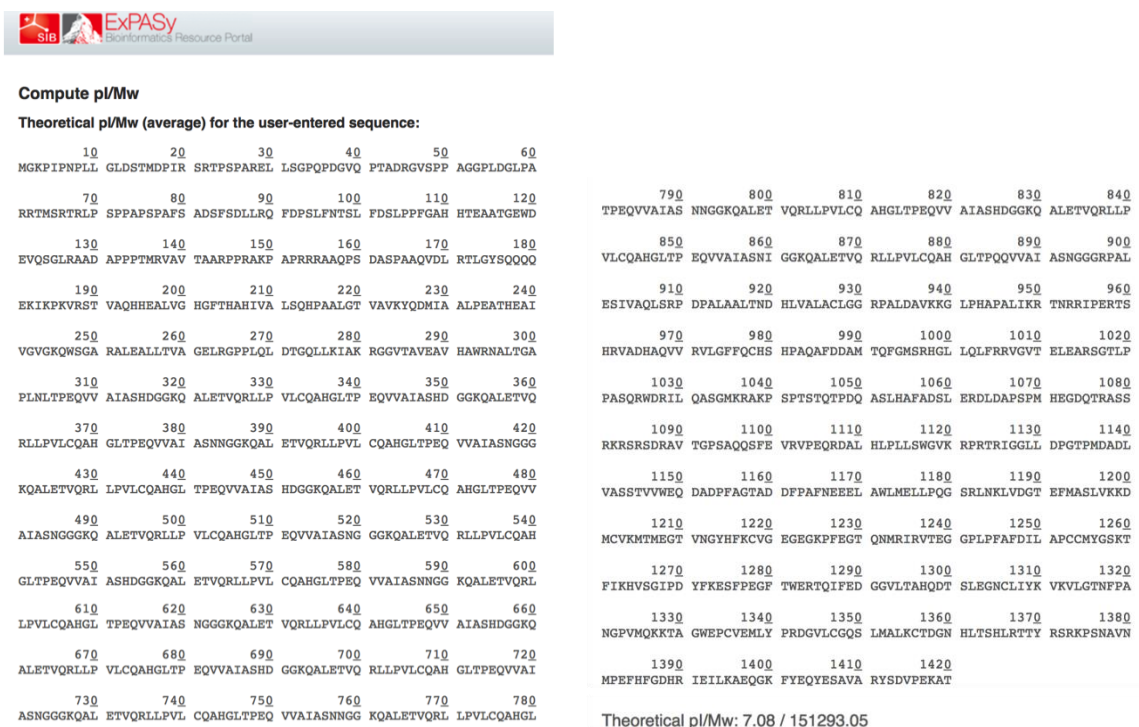
Figura 31 – Fotos dos *pellets* das culturas durante indução da expressão da proteína quimérica biossensora. Amostra 1: Não induzido (sem IPTG); Amostra 2: 4h de indução, Amostras 3 e 4: 24h de indução.



Apesar do clone 3 não ter apresentado coloração roxa, os clones 10 e 19 apresentaram a mudança de coloração no pellet da cultura, demonstrando que houve expressão eficiente da proteína quimérica. Então, foi realizado o SDS-PAGE para visualização dos extratos proteicos somente para os clones 10 e 19, que já haviam indicado resultado positivo visualmente.

O peso molecular da proteína quimérica foi estimado através da ferramenta Compute pI/Mw – ExPASy (Swiss Institute of Bioinformatics). Foi calculado um PM de aproximadamente 151 kDa e um ponto isoelétrico de 7,08 (Figura 32).

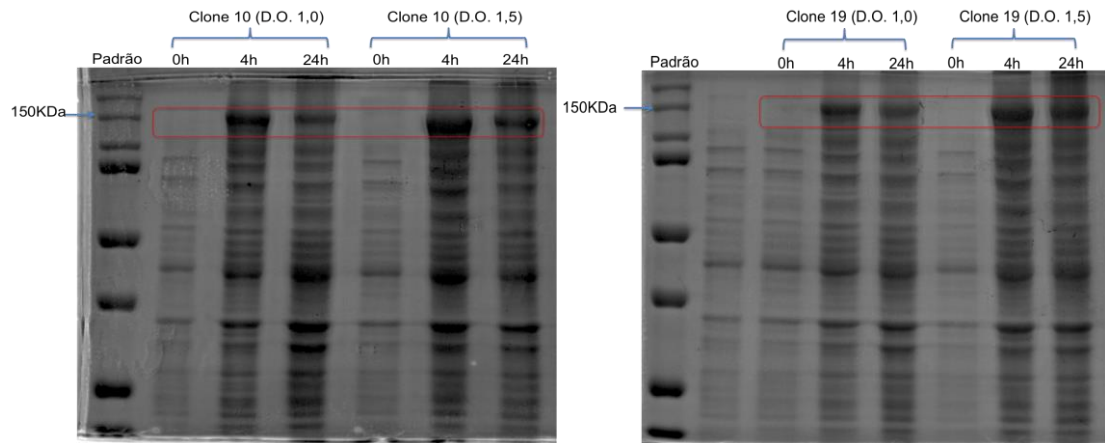
Figura 32 – Sequência de aminoácidos da proteína quimérica biossensora, ponto isoelétrico e peso molecular estimado.



Fonte: Compute pI/Mw – ExPASy – Swiss Institute of Bioinformatics
 (http://web.expasy.org/compute_pi/)

No gel de poliacrilamida (Figura 33) foi possível observar a super-expressão de proteínas com o peso molecular esperado (151 kDa) em ambos os tempos de indução 4h e 24h, tanto para o clone 10 como para o 19, confirmando mais uma vez a expressão da proteína quimérica.

Figura 33 – SDS-PAGE 10% das amostras de indução dos clones 10 (à esquerda) e 19 (à direita). Amostras coletadas nos tempos 0h (pré-IPTG), 4h e 24h pós adição de IPTG nas densidades óticas de 1,0 e 1,5.



Foi feita a análise *in silico* da sequência protéica nas ferramentas ProtParam (Swiss Institute of Bioinformatics) e Recombinant Protein Solubility Prediction da Universidade de Oklahoma. Os dados são apresentados abaixo.

Número total de resíduos negativamente carregados (Asp + Glu): 130

Número total de resíduos positivamente carregados (Arg + Lys): 127

Quantidade de cada resíduo de aminoácido (apolares: vermelho; polares: azul):

161 Alanine (A)

76 Arginine (R)

53 Aspartate (D)

31 Asparagine (N)

27 Cysteine (C)

109 Glutamine (Q)

77 Glutamate (E)

123 Glycine (G)

51 Histidine (H)

39 Isoleucine (I)

163 Leucine (L)

51 Lysine (K)

21 Methionine (M)

31 Phenylalanine (F)

108 Proline (P)

73 Serine (S)

86 Threonine (T)

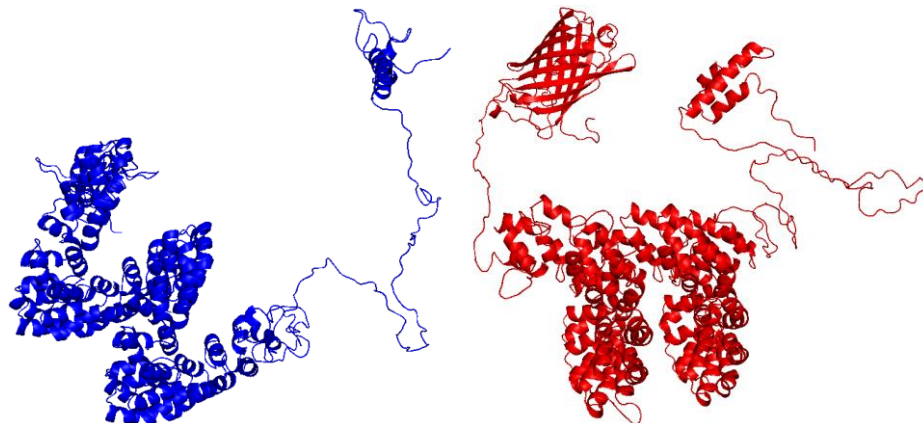
9 Tryptophan (W)

11 Tyrosine (Y)

120 Valine (V)

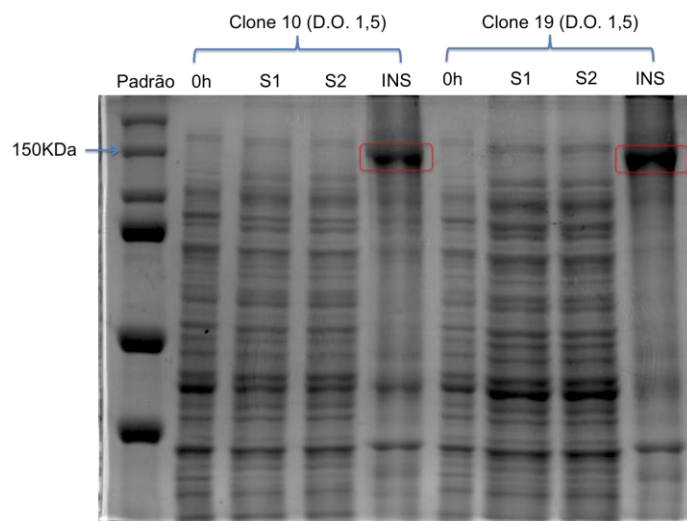
Analisando a sequência de aminoácidos predita da proteína quimérica, foi visto que a maioria dos resíduos são apolares. Além disso, foi obtido através da ferramenta Recombinant Protein Solubility Prediction da Universidade de Oklahoma que há baixa probabilidade da proteína ser produzida na forma solúvel quando super expressa em *Escherichia coli*. Na Figura 34 pode ser visualizada a predição das estruturas tridimensionais da proteína TALE e da proteína quimérica feita no programa RaptorX. Essa plataforma faz uma predição das estruturas secundária e terciária, contatos, solubilidade, regiões de desordem e sítios de ligação da proteína.

Figura 34 – Predição da estrutura tridimensional da proteína TALE (azul) e da proteína resultante da fusão da TALE com a Tinsel Purple (vermelho).



Por fim, foi realizado o teste de solubilidade da proteína quimérica produzida (Figura 35) a fim de subsidiar os experimentos futuros de obtenção da proteína purificada para aplicação diagnóstica. Foi então observado que a proteína quimérica é recuperada totalmente na fração insolúvel, após lise das linhagens recombinantes de *Escherichia coli*.

Figura 35 – Teste de solubilidade das amostras de indução dos clones 10 e 19, obtidos após recombinação.



Levando-se em consideração a insolubilidade da proteína produzida, é possível entender o motivo pelo qual as bandas de 24h de indução ficaram um pouco mais fracas do que as bandas de 4h de indução no gel de SDS PAGE. A alta concentração de proteína com 24h de indução, como foi visto a coloração forte na foto da Figura 33, pode ter levado a formação de agregados, não sendo então possível visualizar no gel, no qual foi utilizado alíquotas dos precipitados dissolvidos em solução aquosa. Então, no tempo de 4h de indução, foi observado uma maior concentração de proteína capaz de ser solubilizada (GUSTAFSSON et al., 2012).

Normalmente, genes sintéticos são otimizados para simplificar as etapas de clonagem e expressão. Porém, um motivo mais importante para essas modificações é que geralmente, os genes de origem natural não são expressos como esperado nos seus hospedeiros heterólogos. Inúmeros fatores podem se tornar empecilhos durante a produção de proteínas recombinantes em sistemas heterólogos, especialmente, durante a etapa de transcrição (GUSTAFSSON et al., 2012; GUSTAFSSON; GOVINDARAJAN; MINSHULL, 2004; WELCH et al., 2009).

Dessa forma, estudos adicionais se fazem necessários a fim de tentar entender o motivo da insolubilidade da proteína produzida. A insolubilidade pode ser devido a escolha do hospedeiro, a linhagem selecionada, ou até mesmo as condições de indução. Todos esses fatores precisam ser estudados um a um para que seja possível determinar se a proteína pode ser produzida na sua forma solúvel no sistema bacteriano. A produção em sistema bacteriano é uma estratégia com inúmeras vantagens (WELCH et al., 2009). A produção de proteínas em *E. coli* é simples, rápido e de baixo custo. Todas essas vantagens são importantes quando for considerados o custo total operacional da produção do sistema biossensor (HU et al., 2014; YING et al., 2017).

Existem alguns trabalhos com o propósito de identificação de sequências de DNA específicas com uso de proteínas ligadoras de DNA. Kim e colaboradores, em 2011, publicaram um estudo do uso de zinc fingers para detecção de DNA num modelo de imobilização em lâmina com polietilenoglicol (PEG). Os resultados obtidos foram bastante positivos. Contudo os domínios zinc fingers são mais difíceis de aplicar, devido a sua alta taxa de ligação *off-target* e sua baixa estabilidade devido ao íon zinco.

Outro trabalho que aborda o diagnóstico de um agente viral foi o de Pardee e colaboradores. Eles conseguiram criar um módulo baseado em CRISPR/Cas9 capaz de distinguir genótipos de Zika vírus com uma resolução de uma base. O sucesso desses estudos tem levado a crer que a Biologia Sintética irá cada vez mais desempenhar um importante papel para a medicina. Apesar de não existir nenhum relato na literatura até o presente momento de estudos de proteínas TALE para aplicação em diagnóstico, os resultados positivos obtidos até o momento levam a crer que a utilização da proteína quimérica construída tem potencial como os estudos previamente mencionados, já que muitos estudos apontam diversas vantagens da TALE em detrimento das zinc fingers e do sistema CRISPR (JANKELE; SVOBODA, 2014b; STREUBEL et al., 2012; YANG et al., 2014).

5 CONSIDERAÇÕES FINAIS

A descoberta de novas estratégias para diagnóstico rápido e de baixo custo são de grande importância para locais de difícil acesso a tecnologias de ponta. Por isso, estão cada vez mais sendo desenvolvidos novos biossensores e as metodologias sendo aprimoradas. Todos os dias novas moléculas são descobertas, entretanto, vale a pena explorar mais as moléculas com potencial para métodos moleculares e de engenharia genética. As proteínas do tipo TALE tem demonstrado uma especificidade ao alvo alta e por esse motivo, tem sido utilizadas com sucesso na área de edição gênica. Porém, o uso em diagnóstico ainda não foi descrito, o que reforça a importância do presente estudo.

A utilização de cromoproteínas como estratégia de detecção visual do biossensor é uma ótima opção para reduzir o custo operacional do dispositivo. As cromoproteínas produzem cor naturalmente e é possível visualizar sem necessidade de equipamentos especiais, como é o caso da fluorescência. Dessa forma, o estudo traz uma proposta interessante de dispositivo biossensor, que se baseia numa proteína TALE para detecção de uma sequência de DNA alvo fusionada a uma cromoproteína. Assim, espera-se que a fusão detecte o alvo e o resultado positivo seja visualizado com a emissão de uma coloração.

Com o presente estudo foi possível construir uma proteína quimérica com eficiência que demonstrou atividade funcional como esperado. Esses resultados são bastante promissores para conduzir o trabalho para etapas posteriores, como a avaliação da ligação da proteína quimérica ao alvo, sua estabilidade e montagem de um sistema biossensor.

Como perspectivas futuras, pode ser feita ainda a purificação da proteína produzida por cromatografia de afinidade, bem como avaliação da sua meia-vida e o teste de identificação da sequência alvo. Experimentos adicionais de Electrophoretic Mobility Assay (EMSA) estão previstos para validar a aplicação da proteína quimérica em sistema biossensor.

REFERÊNCIAS

- ANDRIANANTOANDRO, E. et al. Synthetic biology: New engineering rules for an emerging discipline. **Molecular Systems Biology**, v. 2, 2006.
- BANEYX, F. Recombinant protein expression in Escherichia coli. **Current Opinion in Biotechnology**, v. 10, p. 411–421, 1999.
- BERG, J. M.; SHI, Y. The Galvanization of Biology: Appreciation for the Roles of Growing Zinc. **Science**, v. 271, n. 23 February, p. 1081–1085, 1996.
- BERNARD, K. The genus *Corynebacterium* and other medically relevant coryneform-like bacteria. **Journal of Clinical Microbiology**, v. 50, n. 10, p. 3152–3158, 2012.
- BOCH, J. et al. Breaking the code of DNA binding specificity of TAL-type III effectors. **Science**, v. 326, n. 5959, p. 1509–1512, 2009.
- BOCH, J.; BONAS, U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. **Annual review of phytopathology**, v. 48, p. 419–436, 2010.
- BOGDANOVA, A. J.; SCHORNACK, S.; LAHAYE, T. TAL effectors: Finding plant genes for disease and defense. **Current Opinion in Plant Biology**, v. 13, n. 4, p. 394–401, 2010.
- BOGDANOVA, A. J.; VOYTAS, D. F. TAL Effectors: Customizable Proteins for DNA Targeting. **Science**, v. 333, p. 1843–1847, 2011.
- CAMERON, D. E.; BASHOR, C. J.; COLLINS, J. J. A brief history of synthetic biology. **Nature reviews. Microbiology**, v. 12, n. 5, p. 381–390, 2014.
- CASPER-LINDLEY, C. et al. Direct biochemical evidence for type III secretion-dependent translocation of the AvrBs2 effector protein into plant cells. **PNAS**, v. 99, n. 12, p. 8336–8341, 2002.
- CERMAK, T. et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. **Nucleic Acids Research**, v. 39, n. 12, p. 1–12, 2011.
- COOLING, M. T. et al. Standard virtual biological parts: A repository of modular modeling components for synthetic biology. **Bioinformatics**, v. 26, n. 7, p. 925–931, 2010.
- CORREA, A.; OPPEZZO, P. Tuning different expression parameters to achieve soluble recombinant proteins in E. coli: advantages of high-throughput screening. **Biotechnology journal**, v. 6, n. 6, p. 715–30, jun. 2011.
- DENG, D. et al. Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors. **Science**, v. 335, n. 6069, p. 720–723, 2012.
- ELLERBEE, A. K. et al. Quantifying Colorimetric Assays in Paper-Based Microfluidic Devices by Measuring the Transmission of Light through Paper. **Analytical Chemistry**, v. 81, n. 20, p. 8447–8452, 2009.
- ENDY, D. Foundations for engineering biology. **Nature Reviews**, v. 438, n. 7067, p. 449–453, 2005.
- GALAZKA, A. The changing epidemiology of diphtheria in the vaccine era. **The Journal of infectious diseases**, v. 181 Suppl, n. Supplement 1, p. S2-9, 2000.
- GAO, H. et al. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. **Cell research**, v. 22, n. 12, p. 1716–20, 2012.
- GAO, X. et al. Visual detection of microRNA with lateral flow nucleic acid biosensor. **Biosensors and Bioelectronics**, v. 54, p. 578–584, 2014.
- GHOSH, I. et al. Direct detection of double-stranded DNA: molecular methods and applications for DNA diagnostics. **Molecular Biosystems**, v. 2, n. 11, p. 551–560, 2006.

- GLICK, B. R. Metabolic load and heterologous gene expression. **Biotechnology advances**, v. 13, n. 2, p. 247–261, 1995.
- GUSTAFSSON, C. et al. Engineering genes for predictable protein expression. **Protein expression and purification**, v. 83, n. 1, p. 37–46, maio 2012.
- GUSTAFSSON, C.; GOVINDARAJAN, S.; MINSHULL, J. Codon bias and heterologous protein expression. v. 22, n. 7, 2004.
- HAHN, W. O. et al. Multidrug-Resistant Associated with Increased Use of Parenteral Antimicrobial Drugs. v. 22, n. 11, 2017.
- HU, J. et al. Advances in paper-based point-of-care diagnostics. **Biosensors and Bioelectronic**, v. 54, p. 585–597, 2014.
- JANKELE, R.; SVOBODA, P. TAL effectors: tools for DNA Targeting. **Briefings in Functional Genomics**, v. 13, n. 5, p. 409–419, 2014a.
- JANKELE, R.; SVOBODA, P. TAL effectors: tools for DNA Targeting. **Briefings in Functional Genomics**, v. 13, n. 5, p. 409–419, 2014b.
- JINEK, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. **Science**, v. 337, p. 816–821, 2012.
- KABADI, A. M.; GERSBACH, C. A. Engineering Synthetic TALE and CRISPR/Cas9 Transcription Factors for Regulating Gene Expression. **Methods**, v. 69, n. 2, p. 188–197, 2014.
- KHWANNIMIT, B.; BHURAYANONTACHAI, R. The direct costs of intensive care management and risk factors for financial burden of patients with severe sepsis and septic shock ☆. **Journal of Critical Care**, v. 30, n. 5, p. 929–934, 2015.
- KIM, M. et al. A zinc finger protein array for the visual detection of specific DNA sequences for diagnostic applications. **Nucleic acids research**, v. 39, n. 5, p. 1–9, 2011.
- KOBAYASHI, H. et al. Programmable cells: interfacing natural and engineered gene networks. **Proceedings of the National Academy of Sciences of the United States of America**, v. 101, n. 22, p. 8414–8419, 2004.
- KONRAD, R.; HÖRMANSDORFER, S.; SING, A. Possible human-to-human transmission of toxigenic *Corynebacterium ulcerans*. **Clinical Microbiology and Infection**, v. 21, n. 8, p. 768–771, 2015.
- KOSER, C. U. et al. Routine Use of Microbial Whole Genome Sequencing in Diagnostic and Public Health Microbiology. **PLoS Pathogens**, v. 8, n. 8, p. 1–9, 2012.
- KRISHNA, S. S.; MAJUMDAR, I.; GRISHIN, N. V. Structural classification of zinc fingers. **Nucleic Acids Research**, v. 31, n. 2, p. 532–550, 2003.
- LAMB, B. M.; MERCER, A. C.; BARBAS, C. F. Directed evolution of the TALE N-terminal domain for recognition of all 5' bases. **Nucleic acids research**, v. 41, n. 21, p. 9779–9785, 2013.
- LANGE, O. DE; BINDER, A.; LAHAYE, T. From dead leaf , to new life: TAL effectors as tools for synthetic biology. **The Plant Journal**, v. 78, p. 753–771, 2014.
- MAHFOUZ, M. M. et al. De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. **PNAS**, v. 108, n. 6, p. 2623–2628, 2011.
- MAHFOUZ, M. M.; PIATEK, A.; STEWART JR, C. N. Genome engineering via TALENs and CRISPR / Cas9 systems: challenges and perspectives. **Plant Biotechnology Journal**, v. 12, p. 1006–1014, 2014.
- MAK, A. N. et al. The crystal structure of TAL effector PthXo1 bound to its DNA target.

Science, v. 335, n. 6069, p. 716–719, 2012.

MALI, P. et al. RNA-guided Human Genome Engineering via Cas9. **Science**, v. 339, p. 823–826, 2013.

MATTOS-GUARALDI, A. L. et al. a Multidrug-Susceptible Strain Isolated from the Urine of a Hospitalized 37-Year-Old Female Patient. **Genome Announcements**, v. 3, n. 4, p. 1–2, 2015.

MECKLER, J. F. et al. Quantitative analysis of TALE-DNA interactions suggests polarity effects. **Nucleic Acids Research**, v. 41, n. 7, p. 4118–4128, 2013.

MERCANTE, F. M.; RUMJANEK, N. G.; FRANCO, A. A. Aplicação de genes marcadores em estudos de ecologia microbiana com ênfase no sistema GUS. **Ciência Rural**, v. 30, n. 3, p. 533–539, 2000.

MILLER, J. C. et al. A TALE nuclease architecture for efficient genome editing. **Nature Biotechnology**, v. 29, n. 2, p. 143–148, 2011.

MOSCOU, M.; BOGDANOVA, A. A simple cipher governs DNA recognition by TAL effectors. **Science**, v. 326, p. 1501, 2009.

MUÑOZ BODNAR, A. et al. Tell me a tale of TALEs. **Molecular Biotechnology**, v. 53, n. 2, p. 228–235, 2013.

NAYLOR, L. H. Reporter gene technology: the future looks bright. **Biochemical pharmacology**, v. 58, n. 5, p. 749–757, 1999.

NGA-SZE MAK, A. et al. TAL effectors: function, structure, engineering and applications. **Current opinion in structural biology**, v. 23, n. 1, p. 93–99, 2013.

PACHECO, L. G. C. et al. Draft Genome Sequences of Two Species of “Difficult-to-Identify” Human-Pathogenic Corynebacteria: Implications for Better Identification Tests. **Journal of Genomics**, v. 3, p. 82–84, 2015.

PADILLA-VACA, F.; ANAYA-VELÁZQUEZ, F.; FRANCO, B. Synthetic biology: Novel approaches for microbiology. **International Microbiology**, v. 18, p. 71–84, 2015.

PAVLETICH, N. P.; PABO, C. O. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å. **Science**, v. 252, n. May, p. 809–817, 1991.

POLITZ, M. C.; COPELAND, M. F.; PFLEGER, B. F. Artificial repressors for controlling gene expression in bacteria. **Chem. Commun.**, v. 49, n. 39, p. 4325–4327, 2013.

RICE, W. G.; TURPIN, J. A. Virus-encoded Zinc Fingers as Targets for Antiviral Chemotherapy. **Reviews in Medical Virology**, v. 6, p. 187–199, 1996.

ROSANO, G. L.; CECCARELLI, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. **Frontiers in microbiology**, v. 5, n. April, p. 172, jan. 2014.

SANTOS, C. S. ; MOREIRA, C. A. ; MEYER, R. ; PINHEIRO, CARINA SILVA ; NEVES, N. M. A. ; PACHECO, LGC . **Oligonucleotídeos iniciadores específicos para identificação molecular e detecção de bactérias patogênicas emergentes do gênero Corynebacterium**. 2017, Brasil. Patente: Privilégio de Inovação. Número do registro: BR1020170002373, título: "Oligonucleotídeos iniciadores específicos para identificação molecular e detecção de bactérias patogênicas emergentes do gênero Corynebacterium" , Instituição de registro: INPI - Instituto Nacional da Propriedade Industrial. Depósito: 05/01/2017

SCHOLZE, H.; BOCH, J. TAL effectors are remote controls for gene activation. **Current Opinion in Microbiology**, v. 14, n. 1, p. 47–53, 2011.

SEZONOV, G.; JOSELEAU-PETIT, D.; D'ARI, R. Escherichia coli physiology in Luria-Bertani broth. **Journal of bacteriology**, v. 189, n. 23, p. 8746–9, dez. 2007.

- SHIH, C.-H. et al. Purple Chromoprotein Gene Serves as a New Selection Marker for Transgenesis of the Microalga *Nannochloropsis oculata*. **PloS one**, v. 10, n. 3, p. e0120780, 2015.
- SLOMOVIC, S.; PARDEE, K.; COLLINS, J. J. Synthetic biology devices for in vitro and in vivo diagnostics. **Proceedings of the National Academy of Sciences**, v. 112, n. 47, p. 201508521, 2015.
- STRAUS, A.; LAHAYE, T. Zinc fingers, TAL effectors, or Cas9-based DNA binding proteins: What's best for targeting desired genome loci? **Molecular Plant**, v. 6, n. 5, p. 1384–1387, 2013.
- STREUBEL, J. et al. TAL effector RVD specificities and efficiencies. **Nature Biotechnology**, v. 30, n. 7, p. 593–595, 2012.
- SUN, N. et al. Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. **Molecular BioSystems**, v. 8, n. 4, p. 1255, 2012.
- SUWANTARAT, N. et al. Practical Utility and Accuracy of Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry for Identification of *Corynebacterium* Species and Other Medically Relevant Coryneform-Like Bacteria. **American Journal of Clinical Pathology**, v. 145, n. 1, p. 22–28, 2016.
- TELES, F. R. R.; FONSECA, L. P. Trends in DNA biosensors. **Talanta**, v. 77, n. 2, p. 606–623, 2008.
- TERPE, K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. **Applied microbiology and biotechnology**, v. 72, n. 2, p. 211–22, set. 2006.
- URNOV, F. D. et al. Genome editing with engineered zinc finger nucleases. **Nature Reviews Genetics**, v. 11, n. 9, p. 636–646, 2010.
- VILA, J. et al. Identification of clinically relevant *Corynebacterium* spp., *Arcanobacterium haemolyticum*, and *Rhodococcus equi* by matrix-assisted laser desorption ionization-time of flight mass spectrometry. **Journal of Clinical Microbiology**, v. 50, n. 5, p. 1745–1747, 2012.
- WANG, J. From DNA biosensors to gene chips. **Nucleic acids research**, v. 28, n. 16, p. 3011–3016, 2000.
- WANG, Y.; WEI, K. Y.; SMOLKE, C. D. Synthetic biology: advancing the design of diverse genetic systems. **Annual Review of Chemistry and Biomolecular Engineering**, n. 1, p. 69–102, 2013.
- WEI, T.; CHENG, C. Synthetic Biology-Based Point-of-Care Diagnostics for Infectious Disease. **Cell Chemical Biology**, v. 23, n. 9, p. 1056–1066, 2016.
- WELCH, M. et al. You're one in a googol: optimizing genes for protein expression. **The royal society interface**, v. 6, n. March, p. S467–S476, 2009.
- WIEDENHEFT, B.; STERNBERG, S. H.; DOUDNA, J. A. RNA-guided genetic silencing systems in bacteria and archaea. **Nature**, v. 482, n. 7385, p. 331–338, 2012.
- YAGER, P. et al. Microfluidic diagnostic technologies for global public health. **Nature Insight Review**, v. 442, n. July, 2006.
- YANG, J. et al. Complete decoding of TAL effectors for DNA recognition. **Cell Research**, v. 24, n. 5, p. 628–631, 2014.
- YING, N. et al. Visual detection of nucleic acids based on lateral flow biosensor and hybridization chain reaction amplification. **Talanta**, v. 164, n. October 2016, p. 432–438, 2017.

APÊNDICE I – SEQUÊNCIA DO GENE *efeU* de *Corynebacterium striatum*

Gene *efeU* (228pb):

```
CTTCGAAGAACATGAAGGCAGAACTCGAGGGAAAGCTGGAGGCGGGGCCATCGCAGT  
GGGCGCGGGTTCCGTCTTCGTCCTGGCATTTCATGGCCGCGGTCCGCGAGGGCATCGA  
GACAGCGCTGCTGGTGTTTCGATACCTTTGCCTACGGCTCCACCATGACGCCGGCGCTC  
GGCCTTTCCCTAGGCATCCTCATCGCCGTAGCCGTAGCCGTAGCGATGTACTACGG
```

APÊNDICE II – SEQUÊNCIA GÊNICA DA CROMOPROTEÍNA TINSEL PURPLE

>BBa_K1357008 Part-only sequence (690 bp)

ATGGCGAGCTTGGTTAAGAAAGATATGTGTGTTAAGATGACGATGGAAGGTA CTGTGAA
CGGTTATCACTTTAAGTGC GTTGGCGAGGGTGAAGGCAAGCCGTTTCGAGGGCACGCAG
AACATGCGCATTTCGTGTCACCGAGGGCGGTCCGCTGCCTTTTGCATTTCGACATCCTGG
CCCCGTGCTGTATGTACGGCTCTAAGACCTTCATTAACACGTGAGCGGTATCCCGGAT
TACTTTAAAGAGTCCTTTCCAGAGGGCTTCACTTGGGAACGTACCCAGATTTTTTGAGGA
CGGTGGTGTCTGACCGCGCACCAAGACACCAGCCTGGAAGGTAATTGCCTGATCTAT
AAAGTGAAGGTTCTGGGTACCAATTTCCCGGCGAATGGTCCGGTGATGCAAAAGAAAAC
CGCGGGTTGGGAGCCGTGCGTTCGAGATGCTGTATCCGCGTGACGGCGTCTTGTGTGG
TCAGAGCTTGATGGCGCTGAAGTGCACCGATGGCAATCATCTGACCAGCCACCTGCGC
ACGACGTATCGTAGCCGTAACCGAGCAACGCCGTTAACATGCCGGAGTTCATTTTGG
TGACCATCGCATCGAAATCCTGAAAGCTGAGCAGGGCAAATTCTACGAACAATACGAAT
CGGCTGTCGCACGTTACAGCGATGTGCCGGAAAAAGCGACGTAATAA

APÊNDICE III – SEQUÊNCIA DOS PRIMERS UTILIZADOS NO ESTUDO

tsPurple Foward	GCGAATTCATGGCGAGCTTGGTTAAGAA
tsPurple Reverse	GCGGATCCTTATTACGTCGCTTTTTCCG
T7 Foward	TAATACGACTCACTATAGGG
M13 Foward (-20) Universal	GTAAAACGACGGCCAG
M13 Reverse Universal	CAGGAAACAGCTATGAC

APÊNDICE IV – SEQUÊNCIA DO PLASMÍDEO pTALE

pTALE (6181pb):

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGA
TACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGA
AGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGC
TGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATACGC
GTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAAAAGGCCATCCGTCAGGATGGCC
TTCTGCTTAGTTTTGATGCCTGGCAGTTTATGGCGGGCGTCCTGCCCGCCACCCTCCGG
GCCGTTGCTTACAACGTTCAAATCCGCTCCCGGCGGATTTGTCCTACTCAGGAGAGC
GTTACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTCCGACTGAGCCTTTTCGTT
TTATTTGATGCCTGGCAGTTCCTACTCTCGCGTTAACGCTAGCATGGATGTTTTCCAG
TCACGACGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCCAAATAATGATTTTATTT
TGA CTGATAGTGACCTGTTGTTGCAACAAATTGATGAGCAATGCTTTTTTATAATGCCA
ACTTTGTACAAAAAGCAGGCTGCGGCCGCGCCACCATGGGAAAACCTATTCCTAATCC
TCTGCTGGGCCTGGATTCTACCATGGACCCTATTAGAAGCAGAACACCCTCTCCAGCCA
GAGAACTGCTGTCTGGACCTCAGCCTGATGGAGTGCAGCCTACAGCCGATAGAGGAGT
GTCTCCTCCTGCCGGAGGACCTCTGGATGGACTGCCTGCCCGGAGAACAATGAGCAGA
ACAAGACTGCCTTCTCCTCCAGCCCCATCTCCTGCCTTTTCTGCCGATTCTTTTAGCGAC
CTGCTGAGACAGTTTGACCCCAGCCTGTTTAATACCAGCCTGTTTCGATAGCCTGCCTCC
TTTTGGAGCCCACCACACAGAGGCCGCCACAGGCGAATGGGATGAAGTGCAGTCTGGA
CTGAGAGCCGCCGATGCCCTCCTCTACAATGAGAGTGGCCGTGACAGCCGCCAGA
CCTCCTAGAGCCAAACCTGCCCTAGAAGGAGAGCCGCCAGCCTTCTGATGCCTCTC
CTGCCGCCAGGTGGACCTGAGAACACTGGGATATTCTCAGCAGCAGCAGGAGAAGAT
CAAGCCCAAGGTGAGGTCTACAGTGGCCCAGCACCACGAAGCCCTGGTGGGACACGG
ATTTACACACGCCACATTGTGGCCCTGTCTCAGCACCCCTGCCGCCCTGGGAACAGTG
GCCGTGAAATATCAGGATATGATTGCCGCCCTGCCTGAGGCCACACACGAAGCCATTG
TGGGAGTGGGAAAACAGTGGTCTGGAGCCAGAGCCCTGGAAGCCCTGCTGACAGTGG
CCGGAGA ACTGAGAGGACCTCCTCTGCAGCTGGATACAGGACAGCTGCTGAAGATTGC
CAAAGGGGGCGGAGTGACCGCGGTGGAAGCCGTGCACGCCTGGAGAAATGCCCTGAC
AGGAGCCCCTCTGAACCTGACCCCCGAACAGGTGGTGGCCATTGCCAGCCACGACGG
CGGCAAGCAGGCCCTGGAAACCGTGCAGAGACTGCTGCCCGTGCTGTGCCAGGCCCA
TGGCCTGACACCTGAACAGGTGGTGGCTATCGCCTCTCACGACGGAGGAAAACAGGCT
CTGGAAACAGTGCAGCGGCTGCTGCCTGTGCTGTGTCAGGCTCACGGCTTGA CTCCAG
AACAGGTGGTGGCTATTGCTTCCAACAACGGGGGGAAACAGGCCCTGGAAACTGTGCA
GCGCCTGCTGCCAGTGCTGTGCCAGGCTCACGGACTGACCCCCGAACAGGTGGTGGC

CATTGCCAGCAACGGCGGGCGGCAAGCAGGCCCTGGAAACCGTGCAGAGACTGCTGCC
CGTGCTGTGCCAGGCCCATGGCCTGACACCTGAACAGGTGGTGGCTATCGCCTCTCAC
GACGGAGGAAAACAGGCTCTGGAAACAGTGCAGCGGCTGCTGCCTGTGCTGTGTCAG
GCTCACGGCTTGACTCCAGAACAGGTGGTGGCTATTGCTTCCAACGGCGGGGGGAAAC
AGGCCCTGGAACTGTGCAGCGCCTGCTGCCAGTGCTGTGCCAGGCTCACGGGCTGA
CCCCGAACAGGTGGTGGCCATTGCCAGCAACGGCGGGCGGCAAGCAGGCCCTGGAAA
CCGTGCAGAGACTGCTGCCCGTGCTGTGCCAGGCCCATGGCCTGACACCTGAACAGG
TGGTGGCTATCGCCTCTCACGACGGAGGAAAACAGGCTCTGGAAACAGTGCAGCGGCT
GCTGCCTGTGCTGTGTCAGGCTCACGGCTTGACTCCAGAACAGGTGGTGGCTATTGCT
TCCAACAACGGGGGGAAACAGGCCCTGGAACTGTGCAGCGCCTGCTGCCAGTGCTG
TGCCAGGCTCACGGCCTCACTCCCGAACAGGTGGTGGCCATTGCCAGCAACGGCGGC
GGCAAGCAGGCCCTGGAAACCGTGCAGAGACTGCTGCCCGTGCTGTGCCAGGCCCAT
GGCCTGACACCTGAACAGGTGGTGGCTATCGCCTCTCACGACGGAGGAAAACAGGCTC
TGAAACAGTGCAGCGGCTGCTGCCTGTGCTGTGTCAGGCTCACGGCTTGACTCCAGA
ACAGGTGGTGGCTATTGCTTCCCACGACGGGGGGAAACAGGCCCTGGAACTGTGCA
GCGCCTGCTGCCAGTGCTGTGCCAGGCTCACGGACTGACCCCCGAACAGGTGGTGGC
CATTGCCAGCAACGGCGGGCGGCAAGCAGGCCCTGGAAACCGTGCAGAGACTGCTGCC
CGTGCTGTGCCAGGCCCATGGCCTGACACCTGAACAGGTGGTGGCTATCGCCTCTAAC
AACGGAGGAAAACAAGCACTCGAGACAGTGCAGCGGCTGCTGCCTGTGCTGTGTCAGG
CTCACGGCTTGACTCCAGAACAGGTGGTGGCTATTGCTTCCAACAACGGGGGGAAACA
GGCCCTGGAACTGTGCAGCGCCTGCTGCCAGTGCTGTGCCAGGCTCACGGGCTGAC
CCCCGAACAGGTGGTGGCCATTGCCAGCCACGACGGCGGGCAAGCAGGCCCTGGAAAC
CGTGCAGAGACTGCTGCCCGTGCTGTGCCAGGCCCATGGCCTGACACCTGAACAGGT
GGTGGCTATCGCCTCTAATATCGGAGGAAAACAGGCTCTGGAAACAGTGCAGCGGCTG
CTGCCTGTGCTGTGTCAGGCTCACGGCTTGACTCCACAGCAGGTCGTGGCAATTGCTA
GCAACGGCGGGCGGACGGCCCGCCCTGGAGAGCATTGTGGCCCAGCTGTCTAGACCTG
ATCCTGCCCTGGCCGCCCTGACAAATGATCACCTGGTGGCCCTGGCCTGTCTGGGAGG
CAGACCTGCCCTGGATGCCGTGAAAAAAGGACTGCCTCACGCCCTGCCCTGATCAAG
AGAACAATAGAAGAATCCCCGAGCGGACCTCTCACAGAGTGGCCGATCACGCCCAGG
TGGTGAGAGTGCTGGGATTTTTTCAAGTGTCACTCTCACCTGCCAGGCCCTTTGATGAT
GCCATGACACAGTTTGGCATGAGCAGACACGGACTGCTGCAGCTGTTTAGAAGAGTGG
GAGTGACAGAACTGGAGGCCAGAAGCGGAACACTGCCTCCAGCCTCTCAGAGATGGG
ATAGAATCCTGCAGGCCAGCGGAATGAAGAGAGCCAAACCTTCTCCTACCAGCACCCA
GACACCTGATCAGGCCAGCCTGCACGCCTTTGCCGATTCTCTGGAAAGGGATCTGGAC
GCCCTTCTCCTATGCACGAAGGAGATCAGACAAGAGCCAGCAGCAGAAAGAGAAGCA
GGTCTGATAGAGCCGTGACAGGACCTTCTGCCCAGCAGTCTTTTGAAGTGAGAGTGCC
TGAACAGAGAGATGCCCTGCATCTGCCTCTGCTGTCTTGGGGAGTGAAAAGACCTAGA

ACAAGAATCGGAGGACTGCTGGACCCCGGGACACCTATGGATGCCGATCTGGTGGCCT
CTTCTACAGTGGTGTGGGAACAGGATGCCGATCCTTTTGCCGGAACAGCCGATGATTTCC
CCTGCCTTTAATGAGGAAGAACTGGCCTGGCTGATGGAAGTCTGCCTCAGGGTTCCC
GTTTAAACAAGCTTGTGACGGTACCGAATTCATCGATAGTACTCTCGAGGGATCCGAG
CTCAAGATCTAGCTAAGTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCA
TTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAATAAAAATCATTATTTGC
CATCCAGCTGATATCCCCTATAGTGAGTCGTATTACATGGTCATAGCTGTTTCCTGGCA
GCTCTGGCCCGTGTCTCAAATCTCTGATGTTACATTGCACAAGATAAAAATATATCATC
ATGAACAATAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATAT
TCAACGGGAAACGTGCGAGGCCGCGATTAAATCCAACATGGATGCTGATTTATATGGGT
ATAAATGGGCTCGCGATAATGTGCGGCAATCAGGTGCGACAATCTATCGCTTGTATGGG
AAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGT
TACAGATGAGATGGTCAGACTAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCA
AGCATTTTATCCGTAATCCTGATGATGCATGGTTACTCACCCTGCGATCCCCGGAAAA
ACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTG
GCAGTGTTCCTGCGCCGGTTGCATTGATTCTGTTTGTAAATTGTCCTTTTAAACAGCGAT
CGCGTATTTGCTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAG
TGATTTTGTATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATA
AACTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACC
TTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCA
GACCGATAACCAGGATCTTGCCATCCTATGGAAGTGCCTCGGTGAGTTTTCTCCTTCATT
ACAGAAACGGCTTTTTTCAAATAATGGTATTGATAATCCTGATATGAATAAATTGCAGTTT
CATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAACACTGGCAG
AGCATTACGCTGACTTGACGGGACGGCGCAAGCTCATGACCAAATCCCTTAACGTGA
GTTACGCGTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGA
GATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACCACCGCTACCAGC
GGTGGTTTGTGTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACCTGGCTTCA
GCAGAGCGCAGATAACCAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTC
AAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGC
TGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATA
AGGCGCAGCGGTGCGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAA
CGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCC
CGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGC
GCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCTGTGCGGGTTTCG
CCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGG
AAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCA
CATGTT

APÊNDICE V – SEQUÊNCIA DO PLASMÍDEO pDEST™17

pDEST™17 (6354pb):

AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCT
CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGTCGTACTACCATCACCA
TCACCATCACCTCGAATCAACAAGTTTGTACAAAAAAGCTGAACGAGAAACGTA AAATGA
TATAAATATCAATATATTA AATTAGATTTTGCATAAAAAACAGACTACATAACTGTAAAA
CACACATATCCAGTCACTATGGCGGCCGATTAGGCACCCCAGGCTTTACACTTTATG
CTTCCGGCTCGTATAATGTGTGGATTTTGTAGTTAGGATCCGTCGAGATTTTCAGGAGCT
AAGGAAGCTAAAATGGAGAAAAAATCACTGGATATACCACCGTTGATATATCCCAATG
GCATCGTAAAGAACATTTTGTAGGCATTTTCAGTCAGTTGCTCAATGTACCTATAACCAGAC
CGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAAGTTTTA
TCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTCCGTATGG
CAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTACCCTTGTTACACCGTTTTTC
CATGAGCAAACCTGAAACGTTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCA
GTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCC
TAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAG
TTTTGATTTAAACGTGGCCAATATGGACAACCTTCTTCGCCCCCGTTTTTCACCATGGGCAA
ATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTACAGTTTCATCATGCC
GTCTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGA
GTGGCAGGGCGGGGCGTAAAGATCTGGATCCGGCTTACTAAAAGCCAGATAACAGTAT
GCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTGATATGTATACCCGAAGT
ATGTCAAAAAGAGGTGTGCTATGAAGCAGCGTATTACAGTGACAGTTGACAGCGACAGC
TATCAGTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAAACCATGC
AGAATGAAGCCCGTCGTCTGCGTGCCGAACGCTGGAAAGCGGAAAATCAGGAAGGGAT
GGCTGAGGTCGCCCGGTTTATTGAAATGAACGGCTCTTTTGTGACGAGAACAGGGAC
TGGTGAAATGCAGTTTAAAGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGT
GGATGTACAGAGTGATATTATTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCC
AGTGACAGTCTGCTGTCAGATAAAGTCTCCCGTGAAC TTTACCCGGTGGTGCATATCGG
GGATGAAAGCTGGCGCATGATGACCACCGATATGGCCAGTGTGCCGGTCTCCGTTATC
GGGGAAGAAGTGGCTGATCTCAGCCACCGCGAAAATGACATCAAAAACGCCATTAACC
TGATGTTCTGGGGAATATAAATGTCAGGCTCCCTTATACACAGCCAGTCTGCAGGTCGA
CCATAGTGA CTGGATATGTTGTGTTTTACAGTATTATGTAGTCTGTTTTTTATGCAAATC
TAATTTAATATATTGATATTTATATCATTTTACGTTTCTCGTTCAGCTTTCTTGTACAAAGT
GGTTGATTCGAGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCG
CTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG
CTGAAAGGAGGAACTATATCCGGATATCCACAGGACGGGTGTGGTCCCATGATCGCG

TAGTCGATAGTGGCTCCAAGTAGCGAAGCGAGCAGGACTGGGCGGCGGCCAAAGCGG
TCGGACAGTGCTCCGAGAACGGGTGCGCATAGAAATTGCATCAACGCATATAGCGCTA
GCAGCACGCCATAGTACTGGCGATGCTGTCGGAATGGACGATATCCCGCAAGAGGCC
CGGCAGTACCGGCATAACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAG
GATGACGATGAGCGCATTGTTAGATTTACATACACGGTGCCTGACTGCGTTAGCAATTTA
ACTGTGATAAACTACCGCATTAAAGCTTATCGATGATAAGCTGTCAAACATGAGAATTCT
TGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATG
GTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTT
ATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCCTGATAAATGCTT
CAATAATATTGAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCC
TTTTTTCGGCATTTCCTTCTGTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAA
GATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCG
GTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAG
TTCTGCTATGTGGCGCGGTATTATCCCGTGTGACGCCGGGCAAGAGCAACTCGGTCCG
CCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCT
TACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACA
CTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTG
CACAAATGCGGGGATCATGTAACCTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAG
CCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGTTGCG
CAAATATTAACCTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAAGACTGGAT
GGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTT
ATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGG
GGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAAC
TATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGT
AACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACCTTCATTTTTAATT
AAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAATCCCTTAACGTGAG
TTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC
TTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACCACCGCTACCAGCGGTGG
TTGTTTGCCGGATCAAGAGCTACCAACTTTTTTCCGAAGGTAACCTGGCTTCAGCAGA
GCGCAGATACCAAATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAA
CTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCA
GTGGCGATAAGTCGTGTCTTACCGGGTTGACTCAAGACGATAGTTACCGGATAAGGC
GCAGCGGTCCGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGAC
CTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAA
GGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAC
GAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCAC
CTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA

ACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATG
TTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCT
GATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCG
GAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTACACCCGCAT
ATATGGTGCACCTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACTC
CGCTATCGCTACGTGACTGGGTTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGA
CGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTC
TCCGGGAGCTGCATGTGTGAGAGTTTTTACCAGTCATCACCGAAACGCGCGAGGCAGC
TGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTACAGATGTCTGCCTGTTTCATCC
GCGTCCAGCTCGTTGAGTTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGC
CATGTTAAGGGCGGTTTTTTCTGTTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTC
TGTTTCATGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACT
GATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACACTGGCGGTATGGA
TGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGA
TGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATG
GTGCAGGGCGCTGACTTCCGCGTTTTCCAGACTTTACGAAACACGGAAACCGAAGACCA
TTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTACGTTTCGCTCG
CGTATCGGTGATTATTCTGCTAACCGATAAGGCAACCCCGCCAGCCTAGCCGGGTCC
TCAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGACCCAACGCTGCCCGAGA
TGCGCCCGGTGCGGCTGCTGGAGATGGCGGACGCGATGGATATGTTCTGCCAAGGGT
TGGTTTGCGCATTACAGTTCTCCGCAAGAATTGATTGGCTCCAATTCTTGGAGTGGTG
AATCCGTTAGCGAGGTGCCGCCGGCTTCCATTAGGTCGAGGTGGCCCGGCTCCATGC
ACCGCGACGCAACGCGGGGAGGCAGACAAGGTATAGGGCGGGCGCCTACAATCCATGC
CAACCCGTTCCATGTGCTCGCCGAGGCGGCATAAATCGCCGTGACGATCAGCGGTCCA
GTGATCGAAGTTAGGCTGGTAAGAGCCGCGAGCGATCCTTGAAGCTGTCCCTGATGGT
CGTCATCTACCTGCCTGGACAGCATGGCCTGCAACGCGGGCATCCCGATGCCGCCGG
AAGCGAGAAGAATCATAATGGGGAAGGCCATCCAGCCTCGCGTCGCGAACGCCAGCAA
GACGTAGCCCAGCGCGTCCGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAA
ACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAA
TACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCTCGCCGAA
AATGACCCAGAGCGCTGCCGGCACCTGTCTACGAGTTGCATGATAAAGAAGACAGTC
ATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTG
AAGGCTCTCAAGGGCATCGGTCGATCGACGCTCTCCCTTATGCGACTCCTGCATTAGG
AAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCA
TGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACG
CCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCATCGGTGATGT

CGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGC
GTCCGGCGTAGAGGATCG

APÊNDICE VI – SEQUÊNCIA DO PLASMÍDEO pTSP1

pTSP1 (6853pb):

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGA
TACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGA
AGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGC
TGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATACGC
GTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAAAAGGCCATCCGTCAGGATGGCC
TTCTGCTTAGTTTGATGCCTGGCAGTTTATGGCGGGCGTCCTGCCCGCCACCCTCCGG
GCCGTTGCTTCACAACGTTCAAATCCGCTCCCGGCGGATTTGTCCTACTCAGGAGAGC
GTTACCGACAAACAACAGATAAAACGAAAGGCCAGTCTTCCGACTGAGCCTTTTCGTT
TTATTTGATGCCTGGCAGTTCCTACTCTCGCGTTAACGCTAGCATGGATGTTTTCCAG
TCACGACGTTGTAACGACGCGCCAGTCTTAAGCTCGGGCCCCAAATAATGATTTTATTT
TGACTGATAGTGACCTGTTTCGTTGCAACAAATTGATGAGCAATGCTTTTTTATAATGCCA
ACTTTGTACAAAAAGCAGGCTGCGGCCGCGCCACCATGGGAAAACCTATTCCTAATCC
TCTGCTGGGCCTGGATTCTACCATGGACCCTATTAGAAGCAGAACACCCTCTCCAGCCA
GAGAACTGCTGTCTGGACCTCAGCCTGATGGAGTGCAGCCTACAGCCGATAGAGGAGT
GTCTCCTCCTGCCGGAGGACCTCTGGATGGACTGCCTGCCCGGAGAACAATGAGCAGA
ACAAGACTGCCTTCTCCTCCAGCCCCATCTCCTGCCTTTTCTGCCGATTCTTTTAGCGAC
CTGCTGAGACAGTTTGACCCAGCCTGTTTAATACCAGCCTGTTTCGATAGCCTGCCTCC
TTTTGGAGCCCACCACACAGAGGCCGCCACAGGCCGAATGGGATGAAGTGCAGTCTGGA
CTGAGAGCCGCGATGCCCTCCTCCTACAATGAGAGTGGCCGTGACAGCCGCCAGA
CCTCCTAGAGCCAAACCTGCCCTAGAAGGAGAGCCGCCAGCCTTCTGATGCCTCTC
CTGCCGCCAGGTGGACCTGAGAACACTGGGATATTCTCAGCAGCAGCAGGAGAAGAT
CAAGCCCAAGGTGAGGTCTACAGTGGCCCAGCACCACGAAGCCCTGGTGGGACACGG
ATTTACACACGCCACATTGTGGCCCTGTCTCAGCACCTGCCGCCCTGGGAACAGTG
GCCGTGAAATATCAGGATATGATTGCCGCCCTGCCTGAGGCCACACACGAAGCCATTG
TGGGAGTGGGAAAACAGTGGTCTGGAGCCAGAGCCCTGGAAGCCCTGCTGACAGTGG
CCGGAGAACTGAGAGGACCTCCTCTGCAGCTGGATACAGGACAGCTGCTGAAGATTGC
CAAAAGGGGCGGAGTGACCGCGGTGGAAGCCGTGCACGCCTGGAGAAATGCCCTGAC
AGGAGCCCCTCTGAACCTGACCCCCGAACAGGTGGTGGCCATTGCCAGCCACGACGG
CGGCAAGCAGGCCCTGAAACCGTGCAGAGACTGCTGCCCGTGCTGTGCCAGGCCCA
TGGCCTGACACCTGAACAGGTGGTGGCTATCGCCTCTCACGACGGAGGAAAACAGGCT
CTGGAAACAGTGCAGCGGCTGCTGCCTGTGCTGTGTCAGGCTCACGGCTTGACTIONAG
AACAGGTGGTGGCTATTGCTTCCAACAACGGGGGGAAACAGGCCCTGGAAACTGTGCA
GCGCCTGCTGCCAGTGCTGTGCCAGGCTCACGACTGACCCCCGAACAGGTGGTGGC
CATTGCCAGCAACGGCGGGCGGCAAGCAGGCCCTGGAAACCGTGCAGAGACTGCTGCC

CGTGCTGTGCCAGGCCCATGGCCTGACACCTGAACAGGTGGTGGCTATCGCCTCTCAC
GACGGAGGAAAACAGGCTCTGGAAACAGTGCAGCGGCTGCTGCCTGTGCTGTGTCAG
GCTCACGGCTTGACTCCAGAACAGGTGGTGGCTATTGCTTCCAACGGCGGGGGGAAAC
AGGCCCTGGAAACTGTGCAGCGCCTGCTGCCAGTGCTGTGCCAGGCTCACGGGCTGA
CCCCGAACAGGTGGTGGCCATTGCCAGCAACGGCGGGCGGCAAGCAGGCCCTGGAAA
CCGTGCAGAGACTGCTGCCCGTGCTGTGCCAGGCCCATGGCCTGACACCTGAACAGG
TGGTGGCTATCGCCTCTCACGACGGAGGAAAACAGGCTCTGGAAACAGTGCAGCGGCT
GCTGCCTGTGCTGTGTCAGGCTCACGGCTTGACTCCAGAACAGGTGGTGGCTATTGCT
TCCAACAACGGGGGGAAACAGGCCCTGGAAACTGTGCAGCGCCTGCTGCCAGTGCTG
TGCCAGGCTCACGGCCTCACTCCCGAACAGGTGGTGGCCATTGCCAGCAACGGCGGC
GGCAAGCAGGCCCTGGAAACCGTGCAGAGACTGCTGCCCGTGCTGTGCCAGGCCCAT
GGCCTGACACCTGAACAGGTGGTGGCTATCGCCTCTCACGACGGAGGAAAACAGGCTC
TGAAACAGTGCAGCGGCTGCTGCCTGTGCTGTGTCAGGCTCACGGCTTGACTCCAGA
ACAGGTGGTGGCTATTGCTTCCCACGACGGGGGGAAACAGGCCCTGGAAACTGTGCA
GCGCCTGCTGCCAGTGCTGTGCCAGGCTCACGGACTGACCCCCGAACAGGTGGTGGC
CATTGCCAGCAACGGCGGGCGGCAAGCAGGCCCTGGAAACCGTGCAGAGACTGCTGCC
CGTGCTGTGCCAGGCCCATGGCCTGACACCTGAACAGGTGGTGGCTATCGCCTCTAAC
AACGGAGGAAAACAAGCACTCGAGACAGTGCAGCGGCTGCTGCCTGTGCTGTGTCAGG
CTCACGGCTTGACTCCAGAACAGGTGGTGGCTATTGCTTCCAACAACGGGGGGAAACA
GGCCCTGGAAACTGTGCAGCGCCTGCTGCCAGTGCTGTGCCAGGCTCACGGGCTGAC
CCCCGAACAGGTGGTGGCCATTGCCAGCCACGACGGCGGGCAAGCAGGCCCTGGAAAC
CGTGCAGAGACTGCTGCCCGTGCTGTGCCAGGCCCATGGCCTGACACCTGAACAGGT
GGTGGCTATCGCCTCTAATATCGGAGGAAAACAGGCTCTGGAAACAGTGCAGCGGCTG
CTGCCTGTGCTGTGTCAGGCTCACGGCTTGACTCCACAGCAGGTCGTGGCAATTGCTA
GCAACGGCGGGCGGACGGCCCGCCCTGGAGAGCATTGTGGCCAGCTGTCTAGACCTG
ATCCTGCCCTGGCCGCCCTGACAAATGATCACCTGGTGGCCCTGGCCTGTCTGGGAGG
CAGACCTGCCCTGGATGCCGTGAAAAAAGGACTGCCTCACGCCCTGCCCTGATCAAG
AGAACAATAGAAGAATCCCCGAGCGGACCTCTCACAGAGTGGCCGATCACGCCCAGG
TGGTGAGAGTGCTGGGATTTTTTTCAGTGTCACTCTCACCTGCCCAGGCCTTTGATGAT
GCCATGACACAGTTTTGGCATGAGCAGACACGGACTGCTGCAGCTGTTTAGAAGAGTGG
GAGTGACAGAACTGGAGGCCAGAAGCGGAACACTGCCTCCAGCCTCTCAGAGATGGG
ATAGAATCCTGCAGGCCAGCGGAATGAAGAGAGCCAAACCTTCTCCTACCAGCACCCA
GACACCTGATCAGGCCAGCCTGCACGCCTTTGCCGATTCTCTGGAAAGGGATCTGGAC
GCCCTTCTCCTATGCACGAAGGAGATCAGACAAGAGCCAGCAGCAGAAAGAGAAGCA
GGTCTGATAGAGCCGTGACAGGACCTTCTGCCCAGCAGTCTTTTGAAGTGAGAGTGCC
TGAACAGAGAGATGCCCTGCATCTGCCTCTGCTGTCTTGGGGAGTGAAAAGACCTAGA
ACAAGAATCGGAGGACTGCTGGACCCCGGACACCTATGGATGCCGATCTGGTGGCCT

CTTCTACAGTGGTGTGGGAACAGGATGCCGATCCTTTTGCCGGAACAGCCGATGATTTCC
CCTGCCTTTAATGAGGAAGAACTGGCCTGGCTGATGGAAGTCTGCCTCAGGGTTCCC
GTTTAAACAAGCTTGTGACGGTACCGAATTCATGGCGAGCTTGGTTAAGAAAGATATG
TGTGTTAAGATGACGATGGAAGGTAAGTGTGAACGGTATCACTTTAAGTGCCTGGCGA
GGGTGAAGGCAAGCCGTTTCGAGGGCACGCAGAACATGCGCATTCTGTGCACCGAGGG
CGGTCCGCTGCCTTTTGCATTGACATCCTGGCCCCGTGCTGTATGTACGGCTCTAAGA
CCTTCATTAACACGTGAGCGGTATCCCGGATTACTTTAAAGAGTCCTTTCCAGAGGGC
TTCATTGGGAACGTACCCAGATTTTTGAGGACGGTGGTGTCTGACCGCGCACCAAGA
CACCAGCCTGGAAGGTAATTGCCTGATCTATAAAGTGAAGGTTCTGGGTACCAATTTCC
CGGCGAATGGTCCGGTGTGCAAAAGAAAACCGCGGGTTGGGAGCCGTGCGTCGAGA
TGCTGTATCCGCGTGACGGCGTCTTGTGTGGTCAGAGCTTGTGCGCTGAAGTGCAC
CGATGGCAATCATCTGACCAGCCACCTGCGCACGACGTATCGTAGCCGTAAACCGAGC
AACGCCGTTAACATGCCGGAGTTCCATTTTGGTGACCATCGCATCGAAATCCTGAAAGC
TGAGCAGGGCAAATTCTACGAACAATACGAATCGGCTGTCGCACGTTACAGCGATGTG
CCGGAAAAAGCGACGTAATAAGGATCCGAGCTCAAGATCTAGCTAAGTAGACCCAGCTT
TCTTGTACAAAGTTGGCATTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAACAGG
TCACTATCAGTCAAATAAAATCATTATTTGCCATCCAGCTGATATCCCCTATAGTGAGT
CGTATTACATGGTCATAGCTGTTTCCTGGCAGCTCTGGCCCCGTGTCTCAAATCTCTGA
TGTTACATTGCACAAGATAAAAATATATCATCATGAACAATAAACTGTCTGCTTACATAA
ACAGTAATAACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCGAGGCCGCGATTA
AATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCA
ATCAGGTGCGACAATCTATCGCTTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGA
AACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAACTGG
CTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTAATCCTGATGATGCA
TGGTACTCACCCTGCGATCCCCGGAAAAACAGCATTCCAGGTATTAGAAGAATATCC
TGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGA
TTCCTGTTTGAATTGTCCTTTTAAACAGCGATCGCGTATTTTCGTCTCGCTCAGGCGCAAT
CACGAATGAATAACGTTTTGGTTGATGCGAGTGATTTTGTATGACGAGCGTAATGGCTGG
CCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCAGGATTCAGTC
GTCATCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGT
TGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATG
GAACTGCCTCGGTGAGTTTTCTCCTTATTACAGAAACGGCTTTTTCAAATAATGGTAT
TGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCA
GAATTGGTTAATTGGTTGTAACACTGGCAGAGCATTACGCTGACTTGACGGGACGGCG
CAAGCTCATGACCAAATCCCTTAACGTGAGTTACGCGTCGTTCCACTGAGCGTCAGAC
CCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTCTGCGCGTAATCTGCTGC
TTGCAAACAAAAAACCCAGCTACCAGCGGTGGTTTGTGGCCGATCAAGAGCTACC

AACTCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCT
AGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCG
CTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGG
TTGGACTIONAAGACGATAGTTACCGGATAAGGCGCAGCGGTCTGGGCTGAACGGGGGGT
TCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGC
GTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGG
TAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCT
GGTATCTTTATAGTCCTGTCTGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGAT
GCTCGTCAGGGGGGCGGAGCCTATGGAAAACGCCAGCAACGCGGCCTTTTTACGGTT
CCTGGCCTTTTGCTGGCCTTTTGCTCACATGTT

APÊNDICE VI – SEQUÊNCIA DO PLASMÍDEO pTSP2

pTSP2 (9208pb):

AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTTCCCT
 CTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGTCGTACTIONACCATCACCA
 TCACCATCACCTCGAATCAACAAGTTTGTACAAAAAAGCTGAACGAGAAACGTAAAATGA
 TATAAATATCAATATATTAATTAAGATTTTGCATAAAAAACAGACTACATAACTGTAAAA
 CACAACATATCCAGTCACTATGGCGGCCGCGCCACCATGGGAAAACCTATTCCTAATCC
 TCTGCTGGGCCTGGATTCTACCATGGACCCTATTAGAAGCAGAACACCCTCTCCAGCCA
 GAGAACTGCTGTCTGGACCTCAGCCTGATGGAGTGCAGCCTACAGCCGATAGAGGAGT
 GTCTCCTCCTGCCGGAGGACCTCTGGATGGACTGCCTGCCCGGAGAACAATGAGCAGA
 ACAAGACTGCCTTCTCCTCCAGCCCCATCTCCTGCCTTTTCTGCCGATTCTTTTAGCGAC
 CTGCTGAGACAGTTTGACCCAGCCTGTTTAATACCAGCCTGTTTCGATAGCCTGCCTCC
 TTTTGGAGCCCACCACACAGAGGCCGCCACAGGCGAATGGGATGAAGTGCAGTCTGGA
 CTGAGAGCCGCGGATGCCCTCCTCCTACAATGAGAGTGGCCGTGACAGCCGCCAGA
 CCTCCTAGAGCCAAACCTGCCCTAGAAGGAGAGCCGCCAGCCTTCTGATGCCTCTC
 CTGCCGCCAGGTGGACCTGAGAACACTGGGATATTCTCAGCAGCAGCAGGAGAAGAT
 CAAGCCCAAGGTGAGGTCTACAGTGGCCCAGCACCACGAAGCCCTGGTGGGACACGG
 ATTTACACACGCCACATTGTGGCCCTGTCTCAGCACCCTGCCGCCCTGGGAACAGTG
 GCCGTGAAATATCAGGATATGATTGCCGCCCTGCCTGAGGCCACACACGAAGCCATTG
 TGGGAGTGGGAAAACAGTGGTCTGGAGCCAGAGCCCTGGAAGCCCTGCTGACAGTGG
 CCGGAGAACTGAGAGGACCTCCTCTGCAGCTGGATACAGGACAGCTGCTGAAGATTGC
 CAAAAGGGGGCGGAGTGACCGCGGTGGAAGCCGTGCACGCCTGGAGAAATGCCCTGAC
 AGGAGCCCCTCTGAACCTGACCCCGAACAGGTGGTGGCCATTGCCAGCCACGACGG
 CGGCAAGCAGGCCCTGGAACCGTGCAGAGACTGCTGCCCGTGTGTGCCAGGCCCA
 TGGCCTGACACCTGAACAGGTGGTGGCTATCGCCTCTCACGACGGAGGAAAACAGGCT
 CTGGAACAGTGCAGCGGCTGCTGCCTGTGCTGTGTCAGGCTCACGGCTTGACTCCAG
 AACAGGTGGTGGCTATTGCTTCCAACAACGGGGGGAACAGGCCCTGGAAACTGTGCA
 GCGCCTGCTGCCAGTGTGTGCCAGGCTCACGGACTGACCCCGAACAGGTGGTGGC
 CATTGCCAGCAACGGCGGCGGCAAGCAGGCCCTGGAAACCGTGCAGAGACTGCTGCC
 CGTGCTGTGCCAGGCCCATGGCCTGACACCTGAACAGGTGGTGGCTATCGCCTCTCAC
 GACGGAGGAAAACAGGCTCTGGAACAGTGCAGCGGCTGCTGCCTGTGCTGTGTCAG
 GCTCACGGCTTGACTCCAGAACAGGTGGTGGCTATTGCTTCCAACGGCGGGGGGAAAC
 AGGCCCTGGAAACTGTGCAGCGCCTGCTGCCAGTGTGTGCCAGGCTCACGGGCTGA
 CCCCCGAACAGGTGGTGGCCATTGCCAGCAACGGCGGGCGGCAAGCAGGCCCTGGAAA
 CCGTGCAGAGACTGCTGCCCGTGTGTGCCAGGCCCATGGCCTGACACCTGAACAGG
 TGGTGGCTATCGCCTCTCACGACGGAGGAAAACAGGCTCTGGAAACAGTGCAGCGGCT
 GCTGCCTGTGCTGTGTCAGGCTCACGGCTTGACTCCAGAACAGGTGGTGGCTATTGCT
 TCCAACAACGGGGGGGAAACAGGCCCTGGAAACTGTGCAGCGCCTGCTGCCAGTGTGT
 TGCCAGGCTCACGGCCTCACTCCGAACAGGTGGTGGCCATTGCCAGCAACGGCGGGC
 GGCAAGCAGGCCCTGGAAACCGTGCAGAGACTGCTGCCCGTGTGTGCCAGGCCCAT
 GGCCTGACACCTGAACAGGTGGTGGCTATCGCCTCTCACGACGGAGGAAAACAGGCTC
 TGGAACAGTGCAGCGGCTGCTGCCTGTGCTGTGTGTCAGGCTCACGGCTTGACTCCAGA
 ACAGGTGGTGGCTATTGCTTCCCACGACGGGGGGGAAACAGGCCCTGGAAACTGTGCA
 GCGCCTGCTGCCAGTGTGTGCCAGGCTCACGGACTGACCCCGAACAGGTGGTGGC
 CATTGCCAGCAACGGCGGCGGCAAGCAGGCCCTGGAAACCGTGCAGAGACTGCTGCC
 CGTGCTGTGCCAGGCCCATGGCCTGACACCTGAACAGGTGGTGGCTATCGCCTCTAAC
 AACGGAGGAAAACAAGCACTCGAGACAGTGCAGCGGCTGCTGCCTGTGCTGTGTCAGG
 CTCACGGCTTGACTCCAGAACAGGTGGTGGCTATTGCTTCCAACAACGGGGGGGAAACA
 GGCCCTGGAAACTGTGCAGCGCCTGCTGCCAGTGTGTGCCAGGCTCACGGGCTGAC
 CCCCCAACAGGTGGTGGCCATTGCCAGCCACGACGGCGGCAAGCAGGCCCTGGAAAC
 CGTGCAGAGACTGCTGCCCGTGTGTGCCAGGCCCATGGCCTGACACCTGAACAGGT
 GGTGGCTATCGCCTCTAATATCGGAGGAAAACAGGCTCTGGAAACAGTGCAGCGGCTG

CTGCCTGTGCTGTGTCAGGCTCACGGCTTGACTCCACAGCAGGTCGTGGCAATTGCTA
GCAACGGCGGGCGGACGGCCCGCCCTGGAGAGCATTGTGGCCAGCTGTCTAGACCTG
ATCCTGCCCTGGCCGCCCTGACAAATGATCACCTGGTGGCCCTGGCCTGTCTGGGAGG
CAGACCTGCCCTGGATGCCGTGAAAAAGGACTGCCTCACGCCCTGCCCTGATCAAG
AGAACAAATAGAAGAATCCCCGAGCGGACCTCTCACAGAGTGGCCGATCACGCCAGG
TGGTGAGAGTGCTGGGATTTTTTCAGTGTCACTCTCACCCCTGCCAGGCCTTTGATGAT
GCCATGACACAGTTTGGCATGAGCAGACACGGACTGCTGCAGCTGTTTAGAAGAGTGG
GAGTGACAGAAGTGGAGGCCAGAAGCGGAACACTGCCTCCAGCCTCTCAGAGATGGG
ATAGAATCCTGCAGGCCAGCGGAATGAAGAGAGCCAAACCTTCTCCTACCAGCACCCA
GACACCTGATCAGGCCAGCCTGCACGCCTTTGCCGATTCTCTGGAAAGGGATCTGGAC
GCCCTTCTCCTATGCACGAAGGAGATCAGACAAGAGCCAGCAGCAGAAAGAGAAGCA
GGTCTGATAGAGCCGTGACAGGACCTTCTGCCAGCAGTCTTTTGAAGTGAGAGTGCC
TGAACAGAGAGATGCCCTGCATCTGCCTCTGCTGTCTTGGGGAGTGAAGACCTAGA
ACAAGAATCGGAGGACTGCTGGACCCCGGACACCTATGGATGCCGATCTGGTGGCCT
CTTCTACAGTGGTGTGGGAACAGGATGCCGATCCTTTTGGCGAACAGCCGATGATTTCC
CCTGCCTTTAATGAGGAAGAAGTGGCCTGGCTGATGGAAGTGCCTCAGGGTTCCC
GTTTAAACAAGCTTGTGACGGTACCGAATTCATGGCGAGCTTGGTTAAGAAAGATATG
TGTGTTAAGATGACGATGGAAGGTAAGTGTGAACGGTTATCACTTTAAGTGCCTTGGCGA
GGGTGAAGGCAAGCCGTTTCGAGGGCACGCAGAACATGCGCATTCTGTGCACCGAGGG
CGGTCCGCTGCCTTTTGCATTGCACATCCTGGCCCCGTGCTGTATGTACGGCTCTAAGA
CCTTCATTAACACGAGTGCAGCGGTATCCCGGATTACTTTAAGAGTGCCTTTCCAGAGGGC
TTCATTGGGAACGTACCCAGATTTTTGAGGACGGTGGTGTCTGACCGCGCACCAAGA
CACCAGCCTGGAAGGTAATTGCCTGATCTATAAAGTGAAGGTTCTGGGTACCAATTTCC
CGGCGAATGGTCCGGTGTGCAAAAGAAAACCGCGGGTTGGGAGCCGTGCGTGCAGA
TGCTGTATCCGCGTGACGGCGTCTTGTGTGGTTCAGAGCTTGTGGCGCTGAAGTGCAC
CGATGGCAATCATCTGACCAGCCACCTGCGCACGACGTATCGTAGCCGTAACCCGAGC
AACGCCGTTAACATGCCGGAGTTCCATTTTGGTGACCATCGCATCGAAATCCTGAAAGC
TGAGCAGGGCAAATTCTACGAACAATACGAATCGGCTGTGCGCACGTTACAGCGATGTG
CCGGA AAAAGCGACGTAATAAGGATCCGAGCTCAAGATCTAGCTAAGTAGCATAGTGAC
TGGATATGTTGTGTTTTACAGTATTATGTAGTCTGTTTTTATGCAAAATCTAATTTAAT
ATTGATATTTATATCATTTTACGTTTCTCGTTACGCTTTCTTGTACAAAGTGGTTGATTCC
AGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATA
ACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAG
GAACTATATCCGATATCCACAGGACGGGTGTGGTCCCATGATCGCGTAGTCGATAG
TGGCTCCAAGTAGCGAAGCGAGCAGGACTGGGCGGCGGCCAAAGCGGTCCGACAGTG
CTCCGAGAACGGGTGCGCATAGAAATTGCATCAACGCATATAGCGCTAGCAGCACGCC
ATAGTGACTGGCGATGCTGTGCGAATGGACGATATCCCGCAAGAGGCCCGGCAGTACC
GGCATAACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGA
GCGCATTGTTAGATTTACATACCGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAAC
TACCGCATTAAAGCTTATCGATGATAAGCTGTCAAACATGAGAATTCTTGAAGACGAAAG
GGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACG
TCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTATTTTTCTAAATA
CATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAA
AAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTTCGGGCAT
TTTGCCTTCTGTTTTTGTCTACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATC
AGTTGGGTGCACGAGTGGGTTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGA
GAGTTTTCGCCCCGAAGAAGCTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGG
CGCGGTATTATCCCGTGTGACGCCGGGCAAGAGCAACTCGGTCCCGCATACACTAT
TCTCAGAATGACTTGGTTGAGTACTACCAGTACAGAAAAGCATCTTACGGATGGCAT
GACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACT
TACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCCTTTTTTGCACAACATGGGG
GATCATGTAACCTCGCCTTGTGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGA
CGAGCGTGACACCAGATGCCTGCAGCAATGGCAACAACGTTGCGCAAACCTATTAAC
GGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAAGACTGGATGGAGGCGGATAA
AGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAAT

CTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTA
AGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACG
AAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACC
AAGTTTACTCATATATACTTTAGATTGATTTAAAACCTTCATTTTTAATTTAAAAGGATCTAG
GTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCAC
TGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCG
CGTAATCTGCTGCTTGCAAACAAAAAACCACCGCTACCAGCGGTGGTTTTGTTTGCCGG
ATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCA
AATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACC
GCCTACATACTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGT
CGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGCGG
CTGAACGGGGGGTTCTGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACT
GAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCG
GACAGGTATCCGGTAAGCGGCAGGGTCCGAACAGGAGAGCGCACGAGGGAGCTTCCA
GGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTTCGCCACCTCTGACTTGAGC
GTGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAACGCCAGCAACGC
GGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTT
ATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCC
GCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTG
ATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTACACCCGCATATATGGTGCACCT
TCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACTACTCCGCTATCGCTAC
GTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGAC
GGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTG
CATGTGTCAGAGGTTTTACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGC
TCATCAGCGTGGTTCGTGAAGCGATTACAGATGTCTGCCTGTTTCATCCGCGTCCAGCTC
GTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGG
CGGTTTTTCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTTCATGGGG
GTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACA
TGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACCTGGCGGTATGGATGCGGCGGGAC
CAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCC
ACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCT
GACTTCCGCGTTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTTCATGTTGTTGC
TCAGGTCGCAGACGTTTTGACAGCAGCAGTCGCTTACGTTCCGCTCGCGTATCGGTGATT
CATTCTGCTAACCCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCTCAACGACAGGAG
CACGATCATGCGCACCCGTGGCCAGGACCCAACGCTGCCCGAGATGCGCCGCGTGCG
GCTGCTGGAGATGGCGGACGCGATGGATATGTTCTGCCAAGGGTTGGTTTGCGCATTC
ACAGTTCTCCGCAAGAATTGATTGGCTCCAATTCTTGGAGTGGTGAATCCGTTAGCGAG
GTGCCCGCGGCTTCCATTCAGGTGAGGTGGCCCGGCTCCATGCACCGCGACGCAAC
GCGGGGAGGCAGACAAGGTATAGGGCGGCGCCTACAATCCATGCCAACCCGTTCCAT
GTGCTCGCCGAGGCGGCATAAATCGCCGTGACGATCAGCGGTCCAGTGATCGAAGTTA
GGCTGGTAAGAGCCGCGAGCGATCCTTGAAGCTGTCCCTGATGGTTCGTCATCTACCTG
CCTGGACAGCATGGCCTGCAACGCGGGCATCCCGATGCCGCGGAAGCGAGAAGAAT
CATAATGGGGAAGGCCATCCAGCCTCGCGTCGCGAACGCCAGCAAGACGTAGCCAG
CGCGTCGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGC
GGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGA
CAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCTCGCCGAAAATGACCCAGAG
CGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCG
ACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAG
GGCATCGGTGATCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCAGT
AGTAGGTTGAGGCCGTTGAGCACCGCCGCGCAAGGAATGGTGCATGCAAGGAGATG
GCGCCCAACAGTCCCCGGCCACGGGGCCTGCCACCATAACCACGCCGAAACAAGCG
CTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCATCGGTGATGTCGGCGATATAGG
CGCCAGCAACCGCACCTGTGGCGCCGGTATGCCGGCCACGATGCGTCCGGCGTAGA
GGATCG

