

Universidade Federal do Rio Grande do Sul
Instituto de Biociências
Curso de Graduação em Biotecnologia
Bacharelado em Biotecnologia - Ênfase em Biotecnologia Molecular

José André Della Giustina Neto

Revisão das Principais Técnicas e Proposição de Metodologia para Produção Heteróloga da Proteína Inibidora de Tripsina do Tipo Kunitz, Rica em Metioninas (KTIMet), Derivada de Soja [*Glycine max* (L.) Merrill] em *Escherichia coli*

Porto Alegre

2023

José André Della Giustina Neto

Revisão das Principais Técnicas e Proposição de Metodologia para Produção Heteróloga da Proteína Inibidora de Tripsina do Tipo Kunitz, Rica em Metioninas (KTIMet), Derivada de Soja [*Glycine max* (L.) Merrill] em *Escherichia coli*

Orientador: Prof. Dr. Giancarlo Pasquali

Coorientadora: Dra. Fernanda Cortez Lopes

Porto Alegre

2023

CIP - Catalogação na Publicação

Della_Giustina_Neto, José André

Revisão das Principais Técnicas e Proposição de Metodologia para Produção Heteróloga da Proteína Inibidora de Tripsina do Tipo Kunitz, Rica em Metioninas (KTIMet), Derivada de Soja [Glycine max (L.) Merrill] em Escherichia coli / José André Della_Giustina_Neto. -- 2023.

58 f.

Orientador: Giancarlo Pasquali.

Coorientadora: Fernanda Cortez Lopes.

Trabalho de conclusão de curso (Graduação) -- Universidade Federal do Rio Grande do Sul, Instituto de Biociências, Curso de Biotecnologia: Biotecnologia Molecular, Porto Alegre, BR-RS, 2023.

1. Proteína rica em metionina. 2. Soja. 3. Proteína recombinante. 4. Proteína heteróloga. 5. Kunitz. I. Pasquali, Giancarlo, orient. II. Cortez Lopes, Fernanda, coorient. III. Título.

FOLHA DE APROVAÇÃO

José André Della Giustina Neto

Revisão das Principais Técnicas e Proposição de Metodologia para Produção Heteróloga da Proteína Inibidora de Tripsina do Tipo Kunitz, Rica em Metioninas (KTIMet), Derivada de Soja [*Glycine max* (L.) Merrill] em *Escherichia coli*

Trabalho de Conclusão de Curso apresentado como requisito parcial à obtenção do título de Bacharel em Biotecnologia - Ênfase em Biotecnologia Molecular - do Instituto de Biociências da Universidade Federal do Rio Grande do Sul.
Orientador: Prof. Dr. Giancarlo Pasquali
Coorientadora: Dra. Fernanda Cortez Lopes

Aprovado em: Porto Alegre, 12 de abril de 2023

BANCA EXAMINADORA:

Prof. Dr. Giancarlo Pasquali
Departamento de Biologia Molecular e Biotecnologia, Instituto de Biociências - UFRGS

Dra. Fernanda Lazzarotto
Membro (pós-doutoranda) do Núcleo de Genômica Funcional de Plantas do Laboratório de Genética Vegetal da UFRGS

Dra. Luisa Abruzzi de Oliveira-Busatto
Pesquisadora do Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS

RESUMO

A expressão heteróloga de genes em microrganismos é ferramenta importante da biotecnologia pois representa ciclo produtivo mais rápido, econômico e eficaz quando comparado a métodos para obtenção das proteínas codificadas a partir dos organismos de origem. No entanto, a diversidade de técnicas disponíveis na literatura evidencia a necessidade de utilizar critérios para a seleção do sistema de expressão microbiano mais adequado em cada caso. Neste trabalho, foram estudados alguns dos métodos descritos na literatura científica e apresentados os critérios que permitiram a seleção da metodologia proposta. Para exemplificar as escolhas realizadas durante a elaboração desta monografia, foi utilizada uma proteína recombinante concebida em estágios de graduação e durante o trabalho de mestrado do autor. Trata-se da denominada KTIMet, versão sintética de um inibidor de tripsina do tipo Kunitz enriquecida em metioninas originalmente derivada de soja [*Glycine max* (L.) Merrill]. As principais considerações para a definição da melhor metodologia de produção de KTIMet foram: (I) viabilidade da síntese de proteína recombinante enriquecida em metioninas por bactérias; (II) viabilidade da expressão heteróloga de gene codificador de proteína vegetal inibidora de protease em bactérias; e (III) testar a hipótese levantada em análise *in silico* quanto à redução da atividade inibitória em comparação com a proteína nativa. Foram, assim, propostos métodos para a expressão gênica em *Escherichia coli*, a extração de proteínas e a purificação de KTIMet. Esta monografia poderá auxiliar na continuidade das pesquisas referentes à esta proteína, validando os pressupostos originais relacionados às vantagens de sua incorporação no conjunto de proteínas da soja quanto aos ganhos nutricionais deste grão, os quais são de grande relevância econômica no cenário da indústria de alimentos.

Palavras-chave: Proteína rica em metionina; Soja; Proteína recombinante; Proteína heteróloga; inibidor de protease; Kunitz; KTIMet.

ABSTRACT

The heterologous expression of genes in microorganisms is an important tool in biotechnology because it represents a faster, more economical and more efficient production cycle when compared to methods for obtaining encoded proteins from the original organisms. However, the diversity of techniques available in the literature highlights the need to use criteria for selecting the most appropriate microbial expression system in each case. In this work, some of the methods described in the scientific literature were studied and the criteria that allowed the selection of the proposed methodology were presented. To exemplify the choices made during the preparation of this monograph, a recombinant protein designed in undergraduate stages and during the author's master's work was used. This is the so-called KTIMet, a synthetic version of a Kunitz-type trypsin inhibitor enriched in methionines originally derived from soy [*Glycine max* (L.) Merrill]. The main considerations for defining the best methodology for producing KTIMet were: (I) viability of the synthesis of recombinant protein enriched in methionines by bacteria; (II) viability of heterologous expression of the gene coding for a protease inhibitor plant protein in bacteria; and (III) test the hypothesis raised in an *in silico* analysis regarding the reduction of the inhibitory activity compared to the native protein. Thus, methods for gene expression in *Escherichia coli*, protein extraction and purification of KTIMet were proposed. This monograph will be able to help in the continuity of the research related to this protein, validating the original assumptions related to the advantages of its incorporation in the pool of soy proteins regarding the nutritional gains of this grain, which are of great economic relevance in the food industry scenario.

Keywords: Methionine-rich protein; Soy; Recombinant protein; Heterologous protein; protease inhibitor; Kunitz; KTIMet.

LISTA DE ILUSTRAÇÕES

Figura 1 – Sequência completa do gene <i>ktimet</i>	29
Figura 2 – Estruturas primárias das proteínas nativa e recombinante.	29
Figura 3 – Mapa do plasmídeo pET23a (+).	34
Figura 4 – Mapa do plasmídeo pUC57- <i>ktimet</i>	35
Figura 5 – Mapa do plasmídeo pET23a-KTIMet.	37
Figura 6 – Sítio Ativo de KTi e domínio correspondente de KTIMet.....	44

LISTA DE ABREVIATURAS E SIGLAS

Da	Dalton, unidade de massa atômica (u.m.a $\approx 1,660540 \times 10^{-27}$ kg)
DNA	Ácido desoxirribionucleico, do inglês, <i>deoxyribonucleic acid</i>
GH	Hormônio do crescimento humano, do inglês, <i>Human Growth Hormone</i>
GST	Glutathione S-transferase, do inglês <i>Glutathione S-transferase</i>
His	Histidina
Ile	Isoleucina
INPI	Instituto Nacional da Propriedade Industrial
IPTG	Isopropil β -tiogalatosídeo, do inglês, <i>isopropyl β-thiogalactoside</i>
ITK	Inibidor de Tripsina tipo Kunitz de Soja grau P.A., reagente adquirido da SIGMA
KTi	Proteína nativa de soja inibidora de tripsina do tipo Kunitz, do inglês, <i>Kunitz trypsin Inhibitor protein</i>
<i>KT13</i>	Gene nativo de soja codificador de KTi
<i>ktimet</i>	Gene sintético codificador da proteína KTIMet
KTIMet	Proteína sintética derivada de KTi e enriquecida em metioninas
LB	Meio de Luria Bertani
Leu	Leucina
MALDI-TOF	Ionização e Dessorção a Laser Assistida por Matriz com Analisador de Tempo de Voo, do inglês, <i>Matrix-Assisted Laser Desorption/Ionization and Time of Flight</i>
MCS	Sítio de multiclonação, do inglês, <i>Multiple Cloning Site</i>
mRNA	RNA mensageiro, do inglês, <i>messenger RNA</i>
tRNA	RNA transportador ou de transferência, do inglês, <i>transfer RNA</i>
NMR	Ressonância Nuclear Magnética, do inglês, <i>Nuclear Magnetic Resonance</i>
<i>o/n</i>	“Durante a noite”, do inglês, <i>overnight</i>
PCR	Reação em Cadeia da polimerase, do inglês, <i>Polymerase Chain Reaction</i>
PM	Peso molecular
RNA	Ácido ribonucleico, do inglês, <i>ribonucleic acid</i>

SUMÁRIO

1	APRESENTAÇÃO	9
2	PRODUÇÃO DE PROTEÍNAS HETERÓLOGAS	12
2.1	TIPOS DE CÉLULAS OU “CHASSIS” CELULARES DE PRODUÇÃO	13
2.2	CRITÉRIOS PARA SELEÇÃO DO TIPO DE CÉLULA.....	14
2.2.1	Códons raros.....	14
2.2.2	Processamento pós-traducional.....	15
2.2.3	Toxicidade da proteína-alvo.....	18
3	EXTRAÇÃO E PURIFICAÇÃO DE PROTEÍNAS	20
3.1	LOCALIZAÇÃO FINAL DA PROTEÍNA HETERÓLOGA	20
3.2	PROTEÍNAS FUSIONADAS	20
3.3	SOLUBILIDADE DA PROTEÍNA	21
4	CARACTERIZAÇÃO DE PROTEÍNAS.....	22
4.1	ENSAIOS QUALITATIVOS E QUANTITATIVOS	22
4.1.1	Eletroforese em Gel de Poliacrilamida.....	22
4.1.2	<i>Western Blot</i>.....	23
4.2	ENSAIOS PARA DETERMINAÇÃO ESTRUTURAL DA PROTEÍNA	24
4.2.1	MALDI-TOF	24
4.2.2	Dicroísmo Circular	25
4.2.3	Cristalografia por Raios X.....	25
4.2.4	Ressonância Magnética Nuclear	25
4.3	ENSAIOS PARA DETERMINAÇÃO DE ATIVIDADE INIBITÓRIA	26
4.3.1	Ensaio da Benzoil-DL-Arginina-<i>p</i>-Nitroanilida.....	26
4.3.2	Titulação Calorimétrica Isotérmica.....	27
5	PROPOSIÇÃO DE METODOLOGIA PARA PRODUÇÃO DE KTIMET	28
5.1	EXPRESSÃO HETERÓLOGA DO GENE <i>KTIMET</i>	28
5.1.1	Características da proteína <i>KTIMet</i>	28
5.1.2	Seleção do tipo de célula para expressão de <i>ktimet</i>	31
5.1.3	Seleção do sistema de expressão	33
5.1.4	Seleção do vetor de expressão do sistema pET (Novagen).....	33
5.1.5	Produção do <i>amplicon</i> contendo <i>ktimet</i> e ligação a pET23a.....	35
5.1.6	Transformação de <i>E. coli</i> BL21 e <i>E. coli</i> Lemo21 com pET23a-KTIMet.....	38
5.1.7	Cultivo de <i>E. coli</i> BL21:: e LEMO21::pET23a-KTIMet.....	38

5.2	EXTRAÇÃO DA PROTEÍNA.....	39
5.3	PURIFICAÇÃO DA PROTEÍNA KTIMET.....	40
5.4	CARACTERIZAÇÃO DA PROTEÍNA KTIMET	42
5.4.1	Análise das amostras de proteínas não purificadas.....	42
5.4.2	Análise das amostras de proteínas purificadas.....	43
6	CONCLUSÃO	46
	APÊNDICE A — PLANO DE TRABALHO PROPOSTO	55

1 APRESENTAÇÃO

A expressão de transgenes e a produção heteróloga de proteínas em bactérias ou fungos tem sido utilizada como método para produção de quantidades suficientes destas moléculas visando sua caracterização bioquímica. A opção pela expressão de um gene de interesse em organismos diferentes daqueles de onde a molécula se origina tem normalmente razões técnicas e de ordem prática. No entanto, requer cuidados para garantir a quantidade e a qualidade final desejadas do peptídeo de interesse (KAUR *et al.*,2018; CHEN *et al.*,2019; CHOI *et al.*,2020).

Um aspecto a destacar refere-se à adequação dos códons originais do gene de interesse ao padrão de codificação do microrganismo selecionado para o sistema de expressão. Em alguns casos, a presença de códons raros pode impedir ou reduzir a taxa de tradução do ácido ribonucleico (RNA) mensageiro (mRNA) nas células escolhidas para produção heteróloga, levando a um baixo rendimento na produção da proteína desejada (PETI;PAGE,2007; KAUR *et al.*,2018), por exemplo.

Também é preciso analisar a questão do enovelamento correto da proteína após a tradução, com formação de estruturas secundárias e terciárias presentes no organismo de origem, normalmente resultantes da formação de pontes dissulfeto. A formação de eventuais estruturas quaternárias (formação de subunidades), também pode ser necessária para a manutenção da atividade da proteína-alvo (UTSUMI *et al.*,2002; BASTOLLA;DEMETRIUS,2005; SINGH;PANDA,2005)

Outro ponto de atenção deve ser a eventual necessidade de processamento do peptídeo após a tradução, com inserção de moléculas adicionais (glicosilação, fosforilação etc.). Este processamento poderá requerer o trânsito do peptídeo por compartimentos celulares com auxílio de peptídeos sinais. Esta condição, dependendo do tipo de célula utilizada, pode ou não estar inviabilizada. Neste caso, outras estratégias deverão ser combinadas, ou outro tipo de célula deverá ser escolhido de modo a alcançar o resultado mais adequado (HUNT,2005; KHOW;SUNTRARACHUN,2012; CAI *et al.*,2019).

Por fim, é importante considerar, em uma produção heteróloga de proteínas, a massa e a quantidade da proteína-alvo que o organismo selecionado será capaz de lidar. A formação de *clusters* ou agregados de proteínas (*inclusion bodies* ou corpos de inclusão) é comum sempre que o microrganismo tem alguma dificuldade em manter a proteína solúvel em seu citoplasma como relata KAUR *et al.* (2018) . Neste trabalho,

os autores também abordaram a questão da toxicidade da proteína para a célula hospedeira, que normalmente é contornada direcionando a expressão para o periplasma ou via de secreção ou, ainda, controlando fortemente a taxa de expressão do gene e a tradução do transcrito.

Eventualmente, a formação de corpos de inclusão pode auxiliar no isolamento e na purificação da proteína produzida. Porém, neste caso, irá requerer processamento adicional após sua recuperação do extrato de células visando a solubilização por meio da desnaturação e posterior renaturação do polipeptídeo em um tampão adequado, de modo a recuperar as características originais da molécula, como realizado por SINGH e PANDA (2005) para o hormônio do crescimento (GH, do inglês, *Growth Hormone*). Neste trabalho, os autores estudaram em detalhe os mecanismos de formação dos corpos de inclusão em *Escherichia coli*, testando diversos procedimentos de solubilização e renaturação da proteína produzida.

Considerando todos os aspectos citados anteriormente, a primeira etapa recomendada para estruturar um projeto de pesquisa voltado à produção heteróloga de proteínas seria coletar o máximo de informações disponíveis sobre o gene a ser expresso e a proteína que se pretende produzir. Com estas informações em mãos, é possível antever alguns problemas e orientar as escolhas para as próximas etapas do processo de expressão. Na sequência, é recomendado testar diversos sistemas como o tipo de célula, o vetor de expressão, sequências promotoras, proteínas fusionadas etc., buscando determinar aqueles viáveis e com melhor desempenho. Após esta etapa de testes, dever-se-á selecionar o(s) sistema(s) que demonstraram maior viabilidade para obtenção do polipeptídeo alvo e, só então, desenvolver as melhores condições visando aumentar a quantidade e a qualidade final da proteína produzida (ROSANO;CECCARELLI,2014; AHMAD *et al.*,2018).

No segundo capítulo desta monografia é abordada a produção de proteínas heterólogas em geral, apresentando alguns aspectos da seleção do tipo de células e os critérios mais conhecidos para uma escolha adequada. A revisão bibliográfica sobre os métodos existentes foi direcionada, já que o objetivo principal foi subsidiar a construção de uma metodologia a ser aplicada a uma proteína específica.

No terceiro capítulo, descrevemos os métodos de extração e purificação de proteínas voltados para escala laboratorial. A produção em escala industrial não foi abordada no presente trabalho, mas tal objetivo é previsto a mais longo prazo.

No quarto capítulo, tratamos de métodos de caracterização bioquímica das proteínas produzidas. As técnicas quantitativas e qualitativas básicas utilizadas em pesquisa com proteínas foram abordadas, assim como ensaios para determinação das estruturas primária, secundária, terciária e quaternária das proteínas em geral. Técnicas voltadas à avaliação da atividade enzimática de proteínas também foram incluídas no trabalho.

Os conceitos teóricos abordados nos capítulos dois, três e quatro foram aplicados, a título de exemplificação, de forma dirigida, no capítulo cinco, para a proteína KTi enriquecida de metioninas (KTIMet), cujo detalhamento estrutural e concepção geral podem ser encontrados na dissertação de mestrado do autor (DELLA_GIUSTINA_NETO,2015).

No presente trabalho, optamos por eliminar diversas etapas do protocolo típico de expressão gênica e produção de proteínas heterólogas, buscando na literatura protocolos já testados, visando a produção de proteínas similares à nossa proteína alvo quanto aos aspectos de tamanho, função e estrutura. Um ponto importante na escolha de protocolos pré-existentes é a finalidade da síntese proteica. Portanto, os objetivos do trabalho de referência a ser selecionado deverão estar alinhados com a utilização que se quer dar para a proteína a ser produzida. Com base nestas considerações, elegemos o trabalho realizado por MAJOR e CONSTABEL (2008) como referência inicial para desenvolvimento desta monografia e proposição de uma metodologia visando a produção futura de nossa proteína recombinante. No referido trabalho, os autores estudaram os aspectos funcionais da família de genes codificadores do inibidor de tripsina do tipo Kunitz (KTi, do inglês, *Kunitz Trypsin inhibitor*) que servem como fator de proteção das plantas, no caso específico de *Populus trichocarpa* e *Populus deltoides*, contra herbívoros. Estes pesquisadores tinham por objetivo expressar genes e produzir, de forma heteróloga, algumas proteínas desta família, preservando sua atividade bioquímica e visando a comparação entre estas e um dos inibidores de proteases nativo da soja, o qual foi objeto de estudo pioneiro em trabalhos publicados por KUNITZ (1946) e KUNITZ (1947), de onde deriva o nome desta molécula.

Os objetivos gerais deste Trabalho de Conclusão foram: (I) Revisão dos principais aspectos envolvidos com a produção heteróloga de proteínas em geral; (II) Projetar uma metodologia, a ser validada no futuro, para a produção de uma proteína recombinante específica de soja, já conhecida e estudada anteriormente pelo autor.

2 PRODUÇÃO DE PROTEÍNAS HETERÓLOGAS

A expressão de genes em organismos distintos do original (transgenes) para a produção de proteínas heterólogas tem desempenhado um papel importante nos estudos biotecnológicos considerando as possibilidades de controle das etapas e da aceleração do processo, quando comparada com a alternativa de obter estas proteínas a partir das células e tecidos do organismo de origem (LIN *et al.*,2004; SHIH *et al.*,2004; HUNT,2005; KHOW;SUNTRARACHUN,2012). Portanto, contornar as dificuldades inerentes à técnica de produção heteróloga de proteínas é tarefa permanente dos pesquisadores e das empresas de biotecnologia que desenvolvem sistemas de expressão heteróloga disponibilizados ao mercado. O conhecimento dos mecanismos envolvidos em cada uma das etapas e da natureza dos produtos intermediários é fundamental para o resultado do processo como um todo, como abordam os trabalhos de KAUR *et al.* (2018) e KHOW e SUNTRARACHUN (2012).

Uma questão sempre presente quando se pretende produzir proteínas em escala laboratorial é a quantidade de produto necessária para realizar todas as análises para as quais a proteína-alvo está sendo produzida. Assim, um planejamento prévio dos ensaios que serão realizados, considerando a quantidade de repetições necessárias e as perdas de material em cada análise ou processamento, é fundamental para estabelecer esta quantidade mínima (AHMAD *et al.*,2018).

Admitindo que já tenhamos conhecimento prévio de parte significativa das informações disponíveis na literatura especializada sobre o tipo de proteína ou até mesmo do peptídeo específico que pretendemos produzir, o primeiro ponto sobre o qual devemos dedicar atenção é o tipo de célula a ser utilizado e sua compatibilidade com sistemas de expressão heteróloga disponíveis no mercado. Estes sistemas normalmente possuem um vasto ferramental já disponível para os pesquisadores, os quais resolvem ou contornam problemas usualmente encontrados pelos usuários. É evidente que ao trabalhar com proteínas recombinantes com características muito específicas ou ainda não conhecidas, provavelmente será necessário superar problemas não previstos pelos fornecedores destes sistemas. Entre os sistemas de expressão heteróloga mais conhecidos podemos destacar pelo menos dois, que utilizam organismos procariontes: (I) Sistema pET e (II) Sistema pGEX.

O sistema pET da *Novagen* baseia-se na expressão, em *E. coli*, de genes sob o controle transcricional de sequências derivadas do bacteriófago T7, sendo

amplamente utilizado para a produção laboratorial de proteínas recombinantes (YIN *et al.*,2007).

O sistema pGEX originalmente concebido pela *GE Healthcare Life Sciences* fundamenta-se na expressão de genes de interesse fusionados ao gene da Glutathione S-transferase (GST, do inglês *Glutathione S-transferase*) sob controle de um promotor híbrido (*tac*), obtido a partir de dois promotores nativos de *E. coli* que são o promotor *lac* (operon da lactose) e o promotor *trp* (operon do triptofano) (DE BOER *et al.*,1983; HARPER;SPEICHER,2008; GE_HEALTHCARE,2013).

Selecionado o sistema de expressão e respectivo tipo de células, será necessário definir a linhagem de células e o plasmídeo de expressão mais adequados para a produção do peptídeo-alvo, dentro das opções de cada sistema.

2.1 TIPOS DE CÉLULAS OU “CHASSIS” CELULARES DE PRODUÇÃO

A seleção do tipo de células precede a escolha do sistema de expressão e requer uma análise preliminar quanto a diversos aspectos. Assim, devemos avaliar a compatibilidade geral entre a célula hospedeira, o gene a ser expresso e seu produto. Um tipo de célula mal selecionado poderá resultar em uma série de dificuldades decorrentes da fisiologia da própria célula tais como funcionamento inadequado da maquinaria de tradução, deficiência no transporte interno de aminoácidos, enclausuramento ou não do produto em vesículas, secreção via membrana plasmática ou dispersão no citosol, impossibilidade de processamento pós-traducional, degradação da proteína-alvo e, por fim, morte celular (HANNIG;MAKRIDES,1998; PETI;PAGE,2007; KHOW;SUNTRARACHUN,2012; CONTRERAS-GOMEZ *et al.*,2014; ROSANO;CECCARELLI,2014; AHMAD *et al.*,2018; KAUR *et al.*,2018; CAI *et al.*,2019).

A possibilidade de utilizar mais de um tipo de célula ou, ainda, o mesmo tipo com linhagens diferentes para a expressão heteróloga, dividindo as etapas de expressão gênica e produção do peptídeo, pode ser uma alternativa, como relatam em seu trabalho CHEN *et al.* (2019). Neste artigo, os autores apresentaram uma revisão sobre o assunto abordando diversos trabalhos publicados entre 2014 e 2018, os quais introduzem o conceito de “engenharia modular de cocultura”.

Apesar da diversidade de tipos e combinações de células possíveis, os sistemas de expressão mais conhecidos e difundidos utilizam bactérias (organismos procariontes), fungos e leveduras (organismos eucariontes). As linhagens derivadas

da bactéria *Escherichia coli* K12 tem lugar de destaque na expressão heteróloga de proteínas com vistas às típicas análises bioquímicas realizadas em laboratórios. Isto se deve ao conhecimento já adquirido sobre este microrganismo, a facilidade de cultivar e armazenar em ambiente laboratorial e o baixo risco em relação à biossegurança, já que é considerado organismo de classe 1 de risco biológico, a sua flexibilidade e variabilidade genética e fenotípica, a capacidade de se reproduzir na presença ou na ausência de oxigênio (aeróbio facultativo), crescimento rápido, meios de cultivos mais baratos, grande variedade de plasmídeos compatíveis, além de várias outras características estudadas por PAPANEOPHYTOU e KONTOPIDIS (2012), SCHLEGEL *et al.* (2012) e BLOUNT (2015).

Sistemas de produção que utilizam baculovírus associados a células de insetos são bastante conhecidos em escala industrial e tem resultados bastante satisfatórios quando exigem processamento pós-traducional (CONTRERAS-GOMEZ *et al.*,2014).

2.2 CRITÉRIOS PARA SELEÇÃO DO TIPO DE CÉLULA

Os principais critérios utilizados para a seleção do tipo de célula, como já referido anteriormente, derivam da análise de compatibilidade entre a fisiologia da célula e as características relacionadas à produção da proteína-alvo, desde a expressão do gene até o peptídeo final na sua forma ativa (YIN *et al.*,2007; AHMAD *et al.*,2018; KAUR *et al.*,2018). No entanto, a simples constatação de que um determinado tipo de célula não é o mais adequado em função de alguma dificuldade fisiológica frente aos produtos intermediários do processo de expressão não é suficiente para descartar sua utilização. Outros fatores como disponibilidade no laboratório e conhecimento prévio na manipulação do microrganismo, ou vantagens comparativas em relação aos custos de cultivo, podem ser decisivos para sua seleção, sendo necessário, deste modo, criar os mecanismos para superar as dificuldades pela não adequação inicial (BLOUNT,2015; JOSEPH *et al.*,2015). A seguir, passamos a descrever os principais problemas derivados da escolha do tipo de célula e possíveis mecanismos para contornar estas dificuldades.

2.2.1 Códonos raros

A produção de uma proteína conforme o dogma central da biologia molecular inicia na **transcrição** a partir do gene para um mRNA e, deste, a **tradução** para uma

sequência de aminoácidos conforme o código genético. Uma molécula-chave no processo de tradução, o qual ocorre no ribossomo, é o RNA transportador (tRNA), responsável por reconhecer, ligar e transportar cada aminoácido, posicionando-o no ribossomo para a formação da ligação peptídica. Existem moléculas de tRNA específicas para pelo menos um dos códons correspondentes a cada aminoácido. Portanto, a não disponibilidade de moléculas de tRNA em quantidade suficiente para um determinado códon poderá afetar a tradução, reduzindo a taxa de síntese proteica ou até interrompendo o processo. Sendo o código genético degenerado, existe mais de um códon para a maioria dos aminoácidos. Existem, também, diferenças na frequência de uso destes códons de um organismo para outro. Ao inserir um gene para expressão em um organismo diverso do original, é preciso analisar a existência de códons raros no gene a ser expresso e que sejam incompatíveis com o tipo de célula hospedeira selecionada, como abordam os autores KURLAND (1991); HERSHBERG e PETROV (2008), e QUAX *et al.* (2015).

Caso algum códon presente no gene não seja usual na célula hospedeira, é preciso considerar sua alteração para outro códon do mesmo aminoácido. Esta escolha deverá ser realizada de forma antecipada, já na construção do gene, tendo em mente a célula em que será expresso. Na prática, podem existir restrições a adotar um ou outro códon que facilite a tradução do gene e, assim, deve-se então optar por utilizar vetores capazes de suprir a demanda de tRNAs para a produção de uma proteína, como abordaram os autores KIM *et al.* (1998); SØRENSEN *et al.* (2003) e KAUR *et al.* (2018) ou, ainda, trocar o tipo de célula ou linhagem celular, utilizando linhagens de *E. coli* específicas para este fim, como *E. coli BL21CodonPlus-RIL*, *BL21CodonPlus-(DE3)-RIL-RIPL*, *Rosetta2* e *Rosetta-pLysS* (NOVAGEN,2011; JOSEPH *et al.*,2015).

2.2.2 Processamento pós-traducional

A necessidade de processamento pós-traducional pode ser um impeditivo para a produção da proteína-alvo em determinados tipos de células. O processamento de uma proteína após a tradução pode envolver apenas um conjunto de reações aditivas à molécula do peptídeo tais como glicosilação, fosforilação, metilação, etc. (KRISHNA;WOLD,1993; RAMAZI;ZAHIRI,2021). Outro tipo de processamento pós-traducional, muito frequente em proteínas heterólogas, é a formação de pontes dissulfeto entre resíduos de cisteína (Cys). A linhagem de *E. coli* K12 originais não

têm a capacidade de formar este tipo de ligação em seu citoplasma. No entanto, há sistemas de expressão comerciais com linhagens modificadas desta estirpe de *E. coli* com genes capazes de alterar a característica oxi-redutora do citosol, como relata KAUR *et al.* (2018). Na tabela 2, abaixo, transcrita dos dados suplementares deste artigo podemos encontrar várias linhagens com propriedades específicas.

Tabela 2 – Linhagens comerciais disponíveis de *E. coli* (Transcrição da Tabela 3 – Dados Suplementares de KAUR *et al.* (2018))

Strains	Derivation	Features	Company
<i>B834</i>	<i>B-strain</i>	<i>Met auxotroph, 35S-met labelling</i>	<i>Novagen, USA</i>
<i>BL21</i>	<i>B-strain</i>	<i>Expression host for plasmids/vectors that do not use T7 promoter Host used with pGEX, pMAL, pTrcHis vectors</i>	<i>Novagen, USA; NEB, USA</i>
<i>BL21 (DE3)</i>	<i>B834</i>	<i>Basic IPTG-inducible strain containing T7 RNAP (DE3) Deficient in lon and ompT proteases Host used for pET series vectors</i>	<i>Addgene, USA; Stratagene, USA; NEB, USA; Novagen, USA</i>
<i>BL21 (DE3) pLysS</i>	<i>BL21</i>	<i>pLysS expresses T7 lysozyme to reduce basal expression level. Toxic protein expression</i>	<i>Addgene, USA; Stratagene, USA</i>
<i>BL21 (DE3) pLysE</i>	<i>BL21</i>	<i>pLysE has higher T7 lysozyme expression than pLysS Toxic protein expression</i>	<i>Addgene, USA; Novagen, USA</i>
<i>BLR (DE3)</i>	<i>BL21</i>	<i>RecA-deficient; best for plasmids with repetitive sequences. Unstable protein expression Toxic or insoluble protein expression</i>	<i>Addgene, USA</i>
<i>Tuner (DE3)</i>	<i>BL21</i>	<i>Mutated lac permease that allows linear control of expression Toxic or insoluble protein expression</i>	<i>Addgene, USA</i>
<i>Origami (DE3)</i>	<i>K-12</i>	<i>Double mutant in thioredoxin reductase (trxB) and glutathione reductase (gor) Facilitate proper folding by enhancing disulfide bond formation in cytoplasm Insoluble protein expression</i>	<i>Novagen, USA</i>
<i>Origami B</i>	<i>BL21</i>	<i>Double mutant in thioredoxin reductase (trxB) and glutathione reductase (gor) Facilitate proper folding by enhancing disulfide bond formation</i>	<i>Novagen, USA</i>
<i>AD494</i>	<i>K-12</i>	<i>Mutation in thioredoxin reductase (trxB)</i>	<i>Novagen, USA</i>

Strains	Derivation	Features	Company
		<i>Facilitate cytoplasmic disulfide bond formation</i>	
<i>BL21TrxB(D E3)</i>	<i>BL21</i>	<i>Similar to AD494</i>	<i>Novagen, USA</i>
<i>Rosetta (DE3)</i>	<i>BL21</i>	<i>pRARE plasmid contains 7 tRNAs for rare codons Eukaryotic protein expression</i>	<i>Novagen, USA</i>
<i>Rosetta-gami</i>	<i>Origami</i>	<i>pRARE plasmid contains 7 tRNAs for rare codons Eukaryotic protein expression Deficient in lon and ompT proteases Facilitate cytoplasmic disulfide bond formation</i>	<i>Novagen, USA</i>
<i>BL21 (DE3) CodonPlus-RIL</i>	<i>BL21</i>	<i>Expresses rare tRNAs; useful for AT-rich genomes Deficient in lon and ompT proteases</i>	<i>Stratagene, USA</i>
<i>BL21 (DE3) CodonPlus-RP</i>	<i>BL21</i>	<i>Expresses rare tRNAs; useful for GC-rich genomes Deficient in lon and ompT proteases</i>	<i>Stratagene, USA</i>
<i>C41(DE3)</i>	<i>BL21(DE3)</i>	<i>Single mutant Membrane bound protein expression</i>	<i>Avidis, France</i>
<i>C43(DE3)</i>	<i>BL21</i>	<i>Double mutant Membrane bound protein expression</i>	<i>Avidis, France</i>
<i>BL21-SI</i>	<i>BL21</i>	<i>Salt inducible T7 RNA polymerase Expression is induced with NaCl Better/strident control of toxic gene expression</i>	<i>Invitrogen, USA</i>
<i>BL21-AI</i>	<i>BL21</i>	<i>Arabinose inducible T7 RNA polymerase Expression is induced with arabinose Better control of toxic gene expression</i>	<i>Invitrogen, USA</i>
<i>TKB1</i>	<i>BL21 (DE3)</i>	<i>Generates phosphorylated proteins as it carries elk tyrosine kinase gene</i>	<i>Stratagene, USA</i>
<i>Lemo 21 (DE3)</i>	<i>BL21 (DE3)</i>	<i>Rhamnose-tunable T7 RNAP expression alleviates inclusion body formation. Toxic, insoluble or membrane protein expression</i>	<i>Addgene, USA; NEB, USA</i>
<i>T7 Express</i>	<i>BL21</i>	<i>Does not restrict methylated DNA General protein expression</i>	<i>Addgene, USA</i>
<i>M15 pREP4</i>	<i>K-12</i>	<i>T5 promoter efficiently regulated by lac repressor encoded by lacI gene on pREP4 plasmid Toxic protein expression Expression host used mostly with pQE vectors</i>	<i>Addgene, USA; Qiagen, Germany</i>

Strains	Derivation	Features	Company
<i>HMS174 (DE3)</i>	<i>K-12</i>	<i>recA mutant rif resistance</i>	<i>Novagen, USA</i>
<i>JM 83</i>	<i>K-12</i>	<i>Facilitates the secretion of recombinant protein into periplasm</i>	<i>Addgene, USA</i>
<i>JM109</i>	<i>K-12</i>	<i>recA and endA deficient Harbour lacI^q mutation</i>	<i>Novagen, USA</i>
<i>TB1</i>	<i>JM 83</i>	<i>Host for pUC and pMAL vectors</i>	<i>NEB, USA</i>
<i>TOP10</i>	<i>DH-10B</i>	<i>Host for cloning and plasmid propagation Expression host used for vectors containing BAD promoter</i>	<i>Invitrogen, USA; ThermoFisher Scientific, USA; NEB, USA</i>
<i>SG13009</i>	<i>K-12</i>	<i>Carries pREP4 plasmid Constitutive expression of lac repressor at high levels Expression host used with pQE vectors</i>	<i>Qiagen, Germany</i>
<i>SHuffle</i>	<i>B strain</i>	<i>dsbC gene for proper disulphide bond formation Enhanced BL21 derivative</i>	<i>NEB (New England Biolabs), USA</i>
<i>SHuffle Express</i>	<i>B strain</i>	<i>Constitutive expression of disulphide bond isomerase DsbC</i>	<i>NEB, USA</i>
<i>Shuffle T7</i>	<i>K-12</i>	<i>Suitable for T7 promoter-driven protein expression</i>	<i>NEB, USA</i>
<i>LMG194</i>		<i>Host used for vectors containing BAD promoter</i>	<i>Invitrogen, USA</i>
<i>DL41 (DE3)</i>		<i>Labelling of protein with 35S-methionine and selenomethionine</i>	<i>C. Muller, Germany</i>
<i>Arctic express</i>	<i>B strain</i>	<i>Chaperone encoded gene adapted to low temperature lac, tac and trc promoter-driven expression vector</i>	<i>Agilent Technologies, USA</i>
<i>Arctic express (DE3)</i>	<i>B strain</i>	<i>T7 promoter-driven expression vector like pCAL and pET vectors</i>	<i>Agilent Technologies, USA</i>

Quanto maior a complexidade do processamento pós-traducional requerido, mais importante será a seleção das células hospedeiras e do sistema de expressão.

2.2.3 Toxicidade da proteína-alvo

A proteína-alvo deve ser avaliada quanto a possibilidade de ser tóxica para a célula hospedeira. Neste caso, normalmente, devemos buscar o direcionamento da proteína para o periplasma ou reduzir a taxa de expressão gênica evitando, assim,

qualquer ação deletéria para o organismo. Linhagens de células nas quais é possível bioquimicamente controlar a indução da expressão também podem ser utilizadas. Outra possibilidade seria reduzir o número de plasmídeos no interior da célula ou, ainda, utilizar linhagens tolerantes à proteína tóxica. Estas são, entre outras, estratégias propostas por ROSANO e CECCARELLI (2014) para contornar o problema de baixa taxa de expressão devido à toxicidade da proteína para a célula hospedeira.

3 EXTRAÇÃO E PURIFICAÇÃO DE PROTEÍNAS

Selecionamos três aspectos relacionados diretamente à extração e purificação das proteínas produzidas em sistemas heterólogos para uma análise mais detalhada, quais sejam: (I) localização final da proteína; (II) existência ou não de moléculas fusionadas à proteína; e (III) solubilidade da proteína.

3.1 LOCALIZAÇÃO FINAL DA PROTEÍNA HETERÓLOGA

Os sistemas de extração e purificação de proteínas dependem basicamente do local onde a proteína-alvo será armazenada. A proteína poderá acumular-se no citoplasma, no periplasma ou no espaço extracelular, por exemplo. No caso de sistemas heterólogos, normalmente envolvendo células em lugar de tecidos, o espaço extracelular referido anteriormente corresponde ao meio de cultivo. Existem vantagens e desvantagens associadas à localização da proteína ao final do processo de produção. A extração a partir do periplasma é, a princípio, preferível à extração a partir do interior da célula (citoplasma), pois neste último caso teremos, em geral, mais contaminantes e riscos de degradação da proteína produzida ou formação de corpos de inclusão. Quanto ao espaço extracelular (meio de cultivo), é necessário previamente certificar-se da existência efetiva de rotas de secreção no tipo de células selecionadas, como abordam HANNIG e MAKRIDES (1998).

3.2 PROTEÍNAS FUSIONADAS

A utilização de moléculas fusionadas à proteína-alvo é uma das técnicas empregadas para permitir a separação efetiva e posterior purificação da proteína a partir de extratos brutos de células. Esta estratégia proporciona certa padronização nestes processos, considerando que a molécula fusionada passa a ser uma parte conhecida da proteína expressa, direcionando, assim, as etapas de extração e purificação dos sistemas comerciais existentes. Outras propriedades derivadas da adição de uma proteína auxiliar ao peptídeo-alvo justificam eventualmente suas utilizações. Entre elas podemos citar o aumento da estabilidade da proteína, o direcionamento para um compartimento específico da célula, visando algum tipo de processamento pós-traducional, facilitar a posterior identificação da proteína, a inativação bioquímica temporária, o aumento da solubilidade, etc., como abordado

nos artigos publicados por FORD *et al.* (1991); UHLÉN *et al.* (1992), e YOUNG *et al.* (2012).

3.3 SOLUBILIDADE DA PROTEÍNA

A manutenção da solubilidade da proteína durante as etapas de expressão heteróloga é um aspecto de extrema importância para a obtenção da proteína-alvo com rendimento e qualidade adequados. Portanto, devemos sempre buscar, preferencialmente, uma configuração do sistema de expressão que resulte nesta condição. Para isto, é necessário modular a eficiência de expressão gênica e de formação da proteína. Outro aspecto a ser controlado é a temperatura durante o período de indução, a qual deve ser reduzida de modo a desacelerar as interações que produzem a agregação das moléculas. Quando se deseja obter uma proteína na forma solúvel, devemos considerar também o momento exato para iniciar a indução da expressão. Deste modo, é preciso determinar, em cada caso, o ponto ótimo da curva de multiplicação celular (início, meio ou fim da fase Log) para adição de um indutor químico no meio de cultivo. Quanto às possíveis medidas de controle relacionadas ao meio de cultivo, a adição de glicose pode, por exemplo, reprimir o operador *lac*, reduzindo a expressão gênica e a formação de proteínas a um nível basal. Outros aditivos no meio de cultivo também podem ser utilizados com propósitos de permitir o enovelamento correto da proteína, a manutenção de sua estabilidade, funções protetora ou estabilizadora da proteína-alvo. Entre os tipos conhecidos de aditivos podemos citar: (I) Detergentes, tais como Tween, Triton X-100, Sarkosyl (do inglês, *Sodium lauroyl sarcosinate*) e outros; (II) Agentes Caotrópicos, tais como Ureia, Guanidina HCl, CaCl₂, N-Etil-uréia, e outros; (III) Aminoácidos e derivados, tais como Arginina Etil Ester, Glutamato de potássio, e outros; (IV) Açúcares e osmólitos, tais como Trealose, Manitol, inositol e outros; (V) Etanol, ciclohexanol, e outros. Para informações mais completas, tais como a concentração usual, e outras alternativas de aditivos consultar a tabela 5 do artigo de KAUR *et al.* (2018)

Por fim, a produção conjunta de moléculas chaperonas especializadas na preservação da estrutura e funcionalidade da proteína-alvo também é alternativa plausível para obter proteínas na forma solúvel (AHMAD *et al.*,2018; KAUR *et al.*,2018).

4 CARACTERIZAÇÃO DE PROTEÍNAS

A caracterização de proteínas envolve a determinação ou confirmação de sua estrutura assim como o conhecimento do modo como interagem com outras proteínas, com as demais moléculas presentes no organismo e com o meio em que a célula está localizada. Por exemplo, enzimas, fatores transcricionais, de tradução e de replicação, inibidores e ativadores em geral, transportadores de membrana e receptores são proteínas com propriedades bioquímicas que determinam seus papéis fisiológicos nos seres vivos. Diversos ensaios são normalmente utilizados para a caracterização proteica. A seguir, apresentamos resumidamente alguns deles indicando referências para um estudo mais detalhado das técnicas e suas possíveis aplicações.

4.1 ENSAIOS QUALITATIVOS E QUANTITATIVOS

Entre os primeiros ensaios necessários à caracterização proteica estão aqueles que apenas informem: (I) se a proteína-alvo foi realmente produzida (análise qualitativa); e (II) qual a quantidade de peptídeo produzido (análise quantitativa). Para estas primeiras análises, as amostras obtidas após o processo de extração das proteínas totais das células poderão ser utilizadas diretamente sem qualquer outro processamento ou purificação preliminar. A seguir, os ensaios tradicionalmente empregados com estas finalidades.

4.1.1 Eletroforese em Gel de Poliacrilamida

O ensaio de Eletroforese em Gel de Poliacrilamida, com Dodecil Sulfato de Sódio (SDS-PAGE, do inglês, *Sodium Dodecyl Sulfate – Polyacrylamide Gel Electrophoresis*) é uma técnica que promove a desnaturação de proteínas e de separação eletroforética, a qual permite obter estimativas sobre o peso molecular do peptídeo. Conhecida a estrutura primária da proteína-alvo deduzida a partir da sequência gênica (nucleotídica), assim como de qualquer outra proteína ou molécula fusionada, pode-se inferir o conteúdo de uma banda formada no gel comparando sua posição com a posição das bandas formadas por um marcador de peso molecular utilizado (informação qualitativa), assim como obter uma estimativa da quantidade de polipeptídeo (informação quantitativa) comparando a intensidade da banda da amostra com as intensidades das bandas do marcador, cuja quantidade é conhecida. Para visualização direta e registro das imagens das bandas formadas no gel, são utilizadas

soluções contendo agentes corantes do tipo *Coomassie Blue* ou sais de prata que se ligam à proteína revelando a imagem diretamente no gel. Para realização destas análises, podemos utilizar protocolo descrito em AUSUBEL *et al.* (1999).

4.1.2 *Western Blot*

A técnica de *Western Blot* permite maior seletividade ou especificidade na identificação da proteína-alvo ao empregar anticorpos em relação às técnicas eletroforética com corante descritas no item anterior, já que os corantes normalmente utilizados ligam grupos químicos presentes na maioria das proteínas de modo inespecífico.

A técnica de *Western Blot* inicia com um gel de PAGE ou SDS-PAGE e a transferência das proteínas deste para uma membrana-suporte, podendo ser composta de nitrocelulose ou náilon, aplicando-se uma corrente elétrica ou por gravidade. Após a transferência, as superfícies livres da membrana (onde não se depositaram as proteínas transferidas do gel) são recobertas com outra proteína que não tem especificidade para ligação com os anticorpos selecionados como, por exemplo, caseína ou albumina (etapa de bloqueio). Na sequência, ocorre ligação dos anticorpos com as proteínas específicas. Na etapa final de revelação, os anticorpos primários são ligados por uma segunda classe de anticorpos que, por sua vez, estão covalentemente ligados à grupos luminescentes, fluorescentes ou enzimas que catalizam a formação de produtos de coloração conhecida e específica (análise qualitativa). No caso de anticorpos secundários fluorescentes ou luminescentes, a membrana é exposta a filmes autorradiográficos que, posteriormente, são revelados fotograficamente. No caso de anticorpos acoplados a enzimas, a membrana-suporte fica direta e especificamente marcada pela cor dos produtos no local das bandas onde os anticorpos primários ligaram (AUSUBEL *et al.*,1999). Do mesmo modo que na técnica abordada no item anterior, poderemos também estimar a quantidade da proteína-alvo pela intensidade da cor produzida com a ajuda de *softwares* como, por exemplo, *Image-J* (SCHNEIDER *et al.*,2012; MCKENZIE *et al.*,2015; ANTIABONG *et al.*,2016).

4.2 ENSAIOS PARA DETERMINAÇÃO ESTRUTURAL DA PROTEÍNA

O conhecimento e a eventual confirmação das estruturas primária, secundária, terciária e quaternária de uma proteína tem grande relevância no entendimento de sua função bioquímica e do seu comportamento físico-químico. Portanto, para uma completa caracterização de proteínas, é preciso determinar, ainda que *in silico*, estas estruturas para posterior validação por meio de técnicas específicas.

Para confirmação da estrutura primária são utilizados métodos de espectrometria de massas como a Ionização e Dessorção a Laser Assistida por Matriz com Analisador de Tempo de Voo (MALDI-TOF, do inglês, *Matrix-Assisted Laser Desorption/Ionization -and Time of Flight*).

Para obtenção de informações sobre as estruturas secundárias (hélices α , folhas β , e alças), sua dinâmica e estabilidade, pode ser utilizada a técnica de Dicroísmo Circular.

Duas técnicas muito conhecidas e relevantes para estudos estruturais detalhados de proteínas, envolvendo a estrutura terciária e quaternária são a Cristalografia por Raios X e a Ressonância Magnética Nuclear (RMN ou NMR, do inglês, *Nuclear Magnetic Resonance*).

4.2.1 MALDI-TOF

O ensaio MALDI-TOF ou qualquer outro método que envolva espectrometria de massas (MS, do inglês, *Mass Spectrometry*), permite determinar a estrutura primária peptídica, ou seja, a sequência de resíduos de aminoácidos que compõem a cadeia peptídica, assim como eventuais grupos prostéticos ligados a estes resíduos (BONK;HUMENY,2001; KISLINGER *et al.*,2005; CARDAMONE,2010; KAFKA *et al.*,2011; GOMES *et al.*,2014; HOSSAIN;KOMATSU,2014).

A técnica inicia com a vaporização da amostra depositada sobre uma matriz formada por ácidos orgânicos fracos, responsáveis pela ionização das moléculas. Após a dessorção das moléculas pela ação de um laser, as partículas carregadas são aceleradas em um campo elétrico alcançando os detectores. O tempo de deslocamento (tempo de voo) destas moléculas até serem detectadas fornecem os dados referentes a relação entre a massa e a carga da partícula.

4.2.2 Dicroísmo Circular

O dicroísmo circular permite determinar as estruturas secundária da proteína em estudo. Por meio desta técnica, também podemos estudar como é afetado o enovelamento quando a proteína é exposta a variações de pH e temperatura do meio, acrescentando, assim, mais informações sobre a estabilidade da proteína e sua ação bioquímica. A técnica se utiliza de grupos cromóforos presentes nas ligações peptídicas e a submissão da solução a feixes de radiação plano-polarizadas na região de UV próximo (320 a 260 nm) e UV distante (240 a 180 nm) que circulam em torno da amostra no sentido horário e anti-horário. São analisadas as diferenças de absorção entre os dois sentidos de rotação (MCMILLIN;WALTON,1974; JOHNSON,1999; KELLY;PRICE,2000; KELLY *et al.*,2005; BULHELLER *et al.*,2007; LOUIS-JEUNE *et al.*,2012).

4.2.3 Cristalografia por Raios X

A técnica de cristalografia por raios X utiliza a difração sofrida por um feixe de raios X que colide com um cristal da proteína e cujo espectro consequente de espalhamento de radiação é registrado pela sensibilização de filmes autorradiográficos (GUINIER,1952) e, mais modernamente, por detectores digitais. O cristal é, normalmente, obtido por um processo lento decorrente da supersaturação de uma solução a qual contém a proteína a ser analisada e, eventualmente, outras moléculas que se deseja pesquisar quanto às possíveis interações. É um processo oneroso em tempo e recursos que se justifica pela alta qualidade de seus resultados. Podemos obter informações referentes à formação das estruturas secundária, terciária e até quaternária, quando existentes, incluindo outras moléculas ligadas à proteína ou que apresentem algum tipo de interação. A principal limitação desta técnica envolve a dificuldade de obtenção dos cristais. Esta dificuldade eventualmente resulta na necessidade de utilizar soluções que interferem no enovelamento nativo da proteína (MONTELLIONE *et al.*,2000; SMYTH;MARTIN,2000; BILLETER *et al.*,2008).

4.2.4 Ressonância Magnética Nuclear

A técnica de RMN permite que a proteína seja analisada em solução com sua estrutura e atividade integralmente preservadas. É possível realizar análises submetendo a amostra a condições diversas como, por exemplo, alterando o pH para

visualizar os efeitos estruturais (MONTELIONE *et al.*,2000). A limitação quanto ao emprego desta técnica relacionada à massa (tamanho) da proteína vem sendo atualmente superada como nos relata em seu trabalho KARUNANITHY *et al.* (2021) que apresenta os mais recentes avanços metodológicos alcançando estudos com complexos proteicos de até 200 kDa. Assim como a cristalografia por raios X, esta técnica possibilita conhecer a proteína e suas interações com outras moléculas. Existem algumas peculiaridades da técnica que, dependendo do tamanho da proteína, é preciso trabalhar com um percentual mínimo de isótopos ^{13}C e ^{15}N , considerando que eles não são abundantes na natureza. (MONTELIONE *et al.*,2000; STAUNTON *et al.*,2003). A técnica de RMN cumpre uma função complementar importante nas análises estruturais onde não é possível obter cristais das proteínas para análise por difração de raios X, ou quando é necessária a análise sobre a dinâmica da proteína em estudo. Outros autores podem ser consultados para completa compreensão da técnica, suas vantagens e desvantagens e, entre eles, podemos citar ZUIDERWEG (2002); BILLETER *et al.* (2008); MARKWICK *et al.* (2008); WAGNER (2010)

4.3 ENSAIOS PARA DETERMINAÇÃO DE ATIVIDADE INIBITÓRIA

Para caracterização de proteínas com atividade bioquímica, são necessárias técnicas destinadas a determinar a atividade enzimática ou inibitória de enzimas. Para as proteínas do tipo inibidores de protease são utilizados basicamente dois tipos de ensaios, os quais estão descritos resumidamente a seguir.

4.3.1 Ensaio da Benzoil-DL-Arginina-*p*-Nitroanilida

O ensaio que utiliza benzoil-DL-arginina-*p*-nitroanilida (BAPNA, do inglês, *Benzoyl-DL-Arginine-p-NitroAnilide*) descrito no artigo de KAKADE *et al.* (1974) foi originalmente desenvolvido com a finalidade específica de medir a ação inibitória de produtos derivados de soja sobre a tripsina, visando aperfeiçoar o método publicado pelo mesmo pesquisador em 1969, quanto à medição da ação antitriptica de produtos da soja submetidos a tratamento térmico. Um aperfeiçoamento posterior proposto por HAMERSTRAND *et al.* (1981) ainda tem sido utilizado para aplicação desta técnica. A técnica caracteriza-se pela hidrólise triptica do substrato sintético BAPNA, o qual libera a *p*-nitroanilida. Este produto tem cor amarela (absorção na faixa de 410 nm). Quanto mais intensa a cor, menor é a atividade inibitória sobre a tripsina resultante da

presença do analito. Assim, um amarelo intenso representa uma atividade inibitória nula, enquanto um amarelo quase inexistente representa uma atividade inibitória elevada.

4.3.2 Titulação Calorimétrica Isotérmica

O ensaio de titulação calorimétrica isotérmica (ITC, do inglês, *Isothermal Titration Calorimetry*) permite medir o calor liberado durante a reação de ligação entre moléculas em geral e, entre elas, a reação de ligação entre uma enzima e seu substrato. Considerando que o mecanismo de ação de um inibidor de um determinado tipo de enzima caracteriza-se pela ocupação do espaço destinado ao substrato e por algum tipo de interação química, a intensidade desta interação seria o fator determinante para aferir a atividade inibitória. Portanto, se desejarmos comparar a atividade inibitória de proteases entre duas moléculas que apresentam este tipo de comportamento, bastaria medir a liberação de calor quando a enzima (no caso, uma protease) reagir, separadamente, com cada uma das moléculas com efeito inibidor que se deseja comparar. Detalhes da teoria na qual se embasa esta técnica e referências aos protocolos utilizados podem ser encontrados nos trabalhos de WISEMAN *et al.* (1989) e VELÁZQUEZ-CAMPOY *et al.* (2004).

5 PROPOSIÇÃO DE METODOLOGIA PARA PRODUÇÃO DE KTIMET

Neste capítulo, será apresentado e discutido um roteiro para a expressão heteróloga e produção da proteína recombinante KTIMet em *E. coli*, abordando os tópicos apresentados nos capítulos anteriores.

Para uma melhor visualização do conjunto de técnicas sugeridas, que compõem a metodologia proposta, apresentamos, no **Apêndice A**, um diagrama com o resumo de todas as etapas projetadas.

5.1 EXPRESSÃO HETERÓLOGA DO GENE *ktimet*

A definição do sistema de expressão inicia pelo conhecimento da proteína-alvo e suas propriedades principais, tais como estrutura e atividades bioquímicas.

5.1.1 Características da proteína KTIMet

KTIMet é uma proteína recombinante com aproximadamente 21 kDa, resultado da substituição de códons de leucina (Leu) e isoleucina (Ile) por códons de metionina (Met) e cuja sequência nucleotídica original foi obtida do *GenBank* a partir do trabalho de SEVERIN *et al.* (2010).

A ideia que deu origem à KTIMet está relacionada a uma deficiência nutricional característica das leguminosas em geral, entre elas a soja. Esta deficiência refere-se à baixa disponibilidade do aminoácido Met no conjunto geral de proteínas presentes no grão. Sendo este um aminoácido essencial, ou seja, não produzido pelos animais em geral, deve ser obtido por meio da alimentação. Assim, a busca de uma proteína de reserva de metioninas, no conjunto de proteínas da soja, capaz de mitigar esta deficiência tem um valor biotecnológico relevante, considerando que os produtos derivados de soja são, atualmente, a base da indústria mundial de produção de proteína animal, como principal fonte de proteína para formulação de rações animais.

A concepção geral da proteína KTIMet, envolvendo, entre outros, os seguintes aspectos: (I) escolha de um inibidor de tripsina tipo Kunitz para receber as metioninas adicionais e funcionar como proteína de reserva, (II) justificativa para substituição dos resíduos de Leu e Ile, encontram-se detalhados na dissertação de mestrado do autor (DELLA_GIUSTINA_NETO,2015) de onde foi retirada Figura 1 apresentada a seguir.

```

TAACTAATTAATGTGAATCATATCAAATAATGAAAAGTAAATAAAATTGTAATTAACCTCTATATGTATACACACACAAATA
ATAAATAATAGTAAAAAATTTATGATAAATATTTACCATCTCATAAAGATATTTAAAATAATGATAAAAAATATAGATTATT
TTTTATGCAACTAGCTAGCCAAAAAGAGAACACGGGTATATATAAAAAGAGTACCTTTAAATTTCTACTGTACTTCTTTATTC
CTGACGTTTTTATATCAAGTGGACATACGTGAAGATTTTAATTATCAGTCTAAATATTTTCATTAGCACTTAATACTTTTCTGT
TTTATTCCTATCC[TATAA]GTAGTCCCATTCTCCCAACATTGCTTATTCACACAACCTAAGAAAAGTCTTCCATAGCCCCC
CAAAAATG[AAGAGCACCATCTTCTTCTTCTTCTTCTTGTGCCTTCACCACCTCATACCTACCTTCAGCCATCGCTGATTTTC
GTGATGGATAATGAAGGTAACCCATGGAAAATGGTGGCACATATTATATGATGTCAGACATGACAGCATTGGTGGAAATGAG
AGCAGCCCCAACGGGAAATGAAAGATGCCCTATGACTGTGGTCAATCTCGCAATGAGATGGACAAAGGGATGGGAACAATGA
TGTCGTCCCATAT[CGAATGCGT]TTTATGGCCGAAGGCCATCCTATGAGCATGAAGTTCGATTTCATTGTCAGTTATGATGATG
TGTGTTGGAATGCCTACCGAGTGGTCTGTGTGGAGGATATGCCAGAAGGACCTGCTGTTAAAATGGGTGAGAACAAGATGC
AATGGATGGTTGGTTTAGAATGGAGAGAGTTTCTGATGATGAATTCAATAACTATAAGATGGTGTCTGTCCACAGCAAGCTG
AGGATGACAAATGTGGGGATATGGGGATGAGTATGATATGATGATGGAACCAGGCGTATGGTGGTGTCTAAGAACAACCG
ATGGTGGTTCAGTTTCAAAAATGGATAAGAATCAATGGCCAAGAAAAATCATGGCATGCTCCTCGCAGTGAGTGAAGACAAAG
TGTGAGAGTACTAAATAAATGCTTTGGTTGTACGAAATCATTACACTAAATAAAAATAATCAAAGCTTATATATGCTTCTAAG
GCCGAATGCAAAGAAATTGGTTCTTCTCGTTATCTTTGCCACTTACTAGTACGTATTAATTACTACTTAATCATCTTGTTC
GGCTCATTATATCC

```

Figura 1 – Sequência completa do gene *ktimet*.

Na figura, podemos observar, destacados com fundo cinza e letras em negrito (N), os 41 nucleotídeos alterados de um total de 576, o que representa apenas 7% de bases modificadas na região codificante, em relação ao gene nativo. Os nucleotídeos da sequência codificadora do peptídeo-sinal (69 bp) estão destacados somente com fundo cinza- (NNN). O bloco do [TATAA box] está na região promotora à -30 do provável códon iniciador da tradução [ATG]; na região terminadora está o códon terminador [TGA]. A sequência codificadora da proteína KTIMet está em caracteres sublinhados. O local correspondente ao sítio ativo no gene nativo está indicado com uma borda dupla [NNN]. Edição baseada na Figura 10 da dissertação de mestrado de DELLA_GIUSTINA_NETO (2015).

Na figura 2 abaixo, podemos observar a estrutura primária das proteínas nativa e recombinante, destacando todas as metioninas adicionadas, o que poderá transformar este peptídeo, originalmente destinado à proteção da planta, em proteína de reserva de metioninas, aumentando seu valor nutricional. Nesta figura, também podemos constatar que a isoleucina, posicionada no sítio ativo, foi substituída por uma metionina. Esta substituição poderá agregar outra propriedade nutricional a proteína KTIMet ao reduzir ou eliminar a atividade inibitória de protease.

Sequência Nativa (Kti)

```

MKSTIFFLFLFCAFTTSYLPSAIADFVLDNEGNPLENGGTYIILSDITAFGGIRAAPTGNERCPLTVVQSRNELD
KGIGTIISSPY[RIR]FIAEGHPLSLKFDSFAVIMLCVGIPTESVVEDLPEGPAVKIGENKDAMDGWFRLERVSDD
EFNNYKLVFCPQQAEDDKCGDIGSIDHDDGTRRLVVSKNKPLVVQFQKLDKESLAKKNHGLSRSE*

```

Sequência recombinante (KTIMet) - para expressão em soja

```

MKSTIFFLFLFCAFTTSYLPSAIADFVMDNEGNPMENGGTYIMSDMTAFGGMRAAPTGNERCPLTVVQSRNEMD
KGMGTMMSSPY[RMR]FMAEGHPMSMKFDSFAVMMCVGMPTEWSVVEDMPEGPAVKMGENKDAMDGWFRMERVSDD
EFNNYKMVFCPQQAEDDKCGDMGMSMDHDDGTRRMVVSKNKPMVVQFQKMDKESMAKKNHGMSSRSE*

```

Figura 2 – Estruturas primárias das proteínas nativa e recombinante.

Na figura acima, apresentamos a estrutura primária da proteína nativa KTi e da proteína recombinante KTIMet. Nesta figura, podemos observar a sequência de aminoácidos do peptídeo-sinal destacada em cinza. Em negrito, dentro de uma caixa, estão destacados os resíduos de aminoácido pertencentes ao sítio ativo do inibidor, no qual foi substituída uma isoleucina por uma metionina e, em negrito, todas as metioninas adicionadas na molécula de KTIMet. Edição baseada na Figura 11 da dissertação de mestrado de DELLA_GIUSTINA_NETO (2015).

A proteína nativa, da qual se origina KTIMet apresenta um peptídeo-sinal, indicando que a proteína KTi é, em sua célula de origem, provavelmente traduzida na membrana do retículo endoplasmático (RE) e transportada em vesículas para a membrana celular ou armazenada em vesículas no citosol.

A partir desta inferência, podemos considerar dois aspectos relevantes quanto à expressão do gene nativo de onde deriva KTIMet: (I) a proteína nativa não é encontrada de forma dispersa no citosol, reforçando a ideia de sua possível toxicidade até para a célula onde é originalmente produzida; (II) a proteína nativa requer ao menos um processamento pós-traducional que é a remoção do peptídeo-sinal.

Com base em estudos estruturais detalhados da molécula de KTi realizados por DE MEESTER *et al.* (1998), a forma ativa destes inibidores de proteases, em geral, não requer qualquer processamento adicional do tipo glicosilação, fosforilação, etc.

A formação espontânea de corpos de inclusão, observada no trabalho de MAJOR e CONSTABEL (2008) em *E. coli*, proporciona ao mesmo tempo duas vantagens: (I) a eliminação de um possível efeito tóxico para a célula hospedeira; e (II) uma maior facilidade no procedimento de extração. No entanto, representa uma desvantagem para a obtenção da proteína na sua forma ativa. Os autores supracitados conseguiram obter a molécula de KTi na forma ativa a partir dos corpos de inclusão. Para isto, utilizaram a solubilização dos corpos de inclusão em tampões com pH alcalino e submeteram estes à diálise para obtenção das amostras na forma solúvel e com o enovelamento correto. A possibilidade de obter a proteína KTIMet diretamente na forma solúvel não pode ser descartada, apesar dos resultados obtidos por MAJOR e CONSTABEL (2008) indicarem o contrário.

Para remoção do peptídeo-sinal, adotamos a mesma estratégia de MAJOR e CONSTABEL (2008), ou seja, extrair a sequência indesejada por meio de uma reação em cadeia da DNA polimerase (PCR, do inglês, *Polymerase Chain Reaction*), utilizando como referência *primers* projetados durante o mestrado do autor, especificamente para este fim, porém com a introdução de nucleotídeos adicionais nas extremidades dos *primers*, necessárias para um melhor acoplamento das enzimas de restrição. As sequências completas dos *primers* estão apresentadas na Tabela 1 abaixo.

Tabela 1 – *Primers* flanqueadores de *ktimet* projetados para subclonagens em vetores de transformação de *E. coli*.

Designação	Sequência
Primer 5' (Forward primer)	5'- <u>CGGGATCC</u> <u>ATG</u> GATTTCGTGATGGATA-3'
Primer 3' (Reverse primer)	5'- <u>CCCAAGCTT</u> TTCACTCACTGCGAGACAT-3'

Fonte: *Primers* elaborados com base na dissertação de DELLA_GIUSTINA_NETO (2015).

Nas sequências apresentadas na Tabela 1 acima, as bases que se pretende adicionar ao *amplicon* estão sublinhadas, sendo elas: (I) códon iniciador **ATG** adicionado em função da retirada do peptídeo-sinal; (II) sítio de clivagem de BamHI (GGATCC); (III) sítio de clivagem de HindIII (AAGCTT); e (IV) nucleotídeos adicionais CG e CCC a montante dos sítios de restrição nos *primers* 5'- *forward* e 3'-*reverse*, como proposto por MAJOR e CONSTABEL (2008), de modo a facilitar o posicionamento adequado das enzimas de restrição sobre o *amplicon*, quando o mesmo for submetido, em etapa posterior, a uma digestão seguida de uma reação de ligação com o plasmídeo de expressão.

5.1.2 Seleção do tipo de célula para expressão de *ktimet*

Para a seleção do tipo de célula a ser utilizada na produção da proteína recombinante, é preciso conhecer a família de inibidores de proteases do tipo KTi na sua forma nativa, o qual é produzido em diferentes tecidos vegetais, tais como folhas, caules e sementes (KOIWA *et al.*,1997). Logo, os organismos de origem desta proteína são geralmente eucariontes autotróficos. Existem sistemas de expressão que utilizam organismos eucariontes como fungos e leveduras para expressão heteróloga. No caso de KTIMet, considerando que não existe previsão de processamento pós-traducional para adição de moléculas à cadeia polipeptídica, não foi considerado este tipo de célula para a expressão de *ktimet*.

No artigo de referência selecionado no início deste trabalho (MAJOR;CONSTABEL,2008), a célula e as linhagens utilizadas foram derivadas de *Escherichia coli* K12. A partir desta definição, deveremos verificar quais os possíveis problemas para a produção de nossa proteína-alvo para mitigar ou eliminar os mesmos.

Diferentemente das proteínas estudadas por esses pesquisadores, a proteína KTIMet poderá apresentar desafios adicionais para sua produção quando comparada com a expressão de um inibidor de protease nativo. Entre estes desafios temos a quantidade adicional de metioninas na cadeia primária. A quantidade aumentada de metioninas poderá gerar dificuldades durante a tradução do mRNA recombinante por dois aspectos, quais sejam: (I) cada códon de metionina equivale a um códon iniciador de tradução, elevando, assim, a probabilidade de erro de início de tradução (PREIBISCH *et al.*,1988); e (II) possível escassez de tRNA-Met para o transporte destes aminoácidos durante a síntese proteica (SØRENSEN *et al.*,2003).

Outro ponto relevante é a composição do meio de cultivo das bactérias, o qual pode ser limitante em função das exigências de enxofre acima do usual, como nos indica os trabalhos de FUJIWARA *et al.* (1992); AWAZUHARA *et al.* (2002), CHIAIESE *et al.* (2004) e OHKAMA *et al.* (2002). Estes autores abordaram os efeitos da concentração de enxofre no meio sobre a expressão gênica e a acumulação de proteínas de reserva ricas em metioninas e cisteínas, em plantas.

Quanto a possíveis efeitos após a tradução de KTIMet, é importante destacar que o excesso de metioninas pode, também, trazer instabilidade à proteína, mesmo que estruturalmente o enovelamento não seja muito alterado em relação à molécula nativa, como de fato verificou-se nas análises *in silico* durante o mestrado do autor (DELLA GIUSTINA NETO, 2015). No entanto, persiste a expectativa fundamentada em um trabalho publicado de GASSNER *et al.* (2003) onde foram estudados os efeitos sobre o enovelamento e a estabilidade da T4 lisozima após múltiplas substituições de Leucina, Isoleucina, valina (Val) e fenilalanina (Phe) por Metionina na região C-terminal da proteína. Neste trabalho, os autores relataram pouco efeito sobre a estrutura em geral, mas uma perda progressiva de estabilidade com o aumento do número de metioninas devido a prováveis efeitos entrópicos e perdas de estabilidade hidrofóbica.

Quanto à linhagem de *E. coli* a ser utilizada, selecionamos duas bem conhecidas para expressão heteróloga, quais sejam: *E. coli* BL21 (DE3) e *E. coli* Lemo21 (DE3). A Linhagem BL21 é de uso geral e utilizada no trabalho de referência, enquanto a Lemo21 é notadamente mais eficiente quando se trata de expressar proteínas mais desafiadoras por permitir um controle mais apurado da indução da expressão (SCHLEGEL *et al.*,2012; ROSANO;CECCARELLI,2014).

5.1.3 Seleção do sistema de expressão

Com base no artigo de referência (MAJOR;CONSTABEL,2008), na disponibilidade de vetores para este sistema e na experiência prévia nos laboratórios do CBiot/UFRGS, optamos pelo sistema pET da Novagen para a expressão e produção de nossa proteína-alvo. A seguir descreveremos as etapas da metodologia proposta até a obtenção das culturas de células de *E. coli*.

5.1.4 Seleção do vetor de expressão do sistema pET (Novagen)

O sistema pET (Novagen) previamente selecionado está fundamentado na expressão de genes sobre forte controle transcricional do promotor do bacteriófago T7. A expressão é induzida ao fornecer uma fonte adicional de T7 RNA polimerase para a célula hospedeira. O sistema pode ser utilizado de dois modos básicos: (I) a célula hospedeira é infectada pelo fago λ CE6 que carrega o gene da T7 RNA polimerase sob controle dos promotores λ_{PL} e λ_{PI} após o período de multiplicação; (II) a célula hospedeira carrega em seu cromossomo uma cópia do gene da T7 RNA polimerase sob controle do repressor *lacUV5*, sendo a indução realizada pela adição de um análogo sintético da lactose, o isopropil- β -tiogalatosídeo (IPTG), ao meio de cultura. Uma descrição mais detalhada do sistema, dos vetores e das características das células hospedeiras, assim como os protocolos para cada etapa da expressão podem ser encontrados no manual do usuário (NOVAGEN,2011).

O sistema pET oferece uma série de plasmídeos com diferentes recursos. Foi selecionado o plasmídeo pET23a+ cujo mapa está apresentado na Figura 3. As razões que levaram à seleção deste vetor de transformação entre os disponíveis para o sistema pET foram as seguintes: (I) disponibilidade do plasmídeo nos laboratórios do CBiot/UFRGS; (II) o vetor carrega a sequência codificante para expressão de uma cauda de seis histidinas (His), a exemplo do plasmídeo utilizado por MAJOR e CONSTABEL (2008) no artigo de referência; (III) a cauda de His é expressa na região C-terminal, o que reduz a possibilidade de interferir na atividade bioquímica da molécula no caso de não ser removida após a purificação, já que o sítio ativo interage na região N-terminal na proteína nativa.

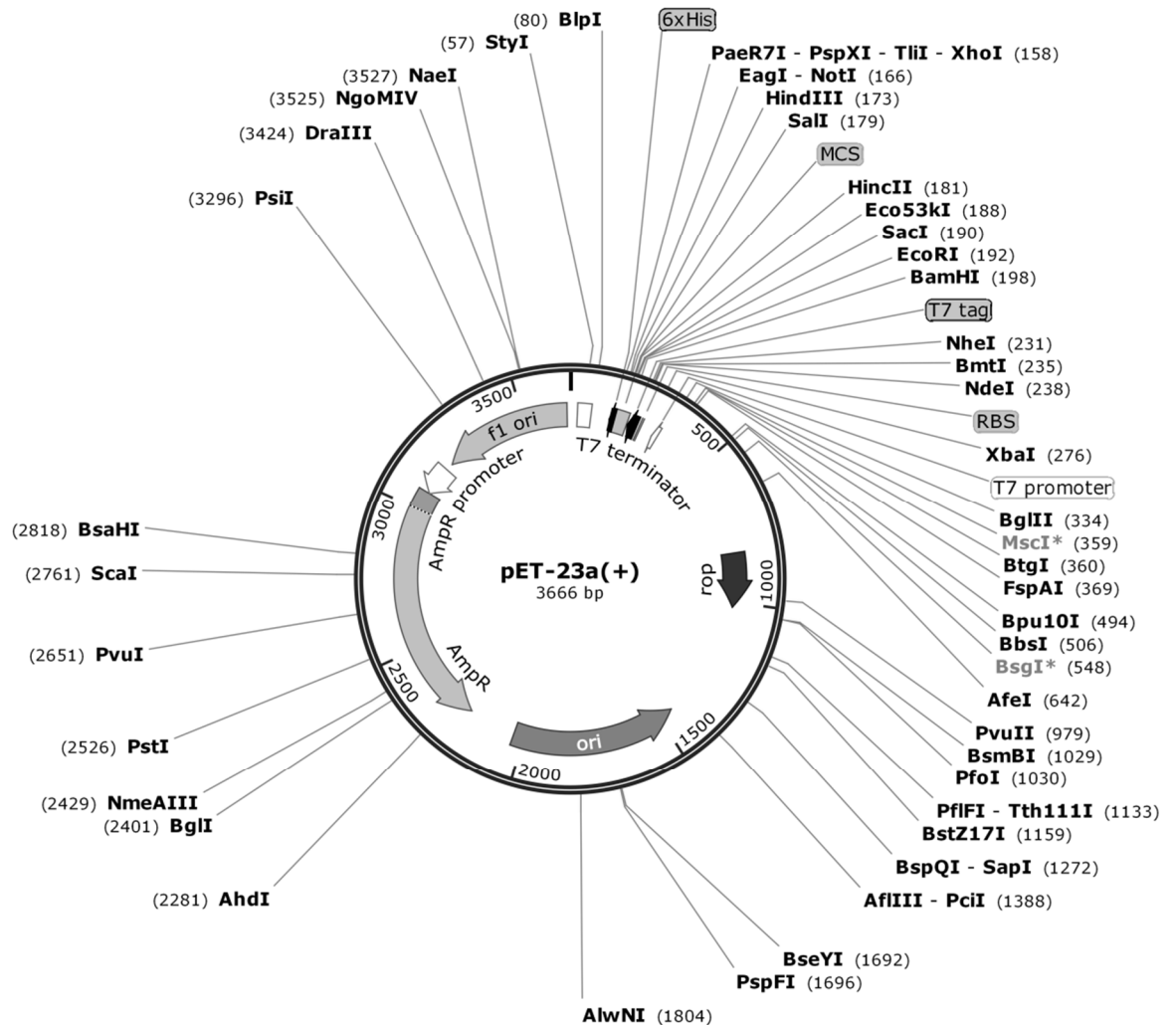


Figura 3 – Mapa do plasmídeo pET23a (+).

O plasmídeo pET23a (+) é o vetor de expressão do sistema pET selecionado para a produção heteróloga de KTIMet. O sinal (+) ao lado do nome do plasmídeo indica que o vetor contém uma origem f1 de replicação que permite a produção de DNA plasmidial de cadeia simples para mutagênese e aplicativos de sequenciamento. Na imagem podemos identificar: AmpR, o gene que confere resistência à ampicilina - produção de β -lactamase (gene bla); AmpR promoter, região promotora do gene bla; T7 terminator, região responsável pelo término da transcrição pela T7 RNA polimerase; T7 promoter, região promotora do gene codificador da T7 RNA polimerase; RBS, região de ligação do ribossomo; f1 ori, origem de replicação do bacteriófago; ori, origem de replicação plasmidial promotora da síntese de número elevado de cópias; 6xHis, região codificadora da cauda de His; MCS, sítio de multiclonagem, do inglês, Multiple Cloning Site. Figura gerada com o emprego do programa SnapGene version 6.2.1.

Na Figura 3 acima, podemos visualizar os sítios de BamHI (posição 198) e HindIII (posição 173), entre os quais será inserido o *amplicon* já orientado no sentido correto, logo após a digestão com as referidas enzimas de restrição.

5.1.5 Produção do amplicon contendo *ktimet* e ligação a pET23a

O ponto de partida da nossa rota de expressão é o plasmídeo projetado durante o mestrado do autor, denominado pUC57-*ktimet*, sintetizado pela empresa *GenScript* a partir de nosso plano de síntese (DELLA GIUSTINA NETO, 2015), o qual contém o gene completo codificador de KTIMet e cuja sequência nucleotídica foi apresentada anteriormente (Figura 1). Na Figura 4, o mapa do plasmídeo pUC57-*ktimet* é apresentado, sendo destacadas e posicionadas as regiões codificadoras de (I) KTIMet, (II) peptídeo-sinal, (III) *primer 5' forward* e (IV) *primer 3' reverse*.

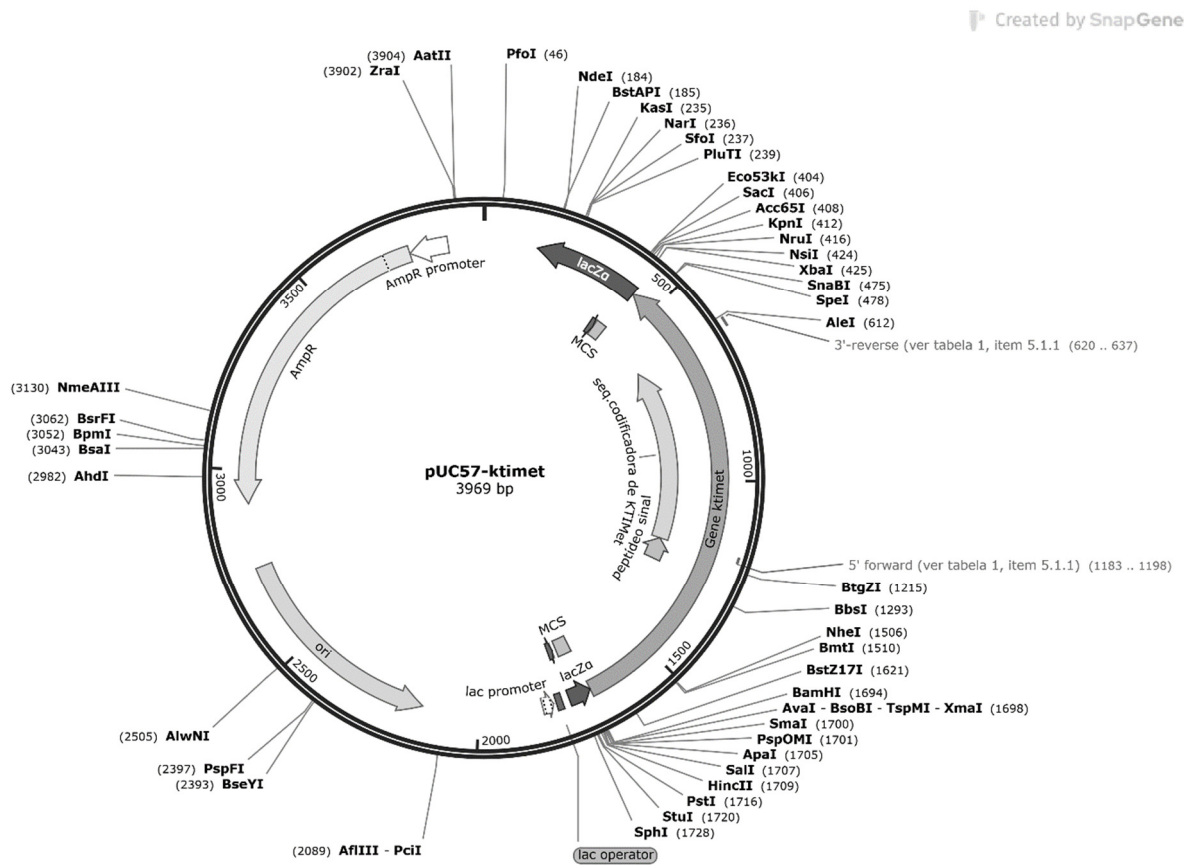


Figura 4 – Mapa do plasmídeo pUC57-*ktimet*.

Vetor de clonagem contendo o gene *ktimet* sintetizado pela *GenScript* durante o mestrado do autor (DELLA GIUSTINA NETO, 2015). Na imagem, pode-se identificar: AmpR, gene *bla* que confere resistência à ampicilina, codificador da β -lactamase; AmpR *promoter*, região promotora do gene *bla*; ori, origem de replicação; MCS, sítio de multiclonagem. Estão representados no mapa o gene *ktimet* completo e, em destaque, as regiões codificadoras do peptídeo-sinal e da proteína KTIMet. Estão posicionados os *primers 5'-forward* e *3'-reverse* que serão utilizados para a produção do amplicons por PCR. Figura gerada pelo emprego do programa *SnapGene version 6.2.1* a partir das sequências de DNA informadas pela *GenScript*.

A partir do plasmídeo pUC57-*ktimet*, pretendemos realizar uma PCR utilizando os *primers* flanqueadores projetados (Tabela 1) e uma DNA polimerase de alta fidelidade. Os parâmetros para ajuste do termociclador e as quantidades dos reagentes deverão seguir as orientações do fabricante da DNA polimerase disponível no momento da realização desta etapa, também é preciso levar em consideração a sequência dos primers para determinar a temperatura de anelamento da reação de PCR,

Finalizada a reação, seguir-se-á eletroforese para resolução em gel de agarose a 0,8% para visualização e obtenção de uma banda contendo o *amplicon*, no qual estão presentes a região codificadora de KTIMet e as sequências adicionais fornecidas pelos *primers* para as etapas posteriores. O *amplicon* deverá ser purificado a partir de um excerto do gel de agarose recortado com ajuda de um transluminador de luz Ultravioleta. Para a purificação, será utilizado o *kit* da Invitrogen denominado *PureLink Quick Gel Extraction Kit*, ou outro protocolo usual para este fim. Obtido o inserto de *ktimet* purificado na etapa anterior, será preciso realizar a digestão deste fragmento e do vetor de expressão pET23a+ utilizando as enzimas de restrição BamHI e HindIII. As condições da reação de digestão deverão seguir as orientações dos fabricantes. Será avaliada a possibilidade de realização da digestão em uma só etapa com as duas enzimas de restrição.

A escolha das enzimas de restrição BamHI e HindIII para a inserção do amplicon no plasmídeo pET23a+ foi resultado da análise de pelo menos dois aspectos, quais sejam: (I) presença de um único sítio de restrição para cada enzima nos fragmentos a serem digeridos; e (II) estarem posicionados no sentido correto de ligação ao plasmídeo de expressão. A ligação na orientação correta será, evidentemente, fundamental para a geração de um transcrito funcional.

A reação de ligação entre o inserto *ktimet* e o plasmídeo pET23a+ digeridos por BamHI e HindIII deverá ser efetivada com o uso de uma T4 DNA ligase. O protocolo da reação, as concentrações dos reagentes, temperaturas e etapas deverão seguir instruções do fabricante da enzima.

Ao final da etapa de ligação, os produtos serão diretamente utilizados na transformação genética de *E. coli* DH5 α por choque térmico (INOUE *et al.* (1990); AUSUBEL *et al.* (1999)). Plasmídeos serão “minipreparados” (AUSUBEL *et al.* (1999)) a partir das colônias resultantes sob seleção do antibiótico ampicilina. A presença do inserto *ktimet* em pET23a será confirmada pela digestão dos plasmídeos com BamHI

e HindIII. As versões positivas deverão apresentar fragmentos de ~3,7 kbp (pET23a) e ~0,6 kbp (*ktimet*) após eletroforese em gel de agarose a 0,8%. Finalmente, a partir da cultura de uma das colônias positivas, teremos o plasmídeo pET23a-KTIMet que será purificado (AUSUBEL *et al.*, 1999) e pronto para transformação das células de *E. coli* selecionadas para expressão da proteína-alvo.

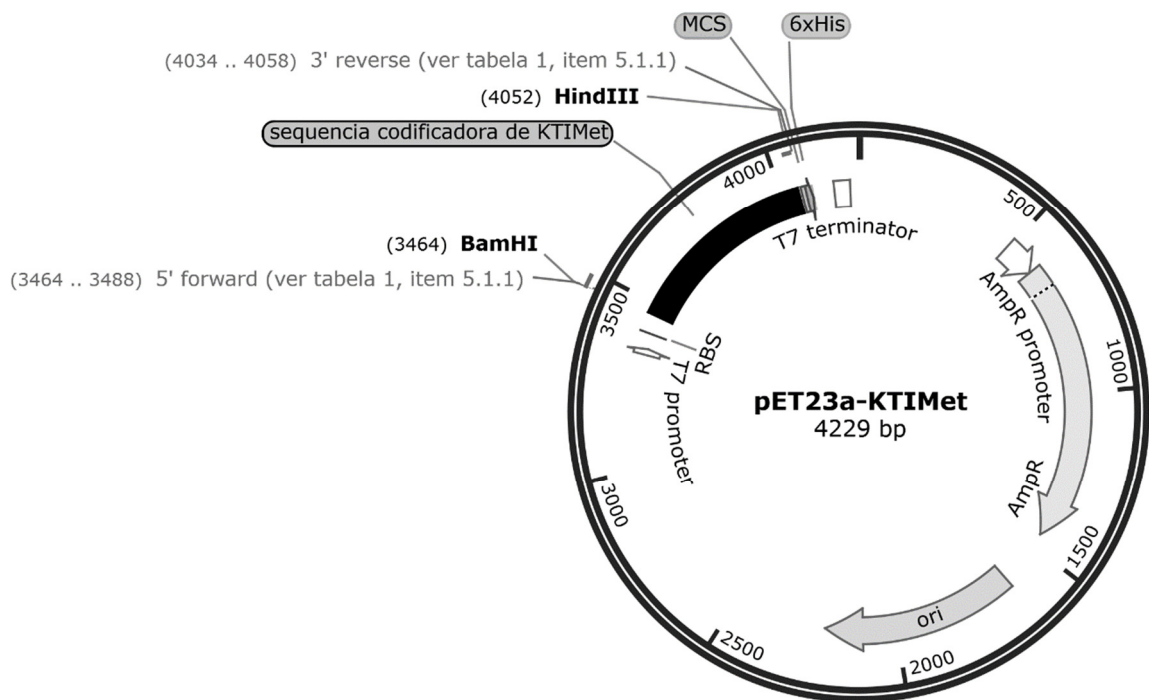


Figura 5 – Mapa do plasmídeo pET23a-KTIMet.

Vetor pET23a(+) contendo a sequência codificadora da proteína KTIMet sem o peptídeo-sinal. Na imagem, podemos identificar AmpR, gene *blasítio* que confere resistência à ampicilina pela produção de β -lactamase; AmpR promoter, região promotora do gene *bla*; T7 terminator, região responsável pelo término da transcrição pela T7 RNA polimerase; T7 promoter, região promotora da T7 RNA polimerase; ori, origem de replicação plasmidial; RBS, região de ligação do ribossomo; 6xHis, região codificadora da cauda de His; MCS, sítio de multiclonagem. Na figura, os primers utilizados na produção do amplicon *ktimet* e que fornecem o códon de início de tradução ATG e códon de fim de tradução TCA estão posicionados, assim como os sítios das enzimas BamHI e HindIII utilizados na digestão antes da reação de ligação. Figura gerada com o emprego do programa *SnapGene version 6.2.1*, a partir da biblioteca de vetores do próprio programa.

É recomendável que, antes da etapa de transformação de *E. coli* BL21 e Lemo21, seja confirmada a integridade da sequência codificadora de KTIMet. Para isto, consideramos utilizar o método de sequenciamento de DNA do tipo Sanger (SANGER; COULSON, 1975; CROSSLEY *et al.*, 2020). Para este sequenciamento, poderão ser utilizados os mesmos primers da Tabela 1.

5.1.6 Transformação de *E. coli* BL21 e *E. coli* Lemo21 com pET23a-KTIMet

Com o plasmídeo pET23a-KTIMet purificado e confirmado, poderemos iniciar o processo de transformação de dois tipos de células com vistas à produção da proteína de interesse. Serão inicialmente utilizadas duas linhagens específicas de *E. coli*, considerando possíveis dificuldades de expressão de *ktimet*, como já referido no item 5.1.2.

O protocolo de preparação de células termocompetentes de *E. coli* deverá ser realizada conforme procedimento descrito por INOUE *et al.* (1990). A transformação de *E. coli* termocompetente deverá ser realizada conforme descrito em AUSUBEL *et al.* (1999).

Desta etapa resultarão duas placas (uma para cada linhagem) de onde serão selecionadas cinco colônias para obtenção de cinco “minipreparações” plasmidiais (AUSUBEL *et al.* (1999)). Após a confirmação da presença dos plasmídeos pET23a-KTIMet por PCR ou digestão com HindIII e BamHI seguido de eletroforese em gel de agarose a 0,8% (AUSUBEL *et al.* (1999)), seguir-se-á com pelo menos duas culturas diferentes uma de cada linhagem para produção da proteína de interesse

5.1.7 Cultivo de *E. coli* BL21:: e LEMO21::pET23a-KTIMet

As linhagens *E. coli* BL21::pET23a-KTIMet e LEMO21::pET23a-KTIMet selecionadas serão inoculadas em meio LB líquido acrescido de ampicilina a 100 mg/L (para BL-21) ou ampicilina e cloranfenicol (para Lemo21) e IPTG a 1 mM. As culturas serão conduzidas *o/n* a 20 °C. A adição de IPTG no meio de cultivo destina-se à inativação do gene repressor *lacUV5* ligado à região promotora do operador *lac* no cromossomo da célula hospedeira, o qual bloqueia a expressão do gene da T7 RNA polimerase. Ao desbloquear esta expressão, a célula hospedeira passará a produzir a T7 RNA polimerase que, por sua vez, passará a transcrever o gene de interesse presente no vetor de expressão pET23a-KTIMet. O mRNA será traduzido pela maquinaria de tradução da célula hospedeira (ribossomos).

As opções quanto aos procedimentos para extração e purificação da proteína a partir das culturas de células estarão condicionadas por outras escolhas realizadas em etapas anteriores. Por exemplo, ao optar por um plasmídeo de expressão que fusiona uma cauda de His à proteína de interesse, direcionaremos uma das etapas de purificação. No caso da extração e purificação da proteína KTIMet, deveremos avaliar,

inicialmente, o que se pretende fazer com o produto. Como já definido anteriormente, pretendemos realizar alguns ensaios com vistas a: (I) confirmar a massa molecular do peptídeo; (II) confirmar sua estrutura primária; (III) confirmar as estruturas secundária e terciária da molécula; e (IV) avaliar a atividade inibitória de protease em comparação com a proteína nativa.

Os objetivos I e II poderiam ser alcançados sem necessidade de grandes processamentos. Para atender os objetivos III e IV, no entanto, pelo menos parte da proteína produzida deverá ser extraída, purificada e processada para eliminar qualquer resíduo do processo de lise celular ou outras proteínas presentes no extrato de células. Após a separação, a proteína-alvo deverá ser desnaturada e renaturada caso seja obtida na forma insolúvel, condição mais provável segundo resultados obtidos por MAJOR e CONSTABEL (2008) para a família de inibidores de protease do tipo Kunitz em álamo (*Populus spp.*).

Analisando os resultados dos estudos estruturais *in silico* apresentados na dissertação de mestrado do autor (DELLA_GIUSTINA_NETO,2015), a remoção da cauda de His poderá, a princípio, ser dispensada para atender aos objetivos III e IV referidos acima. No entanto, somente uma nova análise, adicionando esta sequência de His na cadeia de aminoácidos de KTIMet, poderá antecipar possíveis efeitos estruturais e no sítio ativo.

Com base nos aspectos abordados nos parágrafos anteriores, apresentamos a seguir as etapas da metodologia proposta, desde a cultura de células contendo a proteína alvo até a obtenção da proteína purificada na qualidade e na quantidade necessárias aos ensaios de caracterização pretendidos.

5.2 EXTRAÇÃO DA PROTEÍNA

No caso da produção de KTIMet em *E. coli*, não é esperada a secreção desta proteína para o meio de cultivo. Portanto, toda a proteína produzida deverá estar no interior das células, provavelmente em corpos de inclusão insolúveis. No entanto, considerando as diferenças entre as moléculas de KTi e KTIMet, incluindo a expectativa de redução da atividade inibitória, não podemos descartar a possibilidade real de se obter esta última proteína na forma solúvel. Portanto, o procedimento proposto incluiu uma etapa para verificação da presença da proteína na fase solúvel durante o processo de extração. Neste sentido, será necessário estudar melhor a concentração e pH dos tampões utilizados nesta etapa.

A extração de proteínas totais inicia pela separação das células do meio de cultivo, conforme sequência de etapas listadas abaixo:

- 1º) Centrifugar as culturas em LB, descartando-se o sobrenadante e coletando o sedimento celular;
- 2º) Ressuspender o sedimento celular formado em tampão Tris.HCl 50 mM, pH 8;
- 3º) Lisar com lisozima à temperatura ambiente e após incubar em sonicador de banho;
- 4º) Centrifugar novamente a suspensão em tampão Tris.HCl 50 mM, pH 8;
- 5º) Coletar o sedimento celular e reservar o sobrenadante para verificação de uma possível fração solúvel da proteína-alvo;
- 6º) Ressuspender o sedimento celular formado na 3ª etapa em tampão Tris.HCl, pH8. Neste deverá estar presente a fração insolúvel da proteína.

Ao final do processo de extração, teremos amostras prontas para as análises subsequentes. Uma parte das amostras poderá ser utilizada diretamente para os ensaios desnaturantes como o SDS-PAGE corado com *Comassie Blue* e *Western-Blot*. Outra parte deverá ser reservada para purificação, solubilização, desnaturação e renaturação de modo a garantir que tenhamos a proteína com suas estruturas originais preservadas.

No caso de ocorrer a formação de corpos de inclusão, a separação destes dos demais resíduos das células deverá ser realizada por centrifugação leve. Os agregados proteicos são normalmente mais pesados que as demais partes das células de acordo com SINGH e PANDA (2005).

5.3 PURIFICAÇÃO DA PROTEÍNA KTIMET

Para atender aos objetivos III e IV da caracterização de KTIMet descritos no item 5.2 desta monografia, será necessário processar as amostras brutas de extratos celulares obtidas e purificar a proteína de interesse. No caso de KTIMet, a opção pela fusão com uma cauda de His, fornecida pelo plasmídeo de expressão pET23a, permitirá a utilização da técnica de purificação em coluna por afinidade ao níquel (Chelating Sepharose) (Ni) com a posterior retirada da cauda de His. Assim, poderemos obter uma amostra com pureza suficiente para os ensaios subsequentes de caracterização proteica.

O tratamento das amostras insolúveis requererá um processamento adicional com o objetivo de recuperar a proteína dos corpos de inclusão. Para esta recuperação, serão necessárias três etapas básicas: (I) solubilização, a qual envolverá a desnaturação das proteínas; (II) renaturação; e (III) purificação da proteína KTIMet solubilizada. Entre estas etapas, a solubilização e a renaturação irão requerer maiores cuidados para que, ao final, obtenha-se a proteína na forma ativa, como nos relata SINGH e PANDA (2005). Neste trabalho, os autores analisaram os protocolos tradicionais de recuperação de proteínas a partir de corpos de inclusão.

Considerando que na concepção de KTIMet buscou-se preservar ao máximo as características estruturais de KTi, propomos que seja testado o mesmo protocolo descrito no artigo de referência onde os autores extraíram e purificaram KTi expressa em *E. coli*. (MAJOR;CONSTABEL,2008).

Quanto à remoção da cauda de His, é preciso avaliar, *a priori*, sua interferência nas análises as quais se pretende submeter as amostras de proteína extraída. Para os testes desnaturantes (SDS-PAGE e *Western Blot*) não será necessária a remoção da cauda de His pois os resultados serão apenas qualitativos e quantitativos, sem possibilidade de qualquer avaliação quanto à atividade bioquímica. Na verdade, para o teste *Western Blot* a cauda de His poderá ser utilizada como ferramenta auxiliar na identificação da proteína produzida. Já para os testes destinados a avaliar as estruturas secundária e terciária da proteína, assim como a atividade enzimática, será necessário definir o grau de interferência da cauda de His no enovelamento da molécula e na sua atividade bioquímica, respectivamente.

Analisando-se a estrutura da molécula KTIMet, a extremidade C-terminal da proteína não possui qualquer tipo de interação com o sítio ativo e poderá não interferir no enovelamento da proteína. Portanto, é provável que não seja necessária a remoção da cauda fusionada. Caso estudos complementares orientem para remoção da cauda de His, existem dois métodos básicos e conhecidos na literatura para este procedimento: o enzimático e o químico (CARTER,1990). Os métodos enzimáticos são normalmente muito específicos e exigem condições muito restritivas para ter eficiência. Os métodos químicos, por outro lado, são muito eficientes, porém inespecíficos e muito agressivos, principalmente quando se deseja preservar a atividade enzimática da proteína (FORD *et al.*,1991). Portanto é preciso estudar previamente qual o método e o protocolo mais adequado para cada caso como nos relata UHLÉN *et al.* (1992) em seu artigo.

5.4 CARACTERIZAÇÃO DA PROTEÍNA KTIMET

Os ensaios de caracterização da proteína KTIMet deverão atender aos objetivos iniciais do estudo desta proteína. Como citado anteriormente, pretende-se analisar a manutenção das estruturas secundária e terciária, além de medir a atividade inibitória de proteases para se comparar com a da proteína nativa (KTI).

5.4.1 Análise das amostras de proteínas não purificadas

Iniciaremos a análise com duas amostras de extratos proteicos brutos para cada uma das linhagens de *E. coli* utilizadas na expressão heteróloga: (I) fração contendo as proteínas solúvel; e (II) fração contendo as proteínas insolúveis. Para a confirmação da presença de KTIMet e estimativa da quantidade produzida, será realizado o ensaio SDS-PAGE utilizando o corante *Coomassie Blue* (AUSUBEL *et al.*, 1999).

Com base nas imagens obtidas neste ensaio, processadas pelo *software Image-J* (SCHNEIDER *et al.*, 2012; MCKENZIE *et al.*, 2015; ANTIABONG *et al.*, 2016), teremos resultados que permitirão confirmar, ou não, a produção da proteína (qualitativo) e resultados aproximados quanto à quantidade produzida (quantitativo), pela comparação da intensidade da banda correspondente à proteína-alvo com a da banda do marcador de peso molecular cuja concentração é conhecida. Alternativamente poderá ser utilizada uma amostra de referência com uma proteína padrão cuja concentração seja conhecida.

Para confirmação dos resultados obtidos no SDS-PAGE, optamos pela realização de um *Western Blot* utilizando um protocolo adaptado a partir de HARLOW e LANE (1988) ou AUSUBEL *et al.* (1999), utilizando anticorpo específicos para cauda de His e análise da intensidade da imagem pelo *software Image-J* (SCHNEIDER *et al.*, 2012; MCKENZIE *et al.*, 2015; ANTIABONG *et al.*, 2016). A utilização de anticorpos policlonais específicos para KTI ou KTIMet é outra possibilidade promissora a ser avaliada no futuro, pois poderia fornecer informações importantes sobre a similaridade entre as estruturas primárias das duas proteínas. A preservação desta semelhança estrutural entre as duas proteínas foi um dos objetivos durante a concepção de KTIMet, visando estabelecer uma competição entre a expressão dos genes nativo e recombinante e não uma eliminação total da proteína original, já que KTI pode apresentar valor nutracêutico contra tumores, em concentrações reduzidas, como

referido nos trabalhos de GUILLAMÓN *et al.* (2008) e CRANCIANINOV *et al.* (2005) que relatam a ação de outros inibidores de protease derivados de soja com possível inibição de genes de câncer de mama, pele, bexiga, cólon, pulmão, pâncreas, boca e esôfago. Estudos mais recentes com outros inibidores de proteases tem demonstrado efeitos nutracêuticos destas moléculas, sugerindo a capacidade de induzir apoptose em linhagens de células tumorais, como relatado no trabalho publicado por BAI *et al.* (2015) que realizou análises *in vitro* e *in vivo* referentes aos inibidores de protease presentes no trigo sarraceno.

5.4.2 Análise das amostras de proteínas purificadas

As amostras purificadas serão utilizadas (I) na determinação das estruturas secundária e terciária da proteína-alvo (KTIMet) e (II) na determinação da atividade inibitória comparada à proteína nativa (KTi, não recombinante). Consideramos o uso do dicróismo circular (KELLY;PRICE,2000; KELLY *et al.*,2005) para a determinação das estruturas secundárias presentes na molécula. Esta técnica permitirá estabelecer a presença de hélices alfa, folhas beta (β , do inglês, *beta-sheets*) e regiões desestruturadas em proteínas em seu estado solúvel e ativo. A molécula de KTi, após, enovelamento possui um total de 12 folhas β , as quais são responsáveis pela sua estrutura terciária em forma de cesto, como podemos observar no artigo de DE MEESTER *et al.* (1998).

Na Figura 6 abaixo, retirada da dissertação de mestrado do autor (DELLA_GIUSTINA_NETO, 2015), está ilustrada uma possível alteração no sítio ativo que justifica a expectativa para redução da atividade inibitória que queremos confirmar. Podemos observar que as cadeias laterais dos resíduos da Arg-63, Arg-65, Ile-64/Met-64 próximas da alça responsável pela atividade inibitória estão alteradas em KTIMet quando comparadas com KTi.

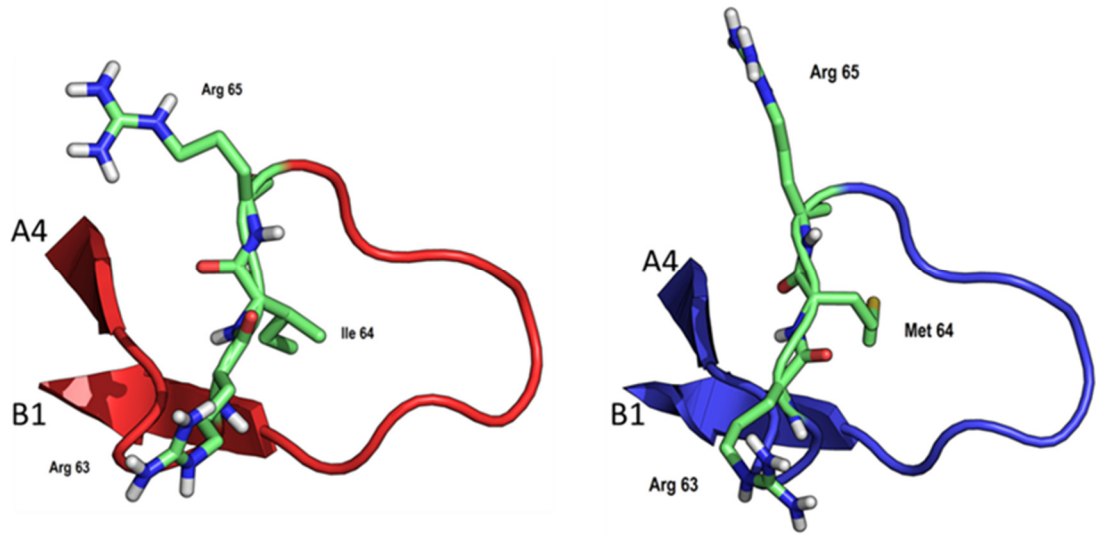


Figura 6 – Sítio Ativo de KTi e domínio correspondente de KTIMet.

Nesta Figura podemos observar os três resíduos Arg-63, Ile-64 e Arg-65 posicionados no sítio ativo do inibidor KTi. Estes resíduos sofrem alterações de posição quando a Ile-64 é substituída pela Met-64 em KTIMet. As imagens foram obtidas dos resultados de uma análise *in silico* que faz parte da dissertação de mestrado do autor (DELLA_GIUSTINA_NETO,2015) e nos estudos estruturais e interações dos inibidores de tripsina do tipo Kunitz elaborados por DE MEESTER *et al.* (1998).

Para a determinação experimental da atividade inibitória de proteases de KTIMet e KTi, poderemos utilizar as técnicas do BAPNA e da ITC. Ambas as técnicas poderão ser empregadas para validar, ou não, a hipótese levantada durante o mestrado do autor, corroborada pela análise *in silico* realizada àquela época, de que a proteína recombinante KTIMet, apesar de se manter estruturalmente semelhante à proteína nativa KTi, sofre uma mudança significativa no sítio ativo e, como consequência, perda da capacidade de inibir proteases.

No ensaio com BAPNA, utilizando protocolo adaptado com base em HAMERSTRAND *et al.* (1981), teremos como controle negativo da atividade inibitória BAPNA + tripsina. Nos testes comparativos, teremos BAPNA + Tripsina + KTIMet. Como controle positivo (atividade inibitória máxima), teremos BAPNA + tripsina + KTi.

A técnica de ITC deverá ser conduzida como descrito em VELÁZQUEZ-CAMPOY *et al.* (2004) e poderá ser utilizada para confirmar os resultados do ensaio anterior e consistirá na medição do calor liberado na reação do inibidor de proteases com a enzima tripsina, que deverá ser proporcional à intensidade desta reação. Deste modo, se a afinidade de KTIMet por tripsina for menor do que a do inibidor nativo KTi, teremos menos calor liberado e mais um resultado em favor da hipótese referida anteriormente, isto é, de que as alterações estruturais promovidas pela substituição

de códon na sequência nucleotídica resultaram em uma proteína com atividade de ligação – e, por conseguinte, inibição – de proteases reduzida ou ausente.

Todas as análises previstas para determinação da alteração da atividade inibitória de KTIMet deverão utilizar a proteína nativa Inibidor de Tripsina tipo Kunitz (ITK) de Soja” (grau P.A.), o substrato BAPNA (grau P.A.) e a enzima tripsina (grau P.A.) adquiridos da Sigma Aldrich. O ITK de soja será a referência para 100% da atividade inibitória na análise comparativa com KTIMet..

Dos resultados desses ensaios, poderemos inferir sobre uma possível redução da atividade inibitória, com evidente ganho nutricional e redução no custo de processamento dos derivados de soja para a indústria. Esta redução dar-se-ia pela redução no consumo de vapor, normalmente utilizado pela indústria para aquecer os produtos derivados de soja (farelo) visando inativar este tipo de fator antinutricional (SILVA;DA_SILVA,2000; CRANCIANINOV *et al.*,2005; CARNEIRO *et al.*,2010). Indiretamente, poderemos esperar um aumento do conteúdo de proteína total do material tratado, considerando que os processos de aquecimento geralmente degradam parte da proteína do produto (DELLA_GIUSTINA_NETO,2015).

6 CONCLUSÃO

Com o desenvolvimento futuro do trabalho aqui proposto, pretendemos trazer informações suficientes para fundamentar a metodologia específica de produção em escala laboratorial da proteína recombinante KTIMet. Não é intenção ou pretensão esgotar os temas da expressão heteróloga de genes e a produção das correspondentes proteína recombinantes, dos métodos de extração e purificação proteica, assim como da caracterização bioquímica destas moléculas de forma geral. Entendemos que o conhecimento prévio dos sistemas existentes de expressão heteróloga e a aplicação dos critérios de seleção abordados nesta monografia são determinantes para o sucesso de quem pretende produzir uma proteína recombinante, em um tipo de célula ou tecido diferente daquele onde ela é produzida naturalmente. Sopesar as vantagens e desvantagens durante a aplicação destes critérios é a chave para o sucesso de qualquer trabalho nesta área como, aliás, o é em qualquer outra área.

Outro ponto importante a ser considerado é o entendimento prévio sobre a proteína a ser produzida, o que envolve, *a priori*, a obtenção de informações sobre o gene, as estruturas da molécula proteica, o tamanho, as propriedades bioquímicas, etc. Conhecida a estrutura primária do peptídeo, a análise *in silico* prévia das estruturas secundária, terciária, quaternária e do comportamento da molécula em diferentes meios poderão auxiliar sobremaneira as decisões a serem tomadas durante o processo de purificação. Nem sempre isto é possível, considerando que o objeto de uma pesquisa poderá ser isolar uma determinada proteína, ou conjunto de proteínas desconhecidas, justamente para permitir seu estudo mais detalhado e determinar suas propriedades ou, até mesmo, descobrir sua origem genética.

Quanto ao plano de trabalho proposto para a expressão heteróloga do gene *ktimet* e produção da proteína KTIMet, esperamos que seja testado e validado no futuro ou adaptado e aperfeiçoado por pesquisadores interessados no estudo dos ganhos nutricionais da soja. Sempre haverá mais de um caminho para expressar transgenes, assim como diferentes pontos de partida e formas de produção de proteínas heterólogas. Portanto a proposição resumida no fluxograma de atividades apresentado no **apêndice A** desta monografia é apenas uma entre as muitas alternativas possíveis, mas que resume, de forma gráfica, as etapas e objetivos pretendidos neste trabalho.

A área de desenvolvimento de novas variedades de sementes com ganhos nutricionais e nutracêuticos encontra-se, ainda, na fronteira tecnológica das pesquisas com a soja e deverá ganhar força à medida que os objetivos de tolerância à seca, a herbicidas e resistência a doenças e pragas, que hoje capturam toda a atenção, estejam plenamente atendidos e não gerem mais ganhos de participação no mercado de sementes (MARINHO *et al.*,2014). Será então necessário apresentar uma novidade para aumentar a participação no mercado de sementes de soja. Esta novidade será o ganho nutricional e nutracêutico, além da redução nos custos de processamento industrial dos derivados da soja (DELLA_GIUSTINA_NETO,2015; PASQUALI;DELLA_GIUSTINA_NETO,2019).

REFERÊNCIAS

- AHMAD, I.; NAWAZ, N. *et al.* Overcoming Challenges for Amplified Expression of Recombinant Proteins Using Escherichia Coli. **Protein Expression and Purification**, 144, p. 12-18, 2018/04/01/ 2018.
- ANTIABONG, J. F.; NGOEPE, M. G.; ABECHI, A. S. Semi-Quantitative Digital Analysis of Polymerase Chain Reaction-Electrophoresis Gel: Potential Applications in Low-Income Veterinary Laboratories. **Veterinary world**, 9, n. 9, p. 935, 2016.
- AUSUBEL, F. M.; BRENT, R. *et al.* **Short Protocols in Molecular Biology**. 4^a ed. New York: 1999.
- AWAZUHARA, M.; KIM, H. *et al.* Composition of Seed Storage Proteins Changed by Glutathione Treatment of Soybeans. **Biosci Biotechnol Biochem**, 66, n. 8, p. 1751-1754, Aug 2002.
- BAI, C. Z.; FENG, M. L. *et al.* Anti-Tumoral Effects of a Trypsin Inhibitor Derived from Buckwheat in Vitro and in Vivo. **Molecular Medicine Reports**, 12, n. 2, p. 1777-1782, 2015.
- BASTOLLA, U.; DEMETRIUS, L. Stability Constraints and Protein Evolution: The Role of Chain Length, Composition and Disulfide Bonds. **Protein Engineering Design and Selection**, 18, n. 9, p. 405-415, September 1, 2005 2005.
- BILLETER, M.; WAGNER, G.; WÜTHRICH, K. Solution Nmr Structure Determination of Proteins Revisited. **Journal of biomolecular NMR**, 42, p. 155-158, 2008.
- BLOUNT, Z. D. The Unexhausted Potential of E. Coli. **Elife**, 4, Mar 25 2015.
- BONK, T.; HUMENY, A. Maldi-Tof-Ms Analysis of Protein and DNA. **The Neuroscientist**, 7, n. 1, p. 6-12, 2001.
- BULHELLER, B. M.; RODGER, A.; HIRST, J. D. Circular and Linear Dichroism of Proteins. **Physical Chemistry Chemical Physics**, 9, n. 17, p. 2020-2035, 2007. 10.1039/B615870F.
- CAI, D.; RAO, Y. *et al.* Engineering Bacillus for Efficient Production of Heterologous Protein: Current Progress, Challenge and Prospect. **Journal of Applied Microbiology**, 126, n. 6, p. 1632-1642, 2019.
- CARDAMONE, J. M. Investigating the Microstructure of Keratin Extracted from Wool: Peptide Sequence (Maldi-Tof/Tof) and Protein Conformation (Ftir). **Journal of Molecular Structure**, 969, n. 1, p. 97-105, 2010/04/22/ 2010.
- CARNEIRO, D. J.; STECH, M. R.; CARVALHO, M. R. B. d. Fatores Antinutricionais E Coeficientes De Digestibilidade Aparente Da Proteína De Produtos De Soja Para O Pacu (Piaractus Mesopotamicus). **Acta Scientiarum. Animal Sciences**, 32, n. 3, 2010.

CARTER, P. Site-Specific Proteolysis of Fusion Proteins. *In: Protein Purification: American Chemical Society*, 1990. v. 427, cap. 13, p. 181-193. (Acs Symposium Series).

CHEN, T.; ZHOU, Y.; LU, Y.; ZHANG, H. Advances in Heterologous Biosynthesis of Plant and Fungal Natural Products by Modular Co-Culture Engineering. **Biotechnology Letters**, 41, n. 1, p. 27-34, 2019/01/01 2019.

CHIAIESE, P.; OHKAMA-OHTSU, N. *et al.* Sulphur and Nitrogen Nutrition Influence the Response of Chickpea Seeds to an Added, Transgenic Sink for Organic Sulphur. **J Exp Bot**, 55, n. 404, p. 1889-1901, Aug 2004.

CHOI, G.-S.; CHOO, H. J.; KIM, B.-G.; AHN, J.-H. Synthesis of Acridone Derivatives Via Heterologous Expression of a Plant Type Iii Polyketide Synthase in Escherichia Coli. **Microbial Cell Factories**, 19, n. 1, p. 73, 2020/03/20 2020.

CONTRERAS-GOMEZ, A.; SANCHEZ-MIRON, A. *et al.* Protein Production Using the Baculovirus-Insect Cell Expression System. **Biotechnol Prog**, 30, n. 1, p. 1-18, Jan-Feb 2014.

CRANCIANINOV, W.; SANTANA, A. *et al.*, 2005, **Determinação Do Inibidor De Tripsina De Kunitz Em Soja Crua E Tratada De Cultivares Brasileiras**. In: JORNADA ACADÊMICA DA EMBRAPA SOJA, 2005, Londrina. Resumos expandidos

CROSSLEY, B. M.; BAI, J. *et al.* Guidelines for Sanger Sequencing and Molecular Assay Monitoring. **Journal of Veterinary Diagnostic Investigation**, 32, n. 6, p. 767-775, 2020.

DE BOER, H. A.; COMSTOCK, L. J.; VASSER, M. The Tac Promoter: A Functional Hybrid Derived from the Trp and Lac Promoters. **Proceedings of the National Academy of Sciences**, 80, n. 1, p. 21-25, 1983.

DE MEESTER, P.; BRICK, P. *et al.* Structure of the Kunitz-Type Soybean Trypsin Inhibitor (Sti): Implication for the Interactions between Members of the Sti Family and Tissue-Plasminogen Activator. **Acta Crystallogr D Biol Crystallogr**, 54, n. Pt 4, p. 589-597, Jul 1 1998.

DELLA_GIUSTINA_NETO, J. A. **Proteínas De Soja (Glycine Max) Modificadas Visando Aumentar O Teor De Metionina Em Sementes**. Orientador: PASQUALI, G. 2015. 107 f. (Mestrado) - CBIOT-Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre.

FORD, C. F.; SUOMINEN, I.; GLATZ, C. E. Fusion Tails for the Recovery and Purification of Recombinant Proteins. **Protein expression and purification**, 2, n. 2-3, p. 95-107, 1991.

FUJIWARA, T.; HIRAI, M. Y. *et al.* Effects of Sulfur Nutrition on Expression of the Soybean Seed Storage Protein Genes in Transgenic Petunia. **Plant Physiol**, 99, n. 1, p. 263-268, May 1992.

GASSNER, N. C.; BAASE, W. A. *et al.* Multiple Methionine Substitutions Are Tolerated in T4 Lysozyme and Have Coupled Effects on Folding and Stability. **Biophysical Chemistry**, 100, p. 325-340, 2003.

GE_HEALTHCARE. Gst Gene Fusion System - Handbook. UK 2013.

GOMES, L. S.; SENNA, R. *et al.* Four Conventional Soybean [Glycine Max (L.) Merrill] Seeds Exhibit Different Protein Profiles as Revealed by Proteomic Analysis. **J Agric Food Chem**, 62, n. 6, p. 1283-1293, Feb 12 2014.

GUILLAMÓN, E.; PEDROSA, M. M. *et al.* The Trypsin Inhibitors Present in Seed of Different Grain Legume Species and Cultivar. **Food Chemistry**, 107, n. 1, p. 68-74, 3/1/ 2008.

GUINIER, A. **X-Ray Crystallographic Technology**. Hilger and Watts, 1952.

HAMERSTRAND, G.; BLACK, L.; GLOVER, J. Trypsin Inhibitors in Soy Products: Modification of the Standard Analytical Procedure. **Cereal chemistry**, 1981.

HANNIG, G.; MAKRIDES, S. C. Strategies for Optimizing Heterologous Protein Expression in Escherichia Coli. **Trends in Biotechnology**, 16, n. 2, p. 54-60, 1998/02/01/ 1998.

HARLOW, E.; LANE, D. **Antibodies: A Laboratory Manual** Cold Spring Harbor Laboratory Press, Plainview. 1988.

HARPER, S.; SPEICHER, D. W. Expression and Purification of Gst Fusion Proteins. **Current Protocols in Protein Science**, 52, n. 1, p. 6.6.1-6.6.26, 2008.

HERSHBERG, R.; PETROV, D. A. Selection on Codon Bias. **Annual Review of Genetics**, 42, n. 1, p. 287-299, 2008.

HOSSAIN, Z.; KOMATSU, S. Soybean Proteomics. *In*: **Methods Mol Biol**, 2014. v. 1072, p. 315-331.

HUNT, I. From Gene to Protein: A Review of New and Enabling Technologies for Multi-Parallel Protein Expression. *In*: **Protein Expr Purif**. United States, 2005. v. 40, p. 1-22.

INOUE, H.; NOJIMA, H.; OKAYAMA, H. High Efficiency Transformation of Escherichia Coli with Plasmids. **Gene**, 96, n. 1, p. 23-28, // 1990.

JOHNSON, W. C. Analyzing Protein Circular Dichroism Spectra for Accurate Secondary Structures. **Proteins: Structure, Function, and Bioinformatics**, 35, n. 3, p. 307-312, 1999.

JOSEPH, B. C.; PICHAIMUTHU, S. *et al.* An Overview of the Parameters for Recombinant Protein Expression in Escherichia Coli. **J Cell Sci Ther**, 6, n. 5, p. 221, 2015.

KAFKA, A. P.; KLEFFMANN, T.; RADES, T.; MCDOWELL, A. The Application of Maldi ToF Ms in Biopharmaceutical Research. **International Journal of Pharmaceutics**, 417, n. 1, p. 70-82, 2011/09/30/ 2011.

KAKADE, M.; RACKIS, J.; MCGHEE, J.; PUSKI, G. Determination of Trypsin Inhibitor Activity of Soy Products: A Collaborative Analysis of an Improved Procedure. 1974.

KARUNANITHY, G.; SHUKLA, V. K.; HANSEN, D. F. Methodological Advancements for Characterising Protein Side Chains by Nmr Spectroscopy. **Current Opinion in Structural Biology**, 70, p. 61-69, 2021/10/01/ 2021.

KAUR, J.; KUMAR, A.; KAUR, J. Strategies for Optimization of Heterologous Protein Expression in E. Coli: Roadblocks and Reinforcements. **International Journal of Biological Macromolecules**, 106, p. 803-822, 2018/01/01/ 2018.

KELLY, S. M.; JESS, T. J.; PRICE, N. C. How to Study Proteins by Circular Dichroism. **Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics**, 1751, n. 2, p. 119-139, 2005/08/10/ 2005.

KELLY, S. M.; PRICE, N. C. The Use of Circular Dichroism in the Investigation of Protein Structure and Function. **Current Protein and Peptide Science**, 1, n. 4, p. 349-384, // 2000.

KHOW, O.; SUNTRARACHUN, S. Strategies for Production of Active Eukaryotic Proteins in Bacterial Expression System. **Asian Pacific Journal of Tropical Biomedicine**, 2, n. 2, p. 159-162, 2012.

KIM, R.; SANDLER, S. J. *et al.* Overexpression of Archaeal Proteins in Escherichia Coli. **Biotechnology Letters**, 20, n. 3, p. 207-210, 1998/03/01 1998.

KISLINGER, T.; HUMENY, A. *et al.* Analysis of Protein Glycation Products by Maldi-ToF/Ms. **Annals of the New York Academy of Sciences**, 1043, n. 1, p. 249-259, 2005.

KOIWA, H.; BRESSAN, R. A.; HASEGAWA, P. M. Regulation of Protease Inhibitors and Plant Defense. **Trends in Plant Science**, 2, n. 10, p. 379-384, 1997.

KRISHNA, R. G.; WOLD, F. Post-Translational Modifications of Proteins. *In*: IMAHORI, K. e SAKIYAMA, F. (Ed.). **Methods in Protein Sequence Analysis**. Boston, MA: Springer US, 1993. p. 167-172.

KUNITZ, M. Crystalline Soybean Trypsin Inhibitor 1. Method of Isolation. **The Journal of General Physiology**, p. 149-154, 1946.

KUNITZ, M. Crystalline Soybean Trypsin Inhibitor li. General Properties. **The Journal of General Physiology**, p. 291-310, 1947.

KURLAND, C. G. Codon Bias and Gene Expression. **FEBS Letters**, 285, n. 2, p. 165-169, 1991/07/01/ 1991.

LIN, J.; FIDO, R. *et al.* The Expression and Processing of Two Recombinant 2s Albumins from Soybean (*Glycine Max*) in the Yeast *Pichia Pastoris*. **Biochim Biophys Acta**, 1698, n. 2, p. 203-212, May 6 2004.

LOUIS-JEUNE, C.; ANDRADE-NAVARRO, M. A.; PEREZ-IRATXETA, C. Prediction of Protein Secondary Structure from Circular Dichroism Using Theoretically Derived Spectra. **Proteins: Structure, Function, and Bioinformatics**, 80, n. 2, p. 374-381, 2012.

MAJOR, I. T.; CONSTABEL, C. P. Functional Analysis of the Kunitz Trypsin Inhibitor Family in Poplar Reveals Biochemical Diversity and Multiplicity in Defense against Herbivores. *In: Plant Physiol.* United States, 2008. v. 146, p. 888-903.

MARINHO, C. D.; MARTINS, F. J. *et al.* Genetically Modified Crops: Brazilian Law and Overview. **Genet Mol Res**, 13, n. 3, p. 5221-5240, 2014.

MARKWICK, P. R.; MALLIAVIN, T.; NILGES, M. Structural Biology by Nmr: Structure, Dynamics, and Interactions. **PLoS computational biology**, 4, n. 9, p. e1000168, 2008.

MCKENZIE, B.; KAY, G. *et al.* The Hen's Egg Chorioallantoic Membrane (Het-Cam) Test to Predict the Ophthalmic Irritation Potential of a Cysteamine-Containing Gel: Quantification Using Photoshop® and Imagej. **International journal of pharmaceuticals**, 490, n. 1-2, p. 1-8, 2015.

MCMILLIN, C. R.; WALTON, A. G. A Circular Dichroism Technique for the Study of Adsorbed Protein Structure. **Journal of Colloid and Interface Science**, 48, n. 2, p. 345-349, 1974/08/01/ 1974.

MONTELIONE, G. T.; ZHENG, D. *et al.* Protein Nmr Spectroscopy in Structural Genomics. **nature structural biology**, 7, n. 11, p. 982-985, 2000.

NOVAGEN. Novagen Pet System Manual. User Protocol TB055 Rev. C 0611JN. 2011.

OHKAMA, N.; GOTO, D. B.; FUJIWARA, T.; NAITO, S. Differential Tissue-Specific Response to Sulfate and Methionine of a Soybean Seed Storage Protein Promoter Region in Transgenic Arabidopsis. **Plant Cell Physiol**, 43, n. 11, p. 1266-1275, Nov 2002.

PAPANEOPHYTOU, C. P.; KONTOPIDIS, G. A. Optimization of Tnf-A Overexpression in *Escherichia Coli* Using Response Surface Methodology: Purification of the Protein and Oligomerization Studies. **Protein Expression and Purification**, 86, n. 1, p. 35-44, 2012/11/01/ 2012.

PASQUALI, G.; DELLA_GIUSTINA_NETO, J. A. Sequência De DNA, Construção Gênica, Sequência De Aminoácidos E Método De Aumento De Valor Nutricional De Planta. INPI. Brasil. BR-102018008975-7 A2: 12 p. 2019.

PETI, W.; PAGE, R. Strategies to Maximize Heterologous Protein Expression in Escherichia Coli with Minimal Cost. *In: Protein Expr Purif.* United States, 2007. v. 51, p. 1-10.

PREIBISCH, G.; ISHIHARA, H.; TRIPIER, D.; LEINWEBER, M. Unexpected Translation Initiation within the Coding Region of Eukaryotic Genes Expressed in Escherichia Coli. **Gene**, 72, n. 1, p. 179-186, 1988/12/10/ 1988.

QUAX, Tessa E. F.; CLAASSENS, Nico J.; SÖLL, D.; VAN DER OOST, J. Codon Bias as a Means to Fine-Tune Gene Expression. **Molecular Cell**, 59, n. 2, p. 149-161, 2015/07/16/ 2015.

RAMAZI, S.; ZAHIRI, J. Post-Translational Modifications in Proteins: Resources, Tools and Prediction Methods. **Database**, 2021, 2021.

ROSANO, G. L.; CECCARELLI, E. A. Recombinant Protein Expression in Escherichia Coli: Advances and Challenges. **Front Microbiol**, 5, p. 172, 2014.

SANGER, F.; COULSON, A. R. A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase. **Journal of Molecular Biology**, 94, n. 3, p. 441-448, 1975/05/25/ 1975.

SCHLEGEL, S.; LOFBLUM, J. *et al.* Optimizing Membrane Protein Overexpression in the Escherichia Coli Strain Lemo21(De3). **J Mol Biol**, 423, n. 4, p. 648-659, Nov 2 2012.

SCHNEIDER, C. A.; RASBAND, W. S.; ELICEIRI, K. W. Nih Image to Imagej: 25 Years of Image Analysis. **Nature Methods**, 9, n. 7, p. 671-675, 2012/07/01 2012.

SEVERIN, A. J.; WOODY, J. L. *et al.* Rna-Seq Atlas of Glycine Max: A Guide to the Soybean Transcriptome. **BMC Plant Biol**, 10, p. 160, 2010.

SHIH, M. D.; LIN, S. C. *et al.* Gene Cloning and Characterization of a Soybean (Glycine Max L.) Lea Protein, Gmpm16. **Plant Mol Biol**, 56, n. 5, p. 689-703, Nov 2004.

SILVA, M. R.; DA_SILVA, M. A. A. P. Fatores Antinutricionais: Inibidores De Proteases E Lectinas. **Revista de Nutrição**, 13, n. 1, p. 3-9, 2000.

SINGH, S. M.; PANDA, A. K. Solubilization and Refolding of Bacterial Inclusion Body Proteins. *In: J Biosci Bioeng.* Japan, 2005. v. 99, p. 303-310.

SMYTH, M.; MARTIN, J. X Ray Crystallography. **Molecular Pathology**, 53, n. 1, p. 8, 2000.

SØRENSEN, H. P.; SPERLING-PETERSEN, H. U.; MORTENSEN, K. K. Production of Recombinant Thermostable Proteins Expressed in Escherichia Coli: Completion of

Protein Synthesis Is the Bottleneck. **Journal of Chromatography B**, 786, n. 1, p. 207-214, 2003/03/25/ 2003.

STAUNTON, D.; OWEN, J.; CAMPBELL, I. D. Nmr and Structural Genomics. **Accounts of chemical research**, 36, n. 3, p. 207-214, 2003.

UHLÉN, M.; FORSBERG, G. *et al.* Fusion Proteins in Biotechnology. **Current opinion in biotechnology**, 3, n. 4, p. 363-369, 1992.

UTSUMI, S.; MARUYAMA, N.; SATOH, R.; ADACHI, M. Structure-Function Relationships of Soybean Proteins Revealed by Using Recombinant Systems. **Enzyme and Microbial Technology**, 30, n. 3, p. 284-288, 3/13/ 2002.

VELÁZQUEZ-CAMPOY, A.; OHTAKA, H. *et al.* Isothermal Titration Calorimetry. **Current Protocols in Cell Biology**, 23, n. 1, p. 17.18.11-17.18.24, 2004.

WAGNER, G. A Topical Issue: Production and Labeling of Biological Macromolecules for Nmr Investigations. Springer. 46: 1-2 p. 2010.

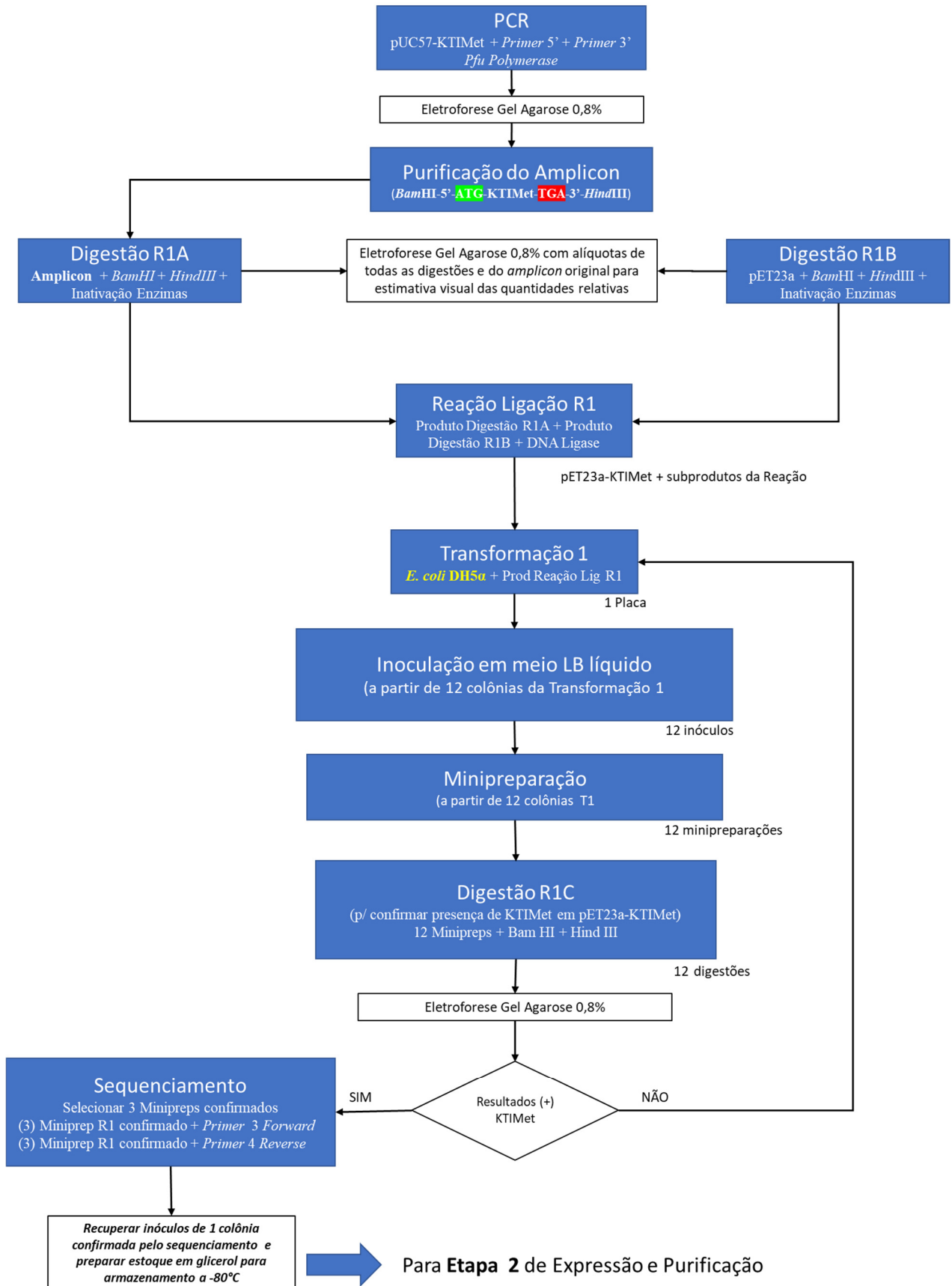
WISEMAN, T.; WILLISTON, S.; BRANDTS, J. F.; LIN, L.-N. Rapid Measurement of Binding Constants and Heats of Binding Using a New Titration Calorimeter. **Analytical Biochemistry**, 179, n. 1, p. 131-137, 1989/05/15/ 1989.

YIN, J.; LI, G.; REN, X.; HERRLER, G. Select What You Need: A Comparative Evaluation of the Advantages and Limitations of Frequently Used Expression Systems for Foreign Genes. **Journal of Biotechnology**, 127, n. 3, p. 335-347, 2007/01/10/ 2007.

YOUNG, C. L.; BRITTON, Z. T.; ROBINSON, A. S. Recombinant Protein Expression and Purification: A Comprehensive Review of Affinity Tags and Microbial Applications. **Biotechnol J**, 7, n. 5, p. 620-634, May 2012.

ZUIDERWEG, E. R. Mapping Protein– Protein Interactions in Solution by Nmr Spectroscopy. **Biochemistry**, 41, n. 1, p. 1-7, 2002.

APÊNDICE A — PLANO DE TRABALHO PROPOSTO



Etapa 2

