



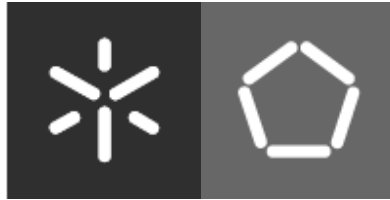
Universidade do Minho

Escola de Engenharia

Simão Daniel Ferreira Lima

**Produção de proteínas recombinantes através de
um novo sistema de fusão em *Escherichia coli***

Outubro 2013



Universidade do Minho

Departamento de Engenharia Biológica

Simão Daniel Ferreira Lima

**Produção de proteínas recombinantes através de
um novo sistema de fusão em *Escherichia coli***

Dissertação de Mestrado

Mestrado Integrado em Engenharia Biológica

Trabalho efectuado sob a orientação

Professora Doutora Lucília Domingues

Doutor André Almeida

Outubro 2013

Agradecimentos

A realização deste trabalho na empresa Hitag Biotechnology contou com o apoio e ajuda de um vasto grupo de pessoas, às quais gostaria de endereçar um sentido agradecimento:

À Doutora Lucília, por me ter despertado o gosto nesta temática e proporcionar a oportunidade de realizar esta dissertação.

Ao Doutor André Almeida, pela oportunidade de realizar a dissertação na sua empresa e por providenciar todos os meios necessários à execução deste trabalho, pela transmissão de conhecimentos, pela disponibilidade, pelo optimismo, pela motivação e, principalmente, pelo entusiasmo ao longo do trabalho.

Ao Hugo Azevedo, por todo o conhecimento que adquiri com ele, pela disponibilidade e paciência na execução das tarefas.

À recente Doutora Sofia Costa, pela constante disponibilidade, dedicação e interesse, pela partilha de conhecimentos e pela boa disposição constante.

A todas as pessoas que tive o prazer de conhecer na instituição do Biocant Park e que proporcionaram momentos de lazer e boa disposição entre tarefas.

À minha colega Daniela pelo carinho, boa disposição e companhia e pelas inúmeras horas de partilha e discussão de ideias e de conhecimentos.

Aos meus amigos Noémia e Rui pela presença e pelas doses de bom humor e de descontração.

Desejava também deixar um agradecimento profundo aos meus pais e ao meu irmão, pelo amor, carinho, motivação e paciência e por acreditarem em mim sempre, permitindo que tivesse todas as condições necessárias, a todos os níveis.

Por fim e de forma igualmente especial, gostaria de agradecer à Cláudia não só pela presença constante, quer nos bons quer nos maus momentos, o carinho e motivação, mas também pelo enorme contributo durante este período.

Sumário

As proteínas desempenham um papel fundamental nos sistemas biológicos de todos os seres vivos, exercendo inúmeras funções que permitem o crescimento e desenvolvimento dos indivíduos. O desenvolvimento da biotecnologia nas últimas décadas expandiu o interesse da indústria para a produção de proteínas em áreas tão diversas como a farmacêutica, no diagnóstico e tratamento de doenças genéticas, a agricultura ou o ambiente.

Numa indústria, cada vez mais rentável, procuram-se novos métodos que permitam maiores eficiências através da otimização de processos e redução de custos. Por outro lado, a produção e purificação de proteínas biologicamente activas, na sua conformação correcta e no estado solúvel são objectivos essenciais para uma produção eficaz que vise a implementação comercial destes compostos.

Estudos anteriores demonstraram a capacidade da proteína Fh8 (8 kDa), uma proteína segregada pelo parasita *Fasciola hepatica* nas fases iniciais da infecção, como *tag* de expressão e de solubilidade em *Escherichia coli*. Assim se formou o novo sistema de fusão utilizado neste estudo (Hitag® - *tag* Fh8) que, pela conjugação do seu pequeno peso molecular com a capacidade de aumentar a expressão e melhorar a solubilidade proteica, se torna numa óptima opção para a produção de proteínas recombinantes se comparado com os outros *fusion tags* existentes actualmente.

A proteína estudada neste sistema foi a hidrolase do molinato (MolA) que é codificada por um gene existente na bactéria *Gulosibacter molinativorax* ON4[†]. Este organismo pertence a um conjunto de cinco membros capazes de mineralizar o molinato, um herbicida sistémico que é amplamente aplicado nos campos de arroz para a eliminação de ervas daninhas e que provoca a contaminação dos aquíferos em todo o mundo.

O presente trabalho teve por objectivo a produção da proteína recombinante MolA em *E. coli*, através do uso do *tag* de fusão Fh8, a avaliação e optimização das suas propriedades a nível da expressão e da solubilidade, bem como a sua purificação.

No estudo aqui apresentado, foi possível clonar e expressar a proteína recombinante (Fh8MolA) com sucesso. A optimização das condições de expressão permitiu concluir que, de modo a reduzir custos de operação e obter melhores níveis de expressão e de solubilidade, a proteína deve ser induzida *overnight* à temperatura de 18°C, no momento em que a densidade óptica (OD₆₀₀) da cultura se situe entre os 0,6 a 0,8 e com uma concentração de IPTG de 0,5mM.

A mini-purificação realizada por cromatografia de afinidade por IMAC permitiu constatar que se conseguiu recuperar a proteína recombinante com maior grau de pureza.

Abstract

Proteins play a fundamental role in biological systems of all living beings, exercising many functions that allow the growth and development of individuals. The development of biotechnology in last decades expanded the interest of industry for the production of proteins in such diverse areas as pharmaceuticals, diagnosis and treatment of genetic diseases, agriculture or the environment.

In an industry, increasingly profitable, it seeks new methods allowing greater efficiency through process optimization and costs reduction. On the other hand, the production and purification of biologically active proteins, in their correct conformation and soluble state, are essential objectives for an efficient production aimed at commercial implementation of these compounds.

Previous studies showed the ability of Fh8 protein (8 kDa), a protein secreted by *Fasciola hepatica* parasite in the early stages of infection, such as expression and solubility fusion tag in *Escherichia coli*. Therefore, the new fusion system used in this study (Hitag® - Fh8 tag) was formed. The combination of their small molecular weight with the ability to increase expression and improve the protein solubility became it a great choice for production of recombinant proteins compared with other currently available fusion tags.

The protein studied in this system was molinate hydrolase (MolA) which is encoded by a gene of the bacteria *Gulosibacter molinivorax* ON4^r. This organism belongs to a five-membered mixed culture able to mineralize molinato, a systemic herbicide which is widely used in rice fields for the elimination of weeds and which causes contamination of aquifers throughout the world.

The present work aimed the production of recombinant protein MolA in *E. coli*, through the use of Fh8 fusion tag, the evaluation and optimization of its expression and solubility properties, as well as its purification.

In the study presented here, it was possible to clone and express the recombinant protein (Fh8MolA) successfully. The optimization of the expression conditions allowed to concluded that, in order to reduce operation costs and to obtain better expression and solubility yields, protein should be induced overnight at 18 °C when the culture optical density (OD₆₀₀) is in a range of 0,6 to 0,8 and with a concentration of 0,5 mM IPTG.

The mini-purification processed through affinity chromatography by IMAC pointed that was able to recover the protein with a greater state of purity.

Índice

Sumário	ii
Abstract.....	iii
Índice de Figuras	vi
Índice de Tabelas	viii
1. Introdução	1
1.1. Objectivos do trabalho	2
1.2. Do DNA à proteína.....	3
1.2.1. Transcrição nos procariotas	6
1.2.2. Tradução nos procariotas.....	7
1.3. Estrutura proteica	10
1.4. Clonagem e expressão de proteínas recombinantes	11
1.4.1. Escherichia coli: o sistema hospedeiro por eleição.....	11
1.4.2. Propriedades do gene e da proteína que influenciam a expressão e a solubilidade proteica	13
1.4.3. Propriedades do vector que influenciam a expressão e solubilidade proteica .	14
1.4.4. Estirpes hospedeiras de E. coli que ajudam a expressão de proteínas heterólogas	18
1.4.5. Condições de expressão que melhoram a solubilidade das proteínas	18
1.4.6. Aumento da solubilidade por coexpressão com outras proteínas.....	19
1.5. A importância do uso de Fusion tags.....	20
1.5.1. O novo sistema de fusão Hitag® - Tag Fh8	23
1.6. Purificação de proteínas recombinantes.....	24
1.7. Hidrolase do molinato (MolA): Descrição e propriedades da proteína de interesse .	27
2. Materiais e Métodos	29
2.1. Esterilização de material, soluções e meios de cultura.....	30
2.2. Produtos químicos, soluções e reagentes.....	30
2.3. Microrganismos e plasmídeos.....	30
2.4. Procedimentos gerais de biologia molecular.....	33
2.4.1. PCR - Reacção em cadeia da polimerase.....	33
2.4.2. Digestão com enzimas de restrição	34
2.4.3. Electroforese de DNA em gel de agarose	35
2.4.4. Extracção e purificação de DNA em gel de agarose.....	36

2.4.5.	Ligação de DNA	36
2.4.6.	Transformação em células competentes E. coli NZY5 α e BL21(DE3)	37
2.4.7.	Crescimento de transformantes E. coli NZY5 α e BL21(DE3)	37
2.4.8.	Extracção de DNA plasmídico de culturas de E. coli	37
2.4.9.	Sequenciação de construções	37
2.5.	Procedimentos gerais de análise da expressão proteica em géis SDS-Page Tris-Tricina.....	38
2.6.	Ensaio de indução e purificação da proteína recombinante Fh8MoIA.....	39
2.6.1.	Ensaio preliminar de crescimento celular e indução proteica	40
2.6.2.	Testes de optimização em pequena escala do crescimento celular e indução proteica.....	41
2.6.3.	Purificação da proteína recombinante Fh8MoIA: Mini-Purificação por IMAC em condições nativas.....	42
3.	Resultados e Discussão	45
3.1.	Estratégias de clonagem e análise de construções	46
3.1.1.	Metodologia adoptada com o gene recebido	48
3.1.2.	Metodologia adoptada com o gene sintetizado	57
3.2.	Controlo da indução e análise da expressão proteica em gel SDS-Page.....	60
3.2.1.	Ensaio preliminar de crescimento celular e indução proteica	61
3.2.2.	Testes de optimização em pequena escala do crescimento celular e indução proteica.....	64
3.3.	Análise da mini-purificação por IMAC em gel SDS-Page	69
4.	Conclusões e Perspectivas de futuros trabalhos	73
5.	Bibliografia	77
Anexos	81
Anexo I	– Produtos químicos, soluções e reagentes.....	82
Anexo II	– Sequência dos primers de amplificação	86
Anexo III	– Composição dos géis de SDS-Page Tris-Tricina	87
Anexo IV	– Sequências codificantes das proteínas	88
	Sequência da construção pGEM NcoI MoIA XhoI	88
	Sequência da construção pETMFh8 MoIA	89
Anexo V	– Histogramas e mapas de optimização do gene sintetizado	90

Índice de Figuras

Figura 1 – Ilustração da estrutura química dos dois açúcares presentes nos ácidos nucleico	3
Figura 2 – Esquema representativo da estrutura de DNA, segundo o modelo de Watson e Crick, e estrutura química dos diferentes nucleótidos do DNA	4
Figura 3 – Ilustração geral das fases de transcrição e tradução no interior das células eucarióticas.....	5
Figura 4 – Esquema e direcção da transcrição num ser procariótico.....	6
Figura 5 – Ilustração do esquema geral da expressão proteica, envolvendo as etapas da transcrição e tradução	7
Figura 6 – Ilustração do código genético com a representação dos 64 codões.....	8
Figura 7 – Esquema de elongação da tradução	9
Figura 8 – Estrutura de um aminoácido e estrutura de ligação de uma cadeia polipeptídica	10
Figura 9 – Ilustração representativa dos parâmetros que frequentemente influenciam a expressão solúvel de proteínas recombinantes em <i>E. coli</i>	13
Figura 10 - Construção de um vector de expressão	15
Figura 11 – Representação esquemática do mecanismo de regulação da expressão usando o sistema de expressão	17
Figura 12 – Princípios da separação na purificação por cromatografia.....	24
Figura 13 – Hidrólise do molinato através da acção da MolA presente no <i>Gulosibacter molinivorax</i> ON4T.	28
Figura 14 – Representação do mapa circular do vector pGEM T-Easy	31
Figura 15 - Representação do mapa circular do vector pETMFh8.....	33
Figura 16 – Representação esquemática das estratégias de clonagem aplicadas neste estudo.	47
Figura 17 – Representação esquemática da reacção de PCR utilizada para amplificação do gene da MolA flanqueado nas suas extremidades pelos locais de reconhecimento para as enzimas de restrição NcoI e XhoI.	48
Figura 18 – Perfil em gel de agarose a 1,2% (p/v) dos produtos obtidos da amplificação por PCR.	49
Figura 19 – Representação do mapa circular da construção pGEM NcoI MolA XhoI.	50
Figura 20 – Esquema representativo do tamanho dos fragmentos gerados através da digestão do constructo pGEM NcoI MolA XhoI com as enzimas de restrição NcoI e XhoI.....	51
Figura 21 – Perfil em gel de agarose a 1,2% (p/v) da digestão do constructo pGEM NcoI MolA XhoI com as enzimas de restrição NcoI e XhoI em simultâneo durante 2h.	52

Figura 22 - Perfil em gel de agarose a 1,2% (p/v) da digestão do constructo pGEM NcoI MoIA XhoI com as enzimas de restrição NcoI e XhoI em simultâneo durante 15 minutos ou 30 minutos.....	53
Figura 23 – Representação esquemática da estratégia de clonagem contemplando digestão por etapas.....	54
Figura 24 - Perfil em gel de agarose a 1,2% (p/v) da digestão do constructo pGEM NcoI MoIA XhoI com a enzima de restrição XhoI durante 2h.....	56
Figura 25 – Representação do mapa circular da construção pMK-RQ NcoI MoIA XhoI.....	57
Figura 26 - Esquema representativo do tamanho dos fragmentos gerados através da digestão do constructo pMK-RQ NcoI MoIA XhoI com as enzimas de restrição NcoI e XhoI.....	58
Figura 27 - Perfil em gel de agarose low melting a 1,4% (p/v) da digestão do pMK-RQ NcoI MoIA XhoI com as enzimas de restrição NcoI e XhoI em simultâneo durante 2h.....	59
Figura 28 – Representação do mapa circular da construção pETMFh8 MoIA.....	60
Figura 29 – Representação gráfica dos valores da densidade óptica medidos em função do tempo de crescimento.....	62
Figura 30 - Perfil em gel SDS-Page Tris-Tricina da análise da indução e da expressão solúvel das culturas de 50 mL de E. coli BL21(DE3) pETMFh8 MoIA em condições nativas.....	63
Figura 31 – Perfil em gel SDS-Page Tris-Tricina da análise da indução e da expressão solúvel das culturas de 10 mL de E. coli BL21(DE3) pETMFh8 MoIA em condições nativas para a temperatura de 37 °C.....	65
Figura 32 – Perfil em gel SDS-Page Tris-Tricina da análise da indução e da expressão solúvel das culturas de 10 mL de E. coli BL21(DE3) pETMFh8 MoIA em condições nativas para a temperatura de 18 °C.....	67
Figura 33 - Perfil em gel SDS-Page Tris-Tricina das fracções recolhidas durante a mini-purificação em condições nativas, do conjunto das fracções solúveis das culturas de 10 mL de E. coli BL21(DE3) pETMFh8 MoIA induzidas no momento OD600: 0,6-0,8.....	69

Índice de Tabelas

Tabela 1: Características dos sistemas de expressão em E. coli	12
Tabela 2: Características dos principais fusion tags usados na expressão e purificação de proteínas.....	21
Tabela 3: Proteases usadas na remoção de tags de fusão	22
Tabela 4: Genótipo das estirpes de E. coli utilizadas.....	31
Tabela 5: Tamanho (pb), peso molecular (kDa) e número de aminoácidos da proteína nativa e recombinante.....	60
Tabela Al. 1: Soluções utilizadas nos procedimentos experimentais.	82
Tabela Al. 2: Produtos químicos e reagentes utilizados nos procedimentos experimentais..	84
Tabela All. 1: Sequência dos primers utilizados para amplificar o gene de interesse	86
Tabela AIII. 1: Composição dos reagentes utilizados num gel de SDS-Page Tris-Tricina	87

1. Introdução

1.1. Objectivos do trabalho

A enzima hidrolase do molinato (MolA) é produzida pela bactéria *Gulosibacter molinivorax* ON4^r sendo responsável pela degradação inicial do molinato, um herbicida sistémico que é amplamente aplicado nos campos de arroz para a eliminação de ervas daninhas e que provoca a contaminação dos aquíferos em todo o mundo.

Estudos anteriores visaram o isolamento e produção desta enzima, que resultaram em expressões maioritariamente insolúveis em *Escherichia coli*.

A empresa Hitag Biotechnology, Lda, patenteou recentemente um novo sistema de fusão (Hitag[®] - tag Fh8) que permite melhorar a expressão e aumentar a solubilidade de proteínas recombinantes neste sistema.

Deste modo, o presente trabalho realizado nas instalações da empresa teve por objectivo produzir a enzima MolA, utilizando-se o sistema de fusão (Hitag[®] - tag Fh8) em *E. coli*, com o intuito de melhorar a solubilidade desta proteína.

Pretendeu-se igualmente avaliar e otimizar as propriedades da proteína, nomeadamente, a toxicidade para o sistema hospedeiro e as condições de expressão e de solubilidade proteica, bem como, realizar a purificação, quantificação da produção e avaliação da extensão da purificação da proteína recombinante.

1.2. Do DNA à proteína

As experiências conduzidas por Gregor Mendel e Friedrich Miescher a meio do século XIX foram a base para a descoberta do DNA e o estudo da genética. A modernização de tecnologias e o contínuo interesse em estudar e descobrir novos genomas levam a que, no século XXI, sejam esperados extraordinários avanços em vários campos da actividade humana, em particular no que diz respeito ao conhecimento sobre a biologia dos organismos e à capacidade de a manipular (Videira 2001).

O DNA (ácido desoxirribonucleico) é o material genético de todos os seres vivos, à excepção de alguns vírus, constituídos por RNA (ácido ribonucleico). Como material hereditário, o DNA deve preencher três requisitos fundamentais: 1) ser reproduzido com alta precisão (replicação), de modo a haver conservação das espécies; 2) ser passível de sofrer alterações (recombinação e mutação), para a evolução das espécies; 3) conter informação para a formação de proteínas, que são o suporte estrutural e funcional de uma célula (Videira 2001).

O DNA é uma macromolécula polimérica que se distingue do RNA pelo respectivo açúcar, a desoxirribose ou a ribose, sendo a única diferença, a presença de um átomo de hidrogénio (-H) no carbono 2' na desoxirribose, em vez de um grupo hidroxilo (-OH) na ribose (Figura 1).

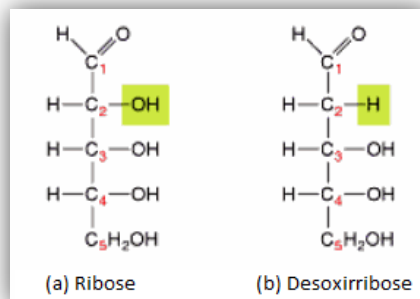


Figura 1 – Ilustração da estrutura química dos dois açúcares presentes nos ácidos nucleicos (adaptado de (Tropp 2012)).

Ligado ao açúcar, uma base azotada (purina ou pirimidina) e um grupo fosfato constituem o nucleótido, ou seja cada unidade monomérica das cadeias polinucleotídicas que constituem o DNA. As bases azotadas ligam-se através do carbono 1' à desoxirribose sendo distinguidas por purinas, adenina (A) e guanina (G), ou pirimidinas, timina (T) e citosina (C). No RNA a timina é substituída pela base pirimídica uracilo (U) (Tropp 2012).

A molécula de DNA é formada por duas cadeias polinucleotídicas enroladas uma em volta da outra no sentido dos ponteiros do relógio, formando uma dupla hélice. As bases azotadas de cada cadeia estão orientadas para o interior da cadeia dupla e estabelecem ligações de hidrogénio apenas entre A e T, com duas ligações de hidrogénio, e entre G e C, com três ligações de hidrogénio, como se observa na Figura 2 (Videira 2001).

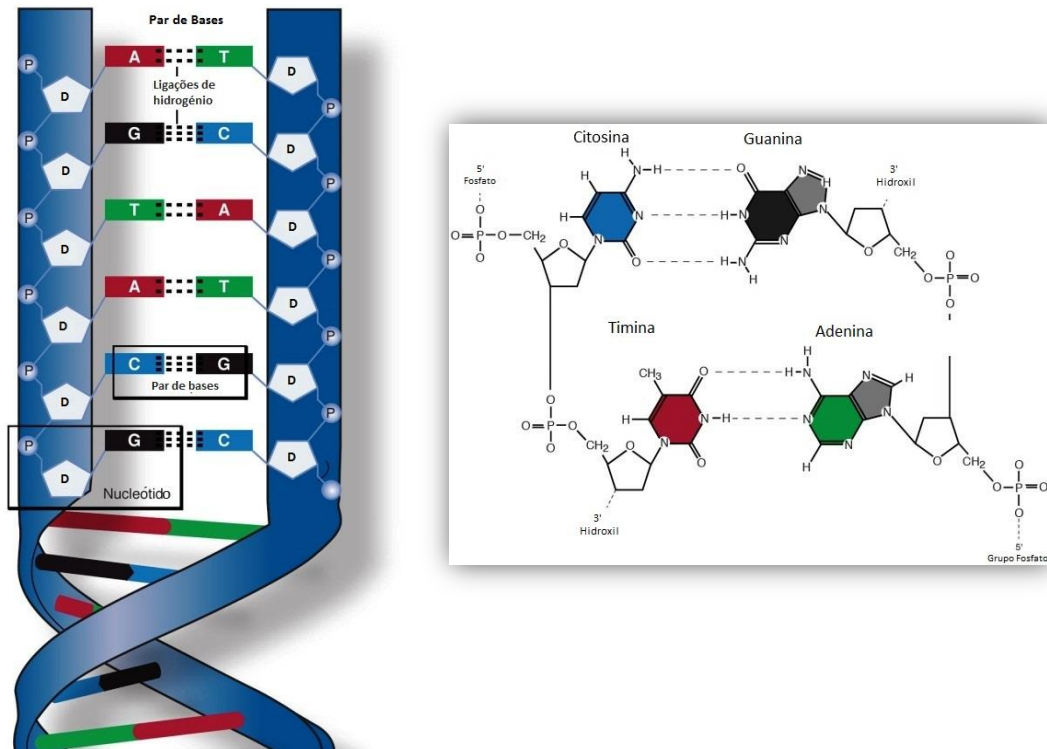


Figura 2 – Esquema representativo da estrutura de DNA, segundo o modelo de Watson e Crick, e estrutura química dos diferentes nucleótidos do DNA (adaptado de (National Human Genome Research Institute 2013)).

O percurso que o DNA faz até à formação de proteínas envolve várias etapas, sendo a primeira, a etapa de transcrição. A transcrição assume-se como o mecanismo universal da expressão de genes, unidades de DNA que contêm a informação necessária à especificidade da síntese de todas as formas funcionais de RNA de cada célula. A nível molecular, a transmissão da informação codificada nos genes é semelhante ao processo de replicação do DNA. Baseia-se numa rigorosa complementaridade estrutural e emparelhamento obrigatório das bases heterocíclicas constituintes dos ácidos nucleicos. Embora obedecendo a mecanismos universais de transcrição de cadeias codificantes do DNA genómico, a síntese de RNA apresenta diferenças que estão relacionadas com o tipo de células em que ocorrem, tal como com o tipo de RNA que é sintetizado (Azevedo 2005). A síntese de RNA é então catalisada na direcção $5' \rightarrow 3'$ pelas RNA polimerases, que utilizam como substrato os ribonucleósidos 5' trifosfato (ATP, GTP, CTP e

UTP). Quanto às cadeias utilizadas para a síntese de RNA, embora num genoma sejam utilizadas as duas cadeias de DNA como molde, para cada gene é usada apenas uma e sempre a mesma cadeia codificadora (Videira 2001).

É através da transcrição dos genes das proteínas que é produzida a população de RNA mensageiros (mRNA), característico de cada espécie e de cada tipo celular (Azevedo 2005).

A transcrição, nos eucariotas, ocorre no núcleo, onde se encontra o DNA genómico (com excepção da transcrição mitocondrial ou cloroplástica), enquanto a tradução tem lugar no citoplasma (Figura 3). Nestas circunstâncias todo o RNA, quer seja mRNA, tRNA ou rRNA tem que migrar do núcleo para o citoplasma, onde desempenham as respectivas funções no processo de síntese de proteínas (Azevedo 2005).

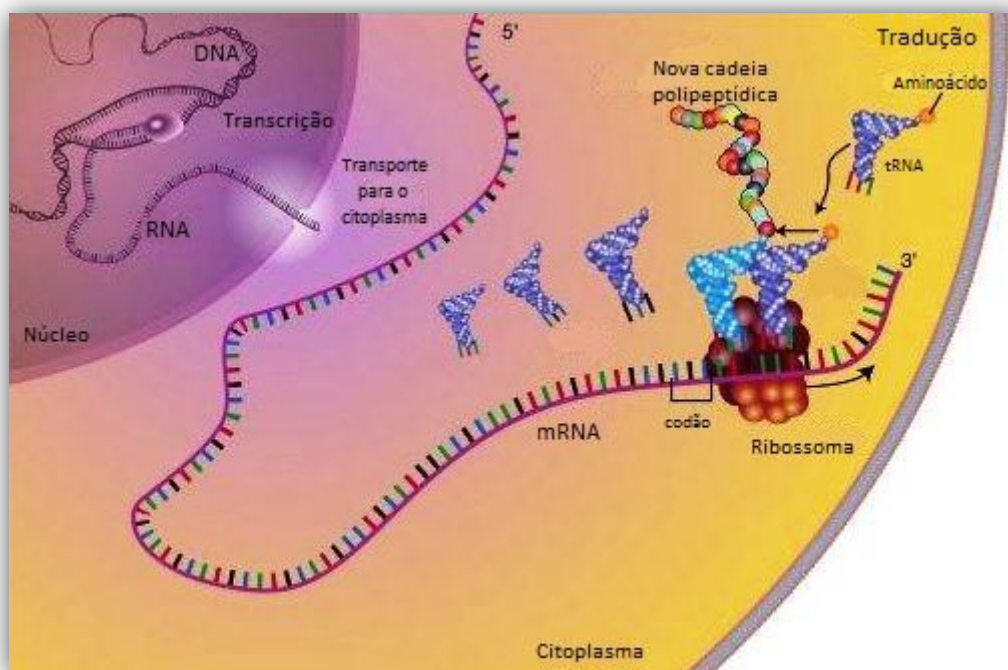


Figura 3 – Ilustração geral das fases de transcrição e tradução no interior das células eucarióticas (adaptado de (National Human Genome Research Institute 2013)).

As células eucarióticas, ao contrário dos procariotas, apresentam uma maior complexidade na biossíntese de formas funcionalmente activas dos diversos tipos de RNA. As dimensões e complexidade da estrutura molecular e supramolecular do genoma conduzem a maiores exigências, implicando a existência de fenómenos de selecção, reconhecimento e transcrição dos diversos genes ou grupos de genes, que estão na base da expressão de cada fenótipo celular (Azevedo 2005).

1.2.1. Transcrição nos procariotas

Nas células procarióticas, a transcrição e a tradução ocorrem no mesmo espaço celular e a transcrição da totalidade dos genes é catalisada por uma única RNA polimerase, independentemente da função e propriedades do RNA a que dá origem (Videira 2001).

A transcrição é, deste modo, um processo bioquímico sequencial que se processa em 3 etapas, nomeadamente, a iniciação, alongação e terminação das cadeias polirribonucleotídicas. Em linhas gerais, a iniciação da reacção consiste no reconhecimento do sitio de DNA genómico, a montante da extremidade 5' do gene (*upstream region*), chamada promotor. O promotor, por sua vez, possui duas regiões importantes para o início da transcrição: a sequência de consenso TTGACA (região -35) e a sequência de consenso TTATAAT (região -10 ou de *Pribnow*). A localização das duas regiões determina qual das cadeias é a codificante e de que modo a RNA polimerase liga ao DNA que, por sua vez, determina o sentido em que ocorre a transcrição (Figura 4). A RNA polimerase reconhece a região -35 e liga fortemente à região -10 provocando a separação das duas cadeias, o que é facilitado pelo facto de a sequencia TTATAAT ser rica em adeninas/timinas (Videira 2001).

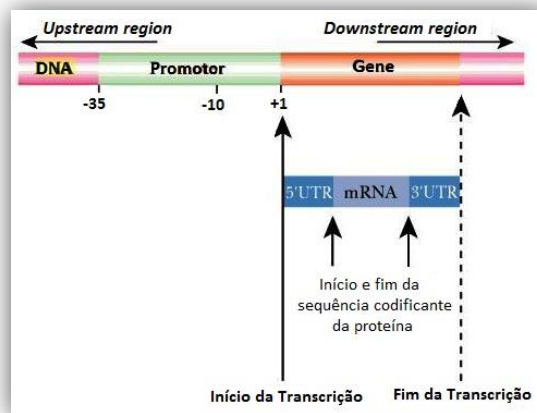


Figura 4 – Esquema e direcção da transcrição num ser procariótico (adaptado de (Clark 2005)).

A alongação das cadeias de RNA é a etapa em que ocorre a polimerização orientada dos nucleótidos, reflectindo a sequência de DNA da cadeia codificante e obedece à regra da complementaridade estrutural das respectivas bases (T-A; A-U; G-C; C-G). À medida que a enzima avança, as duas cadeias desenrolam-se e separam-se pois a RNA polimerase é dotada de propriedade de desnaturar pontualmente o DNA das duas cadeias e de, conseqüentemente, se deslocar ao longo da matriz. Entretanto, a cadeia nascente do RNA neo-sintetizado vai sendo

libertada da cadeia complementar do DNA molde, sendo imediatamente reconstituída na sua conformação nativa, em dupla hélice (Azevedo 2005). A etapa final da transcrição ocorre quando a RNA polimerase encontra sequências nucleotídicas específicas, localizadas a jusante da extremidade 3' da região codificante (*downstream region*), ou associadas a cofactores de transcrição, que actuam como sinais de terminação. A RNA polimerase e a nova cadeia de RNA são então libertadas com a dissociação do complexo de transcrição (Videira 2001).

1.2.2. Tradução nos procariotas

Nestes seres, as cadeias de RNA produzidas, nomeadamente os mRNA, são imediatamente traduzidos em cadeias polipeptídicas, que assim reflectem de forma directa as consequências nucleotídicas do DNA genómico que lhe serviu de molde (Figura 5). A transcrição e a tradução estão assim emparelhadas, ou seja, a tradução de um mRNA é iniciada antes de completada a sua síntese (Azevedo 2005).

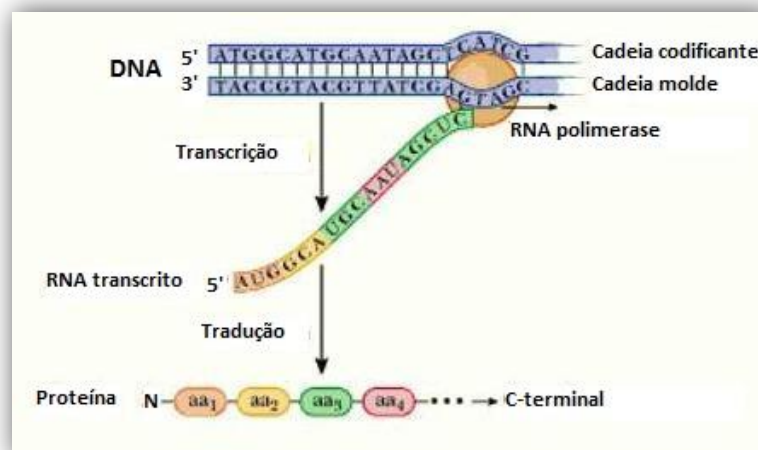


Figura 5 – Ilustração do esquema geral da expressão proteica, envolvendo as etapas da transcrição e tradução (adaptado de (Tropp 2012)).

A conversão de nucleótidos do mRNA em aminoácidos é mediada pelo código genético e consiste num processo complexo de múltiplas reacções em que intervêm, essencialmente, o mRNA, resultante da transcrição do gene codificante para a proteína, o tRNA e factores proteicos. Para que os 20 aminoácidos possam ser codificados por combinação de apenas quatro nucleótidos diferentes, a leitura do mRNA nos ribossomas é conseguida pela utilização de grupos de três (4^3), o que permite a formação de 64 tripletos (codões), que constituem o código

genético completo. Os 20 aminoácidos são codificados por 61 codões e os restantes três são codões de terminação ou codões stop que, à excepção do triptofano, não codificam nenhum aminoácido. Os três codões de terminação são UAA, UAG e UGA e o codão de iniciação é AUG, que codifica a metionina. Apesar de haver preferência pelo uso de certos codões nos diversos organismos, o código genético é considerado universal, salvo algumas excepções. O código genético é também degenerado pois muitos dos aminoácidos podem ser codificados por mais que um codão como se pode observar pela Figura 6 (Videira 2001).

		Segunda base					
		U	C	A	G		
U	UUU	Phe	UCU	UAU	Tyr	UGU	Cys
	UUC		UCC	UAC		UGC	
	UUA		UCA	UAA	STOP	UGA	STOP
	UUG	Leu	UCG	UAG	STOP	UGG	Trp
C	CUU		CCU	CAU	His	CGU	
	CUC	Leu	CCC	CAC		CGC	Arg
	CUA		CCA	CAA	Gln	CGA	
	CUG		CCG	CAG		CGG	
A	AUU	Ile	ACU	AAU	Asn	AGU	Ser
	AUC		ACC	AAC		AGC	
	AUA		ACA	AAA	Lys	AGA	Arg
	AUG	Met	ACG	AAG		AGG	
G	GUU		GCU	GAU	Asp	GGU	
	GUC	Val	GCC	GAC		GGC	
	GUA		GCA	GAA	Glu	GGA	Gly
	GUG		GCG	GAG		GGG	

Figura 6 – Ilustração do código genético com a representação dos 64 codões (adaptado de (Scitable by nature education 2013)).

Tal como na transcrição, a tradução contempla três etapas: iniciação, alongação e terminação. Nos procariotas, a especificidade da etapa da iniciação deve-se à ligação de uma sequência de oito nucleótidos (sequência de Shine-Dalgarno), rica em purinas e localizada perto da extremidade 5' do mRNA, a uma sequência complementar do rRNA da pequena subunidade ribossomal. Esta ligação coloca o codão de iniciação AUG no local do ribossoma denominado P (peptidil). De seguida, o tRNA iniciador suportando a metionina formilada é levado com a ajuda de factores de iniciação para o local P e pode ligar pelo seu anticodão 3'-UAC-5' ao codão 5'-AUG-3'. A etapa da iniciação termina quando se liga à grande subunidade ribossomal, formando o complexo de iniciação e permitindo que o mRNA se mova quando o ribossoma se encontra completo (Videira 2001).

As etapas de alongação e de terminação ocorrem de modo semelhante em procariotas e eucariotas. O processo de alongação corresponde à etapa da formação da cadeia peptídica, ou

seja à formação de uma ligação peptídica entre dois aminoácidos adjacentes, sendo a ordem dos aminoácidos determinada pela ordem dos codões no mRNA (Azevedo 2005). No ribossoma, existe o local A (aminoacil) onde se vão ligar todos os outros tRNAs, com excepção do iniciador. Quando o codão AUG está no local P, está exposto em A outro codão, que determinará qual o tRNA carregado que se vai ligar pelo seu anticodão complementar (Figura 7). Esta fase designada de transpeptidação é catalisada pela enzima peptidil-transferase, que estabelece a ligação peptídica entre o grupo amino do aminoácido do local A e o grupo carboxilo do outro aminoácido (Videira 2001).

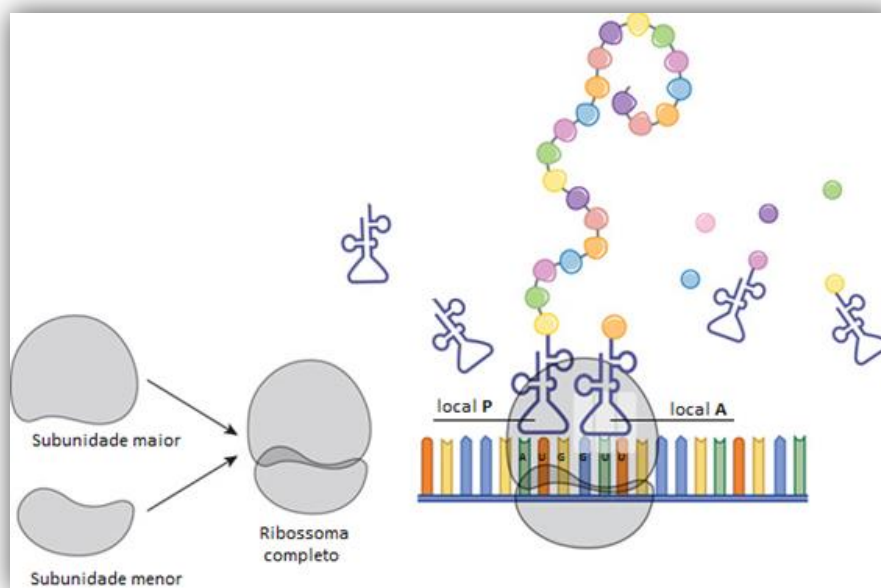


Figura 7 – Esquema de elongação da tradução (adaptado de (Scitable by nature education 2013)).

O mRNA continua o seu movimento ao longo do ribossoma fazendo com que o tRNA iniciador saia, enquanto no local P fica o tRNA que estava em A e que transporta agora um dipeptídeo (translocação). No local A fica exposto um novo codão e repete-se o processo. Ou seja, é introduzido outro tRNA que transporta um novo aminoácido e se liga ao dipeptídeo, formando o tRNA transportador de um tripeptídeo, que passa ao local P e assim sucessivamente, até que é exposto um codão de terminação em A (Videira 2001).

Para os codões de terminação não há tRNAs com anticodões complementares. Porém, factores de terminação reconhecem esses anticodões, ligam-se ao ribossoma e quebram a ligação da cadeia peptídica ao último tRNA, separando ainda o tRNA e o mRNA do ribossoma. Outros factores separam depois as duas subunidades ribossomais, que podem ser reutilizadas

numa nova síntese proteica (Videira 2001). Na *E. coli*, os factores de terminação/dissociação são designados por RF-1, RF-2 e RF-3. O factor RF-1 reconhece os codões de terminação UAA e UAG e o factor RF-2 reconhece UAA e UGA. Já o factor RF-3 liga-se ao GTP de modo a estimular a ligação do RF-1 e RF-2 ao ribossoma (Azevedo 2005).

Após a tradução, as proteínas podem ser modificadas de várias maneiras. Na maior parte dos casos a metionina N-terminal é retirada.

1.3. Estrutura proteica

As proteínas são polímeros essenciais para a função, estrutura, diferenciação e desenvolvimento de um organismo. Uma proteína é constituída por uma ou várias cadeias peptídicas formadas por sequências específicas de aminoácidos. Em cada proteína, os aminoácidos encontram-se ligados pelos seus grupos carboxílicos ao grupo amínico de aminoácido, adjacente por uma ligação peptídica. Numa cadeia polipeptídica, o primeiro aminoácido possui um grupo amínico livre (N-terminal) e o último tem um grupo carboxílico livre (C-terminal), como se pode observar na Figura 8 (Videira 2001).

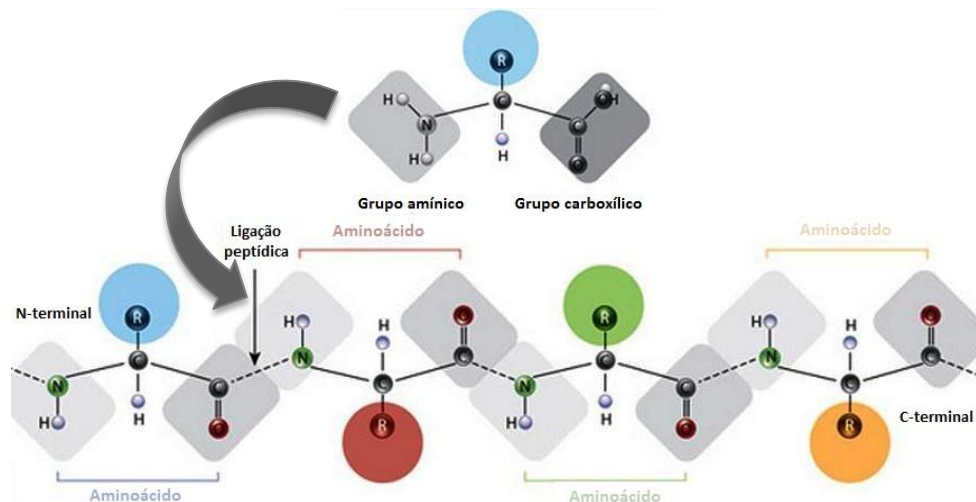


Figura 8 – Estrutura de um aminoácido e estrutura de ligação de uma cadeia polipeptídica (adaptado de (Scitable by nature education 2013)).

A nível de estrutura, as proteínas possuem estruturas complexas com quatro níveis de organização distintas: estrutura primária, constituída pela sequência linear de aminoácidos na cadeia polipeptídica; estrutura secundária, resultante de ligações fracas (hidrogénio, electrostáticas e *van der Waals*) entre os átomos da mesma, ou de diferentes cadeias polipeptídicas, originando configurações específicas das proteínas; a estrutura terciária que

ocorre quando a proteína, na sua estrutura secundária, se dobra sobre si mesma, formando uma estrutura tridimensional; a estrutura quaternária é originada pela associação de duas ou mais estruturas terciárias de cadeias polipeptídicas idênticas ou de diferentes tipos (Videira 2001).

1.4. Clonagem e expressão de proteínas recombinantes

A produção de proteínas constitui uma das mais importantes aplicações da engenharia genética. O aparecimento da biologia molecular, nos anos 70, tornou possível a produção de proteínas heterólogas em diferentes células hospedeiras representando uma alternativa à extracção da proteína original. A extracção clássica permite normalmente concentrações muito baixas tornando-a muito dispendiosa e, muitas vezes, impossível de implementar. Um outro problema associado é a pureza da proteína obtida que pode não ser suficiente para eliminar contaminações tóxicas, pirogénicas ou infecciosas (Teixeira e Fonseca 2007).

Na era pós-genómica são grandes as expectativas das reais potencialidades da expressão e purificação de proteínas recombinantes no diagnóstico e tratamento de doenças genéticas e noutras aplicações na área da saúde, bem como na agricultura, no ambiente e na indústria biotecnológica (Azevedo 2005).

No entanto, a expressão de proteínas recombinantes necessita de um planeamento correcto que vise a produção destas na conformação correcta permitindo o estado solúvel e activo. Para isso são necessários elementos essenciais: Identificação e localização do gene de interesse, inserção do gene alvo num vector adequado, introdução do vector no hospedeiro designado, selecção de células transformantes e multiplicação/expressão do gene escolhido no hospedeiro (Francis e Page 2010).

1.4.1. *Escherichia coli*: o sistema hospedeiro por eleição

A *Escherichia coli* continua a ser o hospedeiro preferido para a expressão de proteínas recombinantes e as razões do seu uso são bem conhecidas: facilidade de manipulação genética, culturas de baixo custo e uma expressão rápida, com as proteínas a serem normalmente produzidas num só dia. Aliada a estas características, a expressão por *E. coli* é um sistema robusto com a proteína expressa a compreender até 50% da totalidade das proteínas celulares (Peti e Page 2007).

Apesar das múltiplas vantagens deste sistema de expressão e do uso generalizado, existem também desvantagens associadas como consta na Tabela 1.

Tabela 1: Características dos sistemas de expressão em *E. coli* (adaptado de (Demain e Vaishnav 2009))

Vantagens	Desvantagens
Expressão rápida	Dificuldade em expressar proteínas com ligações dissulfureto
Rendimentos elevados	Incapacidade de realizar certas modificações pós-traducionais
Fácil cultura e modificação do genoma	Produção de proteínas contendo endotoxinas
Culturas de baixo custo	Densidades elevadas resultam em toxicidade celular
Produção massiva e de baixo custo	Formação de corpos de inclusão

Em contraste com os sistemas de expressão eucarióticos, a transcrição e a tradução acontecem rapidamente, como anteriormente referido, e fortemente ligadas. Visto que, muitas das proteínas eucarióticas requerem tempos mais longos e/ou de assistência dos *chaperones* moleculares para permitir a conformação correcta das proteínas no seu estado nativo, este aumento de velocidade leva, normalmente, a fenómenos conhecidos como *unfolding* ou *misfolding* proteico e, conseqüentemente, à insolubilidade das proteínas (Francis e Page 2010).

Assim, algumas das proteínas, essencialmente proteínas de multidomínios e proteínas membranares, não expressam em *E. coli* ou expressam insolúvelmente como corpos de inclusão. Além disso, a insolubilidade das proteínas não é restrita às proteínas heterólogas pois algumas das próprias proteínas bacterianas podem também não ser expressas na forma solúvel quando são sobre-expressas em *E. coli*. Em adição a estas desvantagens torna-se um desafio produzir eficientemente proteínas contendo ligações dissulfureto no ambiente reduzido do citoplasma da bactéria. Finalmente, a *E. coli* não possui a maquinaria necessária para executar certas modificações pós-traducionais próprias dos eucariotas, tais como a glicosilação, o que pode ser crítico para a formação de proteínas activas (Francis e Page 2010).

De modo a contornar os problemas supra referidos, nos últimos anos têm sido desenvolvidos esforços para maximizar uma produção eficiente de proteínas recombinantes solúveis na bactéria *E. coli*. Um número notável de novos reagentes (vectores, estirpes hospedeiras, entre outros) e novas estratégias (coexpressão usando *chaperones*, indução por baixas temperaturas, entre outros) têm sido desenvolvidas para permitirem que muitas das desvantagens associadas a este sistema sejam superadas (Francis e Page 2010).

Algumas das limitações comumente identificadas durante a expressão em *E. coli* apresentam-se na Figura 9.

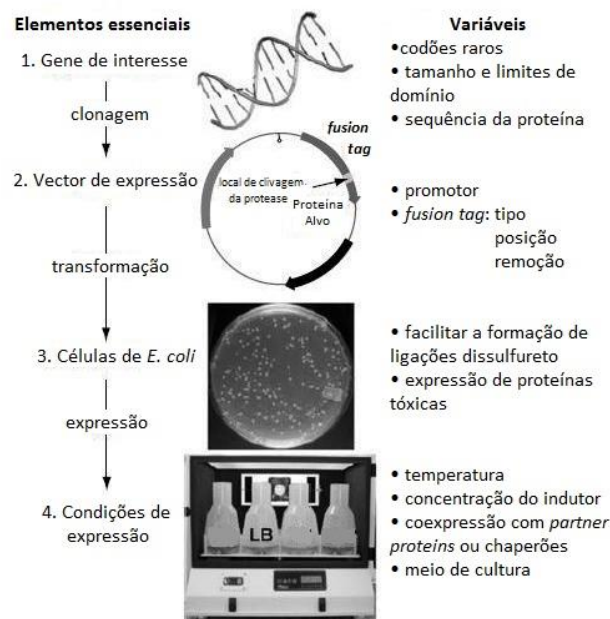


Figura 9 – Ilustração representativa dos parâmetros que frequentemente influenciam a expressão solúvel de proteínas recombinantes em *E. coli* (adaptado de (Francis e Page 2010)).

1.4.2. Propriedades do gene e da proteína que influenciam a expressão e a solubilidade proteica

Uma das razões mais frequentes para as falhas na expressão de proteínas heterólogas neste sistema deve-se à presença de **codões raros** na sequência alvo de mRNA. Genes que incluem um número significativo de codões raros individuais ou um pequeno número de codões raros juntos são mais propícios a que ocorra uma paragem da tradução em *E. coli* e, por isso, a expressão proteica pode falhar completamente, expressar a níveis muito baixos ou levar à produção de proteínas truncadas (Francis e Page 2010).

O **tamanho da proteína** e os **limites do domínio** são outros factores que influenciam na síntese proteica. Estudos mostram que existe uma grande probabilidade da expressão neste sistema diminuir com o aumento do peso molecular das proteínas, especialmente para proteínas maiores que 60 kDa (Canaves, et al. 2004). Assim, quando se utiliza a *E. coli* como hospedeiro é, tipicamente, vantajoso expressar domínios individuais da proteína em vez de expressar proteínas com o seu comprimento total, sempre que tal seja possível. Os resíduos de início e do

fim do domínio alvo podem também afectar o rendimento e a solubilidade da expressão proteica (Francis e Page 2010).

A preferência que *E. coli* tem por certos codões, ou seja o **codon usage** da *E. coli*, é um dos problemas mais comuns na expressão de proteínas recombinantes. Tal como referido anteriormente, os aminoácidos são codificados por mais que um codão (redundância do código genético) e cada organismo apresenta codões preferenciais entre os 61 codões disponíveis. Os tRNAs existentes em cada célula reflectem directamente a preferência de codões do mRNA disponível, o que significa que um tRNA que reconhece um codão pouco utilizado, está presente em pequenas quantidades. A falta de tRNAs adequados pode levar à estagnação ou terminação prematura da tradução, à troca de aminoácidos na sequência ou à incorporação errada de determinados aminoácidos. Teoricamente, a modificação das condições da cultura, como a redução de temperatura ou a alteração da composição do meio, pode ser suficiente para modificar o *codon usage* da *E. coli* e impedir alguns problemas de expressão. Este problema também pode ser ultrapassado através da troca de codões raros, no gene alvo, por codões mais utilizados pela *E. coli* ou pela co-produção dos respectivos tRNAs raros (Terpe 2006).

Adicionalmente, outras propriedades biofísicas presentes na **sequência da proteína** podem influenciar o rendimento, tais como a hidrofobicidade de alguns resíduos, a presença de regiões de baixa complexidade e/ou presença de pontes dissulfureto e de segmentos transmembranares (Francis e Page 2010).

1.4.3. Propriedades do vector que influenciam a expressão e solubilidade proteica

Uma vez determinada a proteína de interesse e a correspondente construção, deve ser subclonada num vector que contenha todos os elementos necessários à transcrição e tradução do gene alvo. Os plasmídeos, pela sua maneabilidade, são os vectores de expressão por excelência que permitem clonar genes ou fragmentos de DNA específicos (Videira 2001). Os elementos integrados nos plasmídeos incluem promotores, sequências de regulação, a caixa de Shine-Dalgarno, terminadores de transcrição, origens de replicação, entre outros. Além disso, os vectores de expressão contêm normalmente uma **marca de selecção**, um local de clonagem múltipla (**multiple cloning site - MCS**) e um local de ligação ao ribossoma (**ribosomal binding site – RBS**). O primeiro é tipicamente um gene de resistência a um antibiótico, que

permite seleccionar as células transformantes após a inserção do vector. O segundo contém sequências de reconhecimento para várias enzimas de restrição e permite clonar directamente DNA digerido com qualquer uma dessas enzimas. O terceiro tem a função de regular a síntese proteica (Sørensen e Mortensen 2005). Um outro elemento relevante que pode ser inserido no vector são **fusion tags** que são essenciais pois são transcritos junto com o gene que codifica a proteína de interesse e permitem ajudar na expressão, solubilidade e, alguns, na purificação da proteína (Francis e Page 2010). O uso de *fusion tags* e a sua importância será abordado mais concretamente na secção 1.5 deste capítulo.

Na Figura 10 está sumariamente representada a arquitectura essencial de um vector de expressão.

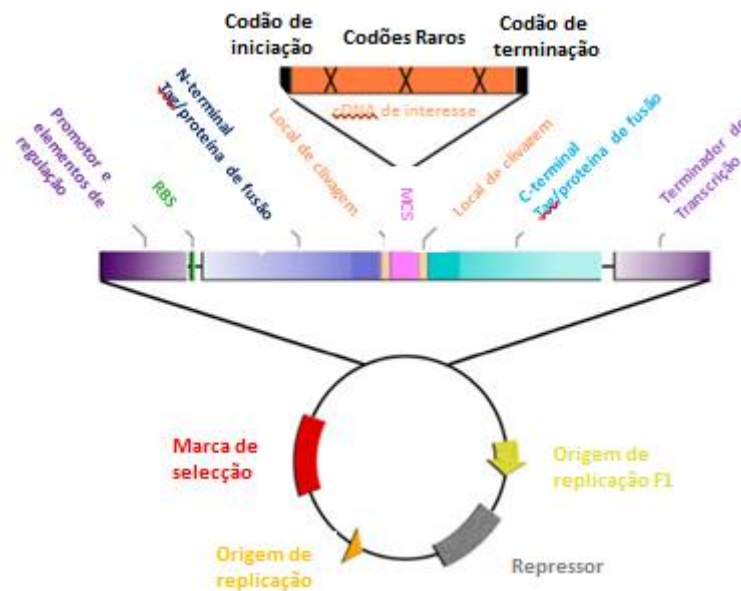


Figura 10 - Construção de um vector de expressão (adaptado de (Domingues 2012)).

A **origem de replicação** de um vector é o local onde a replicação se inicia e que também determina o número de cópias do vector no sistema hospedeiro. Para a produção de proteínas é desejado tipicamente um elevado número de cópias que resultem num rendimento máximo num dado volume de cultura (Francis e Page 2010).

Os **promotores** são outro elemento dos vectores que podem ter um efeito profundo na força e duração da transcrição e, por sua vez, no rendimento proteico. Um promotor eficaz para a expressão de proteínas heterólogas em *E. coli* necessita de possuir três características fundamentais. Primeiro, necessita de ser forte para resultar em expressões fortes do gene alvo (tipicamente de 10% a 50% da proteína total). Segundo, necessita exibir uma actividade na

transcrição baixa de modo a impedir uma transcrição indesejada antes da indução. Terceiro, a indução deve ser simples e de baixo custo (Francis e Page 2010).

Quando se selecciona um sistema promotor deve-se ter em consideração a natureza da proteína desejada e o seu uso posterior. Considerações sobre a natureza tóxica da proteína, da sua tendência para agregar e a obtenção de rendimentos máximos devem ser considerados. Os promotores mais usados em *E. coli* são o promotor T7, o *araBAD*, promotores híbridos e o promotor *cspA*, estando bem caracterizadas as respectivas vantagens e desvantagens de cada um (Francis e Page 2010).

1.4.3.1. O SISTEMA pET E O PROMOTOR T7

O Sistema pET tem sido reconhecido como um dos sistemas mais poderosos desenvolvidos para a clonagem e expressão de proteínas recombinantes em *E. coli*. Os genes de interesse são clonados nos plasmídeos pET sob o controlo de fortes sinais de transcrição e (opcionalmente) de tradução do bacteriófago T7. A expressão é induzida pelo fornecimento de uma fonte da RNA polimerase T7 na célula hospedeira. Esta polimerase é tão selectiva e activa que, quando completamente induzida, a maior parte dos recursos das células são convertidos para a expressão do gene pretendido. Assim, após algumas horas de indução, a proteína desejada pode compreender mais do que 50 % do total das proteínas presentes nas células (Novagen 2010).

O gene que codifica a RNA polimerase T7 tem sido clonado em várias estirpes de *E. coli*, comercialmente disponíveis sob um sistema modificado do operão *lac*. Um desses casos é a estirpe *E. coli* BL21(DE3) na qual, em vez de se encontrar a sequência do promotor T7 antes do operador *lac* (*lacO*), encontra-se a sequência do promotor *lac* (ao qual se ligam as RNA polimerases nativas da *E. coli*). Assim que o repressor *lac* (*LacI*) cai da sequência do operador no DNA do cromossoma do hospedeiro, a RNA polimerase T7 começa a ser transcrita e traduzida.

Este repressor detecta a presença de lactose (um dissacarídeo de galactose-glucose). Tanto o cromossoma do hospedeiro como o plasmídeo possuem cópias do gene do repressor *lac* para garantir que haja sempre proteína *LacI* suficiente para ocupar os sítios de operador de DNA. Na ausência da lactose, o repressor *lac* liga-se à sequência do operador. Isto impede que antes de ocorrer a indução, se desencadeie uma série de reacções, nomeadamente, a transcrição e tradução da RNA polimerase T7, a sua ligação ao promotor T7 e consequentemente a transcrição do gene de interesse. Quando a lactose se liga ao repressor, isso induz uma

alteração conformacional na sua estrutura, que o torna incapaz de se ligar à sequência do operador. O isopropil- β -D-tiogalactopiranosídeo (IPTG) é um mimico estrutural da lactose, que também se liga ao repressor *lac* e induz uma mudança conformacional semelhante, reduzindo significativamente a sua afinidade para o DNA. Ao contrário de lactose, o IPTG não é parte de qualquer via metabólica e, portanto, não vai ser dividido ou consumido pela célula. Isto assegura que a concentração de IPTG adicionada permaneça constante, tornando-se um indutor mais útil do operão *lac* que a própria lactose (Bell e Lewis 2000).

Uma vez que o repressor *lac* já não se pode ligar ao operador, as RNA polimerases nativas da *E. coli* começam a transcrição do gene da RNA polimerase T7, num elevado número de cópias. Assim, quando a proteína da RNA polimerase T7 é expressa, esta liga-se à sequência do promotor T7, a montante do gene de interesse, e ocorre a transcrição deste gene (Bell e Lewis 2000).

A Figura 11 permite elucidar o mecanismo regulado pelo promotor T7.

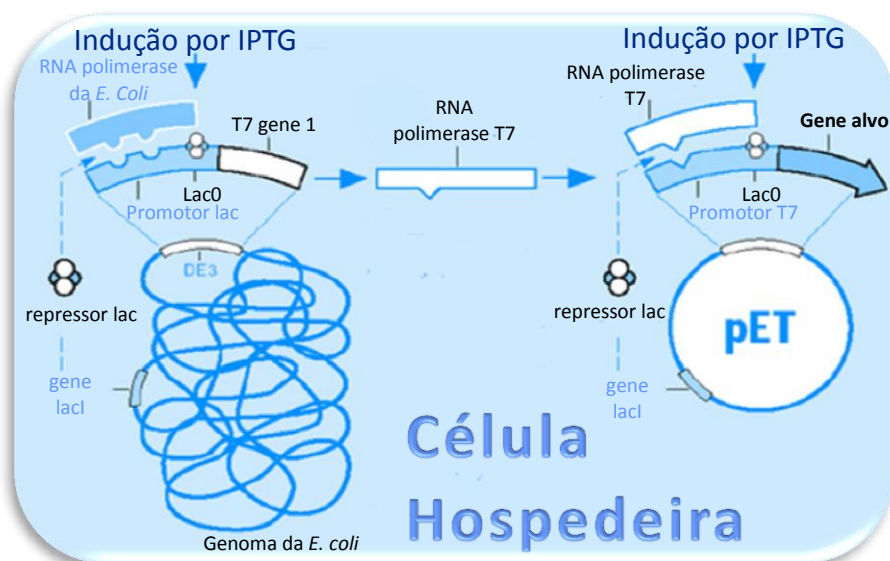


Figura 11 – Representação esquemática do mecanismo de regulação da expressão usando o sistema de expressão pET (adaptado de (Arvani 2009)).

1.4.4. Estirpes hospedeiras de *E. coli* que ajudam a expressão de proteínas heterólogas

Vários elementos das estirpes de *E. coli* utilizadas para a produção de proteínas recombinantes têm um grande impacto no sucesso de uma expressão solúvel. As estirpes não devem conter proteases prejudiciais e devem manter a estabilidade do plasmídeo durante a expressão, conferindo os elementos genéticos importantes para o sistema de expressão (Sørensen e Mortensen 2005). Tal como mencionado anteriormente, estirpes hospedeiras bacterianas foram especificamente desenvolvidas para auxiliar na expressão de proteínas heterólogas de modo a obter expressões específicas, nomeadamente, de proteínas que são susceptíveis à proteólise (e.g. BL21, BL21(DE3) e BL21Star(DE3)) que contêm codões raros (e.g. BL21-RP, BL21-RIL, BL21-RPIL, Rosetta e Rosetta-gami) ou que requerem ligações dissulfureto (e.g. BL21 trxB, Origami e Rosetta gami) (Francis e Page 2010).

1.4.5. Condições de expressão que melhoram a solubilidade das proteínas

O uso de promotores de expressão fortes e de altas concentrações de indutor podem resultar em concentrações da proteína recombinante que, são tão elevadas, que podem agregar antes do *fold*ing. Para isso não ocorrer torna-se necessário reduzir a taxa de transcrição e/ou de tradução (Francis e Page 2010). De forma a aumentar a solubilidade da proteína, podem ser manipulados alguns parâmetros, nomeadamente a temperatura, a concentração do agente indutor e os meios de cultura.

A diminuição da **temperatura** de expressão provoca, normalmente, uma melhoria da solubilidade das proteínas recombinantes expressas. Isto deve-se ao facto de a baixas temperaturas, os processos celulares abrandarem levando não só a taxas reduzidas de transcrição, tradução e divisão celular, mas também a baixas agregações proteicas. Complementariamente, a maior parte das proteases são menos activas a baixas temperaturas, o que resulta numa redução na degradação das proteínas sensíveis à proteólise (Francis e Page 2010).

Tal como a temperatura, uma diminuição da **concentração do agente indutor** pode, por vezes, reduzir a própria taxa de transcrição. Deste modo, reduzindo um pouco a concentração de IPTG, por exemplo, pode aumentar a produção de proteínas solúveis (Francis e

Page 2010). Diferentes mecanismos indutores encontram-se referenciados consoante o promotor usado (Terpe 2006).

Em relação aos **meios de cultura**, o meio *Luria broth* (LB) é o padrão de nutrientes que é usado para a expressão, sendo composto por bacto-triptona que proporciona péptidos, peptonas e aminoácidos essenciais; por extracto de levedura que fornece essencialmente vitaminas; e por cloreto de sódio que fornece iões de sódio de modo a manter o equilíbrio osmótico (Francis e Page 2010).

1.4.6. Aumento da solubilidade por coexpressão com outras proteínas

Algumas proteínas apenas expressam no estado solúvel quando são coexpressas com biomoléculas adicionais, tipicamente outras proteínas, tais como um *binding partner* ou *chaperones* moleculares. Nestes casos a proteína alvo é coexpressa conjuntamente com a segunda proteína que é codificada no mesmo plasmídeo ou num outro em separado (Francis e Page 2010).

A expressão de proteínas cujas actividades quebram o crescimento da *E. coli* e/ou a sinalizam é, frequentemente, tóxico para a célula. Enquanto as células hospedeiras podem ser cultivadas a altas densidades sob condições repressivas, a indução pode resultar numa paragem de crescimento rápida e/ou morte celular. Para algumas proteínas, esta toxicidade pode ser atenuada quando a toxina é coexpressa com uma segunda proteína que, tanto se liga e inibe a toxina, como não se liga e, contudo, neutraliza a sua actividade (Francis e Page 2010).

Para algumas proteínas bacterianas, de menores dimensões, obter uma conformação correcta é um processo normalmente rápido, devido à rápida cinética do *folding*. No entanto, uma grande quantidade de proteínas heterólogas e bacterianas fazem-no mais lentamente requerendo, assim, *chaperones* proteicos e a catálise da etapa de *folding* para prevenir a agregação e facilitar essa mesma etapa (Francis e Page 2010).

1.5. A importância do uso de *Fusion tags*

A insolubilidade, a conformação, a estabilidade e a flexibilidade estrutural das proteínas, tal como, os baixos rendimentos de purificação e a toxicidade para as células hospedeiras, são desafios que se devem ter em conta quando se deseja expressar proteínas recombinantes (Young, Britton e Robinson 2012).

O uso de *tags* de afinidade, também chamados de *fusion tags* ou tags de fusão, conjuntamente com as técnicas de *DNA* recombinante, facilitam a modificação das proteínas de interesse levando a uma eficiente identificação, produção e isolamento do sistema hospedeiro (Young, Britton e Robinson 2012). Estes *tags* são proteínas ou péptidos que se ligam covalentemente à proteína de interesse que, assim se passa a chamar de *fusion protein* ou proteína de fusão. Eles são úteis porque podem melhorar a expressão proteica, promover o *folding*, aumentar a solubilidade e facilitar os processos seguintes, tais como a purificação e a detecção de proteínas (Francis e Page 2010).

No entanto, o *tag* perfeito não existe. Frequentemente, é necessário testar múltiplos *fusion tags* para determinar quais os que resultam em maiores rendimentos, ou uma combinação destes, de modo a facilitar tanto a expressão como a purificação proteica (Francis e Page 2010).

Os *tags* podem ser colocados tanto nas extremidades N- ou C-terminal e esta colocação é devesas importante, uma vez que pode ter um efeito profundo nos níveis de expressão da proteína no estado solúvel. A fusão no N-terminal é a mais comum pois expressa, frequentemente, proteínas solúveis com mais sucesso que a fusão no C-terminal. Uma das razões para este facto deriva da maior parte dos *fusion tags* serem proteínas nativas da *E. coli*. Assim, proporcionam uma sequência 5' eficiente para a iniciação da transcrição e, conseqüentemente, a sequência N-terminal do mRNA e a estrutura são eficientemente reconhecidas pelo ribossoma permitindo uma tradução robusta (Francis e Page 2010).

Existe uma grande diversidade de sistemas de fusão referenciada na bibliografia com vista ao aumento solubilidade e também usados como *tags* de purificação. Na Tabela 2 é possível comparar as principais características entre os *tags* mais usados em processos de expressão e purificação de proteínas.

Tabela 2: Características dos principais *fusion tags* usados na expressão e purificação de proteínas (adaptado de (Francis e Page 2010))

Tag	Proteína	Amino-ácidos	Tamanho (kDa)	Organismo de origem	Tag de purificação	Comentários
Tags com propriedades solubilizadoras						
GST	<i>Glutathione S-transferase</i>	243	28.1	<i>Schistosoma Japonicum</i>	Sim	Formação de dímeros
MBP	<i>Maltose binding protein</i>	390	43.0	<i>Escherichia coli</i>	Sim	Forte solubilizador
DsbA	<i>Disulfide oxidoreductase</i>	228	25.4	<i>Escherichia coli</i>	Não	Facilita a formação de ligações dissulfureto citosólicas
NusA	<i>N-utilizing substance A protein</i>	535	59.3	<i>Escherichia coli</i>	Não	Forte solubilizador
Trx	<i>Thioredoxin</i>	135	14.7	<i>Escherichia coli</i>	Não	Facilita a formação de ligações dissulfureto citosólicas
Z-domain	<i>Protein A IgG ZZ repeat domain</i>	91	10.6	<i>Staphylococcus aureus</i>	Sim	
GB1	<i>Protein G β1 domain</i>	85	9.7	<i>Streptococcus</i> sp.	Sim	Muitas vezes usado com proteínas para análise NMR
SUMO	<i>Small ubiquitin-like modifier</i>	99	11.1	<i>Homo sapiens</i>	Não	
SET	<i>Solubility enhancing tags</i>	<40		T7 phage gene 10B; synthetic	Não	
HaloTag-7	<i>Catalytically inactive derivative of DhaA</i>	296	34.0	<i>Rhodococcus</i> sp.	Sim	Forte solubilizador
Tags sem propriedades solubilizadoras						
His6	<i>Hexahistadine</i>	6	0.8	Sintetizado	Sim	Muitas vezes combinados com tags solubilizadores
Intein	<i>Protein splicing element</i>	128-1650		Variável	Sim	Remover das resinas por auto-clivagem

Na já grande variabilidade de *tags* existentes, destacam-se os sistemas *glutathione-S-transferase* (GST), *maltose-binding protein* (MBP), *thioredoxin A* (TrxA) e *small ubiquitin-related modifier* (SUMO) por serem relativamente conhecidos e usualmente utilizados para aumentar a solubilidade proteica. *Tags* como o *ketosteroid isomerase* (KSI) e o Trp Δ LE, diminuem a solubilidade proteica e têm sido utilizados para conduzir a expressão para a formação de corpos de inclusão (Young, Britton e Robinson 2012). Isto pode ser uma vantagem se, por exemplo, se pretender usar as proteínas como antigénio para a preparação de um anticorpo pois, além de serem fáceis de purificar, são resistentes à degradação de proteases (Videira 2001). Já o *polyhistidine tag* (*His-tag*) não sendo um *tag* de solubilização, adquire enorme importância na área da biologia molecular para purificação de proteínas recombinantes (Young, Britton e Robinson 2012).

A presença de *tags* pode interferir com a actividade biológica da proteína expressa e, nestes casos, torna-se necessário clivar enzimaticamente o *tag* após a purificação da proteína. Isto é alcançado pela utilização de uma protease que possui um sítio de clivagem específico (uma sequência de aproximadamente 7 aminoácidos que são especificamente reconhecidos pela protease) entre o *tag* e a proteína alvo (Francis e Page 2010).

Estão referenciadas diversas proteases, sendo as mais utilizadas a *tobacco etch virus* (TEV), a 3C, o factor Xa, a trombina (Thr) e a enteroquinase (Entk). A descrição e a sequência de corte de cada protease são apresentadas na Tabela 3.

Tabela 3: Proteases usadas na remoção de *tags* de fusão (adaptado de (Francis e Page 2010))

Protease	Descrição	Sequência de corte ^(a)
TEV	Protease da família das cisteínas, que se encontra no vírus <i>Etch</i> do tabaco	EXXYXQ/(G-S) ou mais comum: ENLYFQ/X
3C	Forma recombinante da protease 3C proveniente do rinovírus humano do tipo 14	EVLFQ/GP
Xa	O Factor Xa é uma protease de serina que converte protrombina em trombina	I(E-N)GR/
Thr	A Trombina é uma protease de serina que converte fibrinogénio em fibrina	LVPR/GS
Entk	Subunidade catalítica da enteroquinase bovina	DDDDK/

^(a) O símbolo “/” indica o local de corte da protease e a sequência de corte é dada no código de aminoácido com uma única letra, onde X representa qualquer aminoácido

Das proteases conhecidas, a TEV é claramente a mais utilizada pois possui uma série de vantagens. Em primeiro lugar, é mais rigorosa e, por isso, faz a clivagem de modo eficiente sem ocorrer clivagem secundária não específica. Em segundo lugar, pode acomodar um número de diferentes resíduos no lado C- terminal, eliminando resíduos não nativos no N-terminal da proteína alvo. Esta protease é também activa ao longo de um conjunto diversificado de condições experimentais, tais como o pH, soluções tampão e temperaturas. Por outro lado, a desvantagem é que esta protease é inibida por um grande número de detergentes, tornando-se um desafio a remoção de *fusion tags* a partir de proteínas transmembranares pois requerem complexos proteína-detergente para solubilização (Francis e Page 2010).

1.5.1. O novo sistema de fusão Hitag® - Tag Fh8

O sistema de fusão Hitag (patenteado em 2009: Patente PCT/IB2009/055647) contempla dois *tags* de fusão: H e Fh8. O primeiro trata-se de um *tag* imunopotenciador e o segundo permite aumentar a expressão e a solubilidade (Hitag Biotechnology, Lda 2013).

A proteína Fh8 (acesso do GenBank: AF213970.1) possui apenas 69 aminoácidos e tem peso molecular de cerca de 8 kDa (7,64 kDa) sendo segregada pelo parasita *Fasciola hepatica* nas fases iniciais da infecção (Fraga, et al. 2010). O seu nome deriva precisamente do organismo que a segrega e do seu tamanho. Trata-se de uma protease e de um antigénio que tem importância no mecanismo de infecção do parasita e que, estruturalmente possui duas *EF-hands* que se ligam ao cálcio formando uma estrutura em forma de cubo. Para além do Fh8 ser extremamente solúvel e estável em *E. coli*, pode ser facilmente purificado em cromatografia por níquel e já foi utilizado como ferramenta imunológica para a detecção de infecções pela *F. hepatica* (Fraga, et al. 2010).

Aproveitando esta capacidade, a proteína foi fundida com outras proteínas recombinantes insolúveis e/ou difíceis de expressar, revelando-se como uma ferramenta efectiva na melhoria da expressão e solubilidade proteica em *E. coli*. Assim se formou o novo sistema de fusão (*tag* Fh8) que pela conjugação do seu pequeno peso molecular com a capacidade de melhorar a solubilidade proteica, se torna numa óptima opção para a produção de proteínas recombinantes se comparado com os outros *fusion tags* existentes actualmente (Costa, et al. 2013a).

Recentemente o *tag* Fh8 foi estudado como *tag* de purificação (Costa, et al. 2013b) revelando-se uma alternativa aos sistemas de purificação existentes actualmente. Através da

interacção da proteína Fh8 com resinas hidrofóbicas, este *tag* permite a purificação de proteínas recombinantes por Cromatografia de Interação Hidrofóbica (HIC).

Deste modo, concentra-se num único *tag* a capacidade solubilizadora, a capacidade de purificação que, aliadas ao seu pequeno peso molecular (que não perturba a actividade biológica das proteínas de interesse), resultam na simplicidade e economia de todo o processo de produção de proteínas recombinantes (Costa, et al. 2013b).

1.6. Purificação de proteínas recombinantes

Associado à produção de proteínas recombinantes são comumente aplicados processos de purificação que permitem obter proteínas com elevados níveis de pureza. É possível purificar as proteínas recombinantes recorrendo a técnicas de cromatografia que as separam de acordo com as diferenças das suas propriedades específicas (Figura 12).

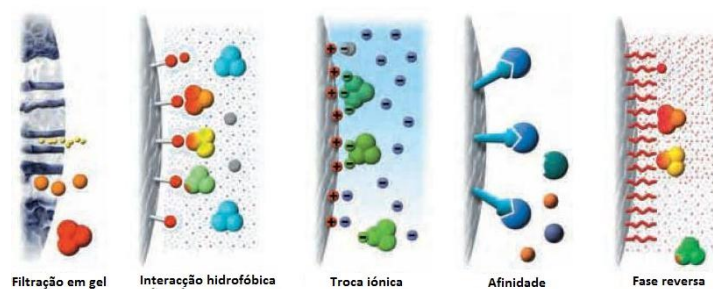


Figura 12 – Princípios da separação na purificação por cromatografia (adaptado de (GE Healthcare Bio-Sciences 2009)).

A cromatografia consiste num conjunto de técnicas através das quais a passagem de solutos por colunas pode ser selectivamente restringida e cujas interacções com a matriz determinam a distribuição de solutos entre as fases móvel e estacionária. As técnicas cromatográficas apresentam a vantagem de se desenrolarem em ambientes suaves, como é exemplo as temperaturas baixas e a ausência de tensões de corte, para produtos com actividade biológica. Os métodos cromatográficos mais utilizados são as cromatografias de afinidade, de troca iónica, de interacção hidrofóbica, de filtração em gel e de fase reversa. A cromatografia de afinidade apresenta muitas vantagens, sendo o mais indicado na purificação de proteínas recombinantes (Rodrigues 2001).

A **Cromatografia de Afinidade** ou *Affinity Chromatography* (AC) é o método mais eficaz de purificar proteínas recombinantes e permite que, quando se usam *fusion tags* ligados a uma

proteína de interesse, a proteína seja capturada com base no reconhecimento do *tag*. Assim, diversas proteínas podem ser purificadas pela mesma técnica de afinidade se todas elas possuírem o mesmo *tag*. Desta forma, os *tags* também permitem a utilização de protocolos comuns para diferentes proteínas recombinantes. As *fusion proteins* são, deste modo, simples e convenientes para se trabalhar e geralmente só requerem uma única etapa de purificação através do uso de uma coluna de cromatografia. De facto, apenas quando se pretende proteínas com altos teores de pureza se torna necessário uma purificação em várias etapas (GE Healthcare Bio-Sciences 2009).

O mecanismo de funcionamento da cromatografia de afinidade baseia-se na separação das proteínas com base numa interacção reversível entre a proteína (ou grupo de proteínas) e o ligando. Esta técnica oferece uma elevada selectividade, alta resolução e, geralmente, elevada capacidade para as proteínas de interesse, permitindo o uso de elevadas velocidades de fase móvel em pequenos volumes de matriz e, por conseguinte, purificações rápidas. A proteína de interesse é, assim, especificamente e reversivelmente ligada por uma substância complementarmente ligada (ligando). Já o material não ligado é removido por lavagem e a proteína alvo é recuperada pela mudança de condições (alteração de pH, força iónica ou polaridade) que favorecem a dessorção. As amostras são concentradas durante a ligação e a proteína é recolhida na purificação sob a forma concentrada (Amersham Pharmacia Biotech 2000).

A **Cromatografia de Troca Iónica** ou *Ion Exchange* (IEX) permite a separação de proteínas com base nas diferenças na sua carga e possibilita a separação com alta resolução para volumes de amostra elevados. Esta separação deve-se à interacção entre grupos carregados na superfície da proteína com grupos de carga oposta immobilizados na matriz. À medida que as proteínas vão sendo carregadas na coluna vão-se estabelecendo essas interacções. Um parâmetro importante é a força iónica do tampão, pois permite controlar a distribuição da proteína entre a fase estacionária e a fase móvel. Este processo engloba assim duas etapas: (1) a aplicação da amostra, sendo retidos os componentes que se ligam ao suporte; (2) a eluição das proteínas por aumento da concentração de um sal ou alteração do pH (Rodrigues 2001).

Quando o pH do meio é superior ao ponto isoeléctrico da proteína (pI), esta apresenta carga negativa e liga-se a um permutador de aniões que se encontra positivamente carregado. O contrário verifica-se para um pH inferior ao ponto isoeléctrico da proteína, onde a proteína se liga

a um permutador de catiões. As proteínas purificadas são então concentradas durante a ligação e recolhidas. A cromatografia por troca iónica é tipicamente usada para ligar a molécula alvo, mas também pode ser usado para ligar as impurezas, se necessário (Amersham Pharmacia Biotech 2000).

A **Cromatografia de Interação Hidrofóbica** ou *Hydrophobic Interaction Chromatography* (HIC) é usada para separar proteínas com base nas diferentes hidrofobicidades das proteínas, ou seja, baseia-se na interacção reversível entre a proteína e a superfície hidrofóbica de um meio cromatográfico (Amersham Pharmacia Biotech 2000).

A maioria das proteínas contém alguns aminoácidos hidrofóbicos (alanina, fenilalanina, valina, triptofano, leucina, isoleucina e metionina), que estão geralmente “enterrados” na estrutura da molécula proteica como resultado do seu ambiente aquoso envolvente. A diversidade de condições de eluição possibilita a resolução de misturas complexas de proteínas, difíceis de purificar por outras técnicas cromatográficas. A presença de contaminantes hidrofóbicos na amostra diminui a eficiência do processo. As proteínas alvo são concentradas durante a ligação e recolhidas na forma purificada (Rodrigues 2001).

A **Cromatografia de Filtração em Gel** ou *Gel Filtration* (GF) *Chromatography* separa as proteínas pela diferença de tamanho molecular. É ideal para a etapa final na purificação quando os volumes de amostra são reduzidos (volume de amostra influencia significativamente a velocidade e a resolução da filtração em gel). As condições do tampão são variadas para se adequarem ao tipo de amostra, aos requisitos para a purificação, à análise ou à etapa de armazenamento, uma vez que a composição do tampão não afecta directamente a resolução. As proteínas são recolhidas na forma purificada no tampão escolhido (Amersham Pharmacia Biotech 2000).

A **Cromatografia de Fase Reversa** ou *Reversed Phase Chromatography* (RPC) pode considerar-se como uma extensão da cromatografia de troca iónica. Esta técnica separa as proteínas e os péptidos com diferentes hidrofobicidades com base na sua interacção reversível com a superfície hidrofóbica de um meio cromatográfico. As amostras ligam-se à medida que são carregadas numa coluna. Devido à natureza das matrizes de fase reversa, a ligação é geralmente muito forte e requer o uso de solventes orgânicos e outros aditivos (agentes de emparelhamento de iões) para a eluição. A eluição é normalmente realizada por aumento da concentração de solvente orgânico, mais comumente acetonitrilo. As amostras, que estão concentradas durante o processo de ligação e separação, são recolhidas na forma purificada.

Esta técnica não é recomendada para a purificação de proteínas se for necessário a recuperação da actividade e o retorno à estrutura terciária na sua conformação correcta, uma vez que muitas proteínas são desnaturadas na presença de solventes orgânicos (Amersham Pharmacia Biotech 2000).

1.7. Hidrolase do molinato (MolA): Descrição e propriedades da proteína de interesse

Os pesticidas têm sido utilizados ao longo das últimas décadas para melhorar os rendimentos na agricultura. No entanto, é discutível se estes atribuem benefícios evidentes à agricultura, uma vez que, estas substâncias químicas podem atingir concentrações relativamente altas no meio ambiente devido à aplicação deliberada e ao uso intensivo (Nunes e Manaia 2007).

O molinato (C₉H₁₇NOS) é um dos pesticidas mais usados em todo o mundo. Trata-se de um herbicida sistémico que é amplamente aplicado nos campos de arroz para a eliminação de ervas daninhas, nomeadamente, as estirpes *Echinochloa spp.* e *Glyceria spp.* (Nunes e Manaia 2007). Quando este herbicida, pertencente à família dos tiocarbamatos, é aplicado aos arrozais alagados, volatiliza e dissipa-se no ambiente contaminando os aquíferos. Para além disso, ocorrem também transformações (foto)químicas e microbiológicas do molinato que provocam a acumulação de metabolitos de oxidação, tais como o oxomolinato e o sulfóxido de molinato, que possuem elevados níveis de toxicidade (Duarte, et al. 2011).

De modo a reduzir os efeitos tóxicos provenientes da utilização deste herbicida e a preservar o meio ambiente, devem ser realizadas metodologias que permitam degradar o molinato e removê-lo das águas contaminadas. A biorremediação tem sido descrita como método de elevada eficiência e de baixo custo para remoção de componentes tóxicos de locais poluídos. No entanto, tal sistema de tratamento requer a utilização de organismos capazes de mineralizar o poluente ou, pelo menos, transformar a molécula em produtos degradados não-tóxicos (Nunes e Manaia 2007).

O único sistema biológico descrito até agora como sendo capaz de mineralizar o molinato e utilizar o herbicida como única fonte de carbono, de energia e de azoto é uma cultura mista de bactérias de cinco membros. Esses cinco membros foram isolados e caracterizados revelando a bactéria *Gulosibacter molinativorax* ON4[†], da família das *Microbacteriaceae*, como responsável

pela degradação inicial de molinato. Esta degradação deve-se ao facto da bactéria *Gulosibacter molinivorax* ON4^r expressar a enzima hidrolase do molinato (MoIA), de 465 aminoácidos e de massa molecular 50,9 kDa, pertencente à subfamília A das hidrolases dependentes de metais (da superfamília das amidohidrolases) (Duarte, et al. 2011).

Esta actua sobre o molinato, clivando a ligação tioéster do molinato, o que liberta etanotiol e azepano-1-carboxilato (ACA), como se pode observar na Figura 13.

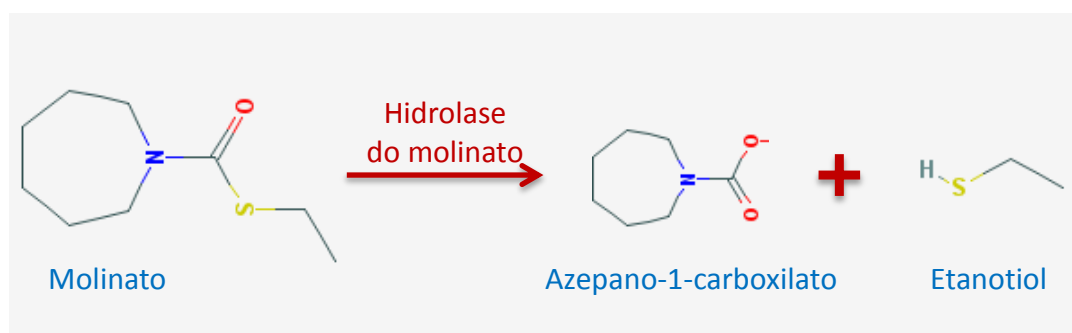


Figura 13 – Hidrólise do molinato através da acção da MoIA presente no *Gulosibacter molinivorax* ON4^r.

Estudos anteriores visaram o isolamento e produção da enzima MoIA, que resultaram em expressões maioritariamente insolúveis em *Escherichia coli* (resultados não publicados).

O presente trabalho teve por objectivo produzir a enzima MoIA, utilizando-se o sistema de fusão (Hitag® - tag Fh8) em *E. coli*, com o intuito de melhorar a expressão solúvel desta proteína.

2. Materiais e Métodos

O trabalho experimental realizado compreendeu as seguintes etapas: aplicação dos procedimentos gerais de biologia molecular na estratégia de clonagem da sequência codificante da proteína de interesse (MolA) no vector de expressão (pETMFh8); transformação do novo constructo na estirpe *E. coli* BL21(DE3); expressão, em pequena escala, da proteína recombinante e optimização das condições de indução e crescimento celular; aplicação dos procedimentos gerais de análise da expressão proteica recorrendo à técnica de SDS-Page e, por último, purificação e análise qualitativa das fracções solúveis da proteína recombinante pelo sistema *His-tag*.

As técnicas e procedimentos desenvolvidos nestas etapas encontram-se descritos nos subcapítulos seguintes.

2.1. Esterilização de material, soluções e meios de cultura

O material, soluções e meios de cultura resistentes à temperatura foram esterilizados por autoclavagem a 121°C e 1 atm, durante 20 minutos. As soluções termolábeis foram esterilizadas por filtração com filtros de celulose estéreis de 0,2 µm.

2.2. Produtos químicos, soluções e reagentes

Todas as soluções, reagentes e produtos químicos utilizados neste trabalho encontram-se na Tabela Al. 1 e Tabela Al. 2 (Anexo I, páginas 82 e 84), respectivamente.

2.3. Microrganismos e plasmídeos

Para a produção da proteína recombinante foram utilizadas duas estirpes de *E. coli*. A estirpe *E. coli* NZY5α¹ (Nzytech), que apresenta propriedades semelhantes às do tipo DH5α, foi utilizada para a manipulação de DNA e obtenção dos vectores recombinantes necessários para a expressão da proteína alvo deste estudo. Designadamente, esta estirpe foi utilizada para a clonagem dos plasmídeos recombinantes pGEM-T Easy¹ (Promega) e o pETMFh8¹. Para a expressão proteica foi utilizada a estirpe *E. coli* BL21(DE3)¹ (Nzytech).

¹ É possível obter uma descrição mais detalhada deste produto na Tabela Al. 2 (Anexo I)

A Tabela 4 apresenta o genótipo de cada estirpe utilizada.

Tabela 4: Genótipo das estirpes de *E. coli* utilizadas

Estirpe	Genótipo
<i>Escherichia coli</i> NZY5 α	fhuA2 Δ (argF-lacZ)U169, phoA, glnV44, Φ 80, Δ (lacZ)M15, gyrA96, recA1, relA1, endA1, thi-1, hsdR17
<i>Escherichia coli</i> BL21(DE3)	Deriva da estirpe <i>E. coli</i> B, F ⁻ , ompT, gal, dcm, lon, hsdSB(rB-mB-), λ (DE3 [lacI, lacUV5-T7, gene 1, ind1, sam7, nin5])

Neste estudo foi usado o sistema pGEM-T Easy (Promega) para clonar os fragmentos obtidos por reacção em cadeia da polimerase (PCR). Este sistema apresenta vantagens relativamente a outros pois permite obter significativamente melhores eficiências de ligação de produtos de PCR (PPCR's), uma vez que, é comercializado contendo o plasmídeo na forma linear que apresenta, nas suas extremidades, resíduos de timinas de cadeia simples. A existência destas extremidades permite que os PPCR's com resíduos de adeninas (gerados por determinadas polimerases termo-estáveis) possam ser ligados directamente por complementaridade ao vector (através de uma ligase) e impede a recircularização do plasmídeo.

Este sistema possibilita, assim, a inserção de fragmentos de DNA numa matriz estável para posteriores manipulações necessárias (digestão com enzimas de restrição, sequenciação para verificação de *inserts*, etc.).

O mapa do vector pGEM T-Easy está representado na Figura 14.

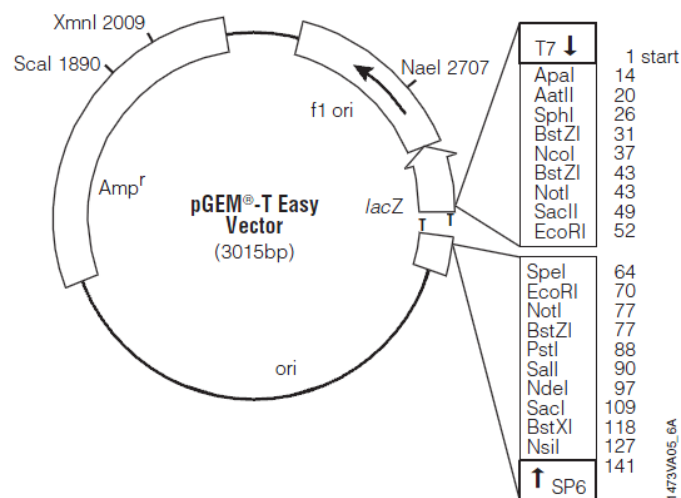


Figura 14 – Representação do mapa circular do vector pGEM T-Easy (adaptado de (Promega 2010)).

Este plasmídeo, com 3015 pb, é um vector de elevado número de cópias que contém um local de clonagem MCS, no qual se encontram, não só os vários locais de corte para diferentes enzimas de restrição, mas também o local de inserção por complementaridade com os resíduos de timina. Neste local pode ainda inserir-se o fragmento de interesse sendo possível remove-lo através da digestão com uma ou duas enzimas de restrição diferentes.

A flanquear o MCS encontra-se o promotor da RNA polimerase T7 e da RNA polimerase SP6. Possui ainda a origem de replicação do fago f1, o gene que confere resistência à ampicilina (Amp^r), o codão inicial do *lacZ* e a sequência do operão *lac*.

O MCS encontra-se no meio do gene que codifica a β -galactosidase (o gene *lacZ*) possibilitando que, quando se consegue clonar com sucesso o *insert*, a transcrição do gene seja interrompida. Assim, quando se procede ao crescimento dos clones em meio LB-agar suplementado com 5-Bromo-4-cloro-3-indolil- β -D-galactopiranosídeo (X-Gal) e IPTG, surgem colónias de cor branca. Caso a clonagem não tenha sucesso e, deste modo, a inserção não ocorra, o vector recirculariza e as colónias passam a apresentar uma cor azul. A diferença de cor ocorre devido à degradação do X-Gal pela actividade enzimática da β -galactosidase, que é induzida pela presença de IPTG no meio e que produz esse precipitado de cor azul.

O vector pETMFh8, com 5549 pb, foi utilizado para a expressão da proteína recombinante na estirpe *E. coli* BL21(DE3). A construção deste vector foi conseguida através da modificação da estrutura do vector de expressão pETM10, no qual foi inserido o *tag* de solubilidade Fh8 e o local de clivagem da protease TEV (Costa, et al. 2013a). Os vectores pETM derivam dos vectores comerciais pET (Novagene), com origem nos plasmídeos pBR322, e partilham algumas características comuns. Entre as mais importantes, destacam-se a inclusão do *tag* de purificação 6xHis-*tag* (o pETM10, especificamente, possui este *tag* tanto no N-terminal como no C-terminal), um local de reconhecimento da protease (excepto o pETM10) e o local de clonagem múltipla (MCS) que começa com um local de reconhecimento da enzima de restrição *Nco*I. A sequência de reconhecimento da *Nco*I tem um codão ATG *in-frame* que pode ser usado para a expressão funcional da proteína de interesse minimizando, assim, o número de aminoácidos não nativos na região N-terminal (Dümmler, Lawrence e Marco 2005).

Assim, o vector pETMFh8 possui, não só as características do vector pETM10 (promotor T7, origem de replicação do fago f1, origem de replicação ColE1, gene que confere resistência à

Canamicina (kan^R), sequência do operão *lac*, entre outras), mas também os mecanismos necessários para a expressão de proteínas recombinantes utilizando o *tag* de solubilização Fh8.

O mapa do vector pETMFh8 está representado na Figura 15.

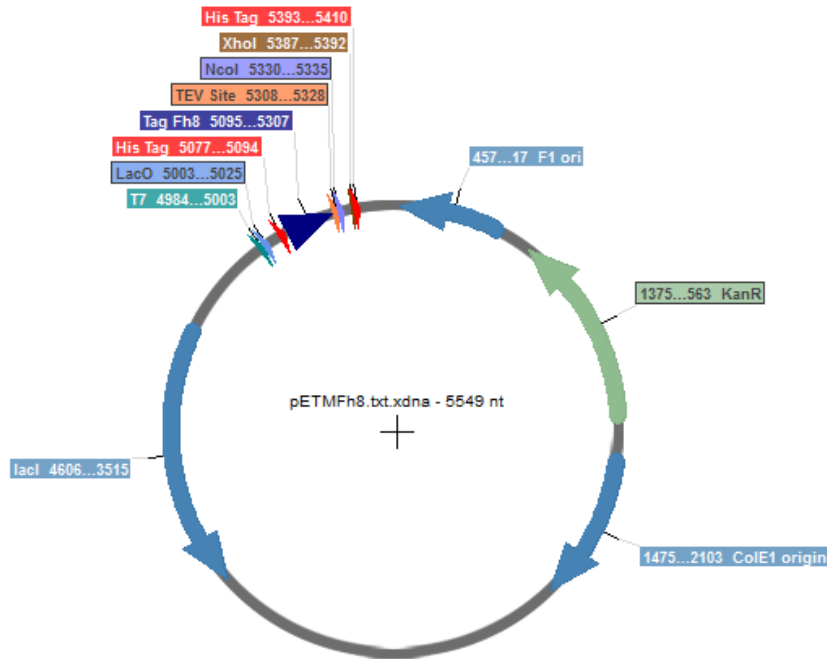


Figura 15 - Representação do mapa circular do vector pETMFh8.

2.4. Procedimentos gerais de biologia molecular

De modo a obter a proteína recombinante foram utilizadas várias metodologias de DNA recombinante, tais como, extração e purificação de DNA em gel de agarose, PCR, digestão de DNA com enzimas de restrição, ligação de DNA, transformação em células competentes, entre outros. As técnicas e os procedimentos utilizados durante a execução do trabalho experimental encontram-se descritos de seguida.

2.4.1. PCR - Reacção em cadeia da polimerase

A reacção em cadeia da polimerase é uma técnica de biologia molecular que permite replicação *in vitro* do DNA de forma extremamente rápida. Durante o PCR são usadas temperaturas elevadas de forma a separar as moléculas de DNA em duas cadeias simples. Isto possibilita a ligação de oligonucleótidos iniciadores (*primers*) que são complementares das sequências que flanqueiam o fragmento de DNA a amplificar, nos seus terminais 3', de modo a

permitir a actuação da DNA polimerase durante a síntese da cadeia complementar e usando como molde cada uma das duas cadeias simples constituintes do DNA a amplificar. Este processo que tem lugar num termociclador envolve três passos, que em conjunto se designam como um ciclo: desnaturação, emparelhamento e polimerização. Este equipamento automaticamente controla e alterna as temperaturas durante períodos programados de tempo para o número apropriado de ciclos de PCR (geralmente entre 25 a 40 ciclos), sendo possível aumentar, em cada ciclo, duas vezes a concentração de DNA pré-existente. Assim, a utilização desta técnica permite que quantidades mínimas de material genético possam ser amplificadas milhões de vezes em pouco tempo.

Neste estudo, a técnica de PCR foi utilizada para amplificar o DNA de interesse e promover a mutação das suas extremidades. Para tal, foram inseridos locais de restrição específicos de duas enzimas de restrição compatíveis com o vector pETMFh8 (*NcoI* e *XhoI*). De igual modo, foram realizadas reacções de PCR para verificação do insert no vector pGEM-T Easy, extraído dos transformantes *E. coli* NZY5 α .

As reacções de amplificação foram realizadas no termociclador *Little Genious* da marca BIOER utilizando-se a mistura reaccional *MyTaq™ Red Mix*¹ (Bioline) de acordo com as instruções do fabricante (reacções para um volume final de 25 μ L contendo 1 μ L de amostra (DNA *template*), 25 μ L da mistura *MyTaq™ Red Mix*, 1 μ L de cada um dos *primers* (*forward* e *reverse*) e água ultrapura até completar o volume final). Em algumas reacções realizou-se um controlo negativo substituindo o volume que corresponderia ao DNA *template* por água ultrapura.

Foi utilizado o seguinte programa de PCR: passo de pré-desnaturação a 94°C durante 5 minutos; 40 ciclos de amplificação compreendendo, cada um, a etapa de desnaturação a 94°C durante 20 segundos, a etapa de emparelhamento a 48°C durante 20 segundos e a etapa de polimerização a 72°C durante 1 minuto; passo adicional de polimerização a 72°C durante 10 minutos e os produtos de PCR foram visualizados em gel de agarose (técnica descrita na secção 2.4.3).

2.4.2. Digestão com enzimas de restrição

Neste estudo realizaram-se digestões de DNA plasmídico com enzimas de restrição de modo a verificar se as clonagens realizadas ao longo deste estudo foram bem sucedidas. Esta

¹ É possível obter uma descrição mais detalhada deste reagente na Tabela A1. 2 (Anexo I)

técnica foi, igualmente, utilizada nas diferentes abordagens na etapa da clonagem (descritas na secção 3.1).

As digestões foram efectuadas a 37°C e de acordo com as instruções do fabricante sendo utilizadas as seguintes soluções e reagentes: enzima(s) de restrição¹, tampão(ões) da(s) enzima(s), DNA e água ultrapura (se necessário). A quantidade de enzima utilizada, bem como a duração de cada reacção variou de acordo com a quantidade de DNA presente e a estratégia de clonagem pretendida (ver secção 3.1.1). O resultado das digestões foi visualizado em gel de agarose.

2.4.3. Electroforese de DNA em gel de agarose

A electroforese em gel de agarose é a forma mais eficaz de separar fragmentos de DNA de uma gama de tamanhos tão variada como 100 pb a 25kb (Lee, et al. 2012). Deste modo, é amplamente usada em técnicas de biologia molecular para separar e estimar o tamanho de fragmentos de ácidos nucleicos. Através da acção de um campo eléctrico é possível fazer a comparação da distância percorrida pelos fragmentos de interesse com a percorrida por fragmentos de peso molecular conhecido (padrões de peso molecular), permitindo estimar o peso molecular de cada fragmento da amostra a analisar. Assim, quando sujeitos a um campo eléctrico, os ácidos nucleicos migram em direcção ao pólo positivo, uma vez que, a pH neutro apresentam carga negativa. Já a matriz de agarose funciona como uma rede cujos poros deixam passar mais facilmente as moléculas mais pequenas, que vão migrar mais do que as de maiores dimensões. Nas preparações com plasmídeos não digeridos, é possível encontrar pelo menos duas formas distintas de DNA, são elas: a forma circular e a forma superenrolada. Sendo que a forma superenrolada migra mais rapidamente do que a circular.

Durante a execução do trabalho experimental realizaram-se electroforeses em gel de agarose¹ para visualizar os PPCR's, o resultado das digestões com enzimas de restrição, bem como, para avaliar a eficácia das extracções e purificações de fragmentos de DNA.

Em todas as análises foi utilizado o marcador de DNA *NZYDNA Ladder III*¹ (Nzytech) e o DNA foi corado com brometo de etídio², um agente mutagénico que intercala nas cadeias de DNA e emite uma fluorescência alaranjada quando exposto a raios UV. A percentagem de agarose presente nos géis (p/v – agarose dissolvida em TAE¹) variou entre 1,2 a 2,4 % de acordo

¹ É possível obter uma descrição mais detalhada deste reagente na Tabela AI. 2 (Anexo I)

² A composição desta solução encontra-se na Tabela AI. 1 (Anexo I)

com o tamanho dos fragmentos de DNA que se pretendia observar (uma maior quantidade de agarose dissolvida permite melhores separações para fragmentos menores e para pesos semelhantes, uma menor quantidade de agarose dissolvida possibilita melhores separações para fragmentos de maior peso molecular).

Sempre que se pretendeu extrair e purificar o DNA proveniente das análises em gel de agarose realizaram-se electroforese em gel de agarose *low melting*¹ a 1,4 % (p/v – agarose dissolvida em TAE) e a banda a ser purificada foi recortada do gel.

2.4.4. Extracção e purificação de DNA em gel de agarose

A extracção e purificação do DNA proveniente das análises em gel de agarose foi efectuada com o kit de purificação *Illustra™ GFX™ PCR DNA and Gel Band Purification* (GE Healthcare), segundo as instruções do fabricante. No final de todas as extracções foram realizadas electroforeses em gel de agarose para confirmação da correcta extracção (de acordo com o procedimento descrito na secção 2.4.3 deste capítulo).

2.4.5. Ligação de DNA

As diferentes abordagens na etapa da clonagem (descritas na secção 3.1) contemplam a ligação de fragmentos de DNA ou a ligação aos vectores utilizados neste estudo.

Assim, realizaram-se ligações utilizando-se, para além do fragmento de interesse e do plasmídeo, duas diferentes ligases (T4 DNA ligase¹ e *Speedy* ligase¹), os respectivos tampões e água ultrapura. As ligações aos plasmídeos pGEM-T Easy e pETMFh8 foram realizadas recorrendo a qualquer uma das ligases, uma vez que, a diferença entre elas é o tempo de reacção. Ambas as ligações foram realizadas à temperatura ambiente e para um volume final de 10 µL seguindo as instruções do fabricante em relação às restantes proporções de reagentes utilizadas. A quantidade do *insert* e DNA plasmídico variou de acordo com a quantidade disponível e da estratégia de clonagem pretendida (ver secção 3.1.1).

¹ É possível obter uma descrição mais detalhada deste reagente na Tabela A1. 2 (Anexo I)

2.4.6. Transformação em células competentes *E. coli* NZY5 α e BL21(DE3)

Uma vez realizadas as ligações de DNA, procedeu-se à transformação em células competentes *E. coli* NZY5 α ou *E. coli* BL21(DE3). Todas as transformações foram realizadas segundo as normas do fabricante à excepção de pequenas alterações descritas de seguida.

Nas transformações em células *E. coli* NZY5 α foi adicionado 10 μ L da reacção de ligação (descrita na secção 2.4.5) a 50 μ L de células competentes enquanto que nas transformações em células *E. coli* BL21(DE3) foi adicionado 10 μ L da reacção de ligação a 100 μ L de células competentes. Em ambos os casos substituiu-se o meio aconselhado (SOC) por meio LB², sendo o volume de 450 μ L adicionado à mistura após choque térmico. Plaqueou-se em triplicado em placas LB agar² com o antibiótico apropriado e, no caso das células transformadas com o vector recombinante pGEM-T Easy, suplementadas com 425 μ L de X-Gal², seguindo as instruções do fabricante.

2.4.7. Crescimento de transformantes *E. coli* NZY5 α e BL21(DE3)

As placas preparadas (secção 2.4.6) foram incubadas *overnight* a 37°C procedendo-se à selecção e crescimento das colónias transformantes (no máximo 6 colónias para cada construção) em Falcons com LB e o antibiótico apropriado.

2.4.8. Extracção de DNA plasmídico de culturas de *E. coli*

O isolamento do DNA plasmídico foi feito através da utilização do kit *Wizard Plus SV Minipreps DNA Purification System* da marca Promega segundo as instruções do fabricante.

2.4.9. Sequenciação de construções

A construção obtida no vector pETMFh8 foi sequenciada na Eurofins MWG Operon (Alemanha).

Para a construção confirmada por sequenciação foram realizados dois stocks distintos: um stock a -20°C do DNA plasmídico enviado para sequenciar e outro de cultura de onde foi extraído o DNA enviado para sequenciar (com 37,5% de glicerol estéril), crioconservado a -80°C.

² A composição desta solução encontra-se na Tabela AI. 1 (Anexo I)

2.5. Procedimentos gerais de análise da expressão proteica em géis SDS-Page Tris-Tricina

A electroforese é uma técnica que tem por objectivo a separação de macromoléculas por aplicação de um campo eléctrico. O desenvolvimento da electroforese em gel de poliacrilamida ou PAGE (*PolyAcrylamide Gel Electrophoresis*) permitiu separar proteínas. Em condições desnaturantes, é possível essa separação através da adição do detergente SDS (*Sodium Dodecyl Sulfate*) às amostras a separar, pelo que a técnica é habitualmente conhecida por SDS-PAGE.

O SDS é um detergente fortemente negativo que é usado para desnaturar proteínas com várias subunidades e para cobrir as moléculas da proteína com cargas negativas. A sua aplicação permite “mascarar” a carga intrínseca à proteína, que é variável em função dos aminoácidos que a constituem e do pH da solução que a contém, tornando a razão carga/massa constante. Tal como as moléculas de DNA, que na electroforese de ácidos nucleicos migram no sentido do eléctrodo positivo devido à carga global negativa (conferida pelo grupo fosfato), as proteínas (agora carregadas negativamente pelo detergente SDS) irão também ser separadas em função do seu tamanho.

O sistema de electroforese desnaturante SDS-Page Tris-Tricina (baseado no sistema dos tampões Tris-Tricina) é frequentemente usado para separar proteínas entre 1-100 kDa sendo o sistema de electroforese ideal para resolução de proteínas menores que 30 kDa (Schägger 2006).

Os géis utilizados neste sistema são essencialmente constituídos por uma mistura de dois polímeros: acrilamida e bis-acrilamida². A sua composição encontra-se na Tabela AIII. 1 (Anexo III, página 87). A mistura destes dois polímeros permite obter uma matriz porosa caracterizada por uma rede com uma malha mais ou menos apertada. Deste modo, quanto maior a concentração de acrilamida, menores serão os poros da matriz. Os géis de poliacrilamida possuem ainda duas fases: a de concentração (*stacking*) e a de resolução ou de corrida (*resolving*). A primeira fase possibilita que as amostras proteicas colocadas em cada poço sejam concentradas antes da corrida no gel. Deste modo, a resolução das proteínas aumenta significativamente. A reacção de polimerização é regulada pelos reagentes persulfato de amónio (PSA)² e tetrametilenodiamina (TEMED).

² A composição desta solução encontra-se na Tabela AI. 1 (Anexo I)

Neste estudo foram realizadas electroforeses em gel de poliacrilamida SDS Tris-Tricina para visualização e análise da proteína produzida e purificada. Os extractos proteicos e as fracções recolhidas durante as purificações foram tratados com base no sistema Laemmli-SDS-Page (Laemmli 1970) e Tris-Tricina (Schägger 2006). Para isso, as amostras proteicas foram tratadas com tampão de amostra Laemmli concentrado 4x¹ (na proporção de 3:1 para um volume final de 40 µL) e com 2-Mercaptoetanol¹ (um agente redutor que destrói as pontes dissulfureto que são responsáveis pela estrutura nativa da proteína) sofrendo um choque térmico a 95°C, durante 20 minutos, antes de serem carregadas no gel, 10 µL por poço.



As electroforeses decorreram no sistema de electroforese *Mini-PROTEAN Tetra Cell* (Bio-Rad) de acordo com as normas do fabricante. Este sistema permite a corrida simultânea das amostras em quatro géis diferentes, sendo constituído por dois reservatórios, um superior (em contacto directo com os géis) onde se colocava o tampão do cátodo², e um inferior, no qual era colocado o tampão de ânodo².

Após corrida, os géis foram corados com uma solução contendo *Coomassie Blue*², um corante que permite visualizar quantidades de proteínas acima de 0,1 µg. De seguida, descorou-se recorrendo a uma solução contendo solventes orgânicos e ácido acético². A solução descorante foi descartada e substituída repetidamente até as proteínas do gel serem reveladas. De seguida, o gel foi fotografado e procedeu-se à análise de resultados.

Em todas as electroforeses em gel de poliacrilamida SDS Tris-Tricina utilizou-se o marcador proteico *NZYS Colour Protein Marker II*¹. As amostras que não foram carregadas imediatamente foram conservadas a -20°C.

2.6. Ensaios de indução e purificação da proteína recombinante Fh8MoIA

De modo a avaliar as propriedades da proteína recombinante, tais como, a toxicidade para a célula e a sua expressão e solubilidade, realizaram-se dois ensaios distintos em células competentes *E. coli* BL21(DE3):

-  Ensaio preliminar do crescimento celular e indução proteica;
-  Ensaio de optimização das condições de densidade óptica, concentração de indutor e de temperatura.

¹ É possível obter uma descrição mais detalhada deste reagente na Tabela AI. 2 (Anexo I)

² A composição desta solução encontra-se na Tabela AI. 1 (Anexo I)

Realizou-se ainda uma mini-purificação de afinidade por IMAC (*Immobilized metal-ion affinity chromatography*) da proteína Fh8MolA.

Todos os manuseamentos celulares foram realizados em gelo ao longo da experiência.

2.6.1. Ensaio preliminar de crescimento celular e indução proteica

Neste estudo realizou-se um ensaio preliminar de crescimento celular em *E. coli* BL21(DE3) onde se procedeu à indução/expressão da proteína recombinante de interesse (Fh8MolA). Para tal, foi preparada uma pré-cultura na qual se inoculou 10 mL de meio LB/Canamicina² (em Falcons de 50 mL) com o stock a -20°C da cultura celular *E. coli* BL21(DE3) contendo o constructo pETMFh8 MolA confirmado por sequenciação. A pré-cultura foi incubada *overnight* à temperatura de 37°C com agitação a 144 rpm.

Em dois frascos (de capacidade 250 mL) inoculou-se em cada um 50 mL de meio de cultura LB/Canamicina com 1/10 de pré-cultura (5 mL) e procedeu-se ao crescimento à temperatura de 37°C com agitação a 200 rpm durante 5 horas. Em ambos os frascos acompanhou-se o crescimento, de hora em hora, a partir da densidade óptica (OD₆₀₀). Após as duas primeiras horas, induziu-se uma das culturas com IPTG² à concentração de 1mM, quando a OD₆₀₀ se situava próxima de 1,2. A indução durou 3 horas tendo-se acompanhado o crescimento bacteriano em ambos os frascos, não induzido (CN) e induzido (I).

No final das 5 horas de crescimento foram recolhidas duas alíquotas de 1 mL, de cada um dos frascos, sendo posteriormente centrifugadas à temperatura refrigerada de 4°C e à velocidade de 14400 rpm durante 10 minutos. Descartou-se o sobrenadante obtendo-se os pellets das células. A restante cultura de células de cada frasco foi processada do mesmo modo que as alíquotas recolhidas mas com centrifugação a 4400 rpm, durante 12 minutos, sendo posteriormente descartados os sobrenadantes e congelados os pellets a -20°C (o congelamento dos pellets promove a lise celular).

2.6.1.1. EXTRACÇÃO PROTEICA EM CONDIÇÕES NATIVAS

Os quatro pellets obtidos a partir das alíquotas recolhidas foram descongelados e ressuspendidos em 200 µL de tampão de lise² para purificação em condições nativas. Efectuou-se a lise das células através de rotura mecânica por sonicação recorrendo ao sonicador modelo

² A composição desta solução encontra-se na Tabela A1. 1 (Anexo I)

VCX130 (130W) Vibra-Cell da marca Sonics®. As condições de sonicação foram as seguintes: 4 ciclos de sonicação de 30 segundos com pausas de 10 segundos a uma amplitude de 40%.

Após sonicação, seleccionaram-se dois tubos (um relativo ao ensaio em que ocorreu a indução e outro em que esta não ocorreu), correspondentes aos lisados totais (LT), para posterior análise em SDS-Page. Os restantes dois tubos foram novamente centrifugados nas mesmas condições que a centrifugação anterior. Posteriormente, recolheram-se os sobrenadantes, ou fracções solúveis (FS) para novos tubos e ressuspendeu-se novamente os precipitados obtidos com 200 µL do mesmo tampão de lise, obtendo-se as fracções insolúveis (FI).




2.6.1.2. DETERMINAÇÃO DE SOLUBILIDADE – ANÁLISE DE SDS-PAGE

As determinações de expressão e de solubilidade foram conseguidas através da análise aos géis de poliacrilamida de SDS-Page Tris-Tricina (técnica descrita na secção 2.5)

Foram analisadas em gel as seguintes amostras: Lisados Totais, Fracções Solúveis e Fracções Insolúveis correspondentes aos ensaios Induzido e Não-Induzido.

2.6.2. Testes de optimização em pequena escala do crescimento celular e indução proteica

Para o crescimento celular e indução proteica da proteína recombinante Fh8MolA em *E. coli* BL21(DE3), foram testadas 18 condições diferentes:

-  OD₆₀₀: 0,4; 0,6-0,8 e >1,0;
-  Concentrações de IPTG: 0,1mM; 0,5mM e 1,0mM;
-  Temperaturas de indução: 37°C e 18°C.

Para tal, preparou-se a pré-cultura nas mesmas condições que a que havia sido preparada para o ensaio preliminar. Em 20 tubos de 50 mL (correspondentes aos 18 ensaios e a dois controlos negativos) inoculou-se em cada um 10 mL de meio de cultura LB/Canamicina com 1/50 de pré-cultura (0,2 mL) procedendo-se ao crescimento a 37°C e com agitação a 200 rpm. Tal como anteriormente, durante o crescimento foi acompanhada a OD₆₀₀ de hora em hora. Uma vez atingidas as OD₆₀₀ pré-estabelecidas, efectuaram-se as induções a diferentes concentrações de IPTG procedendo-se à mudança de temperatura (nos ensaios a 18°C).

As induções a 37°C tiveram a duração de 3 horas enquanto o crescimento induzido a 18°C ocorreu *overnight*³.

No final do crescimento seguiu-se o mesmo procedimento utilizado anteriormente para obter os lisados totais e as fracções solúveis e insolúveis (descrito na secção 2.6.1.1), bem como, para o tratamento das amostras para carregar em gel SDS-Page Tris-Tricina (descrito na secção 2.6.1.2).

Foram analisadas em gel as seguintes amostras: Lisados Totais, Fracções Solúveis e Fracções Insolúveis, correspondentes a cada concentração de IPTG e a cada momento de indução.

2.6.3. Purificação da proteína recombinante Fh8MolA: Mini-Purificação por IMAC em condições nativas

Realizou-se uma mini-purificação de afinidade por IMAC da junção das fracções solúveis obtidas dos ensaios a diferentes concentrações de IPTG correspondentes ao momento em que foi feita a indução a OD₆₀₀: 0,6-0,8 (secção 2.6.2) à temperatura de 18°C. Denominou-se de mini-purificação, uma vez que, as purificações ocorrem a pequena escala (em tubos *Eppendorf*) e se purificam apenas pequenos volumes (de modo a concentrar a proteína).

Os pellets congelados obtidos dos três ensaios a diferentes concentrações de IPTG foram descongelados em gelo e ressuspendidos, cada um, em 100 µL de tampão de lise. Juntaram-se os seus volumes, obtendo-se um volume associado de cerca de 300 µL e seguiu-se o mesmo procedimento utilizado anteriormente para obter apenas as fracções solúveis (descrito na secção 2.6.1.1), descartando-se as insolúveis.

A purificação ocorreu por gravidade⁴, numa micro-coluna improvisada (de 500 µL de capacidade) com 100 µL de resina NI-NTA 50% (*Ni Sepharose™ High Performance*¹) imobilizada em esferas de agarose e previamente equilibrada com tampão de lise. Esta matriz é muito usada nas purificações de proteínas recombinantes que possuam o *tag* de purificação *His-tag* devido à elevada afinidade e selectividade que permite a ligação dos resíduos de histidinas aos iões de níquel imobilizados desta matriz.

³ Foram inicialmente testadas induções à temperatura de 18 °C durante 3 horas, não se obtendo resultados satisfatórios.

⁴ Em alguns momentos da purificação foi necessário centrifugar a coluna variando a velocidade entre 200 a 400 rpm, em períodos de 60 segundos, de modo a permitir a passagem pela coluna.

¹ É possível obter uma descrição mais detalhada deste reagente na Tabela A1. 2 (Anexo I)

As fracções solúveis foram carregadas na coluna, permitindo a ligação à matriz das proteínas recombinantes (que possuem o *His-tag*), e recolheu-se a fracção que passou pela coluna, designando-se de *Flow through* (FT). Esta fracção corresponde às proteínas que passaram através da matriz sem se ligarem.

Foram efectuadas duas lavagens da resina: a primeira (L1) com um volume de 1mL do tampão de lise e, a segunda (L2), com 1mL de tampão de lavagem². Recolheram-se as fracções das duas lavagens que correspondem às proteínas do extracto de *E. coli* que se ligaram fracamente à resina.

De seguida, procedeu-se à eluição (E) da proteína recombinante com 500 µL de tampão de eluição², tendo-se recolhido 5 fracções de 100 µL em 100 µL.

No final da purificação as fracções recolhidas foram guardadas a temperatura refrigerada de 4 °C e seguiu-se o mesmo procedimento utilizado anteriormente para o tratamento das amostras para carregar em gel SDS-Page Tris-Tricina (técnica descrita na secção 2.5)

² A composição desta solução encontra-se na Tabela A1. 1 (Anexo I)



3. Resultados e Discussão

O trabalho desenvolvido pode ser dividido em três etapas principais: estratégias de clonagem da sequência codificante da MolA no vector de expressão (pETMFh8), expressão da proteína de interesse e controlo da indução e purificação proteica. Os resultados obtidos dessas etapas serão discutidos nos subcapítulos seguintes.

3.1. Estratégias de clonagem e análise de construções

Este estudo visava a produção da proteína recombinante MolA através do uso do *tag* de fusão Fh8. Tal como referido anteriormente (secção 2.3), o vector pETMFh8 foi construído de modo a possuir a maquinaria necessária para a expressão solúvel e purificação de proteínas recombinantes.

A clonagem com sucesso neste vector implicou a realização de duas estratégias distintas:

-  Metodologia adoptada com o gene recebido
-  Metodologia adoptada com o gene optimizado e sintetizado

A primeira estratégia era o processo de clonagem ideal por implicar menores custos de projecto sendo, portanto, a estratégia escolhida para a obtenção da construção pETMFh8 MolA que se pretendia. No entanto, por uma série de razões, que se explicam na secção 3.1.1., não foi possível obter a construção desejada. Deste modo, a sequência codificante do gene MolA foi encomendado à empresa Invitrogen.

A Figura 16 esquematiza as metodologias adoptadas na etapa da clonagem.

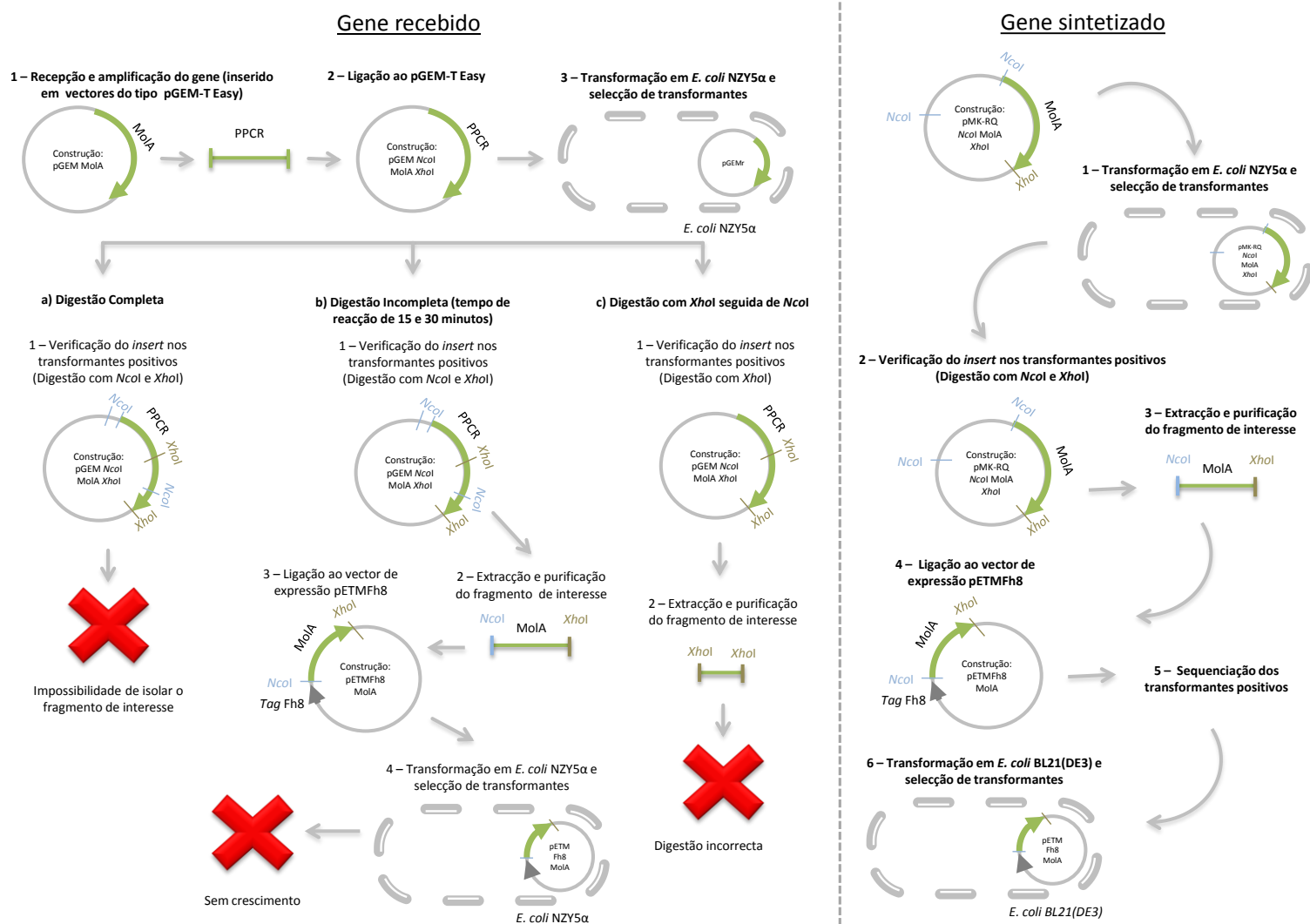


Figura 16 – Representação esquemática das estratégias de clonagem aplicadas neste estudo.

3.1.1. Metodologia adoptada com o gene recebido

A sequência codificante do gene MolA foi fornecida pela Doutora Olga Nunes, Professora Auxiliar no Departamento de Engenharia Química da Faculdade de Engenharia da Universidade do Porto.

Numa primeira fase, a sequência de nucleótidos que codifica o gene MolA foi obtida e amplificada por PCR utilizando como DNA *template* a construção já existente em pGEM-T Easy (acesso do DDBJ/EMBL/GenBank: FN985594). Para fazer a inserção do gene MolA no vector de expressão pETMfH8 foi necessário mutar as suas extremidades. Assim, utilizou-se a técnica de PCR (descrita na secção 2.4.1) para amplificar o DNA de interesse e promover a inserção dos locais de restrição específicos das duas enzimas de restrição compatíveis com o vector pETMfH8 (*Nco*I e *Xho*I), como ilustrado na Figura 17.

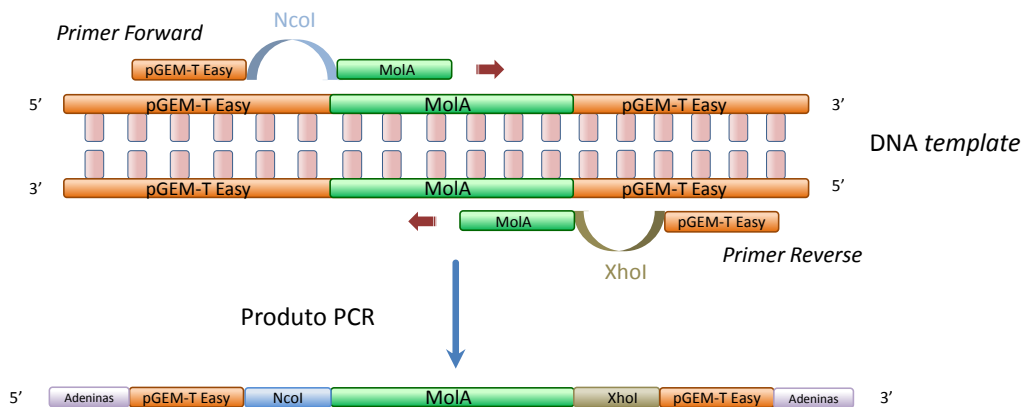


Figura 17 – Representação esquemática da reacção de PCR utilizada para amplificação do gene da MolA flanqueado nas suas extremidades pelos locais de reconhecimento para as enzimas de restrição *Nco*I e *Xho*I.

Através da figura acima é possível verificar que o *primer forward* (Fwd *Nco*I MolA, lado esquerdo) permitiu inserir o local de corte da enzima *Nco*I e o *reverse* (Rev MolA *Xho*I, lado direito) o local de corte da enzima *Xho*I. As sequências dos *primers forward* e *reverse* utilizados nas reacções de PCR encontram-se na Tabela AII. 1 (Anexo II, página 86).

No final da reacção de PCR correu-se um pequeno volume (cerca de 10 μ L) do produto de PCR num gel de agarose (técnica descrita na secção 2.4.3) e comparou-se o tamanho das bandas com o esperado (Figura 18).

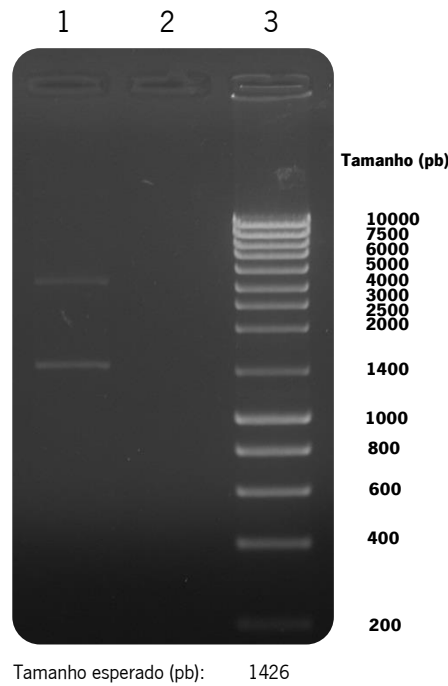


Figura 18 – Perfil em gel de agarose a 1,2% (p/v) dos produtos obtidos da amplificação por PCR.

Legenda: 1 – Produto de PCR; 2 – Controlo negativo da reacção de PCR; 3 – Marcador de DNA

A amplificação do gene permitiu a obtenção de duas bandas (poço 1): uma de cerca de 1400 pb, correspondente ao gene flanqueado nas suas extremidades pelos locais de reconhecimento para as enzimas de restrição *NcoI* e *XhoI* (1426 pb), e outra de cerca de 3000 pb correspondente ao tamanho do plasmídeo do pGEM-T Easy (3015 pb). Esta banda de maior tamanho surgiu muito provavelmente devido ao excesso de DNA molde utilizado na reacção de PCR. Não foi observada, assim, nenhuma amplificação inespecífica. No entanto, de acordo com a intensidade das bandas obtidas, a amplificação não resultou numa quantidade de DNA de interesse muito elevada.

Após a corrida, efectuou-se a extracção e purificação da banda de interesse (de acordo com a técnica descrita na secção 2.4.4).

Uma vez confirmada, em gel de agarose, a obtenção do gene flanqueado nas suas extremidades, efectuou-se a ligação ao vector pGEM-T Easy utilizando-se a ligase T4 (procedimento descrito na secção 2.4.5). Deste modo, obteve-se a construção pGEM *NcoI*/MolA *XhoI*.

Subsequentemente, procedeu-se à transformação em células competentes NZY5 α (seguindo o protocolo descrito na secção 2.4.6). Após incubação das placas *overnight*, seleccionaram-se 7 colónias brancas e confirmou-se a transformação por PCR utilizando-se o

mesmo protocolo usado anteriormente. Uma pequena quantidade dos produtos de PCR (20 µL) foi, mais uma vez, corrida num gel de agarose e o tamanho das bandas foi comparado com o obtido na reacção de PCR anterior.

Selecionou-se um dos transformantes positivos (que apresentou uma banda de tamanho próximo do fragmento de interesse obtido por PCR) e repicou-se a colónia da placa para uma cultura líquida de LB/Ampicilina. A cultura foi incubada *overnight* a 37°C tendo-se riscado, posteriormente, numa placa contendo LB/Ampicilina.

A metodologia utilizada para a obtenção do gene de interesse contemplou a digestão do novo constructo pGEM *Nco*I *Mol*A *Xho*I com as duas enzimas de restrição compatíveis com o vector pETMFh8 (*Nco*I e a *Xho*I). A Figura 19 mostra o mapa circular desta construção exibindo os locais de corte específicos para as enzimas de restrição *Nco*I e *Xho*I.

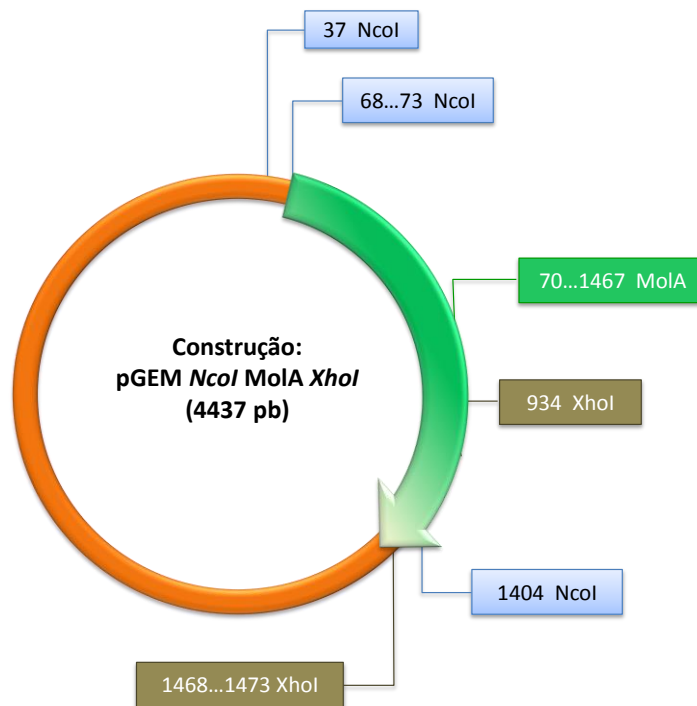


Figura 19 – Representação do mapa circular da construção pGEM *Nco*I *Mol*A *Xho*I.

Observando o mapa desta construção é possível verificar que a enzima de restrição *Xho*I corta a sequência do gene da *Mol*A no par de base 934 e que a enzima *Nco*I corta não só a sequência do gene da *Mol*A no par de base 1404, como também a própria sequência do plasmídeo no par de base 37. Assim, a digestão simultânea com as enzimas de restrição *Nco*I e *Xho*I deverá gerar 5 fragmentos de tamanhos diferentes, tal como representado na Figura 20.

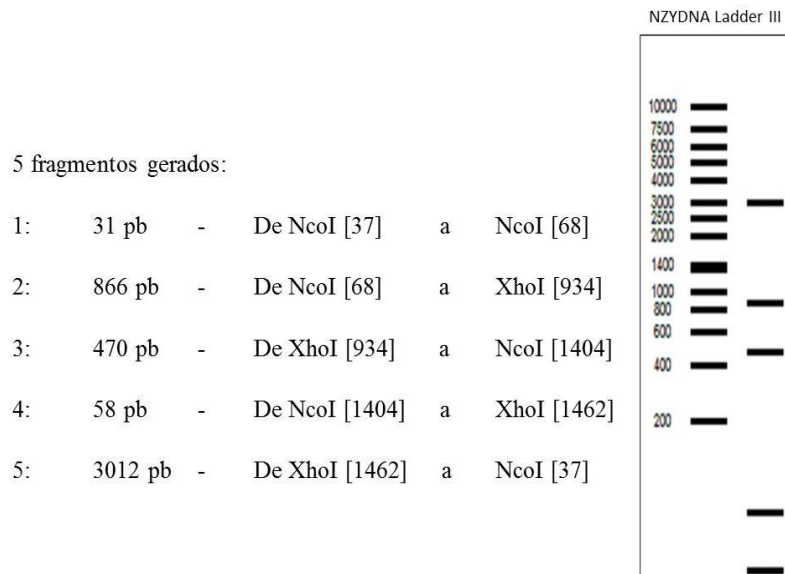


Figura 20 – Esquema representativo do tamanho dos fragmentos gerados através da digestão do constructo pGEM *NcoI* MolA *XhoI* com as enzimas de restrição *NcoI* e *XhoI*.

A obtenção da sequência codificante da MolA, flanqueada nas suas extremidades com as duas enzimas de restrição compatíveis com o vector pETMFh8 tornou-se, assim, na maior entrave à concretização deste trabalho. De modo a contornar este problema foram adoptadas as estratégias que se encontram sumariamente esquematizadas na Figura 16 e que se descrevem de seguida.

3.1.1.1. DIGESTÃO COMPLETA COM ENZIMAS DE RESTRIÇÃO *NCOI* E *XHOI*

A partir das células riscadas na placa seleccionaram-se as colónias mais isoladas (6) e repicou-se para culturas líquidas de LB/Ampicilina. A cultura foi incubada *overnight* a 37°C sendo posteriormente extraído o DNA plasmídico (segundo o protocolo referido na secção 2.4.8). Este foi digerido com as enzimas de restrição *NcoI* e *XhoI* (de acordo com o procedimento descrito na secção 2.4.2) e corrido um pequeno volume (cerca de 10 µL) num gel de agarose (segundo a técnica descrita na secção 2.4.3). A Figura 21 apresenta o perfil da digestão obtida em gel de agarose.

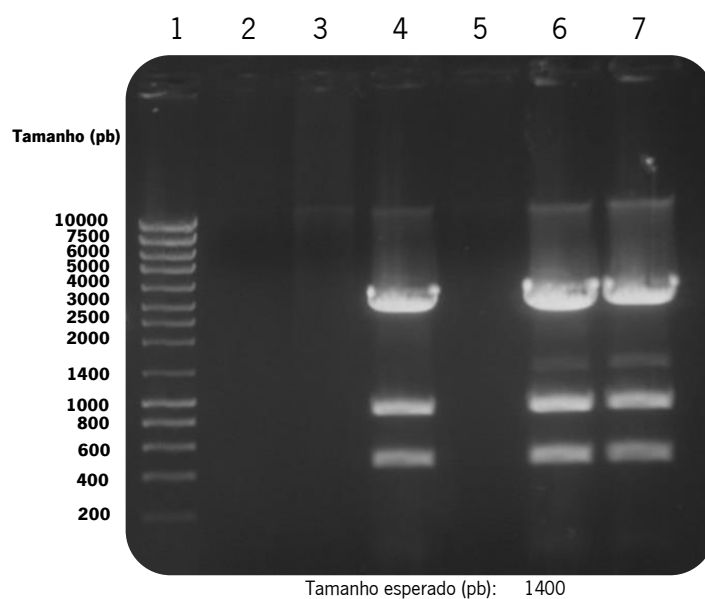


Figura 21 – Perfil em gel de agarose a 1,2% (p/v) da digestão do constructo pGEM *NcoI* MolA *XhoI* com as enzimas de restrição *NcoI* e *XhoI* em simultâneo durante 2h.

Legenda: 1 – Marcador de DNA; 2 a 7 – Digestão sobre o DNA extraído de seis transformantes pGEM *NcoI* MolA *XhoI*.

Pela Figura 21 é perceptível que apenas os clones 3 (poço 4), 5 (poço 6) e 6 (poço 7) apresentam um padrão de restrição. Tal como era esperado, a digestão resultou no aparecimento de três bandas intensas de tamanhos próximos ao previsto (as bandas de 3006 pb, 866 pb e 470 pb). No entanto, a digestão do DNA plasmídico com estas duas enzimas de restrição resultou em 2 clones (5 e 6) que permitem detectar uma banda de pouca intensidade com tamanho próximo da MolA (1400 pb). Isto sugere que possivelmente não ocorreu uma digestão completa do constructo durante o tempo da reacção.

3.1.1.2. DIGESTÃO INCOMPLETA COM ENZIMAS DE RESTRIÇÃO *NcoI* E *XhoI*

Tendo por base os resultados obtidos da digestão do constructo pGEM *NcoI* MolA *XhoI* com as enzimas de restrição *NcoI* e *XhoI*, em simultâneo durante 2h (secção 3.1.1.1), decidiu-se seguir uma nova estratégia que contemplasse a redução dos tempos da reacção da digestão. Denominou-se de digestões incompletas às reacções realizadas durante 15 ou 30 minutos de modo a permitir a obtenção do gene MolA com o tamanho pretendido.

Deste modo, realizaram-se digestões incompletas do DNA plasmídico dos clones positivos 3, 5 e 6, com as enzimas de restrição *NcoI* e *XhoI* (de acordo com o procedimento descrito na secção 2.4.2). Uma pequena quantidade do DNA plasmídico, de cada um destes três clones (20

μL), foi digerida com as enzimas de restrição *NcoI* e *XhoI* em simultâneo durante 15 minutos. De igual modo, efectuou-se outra digestão nas mesmas condições que a primeira durante 30 minutos. Os volumes de tampão e das enzimas utilizados foram os mesmos que na estratégia com a digestão completa do DNA plasmídico de modo a manter as mesmas condições. A Figura 22 apresenta o perfil obtido em gel de agarose das digestões incompletas.

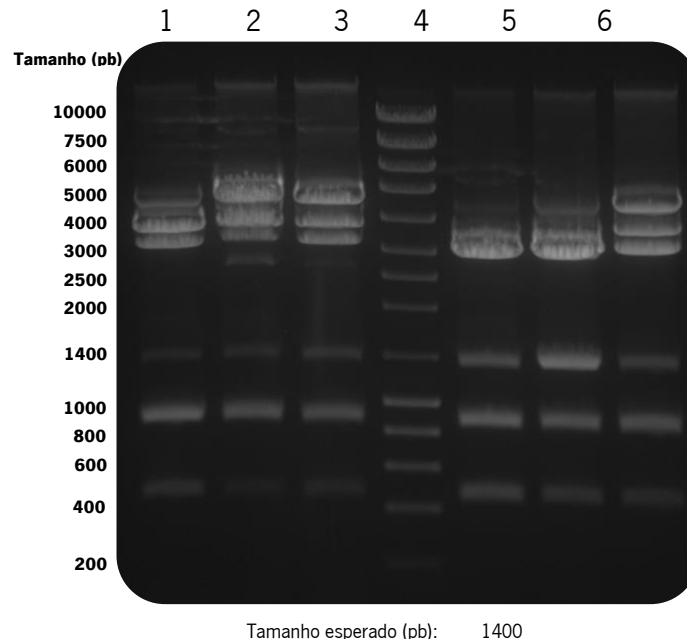


Figura 22 - Perfil em gel de agarose a 1,2% (p/v) da digestão do constructo pGEM *NcoI* MolA *XhoI* com as enzimas de restrição *NcoI* e *XhoI* em simultâneo durante 15 minutos ou 30 minutos.

Legenda: 1 a 3 – Digestão sobre o DNA extraído de três transformantes pGEM *NcoI* MolA *XhoI* durante 15 minutos; 4 – Marcador de DNA; 5 a 7 – Digestão sobre o DNA extraído de três transformantes pGEM *NcoI* MolA *XhoI* durante 30 minutos.

Através da análise da figura anterior, é possível verificar que as digestões incompletas resultaram num maior número de fragmentos. Tal facto era esperado, pois esta digestão não só poderia resultar na digestão completa com o aparecimento dos 5 fragmentos previstos (ver Figura 20) como, também, resultaria em algumas digestões incompletas. Era esse o caso que se pretendia para a obtenção do gene da MolA. Pelo perfil do gel foi possível observar o aparecimento de uma banda em todos os clones digeridos com o tamanho aproximado do pretendido (1400 pb). Deste modo, efectuou-se a extracção e purificação desta banda (de acordo com a técnica descrita na secção 2.4.4) e procedeu-se à ligação ao vector de expressão pETMFh8 utilizando-se a ligase T4 (procedimento descrito na secção 2.4.5) de modo a obter a construção pETMFh8 MolA. Consequentemente procedeu-se à transformação em células competentes NZY5 α (seguindo o protocolo descrito na secção 2.4.6).

Após incubação das placas *overnight* não se verificou o crescimento de nenhuma colónia. Deste modo, foram realizados novos ensaios de ligação ao vector pETMFh8 e, conseqüente, transformação em células competentes NZY5 α , com alterações dos volumes do *insert* e do DNA plasmídico, não resultando, porém, em nenhum crescimento.

Uma justificação para o não crescimento de colónias pode dever-se ao facto dos fragmentos purificados, de tamanho aproximado ao tamanho da MolA (cerca de 1400 pb) possam, na verdade, tratar-se de uma digestão apenas entre enzimas de restrição *Nco*I, uma presente no plasmídico pGEM T-Easy (par de base 37 pb) e outra da MolA mutada (par de base 1404 pb). Assim o fragmento visível em gel de agarose teria um tamanho de 1367 pb não sendo, portanto, simples de distinguir do fragmento pretendido. Se esta hipótese estiver correcta, as extremidades do fragmento não são complementares aos locais de inserção no plasmídeo de expressão pETMFh8, não permitindo o crescimento de transformantes positivos.

Assim, abordaram-se estratégias diferentes de modo a isolar o gene que codifica a MolA.

3.1.1.3. DIGESTÃO COM A ENZIMA DE RESTRIÇÃO *Xho*I SEGUIDA DA DIGESTÃO COM A ENZIMA DE RESTRIÇÃO *Nco*I

Uma vez que as abordagens anteriores não permitiram obter o isolamento do gene de interesse e a transformação em células, decidiu-se testar digestões completas por partes com o objectivo de se voltar a ligar os fragmentos obtidos e restituir o gene da MolA.

A abordagem consistia em seis etapas como se pode observar no esquema da Figura 23.

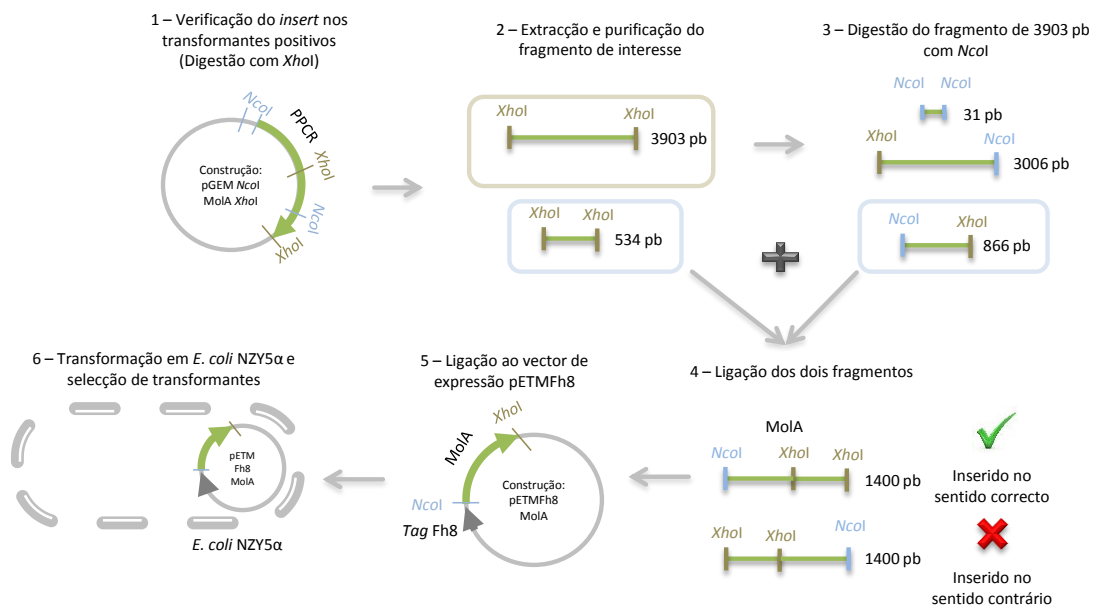


Figura 23 – Representação esquemática da estratégia de clonagem contemplando digestão por etapas.

O primeiro passo seria digerir completamente o constructo pGEM *NcoI* MolA *XhoI* apenas com a enzima de restrição *XhoI*, de modo a ser possível obter-se dois fragmentos distintos. Um de 534 pb, que contém parte da sequência codificante da MolA, e outro de 3903 pb, que abrange o plasmídeo completo do pGEM T-Easy e o restante gene da MolA. No segundo passo, as bandas visíveis em gel de agarose correspondentes a ambos, seriam purificadas e o DNA correspondente à banda maior (3903 pb) seria posteriormente digerido com a enzima de restrição *NcoI* com o objectivo de se obter em gel de agarose os fragmentos correspondentes às seguintes bandas: 31 pb (correspondente à digestão entre os locais de restrição da *NcoI* presente no plasmídeo pGEM-T Easy e o local de restrição da *NcoI* presente no início da MolA), 866 pb (correspondente ao restante gene) e 3006 pb (restante plasmídeo). A banda de interesse (866 pb) seria então purificada e proceder-se-ia à ligação dos dois fragmentos do gene da MolA (o de 866 pb com o de 534 pb), tendo-se em conta que a ligação poderia resultar na incorrecta inserção do fragmento de 534 pb, pois poderia ser inserido no sentido contrário.

Deste modo, seria possível obter-se o gene completo da MolA, flanqueado nas suas extremidades pelos locais de reconhecimento para as enzimas de restrição *NcoI* e *XhoI*, de modo a proceder-se à ligação ao vector pETMFh8 e posterior transformação em células competentes.

No entanto, a estratégia não resultou logo no primeiro passo pois, após a digestão com a enzima de restrição *XhoI* não foi possível obter o perfil desejado em gel de agarose, tal como elucidada a Figura 24.

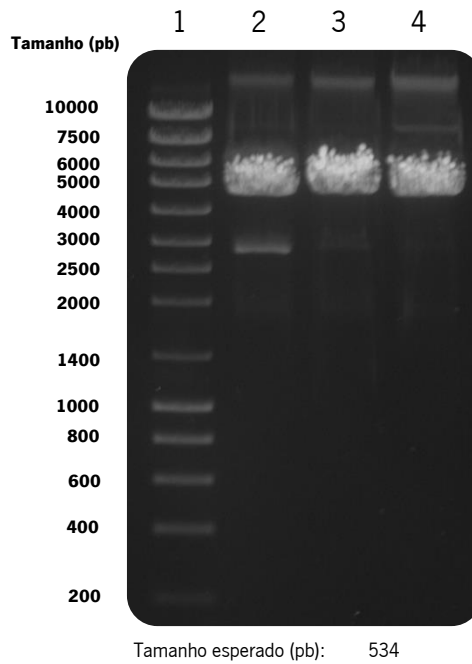


Figura 24 - Perfil em gel de agarose a 1,2% (p/v) da digestão do constructo pGEM *Ncd* MolA *Xho*I com a enzima de restrição *Xho*I durante 2h.

Legenda: 1 – Marcador de DNA; 2 a 4 – Digestão sobre o DNA extraído de três transformantes pGEM *Ncd* MolA *Xho*I.

Era esperado obter-se, tal como referido, uma banda de tamanho 534 pb, que continha a parte da sequência codificante da MolA. No entanto, não foi visível em gel o aparecimento do fragmento pretendido surgindo um perfil inespecífico da digestão.

Após variação nos volumes, mudança de enzima e de tampão utilizado não foi possível obter perfis diferentes deste, pelo que se testaram duas novas abordagens: adição de BSA com o intuito de mimetizar o comportamento da enzima de restrição *Ncd*, uma vez que, ocorre a digestão completa quando a enzima de restrição *Xho*I se encontra na presença da *Ncd*, não ocorrendo quando esta se encontra sozinha; e choque térmico a 90°C no DNA plasmídico de modo a ocorrer a sua desnaturação e proceder-se posteriormente à sua digestão com *Xho*I. Em ambos os casos não se conseguiu obter o perfil desejado.

Deste modo, não se avançou com a digestão com a enzima de restrição *Ncd* e consequentemente abortou-se esta estratégia.

3.1.2. Metodologia adoptada com o gene sintetizado

Uma vez que não foi possível obter a construção pETMFh8 MolA por outro meio, o gene da MolA foi sintetizado. Através do software *GeneOptimizer*® optimizou-se a sequência codificante do gene e inseriram-se os locais de restrição específicos das duas enzimas de restrição compatíveis com o vector pETMFh8 (*NcoI* e *XhoI*). A síntese do gene resultou na optimização de codões (*codon usage*) e do conteúdo de GC, de modo a impedir alguns problemas de expressão (sumariamente referidos na secção 1.4.2). Os histogramas e mapas obtidos da distribuição de qualidade dos codões e de conteúdo de GC encontram-se nas Figura AV. 1 e 2 (Anexo V, páginas 90 e 91).

A sequência de nucleótidos da MolA (1406 bp) foi entregue inserida em vectores do tipo pMK-RQ (portadores do gene que confere resistência à Canamicina) com uma massa de 5 µg na forma liofilizada, ao qual se adicionou 50 µL de água ultrapura perfazendo uma concentração de 0,1 µg/µL. Na Figura 25 está apresentado o mapa circular desta construção, de 3684 pb, exibindo os locais de corte específicos para as enzimas de restrição *NcoI* e *XhoI*.

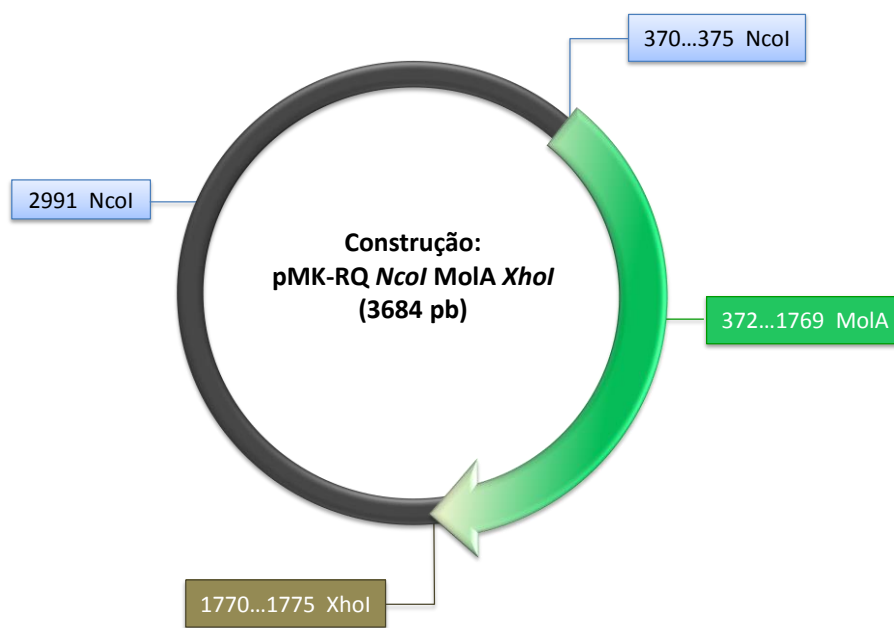


Figura 25 – Representação do mapa circular da construção pMK-RQ *NcoI* MolA *XhoI*.

Comparando o mapa desta construção com o da construção pGEM *NcoI* MolA *XhoI* (Figura 19) é possível verificar que, neste momento, já não existe nenhum local de corte no gene MolA por parte das enzimas de restrição *XhoI* e *NcoI*. Do mesmo modo, constata-se que apesar do plasmídeo pMK-RQ possuir um local de restrição relativo à enzima *NcoI* (par de base 2991), este se encontra numa posição suficiente afastada dos outros locais de restrição. Consegue-se,

portanto, que a corrida do DNA digerido em gel de agarose resulte em fragmentos mais facilmente visíveis.

Assim, a digestão simultânea com as enzimas de restrição *NcoI* e *XhoI* deverá gerar 3 fragmentos de tamanhos distintos, tal como representado na Figura 26.

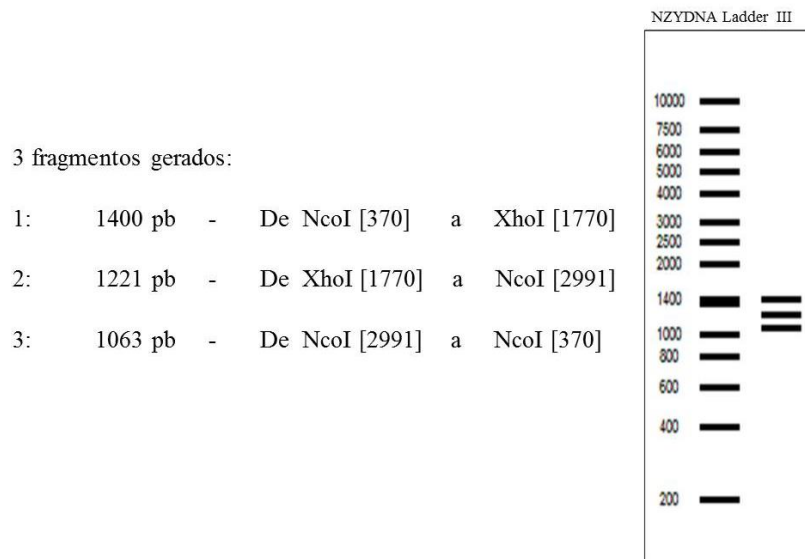


Figura 26 - Esquema representativo do tamanho dos fragmentos gerados através da digestão do constructo pMK-RQ *NcoI* MolA *XhoI* com as enzimas de restrição *NcoI* e *XhoI*.

Deste modo, o constructo pMK-RQ *NcoI* MolA *XhoI* foi transformado em células competentes *E. coli* NZY5 α (segundo o protocolo descrito na secção 2.4.6). Após incubação das placas *overnight*, seleccionaram-se os transformantes mais isolados (4 clones) e repicou-se para culturas líquidas de LB/Canamicina. As culturas foram incubadas *overnight* a 37°C e o DNA plasmídico de cada clone foi extraído (através do kit referido na secção 2.4.8).

Procedeu-se à digestão do DNA plasmídico com as enzimas de restrição *NcoI* e *XhoI* (de acordo com o procedimento descrito na secção 2.4.2) e correram-se as amostras um gel de agarose (segundo a técnica descrita na secção 2.4.3). A Figura 27 apresenta o perfil da digestão obtida em gel de agarose *low melting*.

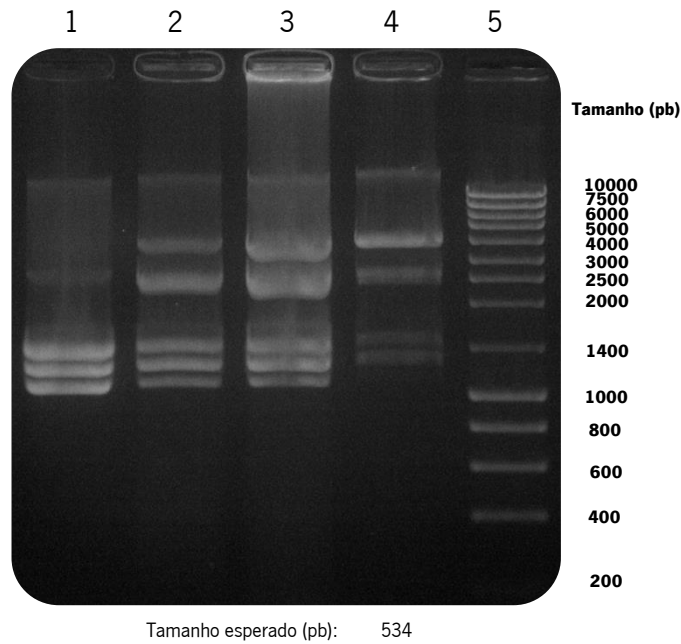


Figura 27 - Perfil em gel de agarose *low melting* a 1,4% (p/v) da digestão do pMK-RQ *NcoI* *MolA* *XhoI* com as enzimas de restrição *NcoI* e *XhoI* em simultâneo durante 2h.

Legenda: 1 a 4 – Digestão sobre o DNA extraído de quatro transformantes pMK-RQ *NcoI* *MolA* *XhoI*; 5 – Marcador de DNA.

Como é perceptível pela figura acima, as digestões dos 4 clones resultaram no aparecimento dos três fragmentos de tamanho próximo ao esperado (consultar Figura 26). No entanto, a visualização das bandas correspondentes a estes fragmentos no clone 4 (poço 4), foi menos intensa. Além destes fragmentos, são visíveis no gel outros fragmentos podendo corresponder a plasmídeo não digerido ou mal digerido.

Deste modo, prepararam-se stocks das células contendo o plasmídeo (com 37,5% de glicerol estéril) que foram crioconservadas a -80°C para posteriores utilizações.

De seguida, procedeu-se à extracção e purificação da banda correspondente ao fragmento do gene da *MolA* (de acordo com a técnica descrita na secção 2.4.4) e promoveu-se a ligação ao vector de expressão pETMFh8 utilizando-se a ligase T4 *Speedy* (procedimento descrito na secção 2.4.5). Obteve-se assim a nova construção, pETMFh8 *MolA*, apresentada na Figura 28.

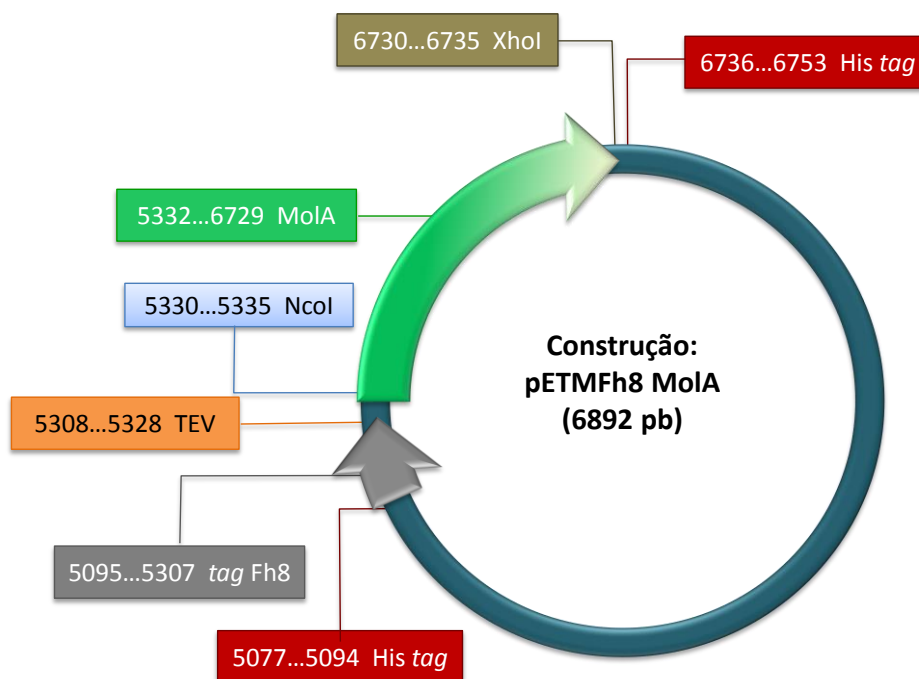


Figura 28 – Representação do mapa circular da construção pETMFh8 MoIA.

A construção obtida foi transformada em células *E. coli* BL21(DE3) (seguindo o protocolo descrito na secção 2.4.6) e sequenciada na Eurofins MWG Operon (de acordo com o procedimento descrito na secção 2.4.9). A sequência codificante desta construção encontra-se no Anexo IV, página 88.

3.2. Controlo da indução e análise da expressão proteica em gel SDS-Page

Uma vez obtido e sequenciado o constructo pETMFh8 MoIA, realizaram-se os ensaios de crescimento celular em células competentes *E. coli* BL21(DE3), onde se iniciou a indução/expressão da proteína recombinante de interesse (Fh8MoIA).

A Tabela 5 apresenta o peso molecular e o número de aminoácidos da nova proteína recombinante em comparação com a proteína nativa.

Tabela 5: Tamanho (pb), peso molecular (kDa) e número de aminoácidos da proteína nativa e recombinante

Proteína	Tamanho (pb)	Peso molecular (kDa)	Número de aminoácidos
MoIA	1398	50,85	465
Fh8MoIA	1659	60,69	552

O peso molecular da nova proteína corresponde à sequência nucleotídica contida no vector de expressão pETMFh8 desde o codão de iniciação ATG (que precede a transcrição das sequências do *His-tag*, do *tag* Fh8 e da protease TEV) até ao codão de terminação TAA (localizado no último codão da sequência da proteína).

Tal como referido na secção 2.6 do capítulo anterior, foram realizados dois ensaios distintos nestas células de modo a avaliar as propriedades da proteína recombinante, tais como, a sua expressão e solubilidade e a toxicidade para a célula. Inicialmente realizou-se um ensaio preliminar do crescimento celular e indução proteica e, a partir da análise dos resultados obtidos neste ensaio, realizou-se um segundo ensaio de optimização das condições de densidade óptica, concentração de indutor e de temperatura. Em ambos, as determinações de expressão e de solubilidade foram conseguidas através da análise aos géis de poliacrilamida de SDS-Page Tris-Tricina (técnica descrita na secção 2.5).

O primeiro ensaio (procedimento descrito na secção 2.6.1), tratando-se de um ensaio preliminar, permitiu sobretudo analisar se, em primeira instância, a indução da proteína recombinante resultou no aumento da toxicidade e consequente morte celular. Pretendeu-se, igualmente, analisar se a indução da proteína recombinante ocorreu e se esta foi expressa na forma solúvel ou se, pelo contrário, a *E. coli* a direccionou para corpos de inclusão. Para analisar a toxicidade celular foi acompanhado o crescimento celular, através da densidade óptica (OD_{600}), antes e durante a indução. A expressão da proteína recombinante induzida (I) e não induzida (CN) foi analisada através da comparação das intensidades das bandas dos extractos celulares correspondentes. Por fim, averiguou-se a solubilidade proteica através da comparação das fracções solúveis (FS) com as insolúveis (FI).

O segundo ensaio (procedimento descrito na secção 2.6.2) consistiu na optimização das condições de indução, pretendendo-se não só melhorar a expressão e solubilidade da proteína recombinante, mas também minimizar custos e tempos de produção (e.g. redução de indutor, tempo de crescimento, entre outros) num futuro *scale-up*.

3.2.1. Ensaio preliminar de crescimento celular e indução proteica

Uma das desvantagens da produção de proteínas recombinantes em *E. coli* é a toxicidade que a proteína recombinante poderá suscitar nas células e consequentemente no crescimento

celular. Enquanto as células hospedeiras podem ser cultivadas a altas densidades debaixo de condições repressivas, a indução pode resultar numa paragem de crescimento rápida e/ou morte celular.

A realização de um ensaio preliminar serviu, não só para analisar a expressão e solubilidade proteica, mas também para analisar se a indução da proteína recombinante poderia resultar em elevados níveis de toxicidade para o hospedeiro que levassem à morte celular. Para tal, acompanhou-se a densidade óptica ao longo do tempo de crescimento, tanto no ensaio negativo (não-induzido), como no ensaio em que se induziu. Este é um método de análise que é normalmente utilizado quando se pretende confirmar se uma dada cultura se encontra em crescimento ou para acompanhar o crescimento microbiano com base no aumento da densidade óptica medida a um comprimento de onda particular.

O resultado da leitura das densidades ópticas ao longo do tempo apresenta-se na Figura 29.

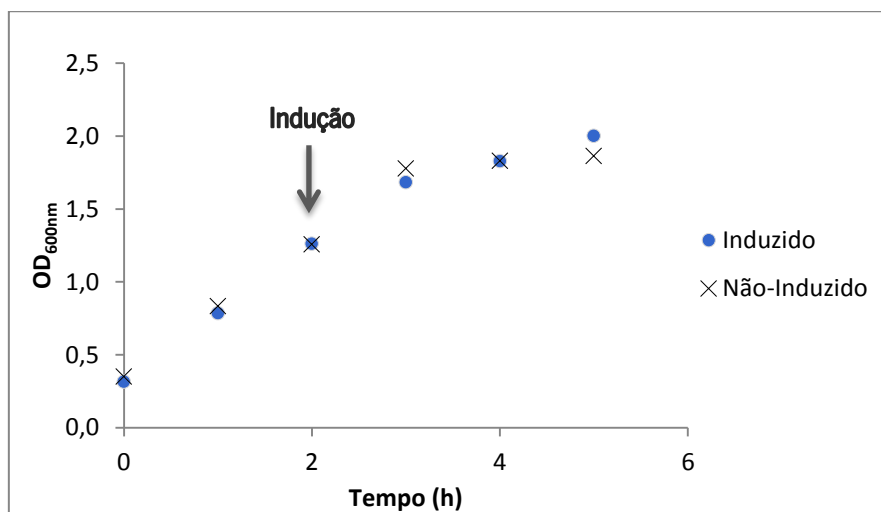


Figura 29 – Representação gráfica dos valores da densidade óptica medidos em função do tempo de crescimento.

Pela análise da figura acima é possível verificar que, em ambos os ensaios, a densidade óptica aumentou ao longo do tempo. Verifica-se igualmente a estagnação do crescimento bacteriano por volta da terceira hora. Analisando apenas os resultados do ensaio induzido, constata-se que, após a indução (na segunda hora), não ocorreu uma queda da densidade óptica, continuando a seguir uma tendência de crescimento.

Apesar deste método de análise não permitir distinguir entre células viáveis e células mortas e não permitir obter directamente valores absolutos da concentração de células, pode no entanto

afirmar-se, com base nos resultados obtidos, que a expressão da proteína recombinante não se revelou tóxica para o organismo hospedeiro pois não provocou paragem do crescimento. Por outro lado, este estudo pode ter erros implícitos, uma vez que não foram feitas leituras da densidade óptica em duplicado ficando subentendido um erro associado a apenas um ensaio, além de que assim não é possível o cálculo do desvio padrão para quantificar esse erro.

O ensaio preliminar tinha também por objectivo avaliar a capacidade de expressão e de solubilidade da nova proteína recombinante.

A Figura 30 apresenta o perfil em gel SDS-Page Tris-Tricina da análise da indução e da expressão solúvel das culturas de 50 mL de *E. coli* BL21(DE3).

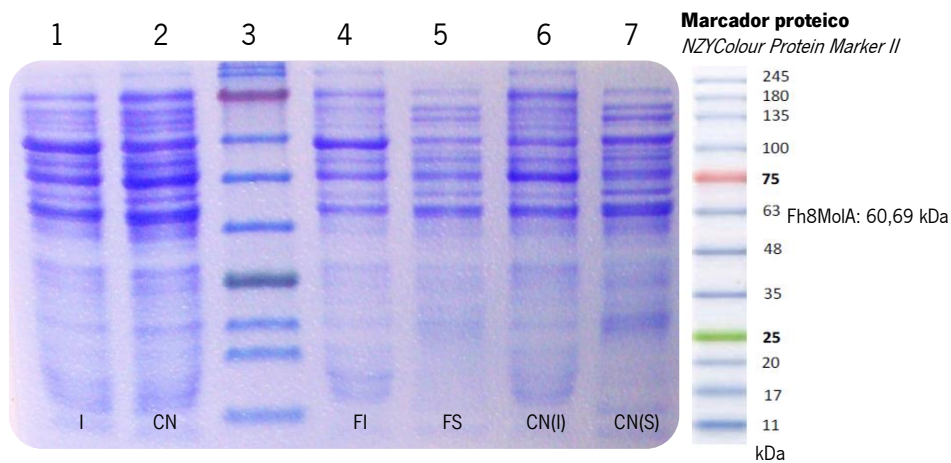


Figura 30 - Perfil em gel SDS-Page Tris-Tricina da análise da indução e da expressão solúvel das culturas de 50 mL de *E. coli* BL21(DE3) pETMFh8 MoIA em condições nativas.

Legenda: 1 – Extracto celular induzido; 2 – Extracto celular do controlo negativo; 3 – Marcador proteico; 4 – Fracção insolúvel do induzido; 5 – Fracção solúvel do induzido; 6 – Fracção insolúvel do controlo negativo; 7 – Fracção solúvel do controlo negativo.

O perfil obtido permitiu verificar o aparecimento de uma banda pouco intensa no extracto celular induzido (poço 1 do gel de SDS-Page), considerando o peso molecular esperado para a proteína recombinante Fh8MoIA (60,69 kDa), sendo assim expectável que essa banda corresponda à proteína recombinante. No extracto celular não-induzido (poço 2) também se verifica o aparecimento de uma banda de tamanho aproximado à proteína de interesse, no entanto, uma vez que o promotor T7 possui mecanismos próprios de regulação (na ausência de IPTG não produz a proteína alvo), a banda visível não deverá corresponder à proteína de interesse mas a outra de tamanho próximo produzida pelo organismo. A utilização de um outro

controlo negativo (e.g. *E. coli* transformada com o vector vazio) poderia ter ajudado a esclarecer esta dúvida.

Contudo, apenas ensaios de *Western Blot* permitiriam concluir se determinada banda corresponde à proteína de interesse.

Relativamente à solubilidade proteica nas células induzidas, a comparação entre as intensidades das bandas das fracções insolúveis (poço 4) e as solúveis (poço 5), permitiu constatar que ocorreu uma maior expressão da proteína recombinante na fracção insolúvel.

Este era um resultado esperado pois estudos anteriores visando o isolamento e produção desta enzima, resultaram em expressões maioritariamente insolúveis em *E. coli*. Por outro lado, o facto de a indução ter sido feita num momento em que a concentração celular já era alta (OD_{600} situava-se próxima de 1,2), poderá não ter potenciado a expressão devidamente. De facto, era esperado que a indução resultasse em aumentos de expressão da proteína recombinante que permitissem facilmente a visualização pela intensidade das bandas no gel SDS-Page.

Deste modo, a temperatura, o momento de indução (densidade óptica OD_{600}) e a concentração de indutor (IPTG) devem ser otimizados de forma a obter conclusões mais significativas sobre a expressão da proteína recombinante.

3.2.2. Testes de optimização em pequena escala do crescimento celular e indução proteica

A partir dos resultados obtidos no ensaio preliminar de crescimento celular e indução proteica, realizaram-se 18 novas condições onde se variaram as concentrações de IPTG utilizadas (0,1mM; 0,5mM e 1,0mM) e o momento de indução (OD_{600} : 0,4; 0,6-0,8 e >1,0) para duas temperaturas de indução diferentes (37°C e 18°C).

Os resultados obtidos dos testes de optimização das culturas de 10 mL de *E. coli* BL21(DE3) induzidas a 37°C apresentam-se na Figura 31.

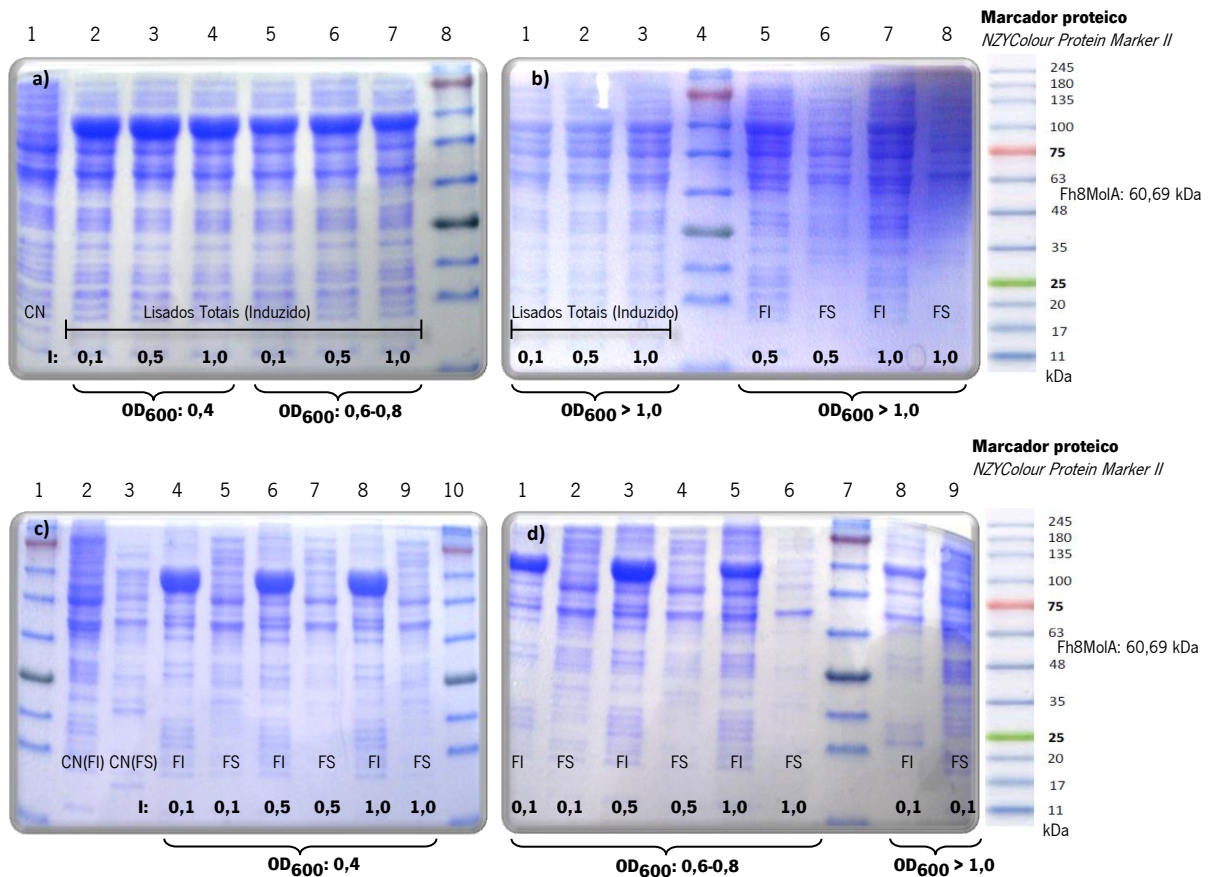


Figura 31 – Perfil em gel SDS-Page Tris-Tricina da análise da indução e da expressão solúvel das culturas de 10 mL de *E. coli* BL21(DE3) pETMfH8 MolA em condições nativas para a temperatura de 37 °C.

Legenda: a1 – Extracto celular do controlo negativo; a2-a4, a5-a7 e b1-b3 – Extractos celulares dos testes induzidos, respectivamente, a OD_{600} : 0,4, 0,6-0,8 e superior a 1,0, para concentrações de 0,1mM, 0,5mM e 1,0mM de IPTG; c2 – Fracção insolúvel do controlo negativo; c4, c6, c8, d1, d3, d5, d8, b5 e b7 – Fracções insolúveis dos testes induzidos, respectivamente, a OD_{600} : 0,4, 0,6-0,8 e superior a 1,0, para concentrações de 0,1mM, 0,5mM e 1,0mM de IPTG; c3 – Fracção solúvel do controlo negativo; c5, c7, c9, d2, d4, d6, d9, b6 e b8 – Fracções solúveis dos testes induzidos, respectivamente, a OD_{600} : 0,4, 0,6-0,8 e superior a 1,0, para concentrações de 0,1mM, 0,5mM e 1,0mM de IPTG; a8, b4, c1, c10 e d7 – Marcador proteico.

Analisando, numa primeira fase, apenas os extractos celulares é possível verificar que, nos ensaios a 37°C, a indução das bactérias resultou consideravelmente num aumento da expressão da proteína Fh8MolA, em contraste com o ensaio em que não foi induzida (poço a1 do gel de SDS-Page). De facto, pela análise dos ensaios em que se induziu no momento em que a densidade óptica se situava próxima de 0,4 (poços a2-a4) e, também, nos ensaios em que se encontrava entre os 0,6-0,8 (poços a5-a7), é facilmente perceptível esse aumento de expressão no perfil do gel pelas intensidades das bandas correspondentes.

Tal como no ensaio preliminar, constata-se que a indução no momento em que a densidade óptica se situava acima de 1,0 (poços b1-b3) resultou num aumento pouco considerável da expressão da proteína recombinante. Por outro lado, é possível verificar que não ocorreram mudanças consideráveis da expressão da proteína de interesse entre as novas condições de densidade óptica a que se induziu (OD_{600} : 0,4 e 0,6-0,8).

Ainda na análise dos extractos celulares induzidos é possível constatar que as induções a diferentes concentrações de IPTG não apresentam variações consideráveis entre si como se pode observar pela semelhança entre as intensidades das bandas.

Numa segunda fase, a comparação dos resultados obtidos das fracções solúveis da proteína Fh8MolA induzida (poços c5, c7, c9, d2, d4, d6, d9, b6 e b8) com as fracções insolúveis (poços c4, c6, c8, d1, d3, d5, d8, b5 e b7) permite verificar que, para todas as condições de densidade óptica diferentes, praticamente a totalidade da proteína Fh8MolA expressa foi direccionada para a fracção insolúvel, ficando retida em corpos de inclusão. Tal como nos extractos celulares, a análise das intensidades das bandas das fracções solúveis e insolúveis revelam que não ocorreram variações consideráveis entre as induções a diferentes concentrações de IPTG. A excepção mais relevante a este resultado é a fracção solúvel da indução à concentração de 1,0mM de IPTG para a OD_{600} : 0,6-0,8 (poço d6), onde se observa que se obteve menor intensidade das bandas em todo o perfil desta condição relativamente às restantes fracções solúveis para esta densidade óptica. Tal facto pode ser justificado por uma incorrecta lise celular na preparação das amostras.

Estudos anteriores realizados com proteínas recombinantes em *E. coli* permitiram formular estratégias de modo a otimizar as condições de expressão. Uma das estratégias resulta da diminuição da temperatura de expressão pois ocorre, normalmente, uma melhoria da solubilidade das proteínas recombinantes expressas. Isto deve-se ao facto de a baixas temperaturas, os processos celulares abrandarem levando a taxas reduzidas de transcrição, tradução e divisão celular e também a baixas agregações proteicas. Além disso, a maior parte das proteases são menos activas a baixas temperaturas, o que resulta numa redução na degradação das proteínas sensíveis à proteólise (Francis e Page 2010).

Com base neste pressuposto realizaram-se testes de optimização das culturas de 10 mL de *E. coli* BL21(DE3) induzidas a 18°C, apresentando-se os resultados obtidos em géis SDS-Page Tris-Tricina na Figura 32.

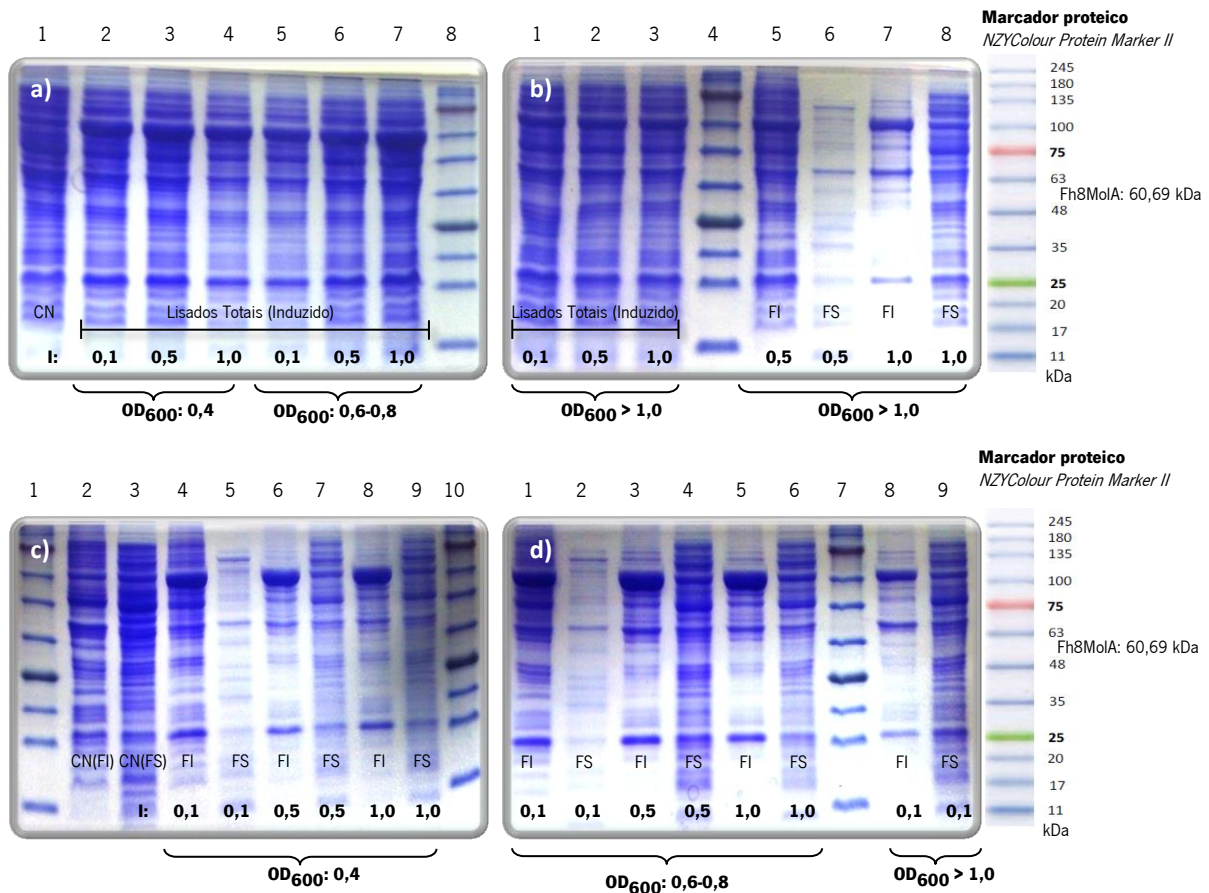


Figura 32 – Perfil em gel SDS-Page Tris-Tricina da análise da indução e da expressão solúvel das culturas de 10 mL de *E. coli* BL21(DE3) pETMfH8 MolA em condições nativas para a temperatura de 18 °C.

Legenda: a1 – Extracto celular do controlo negativo; a2-a4, a5-a7 e b1-b3 – Extractos celulares dos testes induzidos, respectivamente, a OD_{600} : 0,4, 0,6-0,8 e superior a 1,0, para concentrações de 0,1mM, 0,5mM e 1,0mM de IPTG; c2 – Fracção insolúvel do controlo negativo; c4, c6, c8, d1, d3, d5, d8, b5 e b7 – Fracções insolúveis dos testes induzidos, respectivamente, a OD_{600} : 0,4, 0,6-0,8 e superior a 1,0, para concentrações de 0,1mM, 0,5mM e 1,0mM de IPTG; c3 – Fracção solúvel do controlo negativo; c5, c7, c9, d2, d4, d6, d9, b6 e b8 – Fracções solúveis dos testes induzidos, respectivamente, a OD_{600} : 0,4, 0,6-0,8 e superior a 1,0, para concentrações de 0,1mM, 0,5mM e 1,0mM de IPTG; a8, b4, c1, c10 e d7 – Marcador proteico.

Tal como nos ensaios a 37°C, analisando numa primeira fase apenas os extractos celulares é possível verificar que, nos ensaios a 18°C, ocorreu uma maior expressão da proteína Fh8MolA quando induzida, em contraste com o ensaio em que não foi induzida (poço a1 do gel de SDS-Page). Esse aumento de expressão é perceptível através da maior intensidade das bandas, no perfil do gel, dos extractos celulares induzidos (poços a1-a7 e b1-b3) para cada densidade óptica correspondente.

Ao contrário do ensaio a 37°C, a indução no momento em que a densidade óptica se situava acima de 1,0 (poços b1-b3) resultou num aumento semelhante ao que sucedeu com as

outras induções a diferentes densidades ópticas. Deste modo, constata-se que, quando a indução ocorre à temperatura de 18°C, não ocorrem aparentemente mudanças significativas na expressão da proteína recombinante nas diferentes condições de densidade óptica testadas.

Ainda na análise dos extractos celulares induzidos é possível constatar, mais uma vez, que as induções a diferentes concentrações de IPTG não apresentam variações consideráveis entre si como se pode observar pela semelhança entre as intensidades das bandas.

Numa segunda fase, tal como no ensaio a 37°C, a comparação dos resultados obtidos das fracções solúveis da proteína Fh8MolA induzida (poços c5, c7, c9, d2, d4, d6, d9, b6 e b8) com as fracções insolúveis (poços c4, c6, c8, d1, d3, d5, d8, b5 e b7) permite verificar que, para todas as condições de densidade óptica diferentes, praticamente a totalidade da proteína Fh8MolA expressa foi direccionada para a fracção insolúvel, ficando retida em corpos de inclusão.

No entanto, comparando os resultados obtidos nas fracções solúveis do ensaio a 18°C com os do ensaio a 37°C (Figura 31), é possível constatar que se conseguiu, de modo geral, melhorar as expressões na fase solúvel. Esta melhoria de solubilidade é evidenciada principalmente pela intensidade das bandas dos ensaios de OD_{600} : 0,4 para as concentrações de IPTG de 0,5mM (poço c7) e 1,0mM (poço c9) e dos ensaios de OD_{600} : 0,6-0,8 para as concentrações de IPTG de 0,5mM (poço d4) e 1,0mM (poço d6). Analisando estes mesmos ensaios é possível constatar que as induções às concentrações de IPTG de 0,5mM e 1,0mM resultaram em intensidades de bandas semelhantes podendo inferir-se que se pode reduzir custos do reagente indutor num futuro *scale-up*. Do mesmo modo, é possível verificar que os ensaios a OD_{600} : 0,6-0,8 resultaram em maiores intensidades nas bandas correspondentes à proteína recombinante expressa na fracção solúvel, relativamente aos outros momentos de indução. Por outro lado, verifica-se que, tanto para os ensaios de indução a OD_{600} : 0,4 (poço c5) como a OD_{600} : 0,6-0,8 (poço d2), a aplicação da concentração de indutor 0,1mM resultou em menores quantidades de proteína expressa na fracção solúvel.

Os ensaios a $OD_{600} > 1,0$ revelaram resultados díspares para as induções às diferentes concentrações de IPTG, pelo que não se consideraram válidos para posterior purificação devido à instabilidade de resultados.

Por fim, comparando os perfis da fracção insolúvel (poço c2) com a fracção solúvel (poço c3), relativos ao controlo negativo, verifica-se o aparecimento de bandas com tamanho próximo do tamanho correspondente à proteína recombinante. No entanto, tal como referido

anteriormente, uma vez que o promotor T7 possui mecanismos próprios de regulação (na ausência de IPTG não produz a proteína alvo), a banda visível não deverá corresponder à proteína de interesse mas a outra de tamanho próximo produzida pelo organismo. Assim, apenas ensaios de *Western Blot*, ou testes de actividade permitiriam concluir se determinada banda corresponde à proteína de interesse.

3.3. Análise da mini-purificação por IMAC em gel SDS-Page

Um dos objectivos do presente trabalho era produzir a proteína recombinante Fh8MolA com níveis de pureza satisfatórios. Deste modo, foram adoptadas metodologias de tratamento da proteína que permitissem não só obter os níveis de pureza desejados mas principalmente conhecer o perfil de purificação para um posterior *scale-up*. Realizou-se, assim, uma mini-purificação de afinidade por IMAC em condições nativas (procedimento descrito na secção 2.6.3) através da junção das fracções solúveis das culturas de 10 mL, obtidas dos testes de optimização a diferentes concentrações de IPTG correspondentes ao momento em que foi feita a indução a OD_{600} : 0,6-0,8, à temperatura de 18°C.

O perfil obtido após as purificações em gel de poliacrilamida de SDS-Page Tris-Tricina (técnica descrita na secção 2.5) está apresentado na Figura 33.

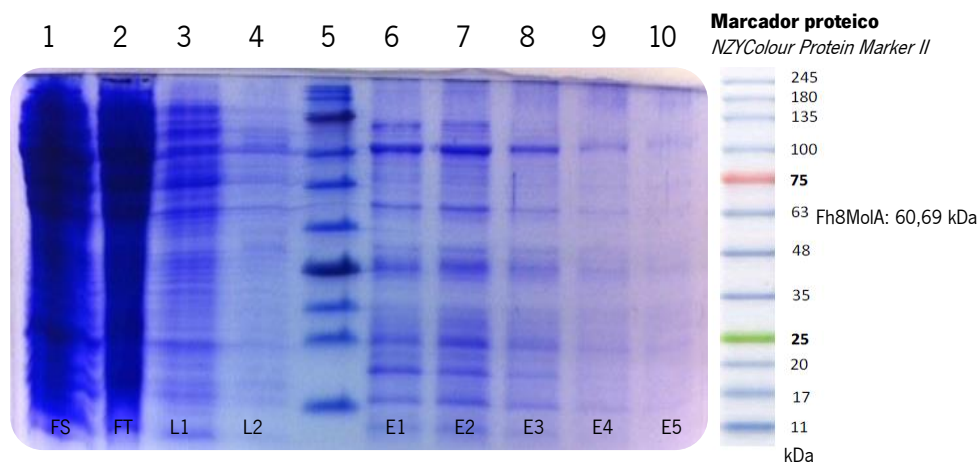


Figura 33 - Perfil em gel SDS-Page Tris-Tricina das fracções recolhidas durante a mini-purificação em condições nativas, do conjunto das fracções solúveis das culturas de 10 mL de *E. coli* BL21(DE3) pETMh8 MolA induzidas no momento OD_{600} : 0,6-0,8.

Legenda: 1 – Fracção solúvel do extracto; 2 – Fracção que atravessou a coluna: *Flow through*; 3 – Lavagem 1; 4 – Lavagem 2; 5 – Marcador proteico; 6 – Eluição 1; 7 – Eluição 2; 8 – Eluição 3; 9 – Eluição 4; 10 – Eluição 5.

A análise do gel SDS-Page Tris-Tricina correspondente à purificação por IMAC das fracções solúveis permite avaliar qualitativamente o resultado da mini purificação realizada.

Assim a sua análise sugere que, tal como era esperado, praticamente toda a fracção solúvel atravessou a coluna. De facto, a comparação entre as intensidades das bandas da fracção solúvel (poço 1) e da fracção que atravessou a coluna (poço 2) permite constatar que a maior fracção proteica não se ligou à matriz e, por isso, foi recolhida após atravessar por gravidade. A forte intensidade das bandas evidenciada nos perfis da fracção solúvel e do *Flow through* deve-se ao facto de se ter concentrado o conjunto das fracções solúveis em apenas 300 µL de tampão de lise. Isso permitiu fortalecer a intensidade das bandas que se pretendiam observar mas, por outro lado, impediu a obtenção de perfis proteicos mais claros nestas duas fracções. De facto, observando novamente a fracção que atravessou a coluna (poço 2), verifica-se a existência de uma banda de maior intensidade com peso molecular próximo do peso da proteína de interesse (60,69 kDa). Pode evidenciar-se assim que, não só a maior fracção proteica não se ligou à matriz, mas também grande parte da proteína de interesse não estabeleceu ligação com a resina, sendo recolhida após atravessar por gravidade (*Flow through*). Uma possível explicação prende-se com o facto de em determinados momentos da purificação ter sido necessário centrifugar a coluna variando a velocidade entre 200 a 400 rpm de modo a permitir a passagem pela coluna. Esta centrifugação poderá ter impedido que parte da proteína recombinante estabelecesse ligação por afinidade com a resina. Além desta razão, a perda de proteína poderá ser explicada por diversos factores, dos quais se destacam os seguintes: difícil acesso do *His-tag* ao níquel (devido à estrutura terciária da própria proteína recombinante que impede a ligação ou devido às restantes biomoléculas presentes na fracção solúvel que interagem e também se ligam à resina); condições de ligação (concentrações elevadas de imidazol no tampão de ligação podem prejudicar a ligação do *His-tag* ao níquel pois o imidazol compete com o *His-tag* pela ligação à resina (Quiagen 2003).

As lavagens permitiram observar que uma quantidade menos considerável da fracção proteica se ligou com menos intensidade à coluna, sendo estes contaminantes removidos após as lavagens. Na segunda lavagem (poço 4) evidencia-se uma diminuição mais significativa de contaminantes com menor afinidade pela resina.

A análise do gel SDS-Page Tris-Tricina das 5 fracções de eluição (poços 6 a 10) recolhidas durante a purificação permitiu, em primeira instância, comparar o posicionamento das bandas (através da comparação dos pesos moleculares) com os géis obtidos dos testes de optimização.

Detectou-se assim uma banda de maior intensidade, próxima do peso molecular da proteína Fh8MolA, que vai perdendo intensidade consoante vai sendo eluída. Pode extrapolar-se assim que, primeiro, as bandas obtidas nos testes de optimização correspondem à proteína de interesse e, segundo, a purificação resultou em níveis de pureza superiores (pela observação da redução de contaminantes). Do mesmo modo, verifica-se que a proteína de interesse, contendo o *tag* de purificação *His-tag*, se ligou efectivamente aos iões de níquel imobilizados da matriz de NI-NTA por afinidade. Porém, tal como referido anteriormente, apenas ensaios de *Western Blot* permitiriam concluir se determinada banda corresponde à proteína de interesse.

Por outro lado, os perfis das fracções eluídas apresentam ainda demasiados contaminantes. A optimização das condições de purificação, em concreto, a optimização dos tampões utilizados e a lavagem da resina com um gradiente crescente de Imidazol, pode permitir um maior nível de pureza da proteína recombinante.

Para além destas opções, uma segunda purificação por interacção hidrofóbica, utilizando a capacidade de purificação do *tag* Fh8, poderá permitir reduzir significativamente a quantidade de contaminantes.

Em suma, a concentração das fracções solúveis num pequeno volume de tampão de lise permitiu a observação em gel das bandas correspondentes à proteína recombinante Fh8MolA (assumindo que as bandas presentes no gel de tamanho próximo ao da proteína recombinante Fh8MolA é efectivamente a proteína de interesse). Uma vez que nos testes de optimização da indução/expressão, a maior parte da proteína tinha sido expressa na fase insolúvel, esta concentração permitiu aumentar a intensidade das bandas para serem visíveis em gel. Se por um lado a mini-purificação resultou em perda da proteína recombinante que não se ligou na totalidade à coluna (e que atravessou a coluna sendo recolhida no *Flow through*), por outro, foi possível verificar que se conseguiu recuperar a proteína em maior estado de pureza (em termos qualitativos).

4. Conclusões e Perspectivas de futuros trabalhos

Este estudo visava a produção da proteína recombinante MolA em *E. coli*, através do uso do *tag* de fusão Fh8, a avaliação e optimização das suas propriedades a nível da expressão e da solubilidade, bem como a sua purificação.

O gene codificante da MolA, com 1398 pb, possui locais de restrição das enzimas de restrição (*NcoI* e *XhoI*) usadas para clonar no vector de expressão (pETMFh8). Por esse motivo, a clonagem directa deste gene no vector de interesse não foi, por várias razões, conseguida com sucesso. Para contornar este problema, o gene foi optimizado e sintetizado de modo a ser clonado neste vector e prosseguir a expressão da proteína recombinante (Fh8MolA).

Para estudar os níveis de expressão e de solubilidade da proteína recombinante foram realizados dois ensaios distintos: um ensaio preliminar de crescimento celular e indução proteica, onde se analisou também a toxicidade da proteína no organismo hospedeiro; e um ensaio de testes de optimização das condições de indução.

O ensaio preliminar permitiu concluir que a indução da proteína recombinante não resultou em níveis de toxicidade elevados que levassem à morte celular. Relativamente à expressão e solubilidade da proteína recombinante, os resultados não foram precisos, pelo que se efectuou o segundo ensaio, em que se optimizaram as condições de indução, nomeadamente, a temperatura, o momento de indução (densidade óptica OD₆₀₀) e a concentração de indutor (IPTG).

Perante os resultados obtidos no ensaio de optimização das condições foi possível concluir que a indução da proteína Fh8MolA permitiu maiores níveis de expressão relativamente ao ensaio não induzido, não apresentando variações consideráveis seja pela indução às diferentes concentrações de IPTG ou às diferentes densidades ópticas testadas. Relativamente à solubilidade da proteína recombinante, praticamente a totalidade da proteína Fh8MolA expressa foi direccionada para a fracção insolúvel, ficando retida em corpos de inclusão. No entanto, os resultados permitiram concluir que a indução *overnight* à temperatura de 18°C permitiu melhorar as expressões na fase solúvel, relativamente à indução à temperatura de 37°C. A indução à temperatura de 18°C não só melhorou a solubilidade, como permitiu verificar que, em qualquer momento de indução, se obtiveram fracções solúveis semelhantes através do uso de concentrações de 0,5mM e de 1,0mM de IPTG. Igualmente concluiu-se que a proteína recombinante foi expressa ainda mais significativamente na fracção solúvel para os ensaios a

OD₆₀₀: 0,6-0,8. Estes dados permitem inferir que se pode reduzir tempos e custos do reagente indutor num futuro *scale-up*.

A proteína recombinante foi purificada através de uma mini-purificação por cromatografia de afinidade por IMAC. Os resultados obtidos desta purificação permitiram concluir que grande parte da proteína recombinante, contendo o *tag* de purificação *His-tag*, se ligou efectivamente aos iões de níquel imobilizados da matriz de NI-NTA por afinidade. Assim, foi possível verificar que se conseguiu recuperar a proteína em maior estado de pureza (em termos qualitativos).

No entanto, a mini-purificação realizada resultou em perda da proteína recombinante que não se ligou na totalidade à coluna (e que atravessou a coluna sendo recolhida no *Flow through*) e os perfis das fracções eluídas apresentaram ainda demasiados contaminantes.

Em suma, os objectivos propostos neste estudo foram conseguidos. Não só se conseguiu expressar a proteína como se realizou, pela primeira vez com este *tag* de fusão, um estudo da influência das condições de indução nesta proteína, nomeadamente, a temperatura, o momento de indução (densidade óptica OD₆₀₀) e a concentração de indutor (IPTG) que poderão servir de base para posteriores estudos com proteínas difíceis de expressar e de solubilizar. Para além disso, este estudo permitiu otimizar as condições para a expressão desta proteína recombinante de modo a reduzir tempos e custos de produção num futuro *scale-up*.

No entanto, após a realização deste trabalho e face aos resultados obtidos são muitas as perspectivas de futuros trabalhos.

A curto prazo será necessário realizar ensaios de *Western Blot* que permitam concluir se a proteína recombinante produzida neste trabalho corresponde de facto à proteína de interesse. É igualmente necessário quantificar a proteína remanescente proveniente das eluições da purificação por IMAC. Só uma análise quantitativa permitirá avaliar sobre o estado de pureza da proteína obtida. Por outro lado, a optimização das condições de purificação, em concreto, a optimização dos tampões utilizados e a lavagem da resina com um gradiente crescente de imidazol, poderá permitir um maior nível de pureza da proteína recombinante. Para além destas opções, uma segunda purificação (e.g. purificação por interacção hidrofóbica) poderá permitir reduzir a quantidade de contaminantes e revelar-se como uma alternativa para se obterem melhores níveis de pureza a pequena escala. Outra alternativa já testada (mas que não foi incluída neste estudo por não se ter conseguido obter resultados satisfatórios) é realizar o *scale-*

up do crescimento celular/indução proteica da proteína recombinante de modo a conseguir purificar a proteína por FPLC para obtenção de melhores rendimentos.

Seria de todo interessante realizar também ensaios de crescimento celular/indução proteica apenas com a proteína nativa MolA, de modo a ser possível comparar o efeito do *tag* de expressão e de solubilização Fh8 nesta proteína.

Por outro lado, dado que a MolA é uma enzima dependente de Cobalto, as caudas de histidina presentes no plasmídeo podem, aquando da expressão, adsorver este metal. De modo a contornar este problema, a proteína poderá ter que ser fornecida sem a presença do *His-tag* do *tag* Fh8 (através da clivagem com a protease TEV) para desenvolvimento de posteriores estudos da bioactividade.

5. Bibliografia

- Amersham Pharmacia Biotech. *The Recombinant Protein Handbook: Protein Amplification and Simple Purification*. Australia: Amersham Pharmacia Biotech, 2000.
- Arvani, S. *Development of a novel T7 RNA polymerase based expression system in phototrophic bacterium Rhodospirillum rubrum*. International Graduate School of Biosciences Ruhr University Bochum, Bochum: Tese de Doutorado, Institute of Molecular Enzyme Technology, 2009.
- Azevedo, C. *Biologia Celular e Molecular*. Lisboa: LIDEL, 2005.
- Bell, C.E, and M. Lewis. "A closer view of the conformation of the Lac repressor bound to operator." *Nature structural biology* 7 (2000): 209-214.
- Canaves, J. M., R. Page, I. A. Wilson, and R. C. Stevens. "Protein biophysical properties that correlate with crystallization success in *Thermotoga maritima*: Maximum clustering strategy for structural genomics." *J. Mol. Biol.* 344 (2004): 977-991.
- Clark, D. *Molecular Biology*. California: Elsevier Academic Press, 2005.
- Costa, S, E. Coelho, L. Franco, A. Almeida, A. Castro, and L. Domingues. "The Fh8 tag: A fusion partner for simple and cost-effective protein purification in *Escherichia coli*." *Protein Expression and Purification* 92 (2013b): 163–170.
- Costa, S. J., A. Almeida, A. Castro, L. Domingues, and H. Besir. "The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology." *Appl Microbiol Biotechnol* 97 (2013a): 6779-6791.
- Demain, A. L., and P. Vaishnav. "Production of recombinant proteins by microbes and higher organisms." *Biotechnology Advances* 27 (2009): 297-306.
- Domingues, L. *Tecnologia de DNA Recombinante*. Universidade do Minho, Braga: Material didático da UC: Biotecnologia Molecular, 2012.
- Duarte, M., et al. "Gulosibacter molinivorax ON4T Molinate Hydrolase, a Novel Cobalt-Dependent Amidohydrolase." *Journal of Bacteriology* 193:10 (2011): 5810-5816.
- Dümmler, A., A. Lawrence, and A. Marco. "Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors." *Microbial Cell Factories* 4:34 (2005): 1-10.
- Fraga, H., T. Q. Faria, F. Pinto, A. Almeida, R. M. Brito, and A. Damas. "FH8 – a small EF-hand protein from *Fasciola hepatica*." *FEBS Journal* 277 (2010): 5072–5085.

- Francis, D. M., and R. Page. "Strategies to Optimize Protein Expression in E. coli." *Current Protocols in Protein Science* 5.24 (2010): 1-29.
- GE Healthcare Bio-Sciences. *Recombinant Protein Purification Handbook: Principles and Methods*. Sweden: GE Healthcare Bio-Sciences, 2009.
- Hitag Biotechnology, Lda. 2013. <http://hitag.pt/index.php/pt/technology/fh8-tag> (accessed 01 27, 2013).
- Laemmli, U. K. "Cleavage of structural proteins during assembly of the head of bacteriophage T4." *Nature* 227 (1970): 680-685.
- Lee, Pei Yun, John Costumbrado, Chih-Yuan Hsu, and Yong Hoon Kim. "Agarose gel electrophoresis for the separation of DNA fragments." *Journal of visualized experiments : JoVE*, 2012: 62.
- National Human Genome Research Institute. 2013. www.genome.gov (accessed 01 22, 2013).
- Novagen. *pET System Manual 11th Edition*. USA: Novagen, 2010.
- Nunes, O. C., and C. M. Manaia. "From Planting to Harvest: Environmental dissipation of the herbicide molinate and proposal of a clean-up methodology." In *Environmental Biodegradation Research Focus*. New York: B. Y. Wang Editor, 2007.
- Peti, W., and R. Page. "Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost." *Protein Expression and Purification* 51 (2007): 1-10.
- Promega. *Technical Manual - pGEM-T and pGEM-T Easy Vector Systems*. USA: Promega, 2010.
- Quiagen. *The QIAexpressionist - A handbook for high-level expression and purification of 6xHis-tagged proteins*. Qiagen, 2003.
- Rodrigues, L. *Valorização da fracção proteica do soro de queijo*. Universidade do Minho, Braga: Tese de Doutoramento, Departamento de Engenharia Biológica, 2001.
- Schägger, H. "Tricine–SDS-PAGE: Protocol." *Nature protocols* 1 (2006): 16-23.
- Scitable by nature education. 2013. <http://www.nature.com/scitable> (accessed 02 02, 2013).

Sørensen, H. P., and K. K. Mortensen. "Advanced genetic strategies for recombinant protein expression in *Escherichia coli*." *Journal of Biotechnology* 115 (2005): 113-128.

Teixeira, J. A., and M. M. Fonseca. *Reactores Biológicos*. Lisboa: LIDEL, 2007.

Terpe, K. "Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems." *Appl Microbiol Biotechnol* 72 (2006): 211-222.

Tropp, B. E. *Molecular Biology: Genes to Proteins*. London: Jones & Barlett Learning, LLC, 2012.

Videira, A. *Engenharia Genética: Princípios e Aplicações*. Lisboa: LIDEL, 2001.

Young, C. L., Z. T. Britton, and A.S. Robinson. "Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications." *Biotechnology Journal* 7 (2012): 620-634.

Anexos

Anexo I – Produtos químicos, soluções e reagentes

As composições, características e respectivas aplicações de cada solução utilizada nos procedimentos experimentais encontram-se indicados na Tabela AI. 1.

Tabela AI. 1: Soluções utilizadas nos procedimentos experimentais.

Solução	Composição e Características	Aplicação
Acrilamida 30%	Solução aquosa 29,1% Acrilamida e 0,8% Bis-Acrilamida	Análise SDS-Page
PSA 10%	Solução aquosa 10% PSA (p/v)	Análise SDS-Page
Tampão do ânodo	0,2M Tris; pH 8,9	Análise SDS-Page
Tampão do cátodo	0,1M Tris; 0,1M Tricina; 0,1% SDS; pH 8,25	Análise SDS-Page
Tampão do gel	3M Tris; 0,3% SDS; pH 8,45	Análise SDS-Page
Solução corante	0,05% corante Comassie Blue R	Análise SDS-Page
Solução descorante	45% Metanol; 10% Ácido acético	Análise SDS-Page
LB/Ampicilina	LB/LB agar; 100 µg/mL Ampicilina	Crescimento celular
LB/Ampicilina/X-Gal	LB/LB agar; 100 µg/mL Ampicilina; 42,5 µg/mL X-Gal	Crescimento celular
LB/Canamicina	LB/LB agar; 1mM Canamicina	Crescimento celular
Solução stock Ampicilina	Ampicilina (aq) 100 mg/mL; filtração 0,20 µm à chama; guardar a 4°C	Crescimento celular
Solução stock Canamicina	Canamicina (aq) 100 mg/mL; filtração 0,20 µm à chama; guardar a 4°C	Crescimento celular
LB	5 g/L NaCl; 5 g/L Extracto de levedura; 10 g/L Triptona	Crescimento celular em meio líquido
LB agar	5 g/L NaCl; 5 g/L Extracto de levedura; 10 g/L Triptona; 15 g/L agar	Crescimento celular em placa
Solução de Brometo de etídio	Brometo de etídio (aq) 5 mg/mL	Electroforese de DNA
Tampão TAE	40mM tris-acetato; 2mM EDTA; pH8,5	Electroforese de DNA
Solução stock IPTG	IPTG (aq) 1M; filtração 0,20 µm à chama; guardar a -20°C	Expressão proteica
Tampão de lise, de ligação e de lavagem 1	20mM Fosfato; 500mM NaCl; 20mM Imidazol; pH 8,0	Expressão proteica e mini-purificação proteica em condições nativas
Água ultrapura	Livre de DNAses e RNAses	Manipulação de DNA

Solução	Composição e Características	Aplicação
Tampão de eluição	(20mM Fosfato; 500mM NaCl e 300mM Imidazol); pH 8,0	Mini-purificação proteica em condições nativas
Tampão de lavagem 2	(20mM Fosfato; 500mM NaCl e 50mM Imidazol); pH 8,0	Mini-purificação proteica em condições nativas
Solução stock X-Gal	X-Gal (aq) 50 mg/mL; filtração 0,20 μ m à chama; guardar a -20°C	Seleccção de crescimento celular

A Tabela Al. 2 apresenta as características e origem dos reagentes e/ou produtos químicos utilizados nos procedimentos experimentais.

Tabela Al. 2: Produtos químicos e reagentes utilizados nos procedimentos experimentais

Reagente/Produto químico	Características	Origem
<u>Biologia Molecular</u>		
PCR		
MyTaq™ Red Mix	Mistura 2x concentrada; contém todos os reagentes necessários para uma reacção de PCR	Bioline
Enzimas de restrição		
NcoI	Local de corte: C [^] CATGG; Condições de incubação: tampão B ou U; 37°C; pode apresentar <i>star activity</i>	Nzytech
XhoI	Local de corte: C [^] TCGAG; Condições de incubação: tampão A ou U; 37°C	Nzytech
Ligases		
T4 DNA ligase	Catalisa a ligação de dois fragmentos de DNA com extremidades coesivas ou rombas. Condições de incubação: Tampão de ligação 10x concentrado; Temperatura ambiente	Nzytech
<i>Speedy</i> ligase	Catalisa de modo muito rápido a ligação de dois fragmentos de DNA com extremidades coesivas ou rombas. Condições de incubação: Tampão de ligação 4x concentrado; Temperatura ambiente	Nzytech
Vectores/plasmídeos		
pGEM-T Easy	Vector de clonagem de produtos de PCR com caudas de adeninas	Promega
pETMFh8	Vector com a maquinaria necessária para a expressão de proteínas recombinantes contendo o <i>tag</i> de solubilização Fh8.	

Reagente/Produto químico	Características	Origem
<u>Biologia Molecular</u>		
Electroforese de DNA		
Agarose – <i>SeaKem® LE</i>	Agarose para análises de rotina de ácidos nucleicos por electroforese em gel.	Lonza
Agarose		
Agarose <i>Low-Melting</i> – LM	Agarose de baixa temperatura de fusão com uma grande capacidade de resolução para pequenos fragmentos de DNA.	CONDA
SIEVE AGAROSE		
NZYDNA Ladder III	Marcador de pesos moleculares definidos que produz um padrão de 14 bandas de pesos moleculares compreendidos entre 200 pb e 10000 pb	Nzytech
<u>Expressão Proteica</u>		
Estirpes		
<i>E. coli</i> NZY5α	Estirpe do tipo (DH5α) utilizada para manipulação de DNA	Nzytech
<i>E. coli</i> BL21(DE3)	Estirpe de expressão proteica	Nzytech
Análise SDS-Page		
Tampão de amostra Laemmli	Tampão 4x concentrado; Permite óptimas resoluções das bandas em géis de poliacrilamida; Constituição: 277,8 mM Tris-HCl a pH 6,8; 44,4% (v/v) glicerol; 4,4% SDS e corante bromofenol a 0,02%	Bio-Rad
2-Mercaptoetanol	Agente redutor que destrói as pontes dissulfureto que são responsáveis pela estrutura nativa da proteína	Bio-Rad
<i>NZYColour Protein Marker II</i>	Marcador de pesos moleculares definidos que produz um padrão de 12 bandas correspondentes a 12 proteínas com pesos moleculares compreendidos entre 11 kDa e 245 kDa	Nzytech
<u>Purificação Proteica</u>		
<i>Resina Ni-NTA para IMAC – Ni</i>	Resina com grânulos de agarose (6%) empacotados e com um grupo quelante acoplado e carregado com Ni ²⁺ ; Tamanho médio: 34 μm; Capacidade de ligação ≥ 40 mg de proteína fundida com <i>His-tag</i> /mL resina; capacidade iónica metálica: aprox. 15 μmol Ni ²⁺ /mL resina.	Ge Healthcare
<i>Sepharose™ High</i>		
<i>Performance</i>		

Anexo II – Sequência dos *primers* de amplificação

Neste trabalho foram desenhados *primers* para amplificar o gene da MolA inserido em vectores do tipo pGEM-T Easy. Recorreu-se à técnica de PCR para amplificar o DNA de interesse e promover a inserção dos locais de restrição específicos das duas enzimas de restrição compatíveis com o vector pETMFh8 (*NcoI* e *XhoI*).

Deste modo, o *primer Forward* hibrida com o alvo pGEM MolA e introduz a enzima de restrição *NcoI*. O início da MolA tem os nucleótidos “ATGG” que coincidem com parte do local de restrição da *NcoI* que, por tal, se mantem. Com este *primer*, é possível obter 7 nucleótidos a hibridar 100% com o vector pGEM, 2 nucleótidos “CC” a introduzir parte do local de restrição da *NcoI*, 4 nucleótidos da MolA que coincidem com a parte restante da *NcoI* e 8 nucleótidos que continuam a ser 100% iguais à MolA.

Relativamente ao *primer Reverse*, este começa por hibridar na MolA, ter o codão STOP (este já existe na sequência codificante da MolA), introduz o local de restrição *XhoI*, e alguns nucleótidos que emparelhem com o restante vector pGEM MolA.

A Tabela AII. 1 apresenta a sequência dos *primers* desenhados para a amplificação do gene codificante da proteína MolA.

Tabela AII. 1: Sequência dos *primers* utilizados para amplificar o gene de interesse

DNA template	Primers	Sequência	Tm (°C)	%GC
pGEM MolA	Fwd <i>NcoI</i> MolA	GAATTCGCCATGGGAGAAACG	61,3	52,4
	Rev MolA <i>XhoI</i>	GAATTCACTCTCGAGCTAGTCGAG	65,2	50%

Anexo III – Composição dos géis de SDS-Page Tris-Tricina

A constituição dos géis de poliacrilamida utilizados neste trabalho é apresentada na Tabela AIII. 1.

Tabela AIII. 1: Composição dos reagentes utilizados num gel de SDS-Page Tris-Tricina

Reagente	Gel <i>Resolving</i> (15 %)	Gel <i>Stacking</i> (4%)
Acrilamida 40%	2000 µL	135 µL
Tampão do Gel	1750 µL	315 µL
Glicerol	500 µL	-
dH ₂ O	925 µL	1170 µL
PSA (10%)	100 µL	50 µL
TEMED	2,5 µL	5 µL

Anexo IV – Sequências codificantes das proteínas

Apresentam-se de seguida as sequências codificantes das duas construções realizadas na execução deste estudo.

Sequência da construção pGEM *NcoI* MolA *XhoI*

A proteína MolA nativa foi amplificada e promoveu-se a inserção dos locais de restrição específicos das duas enzimas de restrição compatíveis com o vector pETMFh8 (*NcoI* e *XhoI*), procedendo-se à sua ligação ao vector de clonagem pGEM-T Easy. Apresenta-se de seguida parte da sequência codificante da construção pGEM *NcoI* MolA *XhoI* de tamanho total 4437 pb (carece de confirmação, uma vez que, não foi confirmado por sequenciação).

```

..GATGAATTCGCCATGGGAGAAACGATCGCGATTGTCGGCGGCACACTCATCGACGAAATGGGGGAGTTCCCGTCCCGA
AACGACCGTGTTTATTGAAGACGGCAGGATCACGAAGGTTGGTTCACGGACCAGATCGAGGTACATCCGAACATCCGGCAG
ATCGATGCCAGGGGAAGTGGATTCTCCCGCCTGGTGAACGAAACGTGCACCTGCTGGACGGAATCATGATGATGGGTG
GGGGCGGCATCGAATACCTCGCCGATTCTGAAGGAACTACTACAAGGTCATCGAGGAAGCCGCCAAATCGCCCTCCGTG
GCGGGGTCACGACCGTGTTCGACACGTGGAACGCGCTCGAACCTGTCACGATCGCGCGGACCGGATCGCATCGGGTGCA
GCCGAAGGCGCCGCATCTTCTCGCGGGAACGCTCATCGGCATGGGCGGACCGTTCACCGGCGACTTCATGCGGCCGTG
ATGCAGGCGCGTACCGTCATGAGTCGCACCTTCGCCGATCGCATGGATGCAATGTTGAGGTCGGAATGGGGCGGCACCTC
AGTACGCTCCCGCCGAGAGGTGCGCCGTTGATCCGCGAGTATTTGGAGCGGGGAGTCGACTTCTGCAAGATCGCCGTC
ACTGATCACCTGGTGGGTCTCCTCGGATTCCGCGCCCGTACTTCACTTCTCGGAGCGGGTGTCTGATGTGCTGGTCGACG
AAGTTCGACGTGCCGGCGTCCCCCTCTCACGCATACGACCTCGTGAAGGCCTCAATACGGCGATCGAGCGGGATGCGG
ACCTGATGATCCACGCCACAATGACCGGGCAGGCCCGATTCCCGAGGAGACGATCGAGAAGCTCCTCGAGAAACAAGTGT
GGAGTGAAGTTCAGCCACGACCATCGCCAGCAGGCGTGGATGGACTCGGTCGACCACCCCTTCGCCGACTTCTCCGGGC
GGGTGCATCATGAGAATGACGTTTCGATGATCAAGGCCGGCGTCCGCTCGTCTGGAACTGACGCCGGGTGCACGGATC
CCGACATCCTCGAAGATATGTCCAGGGTGAATCCACGAGCGGCCGTGGACGCTCGGCGAGGACCACTTCGTGTGGATGC
AGGCAATGGTCGAGAAGGGTATGGATCCGATGGCAGCAATCTTTCGCGGGACTGCAACCCGGCAAGGCATACCGGAAGTT
CGACGAACTCGGCAGCATCGACGTCGGCAAGCTCGGGGACGTCGTGGTGTCTGACCAGGATCCGCTCGCCGACATCACGAA
CATGCGAACGCTGTCTCACGTCGTGAAGGAAGGGCGGGAGATCGACTTCCATGGACTTCCACTGAGTCCGTTGGTCACTGCG
TACCCTCGCACTGCGAACGTCCTCGACTAGCTCGAGAGTGAATTCATC...

```

Legenda:

- Sequência completa do gene que codifica a enzima MolA (inserido com as caudas de “T”)
- Início do local de restrição da *NcoI* que introduz a sequência do gene da MolA
- Local de restrição da *XhoI*

Sequência da construção pETMFh8 MolA

O gene que codifica a proteína MolA foi sintetizado/otimizado e inserido no vector de expressão pETMFh8 obtendo-se, assim, a construção pETMFh8 MolA. Apresenta-se de seguida parte da sequência codificante desta construção de tamanho total 6892 pb (confirmado após sequenciação).

```

...ATGAAACATCACCATCACCATCACCCATGTCCCCTAGTGTTCAAGAGGTTGAAAACTCCTTCATGTTCTCGATCGCAACG
GTGACGGGAAGGTTTCTGCCGAGGAGTTGAAAGCCTTCGCTGATGATTCAAAATGCCTCTGGACTCCAATAAGATCAAGGCT
TTCATTAAGGAACACGATAAAAAACAAGGATGGCAAGCTTGATTTGAAAGAAGCTGTTTCGATTTTGCATCAGAGAATCTTTATT
TTCAGGGCTCCATGGGAGAAACCATTGCAATTGTTGGTGGCACCCCTGATTGATGGTAATGGTGGTGTTCGGTCCGGAAACA
ACCGTTTTTATTGAAGATGGTCGTATTACCAAAGTGGGTAGCACCGATCAGATTGAAGTTCATCGAACATTTCGCAGATTGAT
GCACAGGGTAAATGGATTCTGCCTGGTCTGGTTAATGGTAATGTTTCATCTGCTGGATGGCATTATGATGATGGGTCGTGGTGG
TATTGAATATCTGGCACGTTTTGAAGGCAACTATTACAAAGTATTGAAGAAGCAGCACAGATTGCACTGCGTGGTGGTGTGAC
CACCGTTTTTGATACCTGGAATGCACTGGAACCGGTTACCATTGCACGTGATCGTATTGCAAGCGGTGCAGCAGAAGGTGCAC
GTATCTTTTTGCAGGTACACTGATTGGTATGGGTGGTCCGTTTACCAGTATTGATGCGTCCGAGCATGCAGGCACGTACC
GTTATGAGCCGTACCTTTCGAGATCGTATGGATGCAATGTTTGAAGTGGGTATGGGTGTCATCTGAGCACCCCTGCCTCCGGC
AGAAGTTCGTCCGCTGATCCGTGAGTATCTGGAACGTGGTGTGATTTTTGTAAAATTGCCGTTACCGATCATCTGGTTGGTCT
GCTGGGTTTTCGTGACCGTATTTTACCTTTAGCGAACGTGTTCTGGATGTTCTGGTTGATGAAGTGCCTGCGTCCGGTGTTC
CGCTGCTGACCATAACAACGACCTGGAAGGTCTGAATACCGCAATTGAACGTGATGCAGATCTGATGATTTCATGCAACCATG
ACAGGTCAGGCACCGATTCCGGAAGAAACAATTGAAAACTGCTGAAAAACAGCTGTGGTCAGAAGTTCAGCCGACCACCAT
TGCCAGCAGGCATGGATGGATAGCGTTGATCATCCGTTTGCAGATTTTAGCGGTGCGTTCATCATGAAAATGATGTGCGTA
TGATTAAGCAGGCGTTCCGCTGGTCTGGGCACCGATGCAGGTTGTACCGATCCGGATATTCTGGAAGATATGAGCCAGGG
TGAAGTGCATGAACGTCCGTGGACCCTGGTGAAGATCATTTTGTGGATGCAGGCAATGGTTGAGAAAGGTATGGATCCGA
TGGCAGCAATTCTGGCAGGCACCGCAAATCCGGCAAAGCATATCGTAAATTTGATGAACTGGGCAGCATTGATGTTGGTAAA
CTGGGTGATGTTGTGTGCTGGATCAGGATCCGCTGGCCGATATTACCAATATGCGTACCCTGAGCCATGTTGTTAAAGAAGG
TCGCGAAATTGATTTTCATGGTCTGCCGCTGTACCGCTGGTTACCGCATATCCGCTACCGCAAACGTTCTGGATTAACCTCG
AGCACCACCACCACCACCTGA...

```

- Sequência completa do gene que codifica a enzima MolA (otimizada)
- Início do local de restrição da *Nco*I que introduz a sequência do gene da MolA
- Local de restrição da *Xho*I
- Sequência codificante do *His-tag*
- Local de restrição da protease TEV
- Sequência codificante do *tag* Fh8

Anexo V – Histogramas e mapas de otimização do gene sintetizado

Neste trabalho a sequência codificante do gene *MolA* foi sintetizado. Através do software *GeneOptimizer*® otimizou-se a sequência codificante do gene e inseriram-se os locais de restrição específicos das duas enzimas de restrição compatíveis com o vector pETMFh8 (*Ncd* e *Xhd*). A síntese do gene resultou na otimização de codões (*codon usage*) e do conteúdo de GC, de modo a impedir alguns problemas de expressão.

Na Figura AV. 1 apresentam-se os histogramas e os mapas da distribuição da qualidade de codões, obtidos da otimização do gene sintetizado.

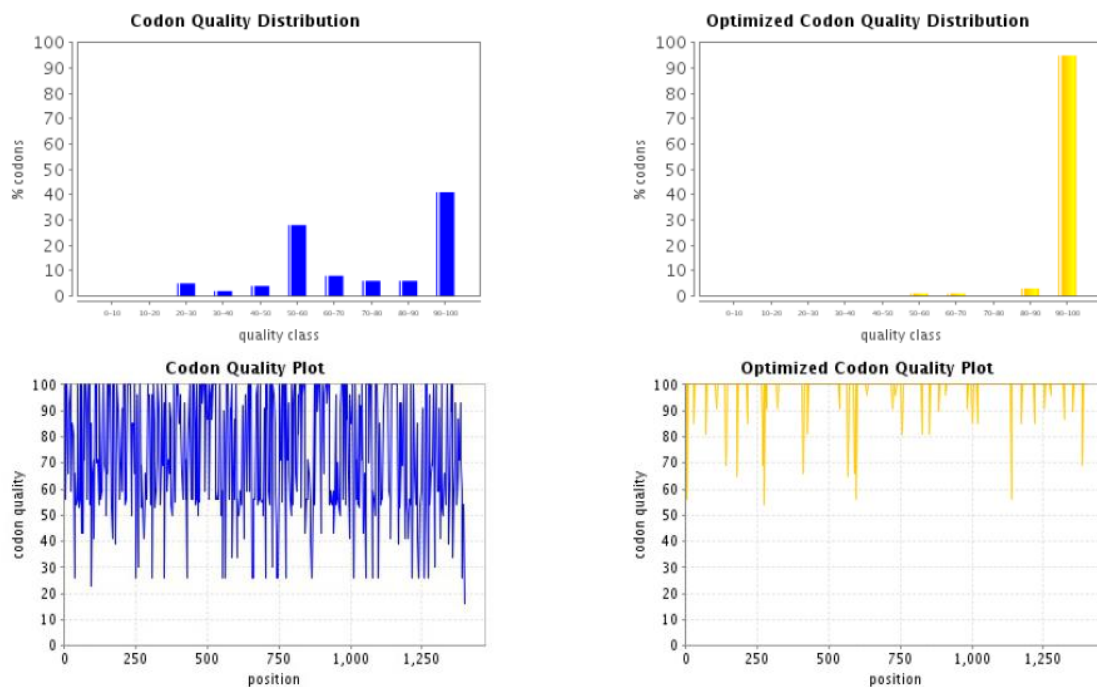


Figura AV. 1 – Ilustração comparativa entre os histogramas (em cima) e os mapas (em baixo) das distribuições de qualidade dos codões para o gene da *MolA* nativo e sintetizado, respectivamente.

Os histogramas mostram a percentagem de codões de seqüências que se enquadram numa certa categoria de qualidade. O valor de qualidade dos codões mais frequentemente utilizados para um determinado aminoácido no sistema de expressão em *E. coli* é definido como 100, os restantes codões são dimensionados em conformidade. Os mapas mostram a qualidade de codões usados em *E. coli* na respectiva posição.

Os mapas de distribuição do conteúdo de GC para o gene da MolA nativo e sintetizado apresentam-se na Figura AV. 2.

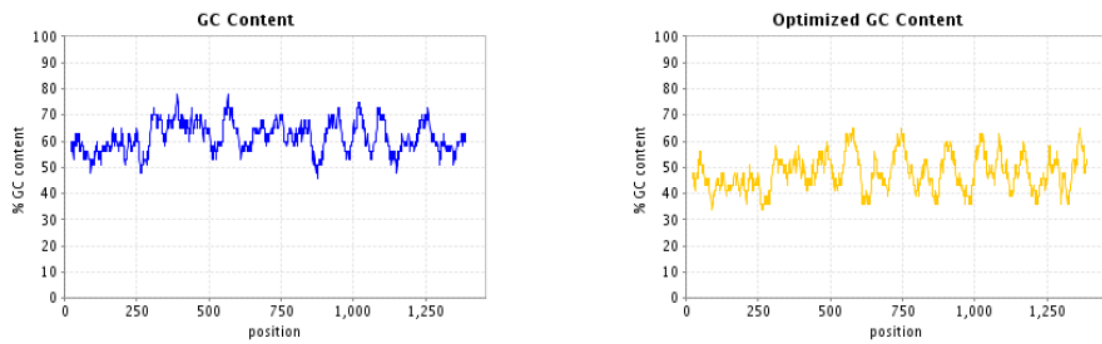


Figura AV. 2 – Ilustração comparativa entre os mapas de distribuição do conteúdo de GC para o gene da MolA nativo e sintetizado.