Universidade do Minho
Escola de Engenharia

Sofia Judite Marques da Costa

Development of a novel fusion system for recombinant protein production and purification in *Escherichia coli*

Janeiro 2013

Development of a novel fusion system for recombinant protein production and purification in *Escherichia coli*
Sofia Judite Marques da Costa
UMinho | 2013

Universidade do Minho
Escola de Engenharia

Sofia Judite Marques da Costa

Development of a novel fusion system for recombinant protein production and purification in *Escherichia coli*

Programa Doutoral em Engenharia Química e Biológica

Trabalho realizado sob a orientação da
Doutora Lucília Domingues
e do
Doutor António Castro

Janeiro 2013

Agradecimentos / Acknowledgments

Quero aqui agradecer a todos que direta ou indiretamente contribuíram de forma imprescindível para a concretização deste trabalho.

Aos meus orientadores Doutora Lucília Domingues e Doutor António Castro. Agradeço o apoio científico e a disponibilidade. Agradeço, principalmente, pela confiança que depositaram em mim, pela paciência e pelo estímulo científico de fazer mais e melhor, sempre! Agradeço todas as conversas e partilha de ideias sobre o trabalho que é aqui apresentado.

Ao meu orientador do período que passei no EMBL, Doctor Hüseyin Besir. I thank you for receiving me in the lab, and for sharing your knowledge with me. I thank you for the availability and motivation. And I thank the EMBL for offering me the opportunity to work in one of the best scientific environments of Europe!

Ao Doutor Frederico Silva (IBMC, Universidade do Porto). Agradeço por me ter recebido no seu laboratório e pela partilha de conhecimentos no capítulo 6.

Agradeço aos meus colegas do antigo LEMM e atual Plataforma de Biologia Molecular e Sintética do Departamento de Engenharia Biológica da Universidade do Minho (DEB-UM). Foram tantos os que por lá passaram nos últimos anos, que vou resumir o meu agradecimento pelo bom ambiente de trabalho e pelo companheirismo. Agradeço de forma muito especial à Tatiana, Diana e Francisco (por todos os *brain stormings*, por toda a ajuda no lab, mas principalmente pela amizade e carinho! E pelos momentos de risota e pelos almoços na cantina. Os dias no DEB sem vocês ficam sempre mais vazios). Agradeço ainda à Carla, à Orquídea, à Vera e à Susana pelo apoio prestado no dia-a-dia do DEB. Quero também agradecer à Xana, pela companhia, pela boa disposição e pelo gosto que me deu a ensinar! Agradeço ainda ao Frederico, à Cláudia, ao Jorge e ao Frank, pela ajuda nas experiências e/ou pelo companheirismo. E também à Dina, pela ajuda no Capítulo 5 e pela companhia.

Agradeço aos meus colegas do Instituto Nacional de Saúde Dr. Ricardo Jorge no Porto. Agradeço ao André, ao Pedro, à Elisabete, à Susana, à Lara, à Claire e à Sara por todos os momentos de partilha científica (e não só!), pela companhia e boa disposição! Agradeço à Lurdes, à Anabela, à Bela, à Rosa, ao Alexandre e ao Doutor José Manuel

por toda a ajuda prestada ao longo destes anos. Agradeço ainda ao Eduardo pela ajuda no meu trabalho, mas especialmente pela companhia e motivação no Capítulo 4.

Agradeço aos colaboradores da Hitag Biotechnology, Lda por acreditarem neste projeto e por levaram o *Fh8 fusion system* mais além.

I thank my colleagues at the EMBL, Ann-Marie, Emmanuel, Ines, Jeanette and Vladimir for the experimental help and their support. I thank all the other colleagues at the EMBL for our scientific talks, lunches and beer sessions.

I specially thank my friends that I met at the EMBL: Tom and Andrew, my dear neighbors, thank you for all the talks, the walking through the woods, and for being there when I needed! Niru, Nayla, Ramesh, Selva, Satya, and Riddhi, thank you *for laughing out loud* so often with the "three of us". You all were the best part of my 2011! Agradeço a todos os portugueses que conheci no EMBL, especialmente, à Andreia e à Raquel pela vossa companhia, pela amizade e por me fazerem sentir "em casa".

Agradeço aos meus amigos, por me fazerem sorrir tantas vezes! Agradeço em especial à Patrícia (obrigada por seres assim!). Agradeço à Ana, à Emília e ao Jorge por me acolherem tão bem e por serem também a minha família!

Agradeço ao Filipe. És a minha "força-metade". Obrigada por seres tu, por me incentivares sempre a ser melhor. Obrigada por toda a paciência nestes últimos anos.

Agradeço à minha família. Mãe, Pai, Carla, Nuno, Susana e Nel, obrigada por me apoiarem todos os dias e me fazerem rir tantas vezes. Beatriz, obrigada por me arrancares um sorriso, sempre! E ainda, obrigada ao mais recente membro, Mariana, por toda a ternura que trouxeste a esta família.

Finalmente, agradeço à Fundação para a Ciência e Tecnologia (FCT), pelo apoio financeiro através da bolsa SFRH/BD/46482/2008 e do projeto PTDC/CVT/103081/2008. Agradeço ainda ao qREN nº 3515.

"It must be a strange world not being a scientist, going through life not knowing - or maybe not caring - about where the air came from, and where the stars at night came from, or how far they are from us. ***I want to know****." (Michio Kaku)*

Dedico esta tese

à minha Família!

Abstract

Development of a novel fusion system for recombinant protein production and purification in *Escherichia coli*

Proteins are now widely produced in diverse microbial cell factories. The *Escherichia coli* is still the dominant host for recombinant protein production but, as a bacterial cell, it presents several issues limiting the efficiency of this production. The aggregation of foreign proteins into insoluble inclusion bodies is perhaps the main limiting problem found when expressing eukaryotic proteins in *E. coli*. Gene fusion technology has been widely used for the improvement of soluble protein expression and/or purification in *E. coli*. Fusion partners/tags are highly soluble and stable proteins that promote target protein solubility, possibly because of the protection of nascent polypeptides from the cytoplasm milieu and the given conditions for their proper folding.

The Fh8 is a calcium-binding protein from excreted-secreted antigens of *Fasciola hepatica* that showed to be highly soluble and stable when produced as a recombinant protein in *E. coli*. This recombinant protein is also a useful tool for the diagnosis of fasciolosis, and its N-terminal sequence of eleven amino acid residues (denominated as H) was suggested to play an important key role in the production and immunological properties of the entire protein. These Fh8 interesting features suggested this protein usage as a fusion partner for enhanced recombinant protein production in *E. coli*. This is the main focus of this thesis work.

The Fh8 and H peptides were analyzed as expression and solubility fusion partners, and at the same time, compared to commonly used fusion tags (Chapter 2). A broad range evaluation was conducted using six target proteins, eight fusion tags, two different induction conditions, and four *E. coli* expression strains. The results showed that Fh8 acts as an effective solubility enhancer tag, being ranked among the best solubility tags. It is also an excellent candidate to be used with other fusion tags in parallel high throughput screenings, presenting advantages over large tags for the evaluation of protein solubility. Results from this work also showed that the H tag did not function as a solubility enhancer tag, but it improves protein expression levels in *E. coli*.

The H tag was then suggested for the recombinant production and adjuvant-free administration of immunogens (Chapter 3). By using a 12-kDa antigen from *Cryptosporidium parvum* as example, the H tag demonstrated to be an attractive tool for the production of polyclonal

antibodies, overcoming several limitations of the process, namely, the availability of antigen, its immunogenicity, adverse effects of adjuvants, and unspecific antibody production.

The Fh8 fusion partner was further investigated as purification handle (Chapter 4). Taking into account its calcium-binding properties, the Fh8 offered a rapid, inexpensive and efficient single-step purification of biologically active target proteins via hydrophobic interaction chromatography (HIC), and under mild conditions. The efficiency of this purification strategy was comparable to that obtained using the His_6 tag, and both purification technologies were also combined into a dual affinity strategy to improve further the purity level of target proteins. Proteins purified by the Fh8-HIC strategy have the extra feature of being free of *E. coli* endotoxins.

Taking into account the versatility of the Fh8 fusion system, a novel strategy for the soluble production of bone morphogenetic protein-2 (BMP-2) and interleukin-10 was developed (Chapter 5). Fh8 fusion proteins were directly soluble produced in *E. coli*, presenting dimeric and oligomeric ordered conformations. The Fh8BMP-2 was, however, not functional, rising intriguing questions about the final structure of the fusion protein. Indeed, the Fh8 might direct the BMP-2 to a soluble folding pathway, but BMP-2 presents a different conformation from that required for its biological function.

Two novel variants of Fh8 were then developed as fusion partners, conducting the mutation of its single cysteine residue to alanine or tyrosine (Chapter 6). Fh8Ala and Fh8Tyr fusion proteins achieved similar solubility as Fh8-fused ones, presenting less calcium-dependent conformational changes and less oligomer forms than the Fh8-fused ones. These two mutations demonstrated the importance of the cysteine residue in Fh8 oligomerization, though other residues may also contribute to this state.

This thesis work reported for the first time the efficient use of the calcium-binding protein Fh8 as a promising gene fusion technology. The Fh8 fusion system is a robust tool for the recombinant protein production in *E. coli*, combining four main skills into such a small partner: protein expression, solubility, purification and immunogenicity.

Resumo

Desenvolvimento de um novo sistema de produção e isolamento de proteínas recombinantes em *Escherichia coli*

As proteínas são atualmente produzidas em diversas fábricas celulares microbianas, sendo a *Escherichia coli* o hospedeiro dominante na produção de proteínas recombinantes. Contudo, este hospedeiro apresenta alguns problemas que limitam a eficiência desta produção, sendo a agregação em corpos de inclusão insolúveis uma das suas principais limitações na expressão de proteínas eucariotas em *E. coli*. A tecnologia de fusão genética tem sido muito usada para melhorar a expressão e purificação de proteínas solúveis em *E. coli*. Os tags /parceiros de fusão são proteínas altamente solúveis e estáveis que promovem a solubilidade de proteínas alvo, possivelmente conferindo-lhes proteção no citoplasma, reunindo as condições necessários para o seu *folding* adequado.

O Fh8 é uma proteína de ligação ao cálcio, excretada-secretada pelos antigénios do *Fascíola hepatica*, que mostrou ser altamente solúvel e estável quando produzida em *E. coli*. O Fh8 é também uma ferramenta útil no diagnóstico da fasciolose, e a sua sequência N-terminal de onze aminoácidos (denominada de H) poderá ser importante na produção e nas propriedades imunológicas da proteína total. Todas estas características do Fh8 realçam o seu possível uso como parceiro de fusão para uma melhor produção de proteínas recombinantes em *E. coli*, sendo esse o objetivo principal do presente trabalho.

Os péptidos Fh8 e H foram analisados como parceiros de expressão e solubilidade, tendo sido também comparados com outros parceiros de fusão regularmente usados (Capítulo 2). Realizou-se uma profunda avaliação com recurso a seis proteínas alvo, oito tags de fusão, duas condições de indução diferentes e quatro estirpes de expressão de *E. coli*. Os resultados mostraram que o Fh8 é de facto um parceiro de fusão que promove a solubilidade, enquadrando-se entre os melhores tags de solubilidade. O Fh8 é também um bom candidato para ser usado com outros tags em extensas avaliações paralelas, apresentando vantagens em relação a tags maiores na avaliação da solubilidade das proteínas. Os resultados deste trabalho também mostraram que o tag H não funcionou como parceiro de solubilidade, mas este parceiro de fusão aumentou os níveis de expressão de proteínas em *E. coli*.

O tag H foi usado para a produção recombinante de imunogénios e para a sua administração sem adjuvantes (Capítulo 3). Usando um antigénio de 12 kDa de *Cryptosporidium parvum* como exemplo, o tag H demonstrou ser uma ferramenta interessante para a produção de

anticorpos policlonais, contornando várias limitações do processo, tais como, a disponibilidade do antigénio, a sua imunogenicidade, os efeitos adversos dos adjuvantes e uma produção de anticorpos pouco específica.

O parceiro de fusão Fh8 foi também estudado como tag de purificação (Capítulo 4). Tendo em conta as suas propriedades de ligação ao cálcio, o Fh8 possibilitou uma purificação rápida, barata e eficiente de proteínas biologicamente ativas numa só etapa, através da cromatografia de interação hidrofóbica (HIC) e em condições moderadas. Esta estratégia de purificação teve uma eficiência semelhante à obtida usando o tag His_6, e estas duas tecnologias de purificação foram ainda combinadas numa dupla purificação para aumentar o nível de pureza das proteínas alvo. A purificação por Fh8-HIC apresenta a mais-valia de obter proteínas livres de endotoxinas de *E. coli*.

Tendo em conta a versatilidade do sistema de fusão Fh8, desenvolveu-se uma nova estratégia para a produção solúvel da *bone morphogenetic protein-2* (BMP-2) e da *interleukin-10* (Capítulo 5). As proteínas de fusão com o Fh8 foram diretamente produzidas de forma solúvel em *E. coli*, apresentando dímeros e oligómeros numa conformação ordenada. No entanto, a Fh8BMP-2 não foi funcional, levantando questões sobre a estrutura final da proteína de fusão. O Fh8 poderá direcionar a BMP-2 para uma via de *folding* solúvel, sendo que, no entanto, a BMP-2 apresenta uma conformação diferente daquela exigida para a sua atividade biológica.

Neste trabalho, foram ainda desenvolvidos dois novos parceiros de fusão variantes do Fh8, tendo-se efetuado uma mutação no único aminoácido cisteína para alanina ou tirosina (Capítulo 6). As proteínas de fusão com Fh8Ala e Fh8Tyr alcançaram uma solubilidade semelhante à obtida pelas proteínas de fusão com Fh8, apresentando uma menor alteração conformacional dependente de cálcio e uma menor oligomerização comparativamente às proteínas de fusão com Fh8. Estes dois mutantes demonstraram a importância do aminoácido cisteína na oligomerização do Fh8; outros aminoácidos podem, porém, contribuir para este estado.

Este trabalho apresentou pela primeira vez o uso eficiente da proteína de ligação ao cálcio, o Fh8, como uma tecnologia de fusão genética promissora. O sistema de fusão Fh8 é uma ferramenta robusta para a produção de proteínas recombinantes em *E. coli*, combinando quatro características principais num pequeno tag: expressão proteica, solubilidade, purificação e imunogenicidade.

Publications

This thesis is based in the following original patents and publications:

Patents

Conceição, M., Costa, S., Castro, A. and Almeida, A. (2010) Fusion proteins, its preparation process and its application on recombinant protein expression systems. In: European Patent Office. WO 2010/082097.

Conceição, M., Costa, S., Castro, A. and Almeida, A. (2011) Immunogens, process for preparation and use in systems of polyclonal antibodies production. In: European Patent Office. WO 2011/071404.

Papers

Costa, S.J., Almeida, A., Castro, A., Domingues, L. and Besir, H. (2012) The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology. Applied Microbiology and Biotechnology. *In press*.

Costa, S.J., Silva, P., Almeida, A., Conceição, A., Domingues, L. and Castro, A. (2013) The novel adjuvant-free H fusion partner for the production of recombinant immunogens in *Escherichia coli*: its application to a 12-kDa antigen from *Cryptosporidium parvum*. *Submitted*

Costa, S. J., Coelho, E., Franco, L. Almeida, A., Castro, A. and Domingues, L. (2013) The Fh8 tag: a novel fusion partner for simple and inexpensive protein purification in *Escherichia coli*. *Submitted.*

Table of contents

List of figures

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

List of tables

CHAPTER 1

CHAPTER 2

CHAPTER 4

CHAPTER 5

CHAPTER 6

Abbreviations

Aa – amino acid(s)

Ala - alanine

ALP - alkaline phosphatase

ANS - 8-anilino-1-naphthalenesulfonic acid ammonium salt

ArsC – stress-responsive arsenate reductase

BB-SpG – albumin-binding region from streptococcal protein G

BMP – bone morphogenetic proteins

BMP-2 – bone morphogenetic protein-2

BSA – bovine serum albumin

CaBP – calcium-binding protein

CaM – calmodulin

CBD – chitin-binding domain

CBM – cellulose-binding module

CBP – calmodulin-binding protein

CD – circular dichroism

Ch8 – *Clonorchis sinensis* 8-kDa protein

CP12 – *Cryptosporidium parvum* oocyst wall 12-kDa protein

C-terminal – carboxyl terminal end

Cys – cysteine

Da – daltons

DLS – dynamic light scattering

DNA – deoxyribonucleic acid

Dock – dockerin domain of *Clostridium josui*

Dsb – disulfide bond proteins

DSF – differential scanning fluorimetry

Ecotin – *E. coli* trypsin inhibitor

EDTA – ethylenediamine tetraacetic acid

ELISA – enzyme-linked immunosorbent assay

EMEA – European Medicines Agency

EntK – Enterokinase

ESP – excreted-secreted antigens

FBS – fetal bovin serum

FDA – U. S. Food and Drug Administration

FITC – fluorescein isothiocyanate

Fh8 – *Fasciola hepatica* 8-kDa protein

Fh22 – *Fasciola hepatica* 22-kDa protein

FhCaM1 or 2 – *Fasciola hepatica* calmodulin-like proteins 1 or 2

FTL – frutalin

GB1 – IgG domain B1 of Protein G

GFP – green fluorescent protein

GST – gluthathione-S-transferase

GRAVY – grand average of hydropathicity

H – eleven amino acid residues of the Fh8 N-terminal sequence

HaloTag – mutated dehalogenase tag

HIC – hydrophobic interaction chromatography

His – histidine

hsHsp – heat shock proteins from Hsp70 and Hsp60 families

HTP – high throughput

HU – hemagglutination units

IEX – ion exchange chromatography

IFA – immunofluorescence assay

IL-10 – interleukin-10

IMAC – immobilized metal affinity chromatography

IMPACT – Intein mediated purification with the chitin-binding domain

IPTG – isopropyl β-D-thiolgalactopyranoside

MBP – maltose binding protein

MCS – multiple cloning site

Mocr – monomeric bacteriophage T7 0.3 protein (Orc protein)

mRNA – messenger ribonucleic acid

MW – molecular weight

Ni-NTA – nickel-nitrilotriacetic acid agarose

nt – nucleotide(s)

N-terminal – amino-terminal end

NusA – N-utilization substance

OD – optical density

o/n – overnight

Ori – origin of replication

PBS – phosphate buffer saline 1x

PBST – PBS with 0.3% (v/v) of Tween 20

PCR – polymerase chain reaction

pI – isolelectric point

PreScission – genetically engineered derivative of human rhinovirus 3C protease

R – Stokes radius

RFU – relative fluorescence units

RLU – relative luminescence units

RNA – ribonucleic acid

Rpm – revolutions per minute

RT – room temperature

RVS167 – reduced viability upon starvation protein 167

SDS-PAGE – sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEC – size exclusion chromatography

SET – solubility-enhancer peptide sequences

Sj8 – *Schistosoma japonicum* 8-kDa protein

Skp – seventeen kilodalton protein

Sm8 – *Schistosoma mansoni* 8-kDa protein

Smt3 – ubiquitin-like protein from *Saccharomyces cerevisiae*

SNUT – Solubility enhancing ubiquitous tag

SOD – superoxide dismutase

SpA – Staphylococcal protein A

SPO14 – phospholipase D1

Strep-II – streptavidin binding peptide

SUMO – small ubiquitin related modifier

Tamavidin – fungal avidin-like protein tag

TEV – tobacco etch virus

TGF-β – transforming growth factor-β

TgOWP – cyst wall protein-1 from *Toxoplasma gondii*

Thr – thrombin

TnC – troponin C

TP – target protein

Trx – thioredoxin

tRNA – transfer RNA

T7PK – phage T7 protein kinase

Tyr – tyrosine

Ub - ubiquitin

Ulp – ubiquitin-like protein(s)

UV – ultraviolet

Xa – activated blood coagulation factor X

YPK1 or 2 – serine/threonine protein kinase 1 or 2

ZZ – IgG repeat domain ZZ of Protein A

Aims and thesis planning

The *Escherichia coli* remains the dominant host for producing recombinant proteins, and its low cost and simplicity of cultivation makes it an unbeatable choice over other expression systems for lab and industrial applications. As a bacterial system, *E. coli* has however its limitations at producing foreign proteins from higher organisms. Here, the development of new strategies that promote the soluble expression of these proteins is essential for an efficient recombinant protein production, and for a successful protein application, for instance, in biopharmaceutical industry.

Gene fusion technology has attained increased attention in the past decades, being widely considered as an attractive tool for the improvement of soluble protein expression and purification of recombinant proteins in *E. coli*.

Fh8 is an 8-kDa excreted-secreted protein from the parasite *Fasciola hepatica*, and it has aroused interest for diagnosis of infections from this parasite, being also a good candidate for vaccine and drug development. The Fh8 is one of the smallest calcium-binding proteins described so far, and it is highly soluble and stable when expressed as a recombinant protein in *E. coli*. Moreover, its N-terminal sequence (herein denominated H) may play a key role in the production and immunological properties of the entire protein. As a result of their features, Fh8 and H peptides have been proposed to function as fusion partners for recombinant protein production in *E. coli*.

The main goal of this work is to develop a novel fusion system for soluble protein expression and purification in *E. coli*, using the Fh8 and H peptides as fusion partners.

To achieve this goal, novel expression vectors were constructed to carry the Fh8 or H peptides as N-terminal fusion partners, and several different proteins of interest were cloned and expressed in *E. coli* as Fh8- or H-fused proteins. These fusion proteins were then evaluated regarding their soluble expression, purification, and/or immunogenicity.

This dissertation is divided into seven chapters, as follows: in the first chapter, a general review of the main subjects of this work is presented; the experimental work and main

results obtained here are described from the second to the sixth chapter; and in the seventh chapter, the main conclusions and future perspectives are presented.

Specific aims of this thesis work:

- To evaluate both Fh8 and H peptides as solubility enhancer partners, and to compare this novel fusion system with the traditionally used solubility tags (Chapter 2);
- To evaluate the H peptide as a fusion partner for the adjuvant-free production of immunogens and corresponding polyclonal antibodies (Chapter 3);
- To study the Fh8 hydrophobic binding properties, and the Fh8 usefulness as a purification handle (Chapter 4);
- To soluble express and purify two proteins difficult-to-express in *E. coli* and with extensive biomedical applications (bone morphogenetic protein-2 and interleukin-10), using the Fh8 fusion system (Chapter 5);
- To characterize and evaluate two novel Fh8 variant fusion partners, and their effect in the Fh8 oligomerization and solubility properties (Chapter 6).

Chapter 1

General introduction

1.1. General introduction

Proteins are key elements of life, constituting the major part of the living cell. They play important roles in a variety of cell processes, namely, cell signaling, immune responses, cell adhesion, and the cell cycle, and their failure is consequently correlated with several diseases.

Amino acids are the basic building block units of proteins that form a polypeptide chain when covalently linked by peptide bonds. This polypeptide sequence constitutes the primary structure of a protein molecule, and it is then folded into a specific functional three-dimensional conformation driven by a large number of non-covalent interactions (as hydrogen bonds and hydrophobic interactions) between amino acids.

With the introduction of the DNA recombinant technology in the seventies, proteins started to be expressed in several host organisms resulting in a faster and easier process compared to their natural sources (Demain and Vaishnav, 2009).

Escherichia coli remains the dominant host for producing recombinant proteins, owing to its advantageous fast and inexpensive, and high yield protein production, together with the well-characterized genetics and variety of available molecular tools (Demain and Vaishnav, 2009).

The recombinant protein production in *E. coli* has greatly contributed for several structural studies, as indicated by the number of solved structures (90%) available at the Protein Data Bank (Nettleship et al., 2010; Bird, 2011). The *E. coli* recombinant production has also boosted the biopharmaceutical industry: 29.8% of the recombinant biopharmaceuticals licensed up to 2009 by the U. S. Food and Drug Administration (FDA) and European Medicines Agency (EMEA) were obtained using this host cell (Ferrer-Miralles et al., 2009; Walsh, 2010).

E. coli recombinant protein-based products can also be found in major sectors of the enzyme industry and the agricultural industry with applications ranging from catalysis (e.g. washing detergents) and therapeutic use (e.g. vaccine development) to functional analysis and structure determination (e.g. crystallography) (Demain and Vaishnav, 2009).

As a bacterial system, the *E. coli* has, however, limitations at expressing more complex proteins due to the lack of sophisticated machinery to perform posttranslational modifications, resulting in poor solubility of the protein of interest that are produced as inclusion bodies (Demain and Vaishnav, 2009; Kamionka, 2011). Previous studies (Bussow et al., 2005; Pacheco et al., 2012) reported that up to 75% of human proteins are successfully expressed in *E. coli* but only 25% are produced in an active soluble form using this host system. Other problems found within this host system include proper formation of disulfide bonds, absence of chaperones for the correct folding, and the miss-match between the codon usage of the host cell and the protein of interest (Terpe, 2006; Demain and Vaishnav, 2009; Pacheco et al., 2012). Moreover, the industrial culture of *E. coli* leads cells to grow in harsh conditions, resulting in cell physiology deterioration (Chou, 2007; Pacheco et al., 2012).

Despite the above-mentioned issues of *E. coli* recombinant protein expression, the benefits of cost and ease of use and scale make it essential to design new strategies directed for recombinant soluble protein production in this host cell.

Several strategies have been made for efficient expression of proteins in *E. coli*, namely, the use of different mutated host strains, co-expression of chaperones and foldases, lowering cultivation temperatures, and addition of a fusion partner (Terpe, 2006; Demain and Vaishnav, 2009). The combination of some of these strategies has improved the soluble protein expression of recombinant proteins in *E. coli*, but the prediction of robust soluble protein production processes is still a "a challenge and a necessity" (Jana and Deb, 2005).

Nowadays, with the aid of genetic and protein engineering, novel tailor-made strategies can be designed to suit user or process requirements.

In this chapter, a brief introduction of the main solubility factors that correlate with successful protein production in *E. coli* is first presented, followed by a comprehensive summary of the available fusion partners for protein expression and purification in the

bacterial host. A general presentation and characterization of the Fh8 as a calcium-binding protein (CaBP) is then described.

1.2. Recombinant protein production in *E. coli*

The above-mentioned advantages of *E. coli* ensured its unbeatable choice as host cell for protein production over other systems. In spite of all the extensive knowledge on genetics and molecular biology of *E. coli*, and despite all the improvements made so far, the efficient expression of recombinant proteins in this host cell is not always guaranteed (Jana and Deb, 2005).

The production of recombinant proteins requires a successful correlation between the protein expression, its solubility and its purification (Esposito and Chatterjee, 2006). The expression levels of recombinant proteins produced in *E. coli* are no longer pointed as a limitation for the success of the overall process, but care should be taken with the protein solubility, which is still a major bottleneck in the field. The downstream processing is deeply associated with an efficient protein production strategy, and thus it must be tailor-designed to maximize the recovery of pure recombinant proteins.

All these three properties – expression, solubility and purification – shall always be considered together as determinants for the effective protein production in *E. coli*. Several aspects are though essential for each individual success, as described below.

1.2.1. Key factors involved in the expression of recombinant proteins

When designing strategies for high-level expression of recombinant proteins in *E. coli*, key factors such as the expression vector, the stability and efficiency of mRNA, and differences in codon usage between the foreign gene and native *E. coli* should be considered, since they play an important role in the process efficiency and regulation (Schumann and Ferreira, 2004; Jana and Deb, 2005; Sorensen and Mortensen, 2005a)

The expression of recombinant proteins is usually induced from a plasmid-harbouring cell. The plasmid, or vector, contains central elements involved in the regulation of a high level of protein synthesis, such as:

i) the **replicon**: it contains the origin of replication (Ori) that controls the plasmid copy number (and consequently the gene dosage), and it prolongs the autonomous plasmid replication. Most of the expression vectors replicate by the ColE1 or the p15A replicon (Sorensen and Mortensen, 2005a).

ii) **transcriptional promoters**: they must exhibit certain features to achieve a desirable gene transcription, such as, strength, tight regulation (presenting a low basal expression level), easily transferable to other *E. coli* strains, and a simple and cost-effective induction. The activity of the promoter is modulated by a suitable repressor, which minimizes basal transcription, and may be present in the vector itself or may be integrated in the host chromosome. Promoter induction can be thermal or chemical, and the most common inducer is the lactose analog isopropyl β-D-thiolgalactopyranoside (IPTG) (Makrides, 1996; Hannig and Makrides, 1998; Jana and Deb, 2005; Sorensen and Mortensen, 2005a; Terpe, 2006).

iii) **antibiotic-resistance markers**: they are genes that confer resistance to the host cell against antibiotics, and are useful for plasmid selection and propagation. Ampicilin, kanamycin, chrolamphenicol or tetracycline are the most used resistance markers in recombinant expression plasmids (Jana and Deb, 2005; Sorensen and Mortensen, 2005a). The use of antibiotic resistance genes is, however, limited in gene therapeutic products. Moreover, antibiotics add extra costs in large-scale cultivation and impose a metabolic burden on the host cells. Antibiotic-free host-plasmid balanced lethal systems start to be used as an alternative to select and maintain the recombinant plasmids (Hägg et al., 2004; Dong et al., 2010; Peubez et al., 2010).

Expression vectors also contain other regions that play an important role in the efficiency of gene expression, such as, the **translational initiation region** (91% of *E. coli* sequenced genes have the initiation codon AUG, and the secondary structure at this region is determinant for a successful gene expression), and **transcriptional and translation terminators**, which are indispensible components for plasmid stability (preventing transcription through the origin of replication) and for translation termination (preferably mediated in *E. coli* by the stop codon UAA), respectively (Jana and Deb, 2005; Sorensen and Mortensen, 2005a).

The work presented in this thesis was mainly conducted using modified pET vectors (pETM), and also the pQE30 vector system (Qiagen). The pETM vectors (EMBL) derive from the pET vector series (Novagen) initially developed by Studier and colleagues, and represent nowadays a powerful system for cloning and expression of recombinant proteins in the *E. coli* host. The pETM plasmids use the same T7 promoter-based transcription-translation system as the pET collection. Some important features of pETM vectors (Figure 1.1.a) are: a conserved multiple cloning site (MCS), the presence of two 6xHis-tags (one before and other after the MCS), and a Tobacco etch virus (TEV) protease recognition site. The pQE-30 plasmid, represented in Figure 1.1.b, uses a transcription–translation system based in T5 promoter/lac operator, which allows high expression levels of recombinant proteins in *E. coli*. The regulation of expression is carried out using the low-copy plasmid pREP4, which confers kanamycin resistance, and constitutively expresses the lac repressor protein encoded by the *lac*I gene.

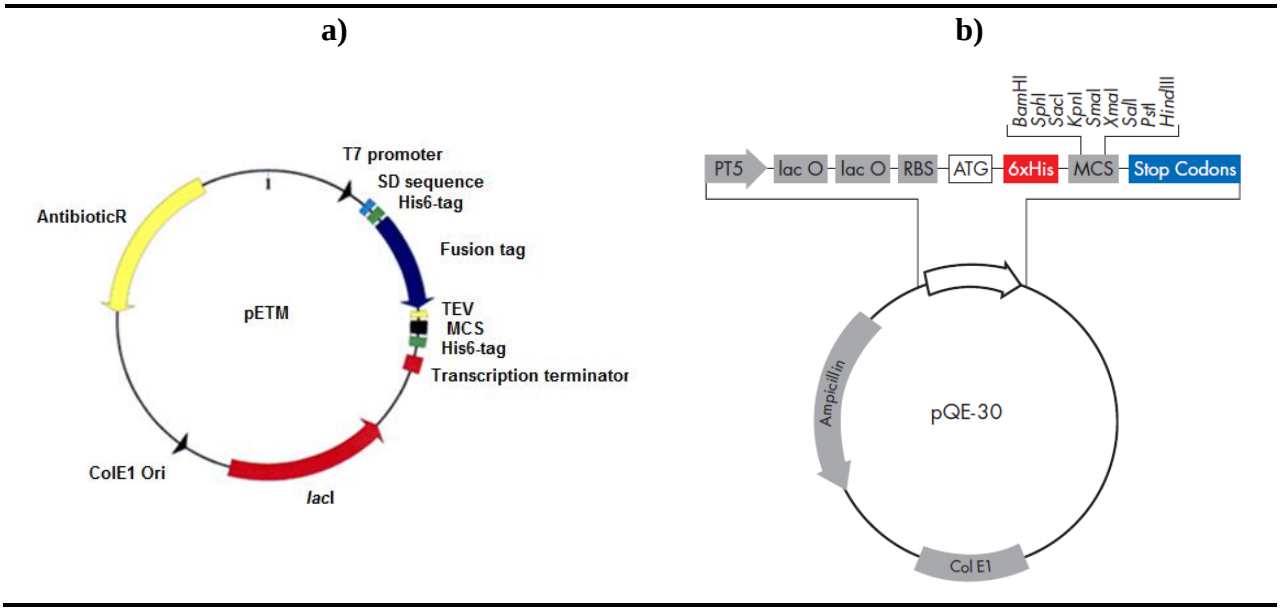


Figure 1.1. Schematic representation of the expression vectors used in this thesis. (**a**) pETM vector characteristics (adapted from EMBL): *ColE1* - origin of replication, *lac*I – repressor, *AntibioticR* – selection resistance marker, *T7 promoter* – transcriptional promoter, *SD sequence* – Shine-Delgarno sequence (required for translation initiation), *His6-tag* – hexahistine affinity tag, *Fusion tag* – fusion protein for soluble production, *MCS* – multiple cloning site, *TEV* – Tobacco Etch Virus protease, *Transcription terminator*. (**b**) pQE-30 vector characteristics (adapted from Qiagen): *ColE1* - origin of replication, *PT5* – T5 promoter, *lacO* – operator (it binds lac repressor), *RBS* – ribosome binding site, *6xHis* – hexahistidine tag, *MCS* - multiple cloning site.

The gene expression levels in *E. coli* are also dependent on the **messenger RNA** (mRNA) stability and its translation efficiency. A rapid degradation of mRNA may, thus, compromise protein production. The mRNA degradation process is mediated by RNases like endonucleases and 3´exonucleases. In spite of some nucleotide sequences have shown to prolong the half-life of several mRNAs, none of these stabilizing sequences works as a "universal stabilizer" (Hannig and Makrides, 1998; Jana and Deb, 2005). Hence, the mRNA stability should be tested on a case-study basis.

The **codon usage** between the host cell and target protein is often correlated with expression problems, affecting the protein production level and quality. *E. coli* presents a non-random usage tendency of synonymous codons. The frequency of use of these codons reveals the abundance of cognate tRNA in the cell cytoplasm (Hannig and Makrides, 1998; Schumann and Ferreira, 2004). The overexpression of heterologous genes enriched with codons that are rarely used in *E. coli* may result in a deficient synthesis of the corresponding protein. The location of rare codons can also interfere with the translational level (Schumann and Ferreira, 2004; Jana and Deb, 2005). Translational errors include mistranslational amino acid substitutions, frameshifting events, or premature translational termination (Kurland and Gallant, 1996; Sorensen and Mortensen, 2005a). The effect of preferential codon bias in *E. coli* can be minimized by genetic alteration of rare codons in the target gene, maintaining the codified protein, and by co-expression of rare tRNAs from additional plasmids, thus increasing the amount of appropriate cognate tRNA (Makrides, 1996; Hannig and Makrides, 1998; Schumann and Ferreira, 2004).

1.2.2. Strategies for the successful and efficient soluble protein production in *E. coli* – prevention of protein aggregation

E. coli recombinant protein expression systems are designed to achieve a high accumulation of soluble protein product in the bacterial cell. However, a strong and rapid protein production can lead to stressful situations for the host cell, resulting in protein misfolding *in vivo*, and consequent aggregation into inclusion bodies (Schumann and Ferreira, 2004; Sorensen and Mortensen, 2005b; Sorensen and Mortensen, 2005a; Sevastsyanovich et al., 2010). For instance, macromolecular crowding of proteins at

high concentrations in the *E. coli* cytoplasm often impairs the correct folding of proteins, leading to the formation of folding intermediates that, when inefficiently processed by molecular chaperones, promote inclusion body formation (Sorensen and Mortensen, 2005b; Sorensen and Mortensen, 2005a).

Apart from the poor quality of production, protein aggregation events are also associated with some diseases, such as Alzheimer's disease, and type II diabetes (Harper and Lansbury, 1997). Strategies that direct the soluble production of proteins in *E. coli* are, thus, envisaged, and become more attractive than protein refolding procedures from inclusion bodies.

Several methods have been shown to prevent or decrease protein aggregation during protein expression in *E. coli* on a trial-and-error basis, including:

i) **Lower expression temperatures**: protein cultivation at reduced temperatures is often used to reduce protein aggregation, since it slows down the rate of protein synthesis and folding kinetics, decreasing the hydrophobic interactions that are involved in protein self-aggregation (Schumann and Ferreira, 2004; Sorensen and Mortensen, 2005b). Low cultivation temperatures can also reduce or impair protein degradation due to a poor activity of heat shock proteases that are usually induced during protein overexpression in *E. coli* (Chesshyre and Hipkiss, 1989). This strategy has, however, some drawbacks as the reduction of temperature can also affect replication, transcription and translation rates, besides decreasing the bacterial growth and protein production yields. Nevertheless, these limitations can be circumvented by the use of cold-inducible promoters that maximize protein expression under low temperature conditions (Mujacic et al., 1999).

ii) ***E. coli*-engineered host strains**: *E. coli* mutant strains are a significant advance towards the soluble production of difficult recombinant proteins. Several targeted strain strategies have been developed through the introduction of DNA mutations that affect protein synthesis, degradation, secretion or folding (reviewed in Makino et al., 2011), including: mutated strains that increase mRNA stability by attenuation of RNases activity, which is responsible for the shorter half-life of mRNA in *E. coli* cells (Lopez et al., 1999); engineered strains that supply extra copies of rare tRNAs, such as the Rosetta strains (Invitrogen) and the BL21 Codon Plus strains (Novagen) (Baca and Hol, 2000; Sorensen et al., 2003b); mutant strains that facilitate disulfide bond formation and protein folding in the *E. coli* cytoplasm by render it oxidizing

due to mutations in glutathione reductase (*gor*) and thioredoxin reductase (*trxB*) genes, and/or by co-expression of Dsb proteins (Bessette et al., 1999; Lobstein et al., 2012), such as the Origami strains (Novagen) or the new SHuffle strain (New England Biolabs) (Lobstein et al., 2012); and C41 and C43 (Avidis) BL21 (DE3) mutant strains that improve the expression of membrane proteins (Miroux and Walker, 1996).

iii) **Cultivation conditions**: protein production can be efficiently improved by the use of high cell-density culture systems like batch, which offers a limited control of the cell growth, and fed-batch, which allows the real time optimization of growth conditions (Sorensen and Mortensen, 2005b). The composition of the cell growth medium and the fermentation variables such as temperature, pH, induction time, and inducer concentration are also essential for the prevention of protein aggregation, whereby a careful optimization improves the yield and quality of soluble protein production (Jana and Deb, 2005).

iv) **Co-expression of molecular chaperones and folding modulators**: The initial folding of proteins can be assisted by molecular chaperones that prevent protein aggregation through binding exposed hydrophobic patches on unfolded, partially folded or misfolded polypeptides, and traffic molecules to their subcellular destination. Protein aggregation is also prevented by folding catalysts that catalyze important events in protein folding such as the disulfide bond formation (Kolaj et al., 2009). A low concentration of these folding modulators in the cell often results in protein folding failures; thereby their co-expression together with the target protein becomes a suitable strategy for the improvement of soluble protein production in *E. coli* (reviewed in Thomas et al., 1997; Schlieker et al., 2002; Baneyx and Palumbo, 2003; Hoffmann and Rinas, 2004; Betiku, 2006; Gasser et al., 2008; Kolaj et al., 2009). Chaperones like trigger factor, DnaK, GroEL, members of the heat shock protein Hsp70 and Hsp60 families (hsHsp proteins), and ClpB assist protein folding in the *E. coli* cytoplasm, and their individual or cooperative activities presents different contributions for target protein solubility (Nishihara et al., 1998; Kuczynska-Wisnik et al., 2002; Schlieker et al., 2002; Deuerling et al., 2003; de Marco and De Marco, 2004; de Marco et al., 2007).

v) **Fusion partner proteins**: in contrast to the above-mentioned strategies, the use of fusion partners involves the target protein engineering. Fusion partners are very

stable peptide or protein molecules soluble expressed in *E. coli* that are genetically linked with target proteins to mediate their solubility and purification. The fusion protein technology is the core strategy applied in the work of this thesis. A summary of the available fusion protein moieties is presented in the next section.

1.2.3. Chromatographic strategies for recombinant protein purification

The protein purification accounts for most of the expenses in recombinant protein production. Hence, the design of a straightforward and cost-effective protein isolation and purification is one of the first steps to be considered in the production strategy.

There is no single or simple way to purify all kinds of proteins because of their diversity and different properties. Therefore, several strategies have been developed in the past decades to address a broad range of samples. With the introduction of recombinant DNA technology in the seventies, novel affinity tagging methodologies have revolutionized protein purification processes and several easy-to-use affinity tags have emerged since then. Besides the isolation of recombinant proteins, the purification process is also used to concentrate the desired protein. The target protein is usually first designed to be affinity tagged, thus facilitating the purification process and allowing the target protein to maintain its properties without interacting directly with a matrix. However, if the target protein cannot be affinity tagged or if further purification is needed, other purification strategies are added to the process.

When designing a purification strategy, one must consider the final goal of the target protein to be purified. For instance, recombinant proteins for therapeutic and biomedical applications require a high-level protein purity and they probably should undergo several subsequent purification steps.

The available protein purification methodologies separate the target proteins according to differences between the properties of the protein to be purified and properties of the rest of the protein mixture. Recombinant proteins are nowadays purified using column chromatography in scales from micrograms or milligrams in research laboratories to kilograms in industrial settings. The purification of a target protein from a crude cell extract is, however, not always easy and even with all the progresses achieved so far, additional physicochemical-based chromatography methods such as size exclusion

(SEC), ion exchange (IEX) and hydrophobic interaction (HIC) are often used to complement the affinity tagging. These methods rely on minor differences between various proteins properties such as size, charge and hydrophobicity, respectively (GE Healthcare, 2010).

In a traditional purification pipeline, the chromatography starts with a capturing step, where the target protein binds to the absorbent while the impurities do not. Then, weakly bound proteins are washed out of the column, and conditions are changed so that the target protein is eluted from the column.

Hydrophobic interaction chromatography (HIC):

HIC separates proteins according to differences in their surface hydrophobicity by using a reversible interaction between non-polar regions on the surface of these proteins and the immobilized hydrophobic ligands of a HIC medium (Queiroz et al., 2001). The proteins are separated according to differences in the amount of exposed hydrophobic amino acids. This technique is ideal for capture and intermediate steps in a multiple-step purification strategy.

The interaction between hydrophobic proteins and a HIC medium is influenced significantly by several parameters (reviewed in Queiroz et al., 2001; Lienqueo et al., 2007), including:

i) **the type of the ligand and degree of substitution**: the type of immobilized ligand (alkyl or aryl) determines the protein adsorption selectivity of the HIC adsorbent. In general, alkyl ligands show more pure hydrophobic character than aryl ligands. The protein binding capacities of HIC adsorbents increase with increased degree of substitution of immobilized ligand. At a reasonably high degree of ligand substitution, the apparent binding capacity of the adsorbent remains constant (the plateau is reached) but the strength of the interaction increases. Solutes bound under such circumstances are difficult to elute due to multi-point attachment (GE Healthcare, 2006).
ii) The **type of base matrix**: the matrix should be neutral to avoid ionic interactions between the protein and the matrix, and it should also be hydrophilic. The two most widely matrices are strongly hydrophilic carbohydrates, such as cross-linked agarose, or synthetic copolymer materials (GE Healthcare, 2006).

iii) **The type and concentration of salt**: a high salt concentration enhances the interaction, while lowering the salt concentration weakens the interaction. The effect of the salt type on protein retention follows the Hofmeister series (see Figure A1.1 in the Apenddix 1.6.2) for the precipitation of proteins from aqueous solutions (Collins and Washabaugh, 1985; Zhang and Cremer, 2006). In Hofmeister series, the chaotropic salts (magnesium sulfate and magnesium chloride) randomize the structure of the liquid water and thus tend to decrease the strength of hydrophobic interactions. In contrast, the kosmotropic salts (sodium, potassium or ammonium sulfates) promote hydrophobic interactions and protein precipitation, due to higher 'salting-out' or molar surface tension increment effects.

iv) **pH**: When pH is close to a protein's pI, net charge is zero and hydrophobic interactions are maximum, due to the minimum electrostatic repulsion between the protein molecules allowing them to get closer. In general, an increase in the pH weakens the hydrophobic interaction probably due to an increased titration of charged groups, thereby leading to an increase of protein hydrophilicity. A decrease of the pH may result in an increase of hydrophobic interactions. However, the effect of pH in HIC is not always straightforward (GE Healthcare, 2006).

v) **Temperature**: the role of temperature in HIC is complex, but in general, increased temperatures enhance the protein retention. Careful should be taken when conducting protein purifications at room temperature as the protein performance in the HIC will probably not be reproducible in a cold room, and *vice-versa*.

vi) **Additives**: Low concentrations of water-miscible alcohols, detergents and aqueous solutions of chaotropic (''salting-in'') salts result in a weakening of the protein-ligand interactions in HIC leading to the desorption of the bound solutes. The non-polar parts of alcohols and detergents compete with the bound proteins for the adsorption sites on the HIC media resulting in the displacement of the latter. Chaotropic salts affect the ordered structure of water and/or that of the bound proteins. Both types of additives also decrease the surface tension of water thus weakening the hydrophobic interactions to give a subsequent dissociation of the ligand-solute complex. The use of additives should be carefully considered as they might compromise the target protein structure and activity (GE Healthcare, 2006; GE Healthcare, 2010).

Proteins bound to HIC media can be eluted using some of the above-mentioned parameters such as reduced salt concentration, increased pH or addition of alcohols or detergents (Lienqueo et al., 2007), but trial-and-error experiments should be conducted to select the best option for each specific target protein.

Besides protein purification, the HIC methodology offers several potentialities in protein production, being described as one of the most used strategies for endotoxin clearance (Wilson et al., 2001; Magalhães et al., 2007; Ongkudon et al., 2012). It can also be used for protein refolding (Hwang et al., 2010).

The HIC methodology has been applied for the purification of calcium-binding proteins (CaBP) (Rozanas, 1998; Shimizu et al., 2003; McCluskey et al., 2007). These proteins expose a large hydrophobic surface in the presence of calcium that can absorb to hydrophobic matrices such as Phenyl Sepharose, even in the presence of low salt concentration. Most of the contaminant proteins will not bind under these conditions, which benefits the recovery of a pure CaBP. The elution step is often achieved by removal of the bound calcium through the use of chelating agents like EDTA (Rozanas, 1998).

Other chromatographic techniques:

i) **Ion exchange chromatography (IEX)** – this technique separates proteins with different surface charges and it offers a high-resolution separation combined with high sample loading capacity. The purification relies on a reversible interaction between a charged protein and an oppositely charged chromatography medium. Proteins purified by IEX are usually obtained in a concentrated form. The net surface of proteins is influenced by the surrounding pH: when the pH is above the protein isoelectric point (pI), the target protein has a negatively charged shield that is used for binding to a positively charged anion exchanger; when the pH is below its pI, the target protein has a positively charged shield that is used for binding to a negatively charged cation exchanger. The IEX purification protocol initiates under low ionic strength, and the conditions are then changed so that the bound substances can be eluted differentially by increasing salt concentration or changing pH using a gradient or stepwise strategy. In general, the IEX is used to bind the target protein, but it can also be used to bind impurities when required. The IEX is

the most common technique used for the capture step in a multiple-step purification strategy, but it can be used in the intermediate step as well (GE Healthcare, 2010).

ii) **Size exclusion chromatography (SEC)** – this technique is a non-binding method that separates protein samples with different molecular sizes under mild conditions. SEC can be used for protein purification, in which it usually dilutes the sample, or for group separation, which is mainly used for desalting and buffer exchange of samples. This technique is ideal for the final polishing in a multiple-step purification strategy. Analytical SEC allows the determination of the hydrodynamic radius of protein molecules and the corresponding molecular weight (GE Healthcare, 2010).

iii) **Affinity chromatography** – this technique separates proteins through a reversible interaction between the target protein and a specific ligand attached to a chromatographic matrix. The interaction can be performed via an antibody (biospecific interaction), or via an immobilized metal ion (nonbiospecific interaction) or dye substance. The affinity chromatography usually offers high selectivity and resolution together with an intermediate-high capacity. The sample is first bound to the ligand using favorable conditions for that binding. Then, the unbound material is washed out of the column and the elution of pure protein is achieved using a competitive ligand or by changing the pH, ionic strength or polarity (GE Healthcare, 2010). This purification strategy can profit from the use of recombinant DNA technology as the affinity tag can be fused to the protein of interest during cloning and it is further presented in the next section.

1.3. Fusion protein technology

Fusion partners or tags are incorporated into *E. coli* expression vectors to improve protein expression yields, solubility and folding, and to facilitate protein purification. They can also confer specific properties for target proteins characterization and study, such as protein immunodetection, quantification, and structural and interactional studies (Malhotra, 2009). Besides the fusion(s) partner(s), *E. coli* expression vectors can contain a protease cleavage site between the fusion partner and passenger protein that allows the tag removal when the latter protein is for using in protein therapies, vaccine development and structural analyses.

Some fusion partners also protect target proteins from degradation by promoting the translocation of the passenger protein to different cellular locations, where less protease content exists (Butt et al., 2005). Both maltose-binding protein (MBP) and small ubiquitin related modifier (SUMO) fusion partners present this feature, passing target proteins from the *E. coli* cytosol for cell membrane and nucleus, respectively (Nikaido, 1994; Kishi et al., 2003).

When designing a fusion strategy, the choice of the fusion partner depends on several aspects (Young et al., 2012), including:

i) **purpose of the fusion**: is it for solubility improvement or for affinity purification? Nowadays, a variety of fusion tags that render different purposes are available, and systems containing both solubility and affinity tags like, for instance, the dual hexahistine (His_6)-MBP tag, can be designed in order to get a rapid "in one step" protein production. Some protein tags can also function in both affinity and solubility roles, as for instance, the MBP or gluthathione-S-transferase (GST) (Esposito and Chatterjee, 2006). If the fusion tag is to be used in protein purification, the cost and buffer conditions are often the criteria for selection. For instance, proteins that require reducing agents as EDTA are not suitable for immobilized metal affinity chromatography (IMAC) via the His_6 tag as nickel ions in the affinity matrix are chelated by EDTA (Malhotra, 2009).
ii) **amino acid composition** and **size**: these two factors should be considered when selecting a fusion partner because target proteins may require larger or smaller tags

depending on their application. Larger tags can present a major diversity in the amino acid content, and will impose a metabolic burden in the host cell different from that imposed by small tags (Malhotra, 2009).

iii) **required expression levels**: structural studies require higher protein expression levels that can be rapidly achieved with a larger fusion tag, which has strong translational initiation signals, whereas the study of physiological interactions demands for lower expression levels and small tags (Malhotra, 2009).

iv) **tag location**: Fusion partners can promote different effects when located at the N-terminus or C-terminus of the passenger protein. Usually, N-terminal tags are advantageous over C-terminal tags because: (1) they provide a reliable context for efficient translation initiation, in which fusion proteins take advantage of efficient translation initiation sites on the tag; (2) they can be removed leaving none or few additional residues at the native N-terminal sequence of the target protein, since most of endoproteases cleave at or near the C-terminus of their recognition sites (Waugh, 2005; Malhotra, 2009).

In spite of all the approaches conducted so far, the choice of a fusion partner is still a trial-and-error experience. Fusion partners do not perform equally with all target proteins, and each target protein can be differentially affected by several fusion tags (Esposito and Chatterjee, 2006). In the past decade, parallel high throughput (HTP) screenings using different fusion partners have developed soluble protein production, and facilitated a rapid, tailored and cost-effective choice of the best fusion partner for each target protein (Hammarstrom et al., 2002; Shih et al., 2002; Dyson et al., 2004; Dummler et al., 2005; Cabrita et al., 2006; Hammarstrom, 2006; Marblestone et al., 2006; Kim and Lee, 2008; Kohl et al., 2008; Ohana et al., 2009; Bird, 2011).

Fusion tags can be incorporated using different strategies: affinity and solubility tags are set individually or together, and sites for protease cleavage are designed between the fusion tags and target proteins.

Figure 1.2 illustrates the schematic pathway from protein expression to purification using solubility and affinity tags.

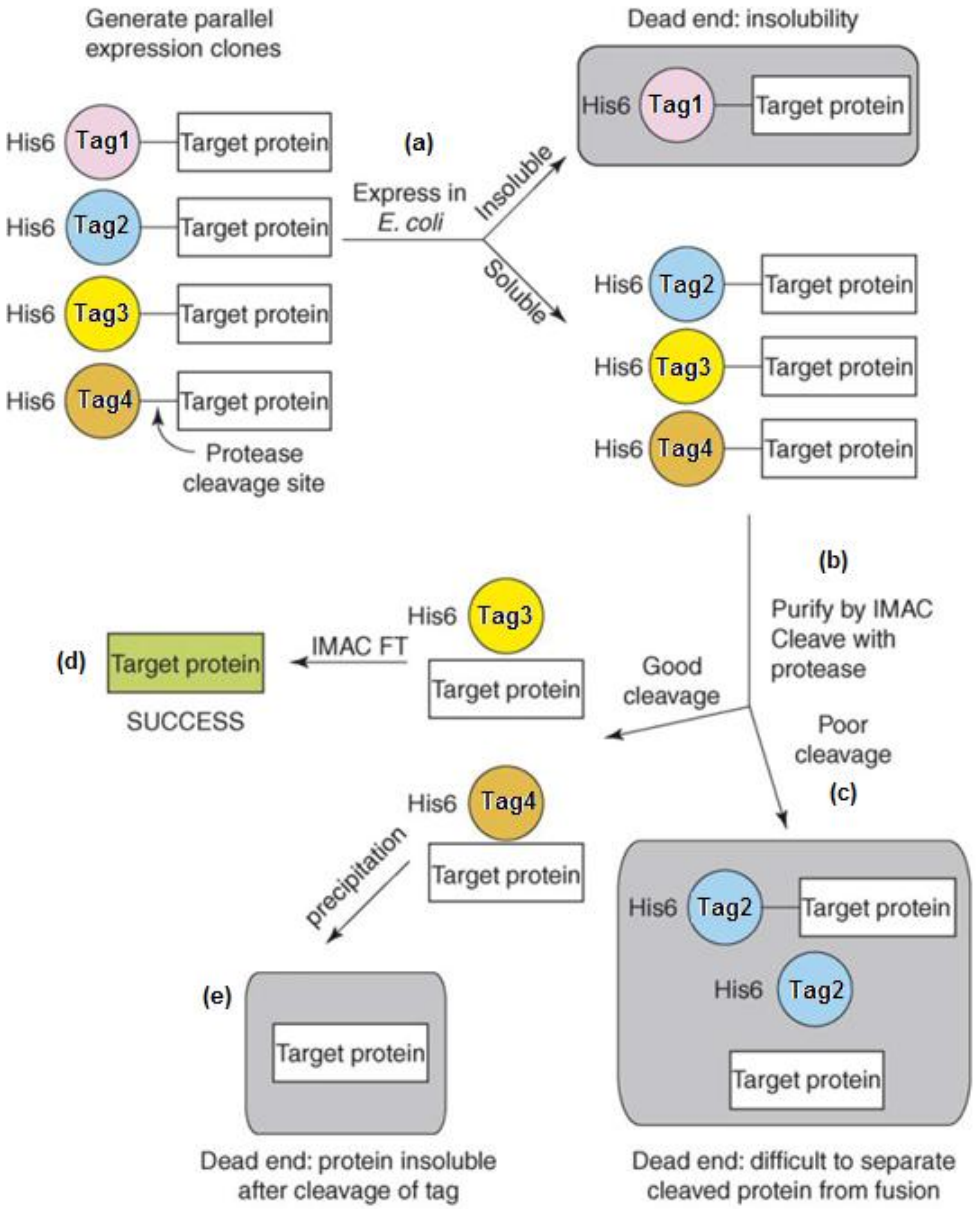


Figure 1.2. Schematic pathway from protein expression to purification using solubility tags and the hexahistidine (His6) affinity tag (adapted from Esposito and Chatterjee, 2006). (**a**) Four tagged versions of the target protein are expressed in *E. coli*, and some fusions will end-up in the insoluble fraction (as occurred with *Tag1*) whereas others remain in the soluble fraction (as occurred with *Tag2*, *Tag3*, and *Tag4*). (**b**) Soluble fusion proteins are then purified by immobilized metal affinity chromatography (IMAC) using the His6 tag, and the fusion tags are removed from the target protein by protease cleavage. (**c**) Some fusions will not cleave efficiently, resulting in a mixture of cleaved and uncleaved proteins that are difficult to separate (as occurred with *Tag2*). (**d**) Other fusions will cleave efficiently, and the target protein will remain in solution, being collected in the flow-through sample of a second IMAC purification step (as occurred with *Tag3*). (**e**) Despite a successful protease cleavage, some target protein becomes insoluble after tag removal (as occurred with *Tag4*).

The mechanisms by which fusion tags enhance the solubility of their partner proteins remain unclear, but several hypotheses have been suggested (Butt et al., 2005; Nallamsetty and Waugh, 2007):

i) fusion proteins form **micelle-like structures**: misfolded or unfolded proteins are sequestered and protected from the solvent and the soluble protein domains face outward;
ii) fusion partners **attract chaperones**: the fusion tag drives its partner protein into a chaperone-mediated folding pathway. MBP and N-utilization substance (NusA) are two fusion tags that present this mechanism, being previously reported to interact with GroEL in *E. coli* (Huang and Chuang, 1999; Douette et al., 2005);
iii) fusion partners have an **intrinsic chaperone-like activity**: hydrophobic patches of the fusion tag interact with partially folded passenger proteins, preventing their self-aggregation, and promoting their proper folding. MBP was previously reported to act also as a chaperone in the fusion context (Kapust and Waugh, 1999; Fox et al., 2001). Solubility enhancer partners may thus play a passive role in the folding of their target proteins, reducing the chances for protein aggregation (Waugh, 2005; Nallamsetty and Waugh, 2006);
iv) fusion partners **net charges**: highly acidic fusion partners were suggested to inhibit protein aggregation by electrostatic repulsion (Zhang et al., 2004; Su et al., 2007).

1.3.1. Solubility enhancer partners

A large variety of solubility enhancer tags are available (Table 1.1), including the well-known MBP, NusA, thioredoxin (TrxA), GST, and SUMO, and several other novel moieties recently discovered.

Table 1.1. Solubility enhancer tags (adapted from Esposito and Chatterjee, 2006; Malhotra, 2009):

Tag	Protein	Size (aa)	Organism	Reference
MBP	Maltose-binding protein	396	*E. coli*	(di Guan et al., 1988; Kapust and Waugh, 1999)
NusA	N-Utilization substance	495	*E. coli*	(Davis et al., 1999)
Trx	Thioredoxin	109	*E. coli*	(Lavallie et al., 1993)
SUMO	Small ubiquitin modified	~100	*Homo sapiens*	(Butt et al., 2005; Marblestone et al., 2006)
GST	Gluthathione-S-transferase	211	*Schistosoma japonicum*	(Smith and Johnson, 1988)
SET	Solubility-enhancer peptide sequences	<20	Synthetic	(Zhang et al., 2004)
GB1	IgG domain B1 of Protein G	56	*Streptococcus sp.*	(Zhou et al., 2001; Cheng and Patel, 2004)
ZZ	IgG repeat domain ZZ of Protein A	116	*Staphylococcus aureus*	(Rondahl et al., 1992; Inouye and Sahara, 2009)
HaloTag	Mutated dehalogenase	~300	*Rhodococcus sp.*	(Ohana et al., 2009)
SNUT	*S*olubility e*N*hancing *U*biquitous *T*ag	147	*Staphylococcus aureus*	(Caswell et al., 2010)
Skp	Seventeen kilodalton protein	161	*E. coli*	(Esposito and Chatterjee, 2006)
T7PK	Phage T7 protein kinase	~240	*Bacteriophage T7*	(Esposito and Chatterjee, 2006)
EspA	*E. coli* secreted protein A	192	*E. coli*	(Cheng et al., 2010)
Mocr	Monomeric bacteriophage T7 0.3 protein (Orc protein)	117	*Bacteriophage T7*	(DelProposto et al., 2009)
Ecotin	*E. coli* trypsin inhibitor	162	*E. coli*	(Malik et al., 2006; Malik et al., 2007)
CaBP	Calcium-binding protein	134	*Entamoeba histolytica*	(Reddi et al., 2002)

ArsC	Stress-responsive arsenate reductase	141	*E. coli*	(Song et al., 2011)
IF2-Domain I	N-terminal fragment of translation initiation factor IF2	158	*E. coli*	(Sorensen et al., 2003a)
Expressivity tag(part of IF2-Domain I)	N-terminal fragment of translation initiation factor IF2	7 (21 nt)	*E. coli*	(Hansted et al., 2011)
RpoA, SlyD, Tsf, RpoS, PotD, Crr	Stress-responsive proteins	329, 196, 283, 330, 348, 169	*E. coli*	(Ahn et al., 2007; Han et al., 2007a; Han et al., 2007c; Han et al., 2007b; Park et al., 2008)
msyB, yjgD, rpoD	*E. coli* acidic proteins	124, 138, 613	*E. coli*	(Su et al., 2007; Zou et al., 2008)

aa – amino acids; nt - nucleotides

MBP is a large (43 kDa) periplasmic and highly soluble protein of *E. coli* that acts as a solubility enhancer tag (Kapust and Waugh, 1999; Fox et al., 2001), and it has a native affinity property to function as a purification handle.

MBP plays an important role in the translocation of maltose and maltodextrins (Nikaido, 1994): it has a natural protein-binding site that it uses to interact with other proteins involved in maltose signaling and chemotaxis, and it has a large hydrophobic cleft close to this site that undergoes conformational changes upon maltose binding (Fox et al., 2001).

When used in the fusion context, MBP promotes target protein solubility by showing chaperone intrinsic activity (Kapust and Waugh, 1999; Bach et al., 2001; Fox et al., 2001), and it is more efficient at the N-terminus of the target proteins rather than at the C-terminus (Sachdev and Chirgwin, 2000). In fact, MBP promotes the proper folding of the target protein by interacting with the latter, and occluding its self-association. This passive role of MBP in protein folding is correlated with the large hydrophobic area exposed on its surface, which is responsible for the contact with other proteins in the maltose transport apparatus (Kapust and Waugh, 1999; Fox et al., 2001). Hence, the MBP hydrophobic cleft is pointed as the site where fused polypeptides interact with the fusion partner (Kapust and Waugh, 1999; Fox et al., 2001; Nallamsetty and Waugh,

2007), similar to what it is reported for GroEL and DnaK molecular chaperones (Buckle et al., 1997; Chatellier et al., 1999; Tanaka and Fersht, 1999). The presence of this cleft can explain why only certain soluble proteins like MBP act as solubilizing agents. Moreover, MBP presents certain conformational flexibility associated with the cleft; thereby it can adjust its shape to accommodate several different polypeptides.

MBP fusion proteins bind to immobilized amylose resins, but this binding is highly dependent on the nature of the passenger protein as it can block or reduce the amylose interaction (Pryor and Leiting, 1997). Difficulties found in the binding of MBP fusion proteins to amylose resins corroborate the hypothesis that target proteins interact with MBP via its binding site (Fox et al., 2001).

Other affinity tags, specific proteases and protein cultivation strategies are being employed together with MBP to improve protein soluble expression, purification and native protein recovery, as for instance, His_6-MBP fusions (Nallamsetty et al., 2005), His_6-MBP-TEV fusions (Rocco et al., 2008), MBP-His_6-Smt3 fusions in which the *Saccharomyces cerevisiae* Smt3 protein is used for protein processing by proteolytic cleavage between the MBP-His_6 tags and the protein of interest (Motejadded and Altenbuchner, 2009), and secretion of MBP fusion protein into the culture medium (Sommer et al., 2009).

Several commercial expression vectors containing the MBP tag are available for cytoplasmic and periplasmic expression of target proteins, including the pMAL series (New England Biolabs) and pIVEX (Roche).

NusA is a transcription anti-termination factor that promotes pauses in DNA transcription by RNA polymerase. NusA (55 kDa) is used as a fusion partner to confer stability and high solubility to its target proteins (De Marco et al., 2004; Dummler et al., 2005; Turner et al., 2005). The NusA ability to improve the soluble production of fusion proteins may be correlated with its intrinsically solubility and biological activity in *E. coli*. NusA slows down translation at the transcriptional pauses, offering more time for protein folding (Davis et al., 1999; De Marco et al., 2004). In contrast to MBP, NusA does not present an intrinsic affinity property, therefore requiring the addition of an affinity tag for efficient protein production, as for instance, the His_6 tag (Davis et al., 1999). As for MBP, several strategies have been exploited to use the NusA solubility

enhancer fusion partner with purification tags and specific proteases like the pETM60 vector (EMBL) (De Marco et al., 2004) that render the expression of a NusA-His_6-TEV fusion protein, or the pET43 (Novagen), that offers the same NusA-His_6 fusion protein but with a thrombin and enterokinase cleavage sites between the fusion tags and target proteins.

In spite of the different physiochemical and structural properties, as well as different biological functions, **MBP** and **NusA** are often reported to promote similar solubility improvements in their target proteins, being ranked as two of the best tags for making soluble proteins (Shih et al., 2002; Kohl et al., 2008; Bird, 2011). Both fusion partners were reported to probably work by similar mechanisms, in which NusA, like MBP, plays a passive role on the target protein folding (Nallamsetty and Waugh, 2006).

TrxA, or **Trx**, is a 12-kDa intracellular thermostable protein of *E. coli* that is highly soluble expressed in its cytoplasm (Young et al., 2012). The *E. coli* Trx can be used for co-expression with a target protein, improving the solubility of the latter (Yasukawa et al., 1995). Trx is also commonly employed as a fusion tag to avoid inclusion body formation in recombinant protein production by taking advantage of its intrinsic oxido-reductase activity responsible for the reduction of disulfide bonds through thio-disulfide exchange (Stewart et al., 1998; LaVallie et al., 2000; Young et al., 2012). The fusion partner Trx can be placed both at the N- or C-terminal of target proteins (LaVallie et al., 2000) but this fusion partner is more effective at the N-terminal of the target protein (Terpe, 2003; Dyson et al., 2004). In some high throughput screenings (Hammarstrom et al., 2002; Dyson et al., 2004; Kim and Lee, 2008), the **Trx** fusion partner improves target protein solubility similar to **MBP** tag, being considered one of the best choices for protein production in *E. coli*.

Unlike MBP, Trx does not have intrinsic affinity properties, thus requiring an additional fusion tag for protein purification such as the His_6 tag. The pET32 (Novagen), one of the commercially available vectors for Trx tagging, carries this dual-fusion partners for protein expression and purification (Austin, 2003).

Trx fusion partner has been useful in protein crystallization of its target proteins because it readily form several crystals itself, and it offers a rigid connection to the target protein, which is an essential feature for blocking conformational heterogeneity usually

found in various attempts of fusion proteins crystallization (Smyth et al., 2003; Corsini et al., 2008).

SUMO is a small protein (~11 kDa) found in yeast (one single gene codifying for Smt3) and vertebrates (three genes codifying for SUMO-1, SUMO-2, and SUMO-3) (Kawabe et al., 2000) that has recently been used as an effective N-terminal solubility enhancer fusion partner, offering advantages over other fusion systems (Marblestone et al., 2006; Bird, 2011).

The SUMO fusion partner offers the production of the target protein or peptide with its native N-terminal amino acid composition. Moreover, the robust SUMO protease (catalytic domains of Ulp1) recognizes the tertiary structure of SUMO, offering a significant advantage over other endoproteases as it does not present unspecific cleavage of the protein linear amino acid sequence (Malakhov et al., 2004; Marblestone et al., 2006).

SUMO promotes the proper folding and solubility of its target proteins possibly by exerting chaperoning effects in a similar mechanism to the described for its structural homologous Ubiquitin (Ub) (Khorasanizadeh et al., 1996). Ub was reported to be the nature's fastest folding protein, and SUMO also present a tight, rapidly folding soluble structure (Marblestone et al., 2006). In addition, Ub and Ub-like proteins (Ulp) have a highly hydrophobic inner core and a hydrophilic surface that, together with such a rapid folding, may explain the SUMO behaviour as a nucleation site for the proper folding of target proteins (Malakhov et al., 2004; Marblestone et al., 2006).

SUMO fusion proteins or peptides are usually purified by affinity chromatography using the His_6 tag (Lee et al., 2008; Gao et al., 2010; Wang et al., 2010; Satakarni and Curtis, 2011). Due to its unique features, SUMO technology has being constantly explored, and novel strategies for a facile and rapid protein production are now available, as the SUMO-intein system (Wang et al., 2012). The SUMO fusion partner is also available for recombinant protein expression in other host cells, namely, insect cells and other eukaryotic cells (Panavas et al., 2009).

GST is a gluthathione-S-transferase from *Schistosoma japonicum* (26 kDa) that has been used as an affinity fusion partner for the single-step purification of its target

proteins (Smith and Johnson, 1988). GST can also promote protein soluble expression in *E. coli*, being more efficient when positioned at the N-terminal rather than at the C-terminal end (Malhotra, 2009). This fusion partner can protect its target protein from the proteolytic degradation, stabilizing it into the soluble fraction (Kaplan et al., 1997; Hu et al., 2008; Young et al., 2012). In spite of performing quite well in some high throughput studies (Dummler et al., 2005; Cabrita et al., 2006; Kim and Lee, 2008), GST is often a poor solubility tag when compared to other commonly fusion partners, rendering the target protein production into inclusion bodies (Hammarstrom et al., 2002; Dyson et al., 2004; Hammarstrom, 2006; Kohl et al., 2008; Ohana et al., 2009).

Glutathione transferases are dimeric enzymes that catalyze the nucleophilic addition of the thiol of glutathione to a wide range of hydrophobic electrophilic molecules (Ketterer, 2001). Taking this feature into account, GST can be useful for monitoring the protein expression and purification via its catalytic activity, and the purification of GST fusion proteins can be easily performed by affinity chromatography using glutathione derivates immobilized into a solid support (Viljanen et al., 2008). GST fusion proteins can be eluted with glutathione under mild conditions (Vinckier et al., 2011).

A major disadvantage for using GST as solubility and affinity tag relies on its oligomerized form: GST has four solvent exposed cysteines that can provide a significant oxidative aggregation (Kaplan et al., 1997), making it a poor choice for tagging oligomeric target proteins (Malhotra, 2009).

As occurs with MBP, GST can be coupled with other affinity strategies, for instance, the His_6 tag, to improve the protein expression and purification (Scheich et al., 2003; Hayashi and Kojima, 2008; Hu et al., 2008). GST expression vectors like the pGEX (Hakes and Dixon, 1992) or pCold-GST (Hayashi and Kojima, 2008) usually contain a protease cleavage site between the fusion tag and target protein for GST tag removal after or during protein purification.

GST has also been applied as a fusion partner in other expression systems apart from the *E. coli*, such as yeast (Mitchell et al., 1993), insect cells (Beekman et al., 1994), and mammalian cells (Rudert et al., 1996). This fusion partner has shown to be useful for protein labeling (Ron and Dressler, 1992; Viljanen et al., 2008), antibody production (Aatsinki and Rajaniemi, 2005), and vaccine development (Mctigue et al., 1995).

In addition to these commonly used fusion partners, new solubility enhancer tags are constantly emerging in literature (see the corresponding references in Table 1.1), as for instance, the **HaloTag** (34 kDa), which uses a modified haloalkane dehalogenase protein that improves protein solubility and can bind to several synthethic ligands, the monomeric mutant of Orc protein of the bacteriophage T7 (**Mocr**), the *E. coli* protein **Skp**, stress-responsive proteins **RpoA**, **SlyD**, **Tsf**, **RpoS**, part of the domain I of IF2 (**Expressivity tag**), the *E. coli* secreted protein A (**EspA**), and the **SNUT** tag, which is a protein derived from a portion of the bacterial transpeptidase sortase A of *Staphylococcus aureus*.

1.3.2. Affinity purification handles

Affinity fusion partners have widely contributed for the development of recombinant protein production studies in basic research and in high throughput structural biology (Waugh, 2011) by simplifying protein purification procedures, and allowing for protein detection, and characterization (Butt et al., 2005; Malhotra, 2009; Young et al., 2012).

Affinity purification handles can be divided into two groups: (**1**) peptides or proteins that bind a small ligand immobilized on a solid support, as for instance, the His_6 tag and nickel affinity resins, and (**2**) tags that bind to an immobilized molecule such as antibodies (Arnau et al., 2006).

The purification of a target protein using an affinity handle offers several **advantages** over the conventional chromatographic methodologies, namely:

i) the target protein never interacts directly with the chromatographic resin (Waugh, 2005);
ii) target proteins can be easily obtained pure after a single-step purification (Terpe, 2003);
iii) affinity purification offers a variety of strategies to bind the target protein on an affinity matrix (Malhotra, 2009);
iv) affinity tags are an economically favorable and time-saving strategy, and they allow different proteins to be purified using a common method in contrast to highly customized procedures used in conventional chromatographic purification (Arnau et al., 2006).

An affinity tag is often chosen taking into account the purification costs whereby different affinity media have different expenses in the resin itself and in the operation process. The buffer requirements are also essential for the designing of an efficient purification strategy (Malhotra, 2009). In addition, the choice of an affinity can also rely on the size: small tags are useful for protein detection and antibody production, as they are not immunogenic as large tags (Terpe, 2003).

Tandem affinity purification (TAP) or **dual-tagging** strategies are now commonly used in recombinant protein production: they offer a highly specific isolation of target proteins with minimal background and under mild conditions, and they are very useful in the study of protein interactions, allowing the separation of different mixed protein complexes (Arnau et al., 2006; Li, 2010).

Table 1.2 lists some of the common purification tags used in recombinant protein production.

Table 1.2. Affinity purification tags (adapted from Esposito and Chatterjee, 2006; Malhotra, 2009):

Tag	Protein	Size (aa)	Affinity matrix	Reference
His_6	Hexahistidine tag	6-10	Immobilized metal ion – Ni, Co, Cu, Zn	(Gaberc-Porekar and Menart, 2001)
GST	Gluthathione-S-transferase	211	Gluthathione	(Smith and Johnson, 1988)
MBP	Maltose-binding protein	396	Amylose	(di Guan et al., 1988; Pryor and Leiting, 1997)
FLAG	FLAg tag peptide	8	Anti-FLAG antibody	(Einhauer and Jungbauer, 2001)
Strep-II	Streptavidin binding peptide	8	Streptavidin	(Schmidt and Skerra, 1994)
CBP	Calmodulin-binding protein	26	Immobilized calmodulin	(Vaillancourt et al., 2000)
HaloTag	Mutated dehalogenase	~300	Chloroalkane	(Ohana et al., 2009)
Protein A	Staphylococcal Protein A	280	Immobilized IgG	(Stahl and Nygren, 1997)
IMPACT (CBD)	Intein mediated purification with the chitin-binding domain	51	Chitin	(Chong et al., 1997; Sheibani, 1999)
CBM	Cellulose-binding module	*	Cellulose	(Tomme et al., 1998)
Dock	Dockerin domain of *Clostridium josui*	22	Cohesin - Cellulose	(Kamezaki et al., 2010)
Tamavidin	fungal avidin-like protein	~140	Biotin	(Takakura et al., 2010)

*Several sizes, from 4-20 kDa

The **polyhistidine affinity** tag or **His** tag consists of a variable number of consecutive histidine residues (usually six) that coordinate, via the histidine imidazole ring, transition metal ions such as Ni^{2+} or Co^{2+} immobilized on beads or a resin for immobilized metal affinity purification (IMAC) (Gaberc-Porekar and Menart, 2001; Terpe, 2003; Kimple and Sondek, 2004; Malhotra, 2009). Commonly used IMAC resins such as nitrilotriacetic acid agarose (Ni-NTA, from Qiagen), or carboxymethylasparte agarose (Talon, from ClonTech) have a high binding capacity, and can be used for purification of fusion proteins directly from crude cell lysates (Terpe, 2003; Kimple and Sondek, 2004; Li, 2010).

The His tag is one of the most widely used purification tags, and it offers several **advantages** (Kimple and Sondek, 2004; Li, 2010):

i) its small size and charge rarely interferes with protein function and structure;
ii) it can be used under native and denaturing conditions
iii) target proteins can be eluted under mild conditions by imidazole competition or low pH

The His tag has been used in several high throughput screenings, placed at the N- or C-terminal end, or even in the middle of the fusion protein (Cabrita et al., 2006; Hammarstrom, 2006; Marblestone et al., 2006; Bird, 2011), and it is also an useful tool in protein crystallization as well as protein detection (Carson et al., 2003; Kimple and Sondek, 2004).

Taking into account the mechanism of protein interaction with the immobilized ions, careful should be taken in IMAC to avoid strong reducing and chelating agents in any of the buffers (as for instance, EDTA), as they will reduce or strip the immobilized metal ions (Carson et al., 2003; Kimple and Sondek, 2004; Li, 2010).

Epitope tags are short sequences of amino acids that serve as the antigen region to which the antibody binds, being suitable for several immunoapplications. These include affinity chromatography on immobilized monoclonal antibodies, and protein trafficking *in vitro* or in cell cultures (Kimple and Sondek, 2004; Young et al., 2012). Epitope tagging engages an expensive purification that often limits its wide application.

The following partners are often used as epitope tags: the FLAG tag (Einhauer and Jungbauer, 2001), the hemaglutinin, and the c-Myc (Fritze and Anderson, 2000). Their

short sequences rarely interfere with structure or function of target proteins, and are very specific for their respective primary antibodies (Kimple and Sondek, 2004; Malhotra, 2009). The **FLAG** tag is a short hydrophilic eight amino-acid peptide, and it was the first tag to be used in the epitope context. This tag works either for protein detection or purification (Hopp et al., 1988; Knappik and Pluckthun, 1994), and it has an intrinsic enterokinase cleavage site at its C-terminus end, allowing its complete removal from the target protein (Einhauer and Jungbauer, 2001; Young et al., 2012).

Strep II tag is a short tag of only eight amino acid residues that possesses a strong and specific binding to streptavidin via its biotin pocket (Schmidt and Skerra, 1994). This affinity partner can be fused at both N- or C-terminal ends, or within the target protein. Strep II-fused proteins elute from streptavidin columns with biotin derivates under gentle conditions (Terpe, 2003; Li, 2010).

The **CBP** tag is a calmodulin-binding peptide derived from the C-terminus of skeletal muscle myosin light chain kinase, and it has been used as an N- or C-terminal affinity tag of target protein purification on a calmodulin immobilized matrix (Terpe, 2003; Malhotra, 2009). The CBP interaction with calmodulin is calcium-dependent, and hence, the addition of calcium-chelating allows the single step elution of target proteins under gentle conditions (Terpe, 2003; Malhotra, 2009; Li, 2010). This tag is an affinity system highly specific for protein purification in *E. coli* but not in eukaryotic systems, as *E. coli* does not contain endogenous proteins that interact with calmodulin (Terpe, 2003; Malhotra, 2009).

In addition to the above-mentioned affinity tags, new affinity purification strategies are now described in literature for protein isolation and detection (see the corresponding references in Table 1.2) such as cellulose-binding domains I, II and III (**CBD**), the **HaloTag**, the dockerin domain **Dock** tag, and the avidin-like protein, **Tamavidin** tag.

1.3.3. Tag removal

The removal of the fusion partner from the final protein is often necessary because the tag can potentially interfere with the proper structure and functioning of the target protein (Waugh, 2005; Malhotra, 2009; Young et al., 2012).

The fusion partners are removed from their target proteins either by **enzymatic cleavage**, in which site specific proteases are used under mild conditions, or by **chemical cleavage** that offers a less expensive tag removal but it is also less specific compared to the enzymatic strategy, besides presenting harsh conditions that can affect the target protein stability and solubility (Malhotra, 2009; Li, 2011).

The efficiency of the **enzymatic removal** of fusion proteins may vary in an unpredicted manner with different proteins (Li, 2011; Vergis and Wiener, 2011; Young et al., 2012), and it often requires the optimization of cleavage conditions through and trial-and-error process (Malhotra, 2009). Two types of proteases can be used for tag removal (reviewed in Waugh, 2011):

i) **endoproteases**: they are divided into *serine proteases* such as the activated blood coagulation factor X (Factor Xa), enterokinase, and α-thrombin, and *viral proteases* like the tobacco etch virus (TEV), and the human rhinovirus 3C protease (Table 1.3). In spite of recognizing a similar number of substrate amino acid residues, viral proteases have usually more stringent sequence specificity than serine proteases, presenting also slower rates than the latter. Endoproteases are useful tools for the removal of N-terminal fusion tags, since they cleave close to the C-terminus end of their recognition sites thus leaving the target protein with its native N-terminal sequence.

ii) **exoproteases**: they are often used together with endoproteases mainly for the removal of C-terminal fusion tags. The available exoproteases include metallocarboxypeptidases, and aminopeptidases.

Table 1.3. Common endoproteases for tag removal (adapted from Malhotra, 2009):

Protease	Source	Cleavage site	Reference
TEV	Tobacco etch virus protease	ENLYFQ/G	(Parks et al., 1994; Kapust et al., 2001)
EntK	Enterokinase	DDDDK/	(Choi et al., 2001)
Xa	Factor Xa	IEGR/	(Jenny et al., 2003)
Thr	Thrombin	LVPR/GS	(Jenny et al., 2003)
PreScission	Genetically engineered derivative of human rhinovirus 3C protease	LEVLFQ/GP	GE Healthcare; (Cordingley et al., 1990)
SUMO protease	Catalytic core of Ulp1	Recognizes SUMO tertiary structure and cleaves at the C-terminal end of the conserved Gly-Gly sequence in SUMO	(Malakhov et al., 2004; Butt et al., 2005; Marblestone et al., 2006)

The removal of a fusion tag is usually accomplished by two purification steps, as follows: after the initial affinity purification step, the purified fusion protein is mixed in solution with the endoprotease to cleave off the tag. The cleaved target protein (TP) is recovered in the flow-through sample after a second affinity purification step, in which the cleaved fusion tag and the added protease are collected in the eluted sample.

Figure 1.3 illustrates an example of a tag removal using the above-mentioned two-step purification protocol.

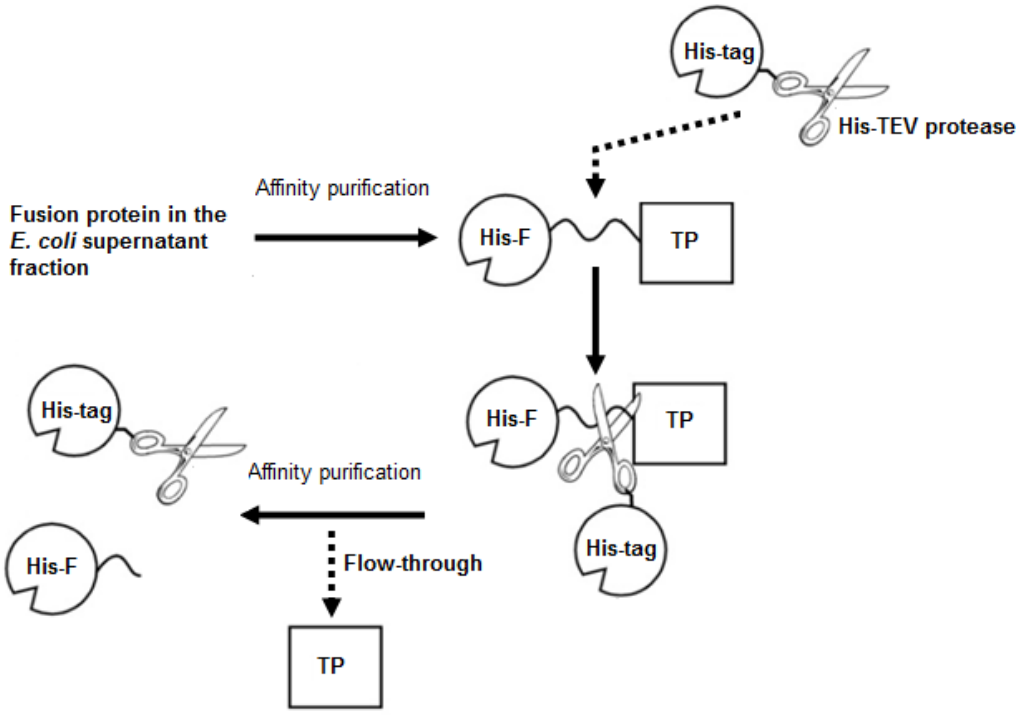


Figure 1.3. Tag removal strategy using the His-TEV endoprotease (adapted from Waugh, 2011). *His*: hexahistidine tag. *F*: solubility fusion partner. *TP*: target protein. See the explanation in the text above.

In spite of widely employed, the removal of fusion partners has always been the Achilles' heel of affinity tagging, presenting several **difficulties** such as:

i) unspecific cleavage due to the recognition of a linear amino acid sequence (except for SUMO protease);
ii) inefficient processing due to steric hindrance or the presence of unfavorable residues around the cleavage site (Li, 2011; Waugh, 2011). The inclusion of extra amino acid residues (a spacer or linker) between the cleavage site and target protein (Esposito and Chatterjee, 2006; Malhotra, 2009) can alleviate this problem;
iii) low protein yields after tag removal, and failure in recover active, structurally organized target proteins due to protein precipitation and aggregation (Butt et al., 2005; Waugh, 2011);
iv) high costs of proteases and tedious optimization of cleavage conditions (Smyth et al., 2003).

Independently of the cleavage type, additional chromatographic steps are often required to purify the target protein from the cleavage mixture. Although conventional affinity technologies have greatly simplified recombinant protein production, resins and buffers are still too expensive. Hence, the tag removal adds another layer of complexity and expense to the recombinant protein production process (Mee et al., 2008; Li, 2011).

Self-cleaving tags are a special group of fusion tags that possess inducible proteolytic activity, therefore being considered an attractive alternative to the existent affinity strategies for simple and costless protein purification and tag removal (Chong et al., 1997; Li, 2011).

The protein splicing is a process in which the intervening sequence (intein) removes itself and binds the flaking residues (exteins) to produce two independent protein products (Perler et al., 1994). Self-cleaving tags undergo specific cleavage upon being triggered by low molecular weight compounds or upon a change of conformation. The available technologies include inteins, the *Staphylococcus aureus* sortase A, the N-terminal protease (N^{pro}), the *Neisseria meningitides* iron-regulated protein FrpC, and the cysteine protease domain secreted by *Vibrio cholerae*, all of them reviewed in Li (2011).

1.4. The Fh8 excreted-secreted protein of *Fasciola hepatica*

The liver fluke *Fasciola hepatica* is a trematode parasite of ruminants (mainly cattle, goat and sheep), but it also infects humans worldwide. Fasciolosis is an anthropozoonotic disease caused by the *F. hepatica*, and it is responsible for a large economic loss in the agricultural industry (Silva et al., 2004; Mas-Coma et al., 2005).

Infection from *F. hepatica* occurs upon ingestion of vegetables contaminated with parasite metacercariae. After occlusion, the worm penetrates into the intestinal wall and migrates through the peritoneal cavity and liver.

The *F. hepatica* secretes a large number of proteins into the host organism that are important for the host-parasite interaction during infection, as highlighted by an integrated transcriptomic and proteomic analysis of the parasite secretome (Robinson et al., 2009).

Despite the low number of human cases of fasciolosis, this parasite infection is considered a public health problem. The diagnosis of human and animal fasciolosis has been improved by the introduction of new tools, using more specific and sensitive antigens, as for instance, the excreted-secreted antigens (ESP) (Berasain et al., 1997; Silva et al., 2004; Hewitson et al., 2009).

Several proteins of the *F. hepatica* ESP have been identified as useful tools for serodiagnosis, and as infection markers, namely the cathepsin L (O'Neill et al., 1999), peroxiredoxin (Salazar-Calderon et al., 2000), and the calcium-binding protein (CaBP) Fh22 (de Eguino et al., 1999).

The **Fh8** (Genbank ID: AF213970.1) is one of the excreted-secreted antigens by the parasite *F. hepatica* in the early stages of infection (Silva et al., 2004). This protein presents a low molecular weight (8 kDa), and it is located on the surface of the parasite, therefore being suggested as a useful tool for the diagnosis, vaccine and drug development against *F. hepatica* infections (Silva et al., 2004). The Fh8 has high homology with 8-kDa CaBPs of *Schistosoma mansoni* (Sm8) (Ram et al., 1989), of *Clonorchis sinensis* (Ch8), and of *Schistosoma japonicum* (Sj8) (Lv et al., 2009), and it belongs to the calmodulin-like EF-hand CaBP family (Fraga et al., 2010).

Besides Fh8 and Fh22, two other CaBPs have been recently identified in *F. hepatica*, presenting high homology with calmodulin: the FhCaM1 and the FhCaM2 (Russell et al., 2007).

Calcium is an important second messenger in eukaryotes that controls vital processes such as bone mineralization and cell signaling (Lewit-Bentley and Rety, 2000). This ion exerts its effects by binding to a very large number of proteins that constitute the calcium-binding proteins family.

CaBPs are structurally organized by EF-hand motifs, which are helix-loop-helix structures that participate in Ca^{2+} coordination (Bhattacharya et al., 2004; Zhou et al., 2006; Chazin, 2011). Proteins of the EF-hand CaBP family are usually classified as Ca^{2+} sensors or Ca^{2+} buffers. Upon calcium binding, **sensor proteins,** like calmodulin (Nelson and Chazin, 1998; Chin and Means, 2000) and troponin C (Nelson and Chazin, 1998), translate the physiological changes in calcium levels by undergoing a conformational change. This then allows the binding of other proteins downstream the process. In EF-hand proteins, the open of the EF-hand structure exposes a hydrophobic surface, which binds the target sequence (Lewit-Bentley and Rety, 2000; Bhattacharya et al., 2004). **Ca^{2+} buffer proteins**, such as calbindin D_{9k} and parvalbumin (Schwaller, 2010), are involved in calcium signal modulation, undergoing minimal conformational changes upon calcium binding.

The **Fh8** (Figure 1.4) presents two EF-hand motifs, and it was characterized as a *calcium sensor protein*: when calcium binds, the Fh8 switches from a closed (apo-state) to an open (calcium-loaded state) conformation due to the reorientation of the four helices, exposing a large hydrophobic region that acts as a target-binding surface (Fraga et al., 2010).

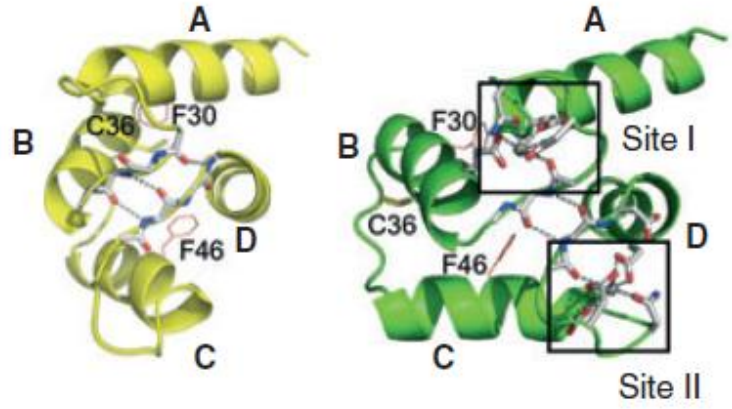


Figure 1.4. Modeled structure of Fh8 in the open (green) and closed (yellow) conformation (from (Fraga et al., 2010)). The Fh8 has two calcium binding sites, located the EF-hands (site I and II). The four EF-hand helices are named A, B, C and D.

The use of recombinant Fh8 produced in *E. coli* led to the development of a novel, rapid and simple immunodetection of *F. hepatica* infections (Silva et al., 2004). Moreover, when produced recombinantly in *E. coli*, the Fh8 revealed to be a highly soluble and unusual thermal stable protein (keeping secondary structure integrity up to 74 ºC) (Silva et al., 2004; Fraga et al., 2010).

Previous studies of the Fh8 by directed mutagenesis (conducted at the National Health Institute Doutor Ricardo Jorge) revealed that EF-hand-mutated Fh8 proteins produced in *E. coli* presented similar expression levels to those obtained for the wild type Fh8. These results suggested that other Fh8 residue sequences (apart from those involved in the calcium-binding) could be critical for Fh8 stability and production. The prediction of the Fh8 three-dimensional structure showed that, except for small sequences in the N-terminal (11 amino acid residues) and C-terminal (6 amino acid residues), all the remaining residues are involved or affected by the calcium-binding. Taking into account that the N-terminal of a protein is very important for its half-life, the N-terminal eleven residues of Fh8 were suggested to play a key role in the stability and production of the entire Fh8 protein. This 11-amino acid sequence (named H) may also be critical for the immunological response of the Fh8 antigen.

Both Fh8 and H peptides improved the *E. coli* expression levels of several target proteins (see patent WO 2010/082097 in the Appendix 1.6.1), in which the Fh8 presented further improvements than the H peptide.

This thesis stems from these findings and due to **Fh8** properties, namely its high solubility, high stability and calcium-binding features, it aims at studying this protein as a **fusion partner** for the soluble overexpression and purification of recombinant proteins in *E. coli*. Taking into account the above-mentioned importance of the Fh8 N-terminal eleven residues, this thesis also aims at studying the **H tag** for the production of immunogenic proteins.

In this work, the novel fusion tag Fh8 is evaluated as a solubility enhancer and compared with the traditionally used solubility tags in Chapter 2. The H tag is investigated in Chapter 3 as a fusion partner for the adjuvant-free production of polyclonal antibodies. The Fh8 tag is also explored as a purification handle, as described in Chapter 4. The utility of Fh8 partner is highlighted in Chapter 5, whereby it is used for the soluble expression and purification of two difficult-to-express proteins. As a solubility and purification tag, the Fh8 partner should interfere as less as possible with its target proteins, but its previously reported oligomerization tendency (Fraga et al., 2010) can add some difficulties in the fusion context. An attempt to reduce this Fh8 characteristic is revealed in the Chapter 6 by studying two novel Fh8 variants as fusion tags.

1.5. References

Aatsinki, J.T. and Rajaniemi, H.J. (2005) An alternative use of basic pGEX vectors for producing both N- and C-terminal fusion proteins for production and affinity purification of antibodies. Protein Expression and Purification 40, 287-291.

Ahn, K.Y., Song, J.A., Hah, K.Y., Park, J.S., Seo, H.S. and Lee, J. (2007) Heterologous protein expression using a novel stress-responsive protein of *E. coli* RpoA as fusion expression partner. Enzyme and Microbial Technology 41, 859-866.

Arnau, J., Lauritzen, C., Petersen, G.E. and Pedersen, J. (2006) Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expression and Purification 48, 1-13.

Austin, C. (2003) Novel approach to obtain biologically active recombinant heterodimeric proteins in *Escherichia coli*. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences 786, 93-107.

Baca, A.M. and Hol, W.G.J. (2000) Overcoming codon bias: A method for high-level overexpression of Plasmodium and other AT-rich parasite genes in *Escherichia coli*. International Journal for Parasitology 30, 113-118.

Bach, H., Mazor, Y., Shaky, S., Shoham-Lev, A., Berdichevsky, Y., Gutnick, D.L. and Benhar, I. (2001) *Escherichia coli* maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies. Journal of Molecular Biology 312, 79-93.

Baneyx, F. and Palumbo, J.L. (2003) Improving heterologous protein folding via molecular chaperone and foldase co-expression. Methods in Molecular Biology 205, 171-197.

Beekman, J.M., Cooney, A.J., Elliston, J.F., Tsai, S.Y. and Tsai, M.J. (1994) A rapid one-step method to purify baculovirus-expressed human estrogen-receptor to be used in the analysis of the oxytocin promoter. Gene 146, 285-289.

Berasain, P., Goni, F., McGonigle, S., Dowd, A., Dalton, J.P., Frangione, B. and Carmona, C. (1997) Proteinases secreted by *Fasciola hepatica* degrade extracellular matrix and basement membrane components. Journal of Parasitology 83, 1-5.

Bessette, P.H., Aslund, F., Beckwith, J. and Georgiou, G. (1999) Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. Proceedings of the National Academy of Sciences of the United States of America 96, 13703-13708.

Betiku, E. (2006) Molecular chaperones involved in heterologous protein folding in *Escherichia coli*. Biotechnology and Molecular Biology Review 1, 66-75.

Bhattacharya, S., Bunick, C.G. and Chazin, W.J. (2004) Target selectivity in EF-hand calcium binding proteins. Biochimica Et Biophysica Acta-Molecular Cell Research 1742, 69-79.

Bird, L.E. (2011) High throughput construction and small scale expression screening of multi-tag vectors in *Escherichia coli*. Methods 55, 29-37.

Buckle, A.M., Zahn, R. and Fersht, A.R. (1997) A structural model for GroEL-polypeptide recognition. Proceedings of the National Academy of Sciences of the United States of America 94, 3571-3575.

Bussow, K., Scheich, C., Sievert, V., Harttig, U., Schultz, J., Simon, B., Bork, P., Lehrach, H. and Heinemann, U. (2005) Structural genomics of human proteins - target selection and generation of a public catalogue of expression clones. Microbial Cell Factories 4.

Butt, T.R., Edavettal, S.C., Hall, J.P. and Mattern, M.R. (2005) SUMO fusion technology for difficult-to-express proteins. Protein Expression and Purification 43, 1-9.

Cabrita, L.D., Dai, W.W. and Bottomley, S.P. (2006) A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production. Bmc Biotechnology 6.

Carson, M., Johnson, D.H., McDonald, H., Brouillette, C. and Delucas, L.J. (2003) His-tag impact on structure. Acta Crystallographica Section D-Biological Crystallography 63, 295-301.

Caswell, J., Snoddy, P., McMeel, D., Buick, R.J. and Scott, C.J. (2010) Production of recombinant proteins in *Escherichia coli* using an N-terminal tag derived from sortase. Protein Expression and Purification 70, 143-150.

Chatellier, J., Buckle, A.M. and Fersht, A.R. (1999) GroEL recognises sequential and non-sequential linear structural motifs compatible with extended beta-strands and alpha-helices. Journal of Molecular Biology 292, 163-172.

Chazin, W.J. (2011) Relating form and function of EF-hand calcium binding proteins. Accounts of Chemical Research 44, 171-179.

Cheng, Y., Gu, J.A., Wang, H.G., Yu, S., Liu, Y.Q., Ning, Y.L., Zou, Q.M., Yu, X.J. and Mao, X.H. (2010) EspA is a novel fusion partner for expression of foreign proteins in *Escherichia coli*. Journal of Biotechnology 150, 380-388.

Cheng, Y. and Patel, D.J. (2004) An efficient system for small protein expression and refolding. Biochemical and Biophysical Research Communications 317, 401-405.

Chesshyre, J.A. and Hipkiss, A.R. (1989) Low temperatures stabilize interferon a-2 against proteolysis in *Methylophilus methylotrophus* and *Escherichia coli*. Applied Microbiology and Biotechnology 31, 158-162.

Chin, D. and Means, A.R. (2000) Calmodulin: a prototypical calcium sensor. Trends in Cell Biology 10, 322-328.

Choi, S.I., Song, H.W., Moon, J.W. and Seong, B.L. (2001) Recombinant enterokinase light chain with affinity tag: Expression from *Saccharomyces cerevisiae* and its utilities in fusion protein technology. Biotechnology and Bioengineering 75, 718-724.

Chong, S.R., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., Perler, F.B., Benner, J., Kucera, R.B., Hirvonen, C.A., Pelletier, J.J., Paulus, H. and Xu, M.Q. (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. Gene 192, 271-281.

Chou, C.P. (2007) Engineering cell physiology to enhance recombinant protein production in *Escherichia coli*. Applied Microbiology and Biotechnology 76, 521-532.

Collins, K.D. and Washabaugh, M.W. (1985) The Hofmeister effect and the behavior of water at interfaces. Quarterly Reviews of Biophysics 18, 323-422.

Cordingley, M.G., Callahan, P.L., Sardana, V.V., Garsky, V.M. and Colonno, R.J. (1990) Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. The Journal of Biological Chemistry 265, 9062-5.

Corsini, L., Hothorn, M., Scheffzek, K., Sattler, M. and Stier, G. (2008) Thioredoxin as a fusion tag for carrier-driven crystallization. Protein Science 17, 2070-2079.

Davis, G.D., Elisee, C., Newham, D.M. and Harrison, R.G. (1999) New fusion protein systems designed to give soluble expression in *Escherichia coli*. Biotechnology and Bioengineering 65, 382-388.

de Eguino, A.D.R., Machin, A., Casais, R., Castro, A.M., Boga, J.A., Martin-Alonso, J.M. and Parra, F. (1999) Cloning and expression in *Escherichia coli* of a *Fasciola hepatica* gene encoding a calcium-binding protein. Molecular and Biochemical Parasitology 101, 13-21.

de Marco, A. and De Marco, V. (2004) Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning. Journal of Biotechnology 109, 45-52.

de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T. and Bukau, B. (2007) Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. Bmc Biotechnology 7.

De Marco, V., Stier, G., Blandin, S. and de Marco, A. (2004) The solubility and stability of recombinant proteins are increased by their fusion to NusA. Biochemical and Biophysical Research Communications 322, 766-771.

DelProposto, J., Majmudar, C.Y., Smith, J.L. and Brown, W.C. (2009) Mocr: A novel fusion tag for enhancing solubility that is compatible with structural biology applications. Protein Expression and Purification 63, 40-49.

Demain, A.L. and Vaishnav, P. (2009) Production of recombinant proteins by microbes and higher organisms. Biotechnology Advances 27, 297-306.

Deuerling, E., Patzelt, H., Vorderwulbecke, S., Rauch, T., Kramer, G., Schaffitzel, E., Mogk, A., Schulze-Specking, A., Langen, H. and Bukau, B. (2003) Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. Molecular Microbiology 47, 1317-1328.

di Guan, C., Li, P., Riggs, P.D. and Inouye, H. (1988) Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. . Gene 67, 21-30.

Dong, W.R., Xiang, L.X. and Shao, J.Z. (2010) Novel antibiotic-free plasmid selection system based on complementation of host auxotrophy in the NAD *de novo* synthesis pathway. Applied and Environmental Microbiology 76, 2295-2303.

Douette, P., Navet, R., Gerkens, P., Galleni, M., Levy, D. and Sluse, F.E. (2005) *Escherichia coli* fusion carrier proteins act as solubilizing agents for recombinant uncoupling protein 1 through interactions with GroEL. Biochemical and Biophysical Research Communications 333, 686-693.

Dummler, A., Lawrence, A.M. and de Marco, A. (2005) Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. Microbial Cell Factories 4.

Dyson, M.R., Shadbolt, S.P., Vincent, K.J., Perera, R.L. and McCafferty, J. (2004) Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. Bmc Biotechnology 4.

Einhauer, A. and Jungbauer, A. (2001) The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. Journal of Biochemical and Biophysical Methods 49, 455-65.

Esposito, D. and Chatterjee, D.K. (2006) Enhancement of soluble protein expression through the use of fusion tags. Current Opinion in Biotechnology 17, 353-358.

Ferrer-Miralles, N., Domingo-Espin, J., Corchero, J.L., Vazquez, E. and Villaverde, A. (2009) Microbial factories for recombinant pharmaceuticals. Microbial Cell Factories 8.

Fox, J.D., Kapust, R.B. and Waugh, D.S. (2001) Single amino acid substitutions on the surface of *Escherichia coli* maltose-binding protein can have a profound impact on the solubility of fusion proteins. Protein Science 10, 622-630.

Fraga, H., Faria, T.Q., Pinto, F., Almeida, A., Brito, R.M.M. and Damas, A.M. (2010) FH8-a small EF-hand protein from *Fasciola hepatica*. Febs Journal 277, 5072-5085.

Fritze, C.E. and Anderson, T.R. (2000) Epitope tagging: general method for tracking recombinant proteins. Methods in Enzymology 327., 3-16.

Gaberc-Porekar, V. and Menart, V. (2001) Perspectives of immobilized-metal affinity chromatography. Journal of Biochemical and Biophysical Methods 49, 335-360.

Gao, X., Chen, W., Guo, C., Qian, C., Liu, G., Ge, F., Huang, Y., Kitazato, K., Wang, Y. and Xiong, S. (2010) Soluble cytoplasmic expression, rapid purification, and characterization of cyanovirin-N as a His-SUMO fusion. Applied Microbiology and Biotechnology 85, 1051-60.

Gasser, B., Saloheimo, M., Rinas, U., Dragosits, M., Rodriguez-Carmona, E., Baumann, K., Giuliani, M., Parrilli, E., Branduardi, P., Lang, C., Porro, D., Ferrer, P., Tutino, M.L., Mattanovich, D. and Villaverde, A. (2008) Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microbial Cell Factories 7.

GE Healthcare (2006) Hydrophobic Interaction and Reversed Phase Chromatography - Principles and Methods, 11-0012-69 AA.

GE Healthcare (2010) Strategies for protein purification - Handbook, 28-9833-31 AA.

Hägg, P., de Pohl, J.W., Abdulkarim, F. and Isaksson, L.A. (2004) A host/plasmid system that is not dependent on antibiotics and antibiotic resistance genes for stable plasmid maintenance in *Escherichia coli*. Journal of Biotechnology 111, 17-30.

Hakes, D.J. and Dixon, J.E. (1992) New vectors for high-level expression of recombinant proteins in bacteria. Analytical Biochemistry 202, 293-298.

Hammarstrom, M. (2006) Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein. Journal of Structural and Functional Genomics 7, 1-14.

Hammarstrom, M., Hellgren, N., Van den Berg, S., Berglund, H. and Hard, T. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. Protein Science 11, 313-321.

Han, K.Y., Seo, H.S., Song, J.A., Ahn, K.Y., Park, J.S. and Lee, J. (2007a) Transport proteins PotD and Crr of *Escherichia coli*, novel fusion partners for heterologous protein expression. Biochimica Et Biophysica Acta-Proteins and Proteomics 1774, 1536-1543.

Han, K.Y., Song, J.A., Ahn, K.Y., Park, J.S., Seo, H.S. and Lee, J. (2007b) Enhanced solubility of heterologous proteins by fusion expression using stress-induced *Escherichia coli* protein, Tsf. Fems Microbiology Letters 274, 132-138.

Han, K.Y., Song, J.A., Ahn, K.Y., Park, J.S., Seo, H.S. and Lee, J. (2007c) Solubilization of aggregation-prone heterologous proteins by covalent fusion of stress-responsive *Escherichia coli* protein, SlyD. Protein Engineering Design & Selection 20, 543-549.

Hannig, G. and Makrides, S.C. (1998) Strategies for optimizing heterologous protein expression in *Escherichia coli*. TRENDS in Biotechnology 16, 54-60.

Hansted, J.G., Pietikainen, L., Hog, F., Sperling-Petersen, H.U. and Mortensen, K.K. (2011) Expressivity tag: A novel tool for increased expression in *Escherichia coli*. Journal of Biotechnology 155, 275-283.

Harper, J.D. and Lansbury, P.T. (1997) Models of amyloid seeding in Alzheimier's disease and scrapie: Mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. Annual Review of Biochemistry 66, 385-407.

Hayashi, K. and Kojima, C. (2008) pCold-GST vector: A novel cold-shock vector containing GST tag for soluble protein production. Protein Expression and Purification 62, 120-127.

Hewitson, J.P., Grainger, J.R. and Maizels, R.M. (2009) Helminth immunoregulation: The role of parasite secreted proteins in modulating host immunity. Molecular and Biochemical Parasitology 167, 1-11.

Hoffmann, F. and Rinas, U. (2004) Roles of heat-shock chaperones in the production of recombinant proteins in *Escherichia coli*. Advances in Biochemical Engineering / Biotechnology 89, 143-161.

Hopp, T.P., Prickett, K.S., Price, V.L., Libby, R.T., March, C.J., Cerretti, D.P., Urdal, D.L. and Conlon, P.J. (1988) A short polypeptide marker sequence useful for recombinant protein identification and purification. Bio-Technology 6, 1204-1210.

Hu, J., Qin, H.J., Sharma, M., Cross, T.A. and Gao, F.P. (2008) Chemical cleavage of fusion proteins for high-level production of transmembrane peptides and protein domains containing conserved methionines. Biochimica Et Biophysica Acta-Biomembranes 1778, 1060-1066.

Huang, Y.S. and Chuang, D.T. (1999) Mechanisms for GroEL/GroES-mediated folding of a large 86-kDa fusion polypeptide in vitro. Journal of Biological Chemistry 274, 10405-10412.

Hwang, S.M., Kang, H.J., Bae, S.W., Chang, W.J. and Koo, Y.M. (2010) Refolding of lysozyme in hydrophobic interaction chromatography: effects of hydrophobicity of adsorbent and salt concentration in mobile phase. Biotechnology and Bioprocess Engineering 15, 213-219.

Inouye, S. and Sahara, Y. (2009) Expression and purification of the calcium binding photoprotein mitrocomin using ZZ-domain as a soluble partner in *E. coli* cells. Protein Expression and Purification 66, 52-57.

Jana, S. and Deb, J.K. (2005) Strategies for efficient production of heterologous proteins in *Escherichia coli*. Applied Microbiology and Biotechnology 67, 289-298.

Jenny, R.J., Mann, K.G. and Lundblad, R.L. (2003) A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. Protein Expression and Purification 31, 1-11.

Kamezaki, Y., Enomoto, C., Ishikawa, Y., Koyama, T., Naya, S., Suzuki, T. and Sakka, K. (2010) The Dock tag, an affinity tool for the purification of recombinant proteins, based on the interaction between dockerin and cohesin domains from *Clostridium josui* cellulosome. Protein Expression and Purification 70, 23-31.

Kamionka, M. (2011) Engineering of Therapeutic Proteins Production in *Escherichia coli*. Current Pharmaceutical Biotechnology 12, 268-274.

Kaplan, W., Husler, P., Klump, H., Erhardt, J., SluisCremer, N. and Dirr, H. (1997) Conformational stability of pGEX-expressed *Schistosoma japonicum* glutathione S-transferase: A detoxification enzyme and fusion-protein affinity tag. Protein Science 6, 399-406.

Kapust, R.B., Tozser, J., Fox, J.D., Anderson, D.E., Cherry, S., Copeland, T.D. and Waugh, D.S. (2001) Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Engineering 14, 993-1000.

Kapust, R.B. and Waugh, D.S. (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Science 8, 1668-1674.

Kawabe, Y., Seki, M., Seki, T., Wang, W.S., Imamura, O., Furuichi, Y., Saitoh, H. and Enomoto, T. (2000) Covalent modification of the Werner's syndrome gene product with the ubiquitin-related protein, SUMO-1. Journal of Biological Chemistry 275, 20963-20966.

Ketterer, B. (2001) A bird's eye view of the glutathione transferase field. Chemico-Biological Interactions 138, 27-42.

Khorasanizadeh, S., Peters, I.D. and Roder, H. (1996) Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. Nature Structural Biology 3, 193-205.

Kim, S. and Lee, S.B. (2008) Soluble expression of archaeal proteins in *Escherichia coli* by using fusion-partners. Protein Expression and Purification 62, 116-119.

Kimple, M.E. and Sondek, J. (2004) Overview of affinity tags for protein purification. In: Current Protocols in Protein Science.

Kishi, A., Nakamura, T., Nishio, Y., Maegawa, H. and Kashiwagi, A. (2003) Sumoylation of Pdx1 is associated with its nuclear localization and insulin gene activation. American Journal of Physiology-Endocrinology and Metabolism 284, E830-E840.

Knappik, A. and Pluckthun, A. (1994) An improved affinity tag based on the Flag® peptide for the detection and purification of recombinant antibody fragments. Biotechniques 17, 754-761.

Kohl, T., Schmidt, C., Wiemann, S., Poustka, A. and Korf, U. (2008) Automated production of recombinant human proteins as resource for proteome research. Proteome Science 6.

Kolaj, O., Spada, S., Robin, S. and Wall, J.G. (2009) Use of folding modulators to improve heterologous protein production in *Escherichia coli*. Microbial Cell Factories 8.

Kuczynska-Wisnik, D., Kedzierska, S., Matuszewska, E., Lund, P., Taylor, A., Lipinska, B. and Laskowska, E. (2002) The *Escherichia coli* small heat-shock proteins IbpA and IbpB prevent the aggregation of endogenous proteins denatured in vivo during extreme heat shock. Microbiology-Sgm 148, 1757-1765.

Kurland, C. and Gallant, J. (1996) Errors of heterologous protein expression. Current Opinion in Biotechnology 7, 489-493.

Lavallie, E.R., Diblasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F. and Mccoy, J.M. (1993) A Thioredoxin gene fusion expression system that circumvents inclusion body formation in the *Escherichia coli* cytoplasm. Bio-Technology 11, 187-193.

LaVallie, E.R., Lu, Z.J., Diblasio-Smith, E.A., Collins-Racie, L.A. and McCoy, J.M. (2000) Thioredoxin as a fusion partner for production of soluble recombinant proteins in *Escherichia coli*. Applications of Chimeric Genes and Hybrid Proteins, Pt A 326, 322-340.

Lee, C.D., Sun, H.C., Hu, S.M., Chiu, C.F., Homhuan, A., Liang, S.M., Leng, C.H. and Wang, T.F. (2008) An improved SUMO fusion protein system for effective production of native proteins. Protein Science 17, 1241-1248.

Lewit-Bentley, A. and Rety, S. (2000) EF-hand calcium-binding proteins. Current Opinion in Structural Biology 10, 637-643.

Li, Y.F. (2010) Commonly used tag combinations for tandem affinity purification. Biotechnology and Applied Biochemistry 55, 73-83.

Li, Y.F. (2011) Self-cleaving fusion tags for recombinant protein production. Biotechnology Letters 33, 869-881.

Lienqueo, M.E., Mahn, A., Salgado, J.C. and Asenjo, J.A. (2007) Current insights on protein behaviour in hydrophobic interaction chromatography. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences 849, 53-68.

Lobstein, J., Emrich, C.A., Jeans, C., Faulkner, M., Riggs, P. and Berkmen, M. (2012) SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. Microbial Cell Factories 11.

Lopez, P.J., Marchand, I., Joyce, S.A. and Dreyfus, M. (1999) The C-terminal half of RNase E, which organizes the *Escherichia coli* degradosome, participates in mRNA degradation but not rRNA processing in vivo. Molecular Microbiology 33, 188-199.

Lv, Z.Y., Yang, L.L., Hu, S.M., Sun, X., He, H.J., He, S.J., Li, Z.Y., Zhou, Y.P., Fung, M.C., Yu, X.B., Zheng, H.Q., Cao, A.L. and Wu, Z.D. (2009) Expression profile, localization of an 8-kDa calcium-binding protein from *Schistosoma japonicum* (SjCa8), and vaccine potential of recombinant SjCa8 (rSjCa8) against infections in mice. Parasitology Research 104, 733-743.

Magalhães, P.O., Lopes, A.M., Mazzola, P.G., Rangel-Yagui, C., Penna, T.C.V. and Pessoa Jr., A. (2007) Methods of endotoxin removal from biological preparations: a review. Journal of Pharmacy & Pharmaceutical Sciences 10, 388-404.

Makino, T., Skretas, G. and Georgiou, G. (2011) Strain engineering for improved expression of recombinant proteins in bacteria. Microbial Cell Factories 10.

Makrides, S.C. (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. Microbiological Reviews 60, 512-538.

Malakhov, M.P., Mattern, M., Malakhova, O.A., Drinker, M., Weeks, S.D. and Butt, T. (2004) SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. Journal of Structural and Functional Genomics 5, 75–86.

Malhotra, A. (2009) Tagging for Protein Expression. Guide to Protein Purification, Second Edition 463, 239-258.

Malik, A., Jenzsch, M., Lubbert, A., Rudolph, R. and Sohling, B. (2007) Periplasmic production of native human proinsulin as a fusion to *E. coli* ecotin. Protein Expression and Purification 55, 100-111.

Malik, A., Rudolph, R. and Sohling, B. (2006) A novel fusion protein system for the production of native human pepsinogen in the bacterial periplasm. Protein Expression and Purification 47, 662-671.

Marblestone, J.G., Edavettal, S.C., Lim, Y., Lim, P., Zuo, X. and Butt, T.R. (2006) Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. Protein Science 15, 182-189.

Mas-Coma, S., Bargues, M.D. and Valero, M.A. (2005) Fascioliasis and other plant-borne trematode zoonoses. International Journal for Parasitology 35, 1255-1278.

McCluskey, A.J., Poon, G.M.K. and Gariepy, J. (2007) A rapid and universal tandem-purification strategy for recombinant proteins. Protein Science 16, 2726-2732.

Mctigue, M.A., Williams, D.R. and Tainer, J.A. (1995) Crystal-structures of a schistosomal drug and vaccine target - Glutathione-S-Transferase from *Schistosoma japonica* and its complex with the leading antischistosomal drug praziquantel. Journal of Molecular Biology 246, 21-27.

Mee, C., Banki, M.R. and Wood, D.W. (2008) Towards the elimination of chromatography in protein purification: Expressing proteins engineered to purify themselves. Chemical Engineering Journal 135, 56-62.

Miroux, B. and Walker, J.E. (1996) Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. Journal of Molecular Biology 260, 289-298.

Mitchell, D.A., Marshall, T.K. and Deschenes, R.J. (1993) Vectors for the inducible overexpression of Glutathione-S-transferase fusion proteins in yeast. Yeast 9, 715-722.

Motejadded, H. and Altenbuchner, J. (2009) Construction of a dual-tag system for gene expression, protein affinity purification and fusion protein processing. Biotechnology Letters 31, 543-549.

Mujacic, M., Cooper, K.W. and Baneyx, F. (1999) Cold-inducible cloning vectors for low-temperature protein expression in *Escherichia coli*: application to the production of a toxic and proteolytically sensitive fusion protein. Gene 238, 325-332.

Nallamsetty, S., Austin, B.P., Penrose, K.J. and Waugh, D.S. (2005) Gateway vectors for the production of combinatorially-tagged His_6-MBP fusion proteins in the cytoplasm and periplasm of *Escherichia coli*. Protein Science 14, 2964-2971.

Nallamsetty, S. and Waugh, D.S. (2006) Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. Protein Expression and Purification 45, 175-182.

Nallamsetty, S. and Waugh, D.S. (2007) Mutations that alter the equilibrium between open and closed conformations of *Escherichia coli* maltose-binding protein impede its ability to enhance the solubility of passenger proteins. Biochemical and Biophysical Research Communications 364, 639-644.

Nelson, M.R. and Chazin, W.J. (1998) An interaction-based analysis of calcium-induced conformational changes in Ca^{2+} sensor proteins. Protein Science 7, 270-282.

Nettleship, J.E., Assenberg, R., Diprose, J.M., Rahman-Huq, N. and Owens, R.J. (2010) Recent advances in the production of proteins in insect and mammalian cells for structural biology. Journal of Structural Biology 172, 55-65.

Nikaido, H. (1994) Maltose Transport-System of *Escherichia coli* - an Abc-Type Transporter. Febs Letters 346, 55-58.

Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H. and Yura, T. (1998) Chaperone coexpression plasmids: Differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2 in *Escherichia coli*. Applied and Environmental Microbiology 64, 1694-1699.

O'Neill, S.M., Parkinson, M., Dowd, A.J., Strauss, W., Angles, R. and Dalton, J.P. (1999) Short report: Immunodiagnosis of human fascioliasis using recombinant *Fasciola hepatica* cathepsin L1 cysteine proteinase. The American Journal of Tropical Medicine and Hygiene 60, 749-51.

Ongkudon, C.M., Chew, J.H., Liu, B. and Danquah, M.K. (2012) Chromatographic removal of endotoxins: a bioprocess engineer's perspective. International Scholarly Research Network - ISRN Chromatography 649746.

Ohana, R.F., Encell, L.P., Zhao, K., Simpson, D., Slater, M.R., Urh, M. and Wood, K.V. (2009) HaloTag7: A genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification. Protein Expression and Purification 68, 110-120.

Pacheco, B., Crombet, L., Loppnau, P. and Cossar, D. (2012) A screening strategy for heterologous protein expression in *Escherichia coli* with the highest return of investment. Protein Expression and Purification 81, 33-41.

Panavas, T., Sanders, C. and Butt, T.R. (2009) SUMO fusion technology for enhanced protein production in prokaryotic and eukaryotic expression systems. Methods in Molecular Biology, 303-17.

Park, J.S., Han, K.Y., Lee, J.H., Song, J.A., Ahn, K.Y., Seo, H.S., Sim, S.J.J., Kim, S.W. and Lee, J. (2008) Solubility enhancement of aggregation-prone heterologous proteins by fusion expression using stress-responsive *Escherichia coli* protein, RpoS. Bmc Biotechnology 8.

Parks, T.D., Leuther, K.K., Howard, E.D., Johnston, S.A. and Dougherty, W.G. (1994) Release of proteins and peptides from fusion proteins using a recombinant plant-virus proteinase. Analytical Biochemistry 216, 413-417.

Perler, F.B., Davis, E.O., Dean, G.E., Gimble, F.S., Jack, W.E., Neff, N., Noren, C.J., Thorner, J. and Belfort, M. (1994) Protein splicing elements - Inteins and Exteins - a definition of terms and recommended nomenclature. Nucleic Acids Research 22, 1125-1127.

Peubez, I., Chaudet, N., Mignon, C., Hild, G., Husson, S., Courtois, V., De Luca, K., Speck, D. and Sodoyer, R. (2010) Antibiotic-free selection in *E. coli*: new considerations for optimal design and improved production. Microbial Cell Factories 9.

Pryor, K.D. and Leiting, B. (1997) High-level expression of soluble protein in *Escherichia coli* using a His_6-tag and maltose-binding-protein double-affinity fusion system. Protein Expression and Purification 10, 309-319.

Queiroz, J.A., Tomaz, C.T. and Cabral, J.M.S. (2001) Hydrophobic interaction chromatography of proteins. Journal of Biotechnology 87, 143-159.

Ram, D., Grossman, Z., Markovics, A., Avivi, A., Ziv, E., Lantner, F. and Schechter, I. (1989) Rapid changes in the expression of a gene encoding a calcium-binding protein in *Schistosoma mansoni*. Molecular and Biochemical Parasitology 34, 167-175.

Reddi, H., Bhattacharya, A. and Kumar, V. (2002) The calcium-binding protein of *Entamoeba histolytica* as a fusion partner for expression of peptides in *Escherichia coli*. Biotechnology and Applied Biochemistry 36, 213-218.

Robinson, M.W., Menon, R., Donnelly, S.M., Dalton, J.P. and Ranganathan, S. (2009) An integrated transcriptomics and proteomics analysis of the secretome of the helminth pathogen *Fasciola hepatica*: proteins associated with invasion and infection of the mammalian host. Molecular & Cellular Proteomics 8, 1891-907.

Rocco, C.J., Dennison, K.L., Klenchin, V.A., Rayment, I. and Escalante-Semerena, J.C. (2008) Construction and use of new cloning vectors for the rapid isolation of recombinant proteins from *Escherichia coli*. Plasmid 59, 231-237.

Ron, D. and Dressler, H. (1992) Pgstag - a versatile bacterial expression plasmid for enzymatic labeling of recombinant proteins. Biotechniques 13, 866-869.

Rondahl, H., Nilsson, B. and Holmgren, E. (1992) Fusions to the 5' end of a gene encoding a 2-domain analog of Staphylococcal Protein-A. Journal of Biotechnology 25, 269-287.

Rozanas, C. (1998) Purification of calcium-binding proteins using hydrophobic interaction chromatography. Life Science News I.

Rudert, F., Visser, E., Gradl, G., Grandison, P., Shemshedini, L., Wang, Y., Grierson, A. and Watson, J. (1996) pLEF, a novel vector for expression of glutathione S-transferase fusion proteins in mammalian cells. Gene 169, 281-282.

Russell, S.L., McFerran, N.V., Hoey, E.M., Trudgett, A. and Timson, D.J. (2007) Characterisation of two calmodulin-like proteins from the liver fluke, *Fasciola hepatica*. The Journal of Biological Chemistry 388, 593-9.

Sachdev, D. and Chirgwin, J.M. (2000) Fusions to maltose-binding protein: Control of folding and solubility in protein purification. In: Applications of Chimeric Genes and Hybrid Proteins, Pt A, Vol. 326, p. 312-321.

Salazar-Calderon, M., Martin-Alonso, J.M., de Eguino, A.D.R., Casais, R., Marin, M.S. and Parra, F. (2000) *Fasciola hepatica*: Heterologous expression and functional characterization of a thioredoxin peroxidase. Experimental Parasitology 95, 63-70.

Satakarni, M. and Curtis, R. (2011) Production of recombinant peptides as fusions with SUMO. Protein Expression and Purification 78, 113-119.

Scheich, C., Sievert, V. and Bussow, K. (2003) An automated method for high-throughput protein purification applied to a comparison of His-tag and GST-tag affinity chromatography. Bmc Biotechnology 3.

Schlieker, C., Bukau, B. and Mogk, A. (2002) Prevention and reversion of protein aggregation by molecular chaperones in the *E. coli* cytosol: implications for their applicability in biotechnology. Journal of Biotechnology 96, 13-21.

Schmidt, T.G.M. and Skerra, A. (1994) One-step affinity purification of bacterially produced proteins by means of the Strep tag and immobilized recombinant core Streptavidin. Journal of Chromatography A 676, 337-345.

Schumann, W. and Ferreira, L.C.S. (2004) Production of recombinant proteins in *Escherichia coli*. Genetics and Molecular Biology 27, 442-453.

Schwaller, B. (2010) Cytosolic Ca^{2+} Buffers. Cold Spring Harbor Perspectives in Biology 2.

Sevastsyanovich, Y.R., Alfasi, S.N. and Cole, J.A. (2010) Sense and nonsense from a systems biology approach to microbial recombinant protein production. Biotechnology and Applied Biochemistry 55, 9-28.

Sheibani, N. (1999) Prokaryotic gene fusion expression systems and their use in structural and functional studies of proteins. Preparative Biochemistry and Biotechnology 29, 77-90.

Shih, Y.P., Kung, W.M., Chen, J.C., Yeh, C.H., Wang, A.H.J. and Wang, T.F. (2002) High-throughput screening of soluble recombinant proteins. Protein Science 11, 1714-1719.

Shimizu, F., Sanada, K. and Fukada, Y. (2003) Purification and immunohistochemical analysis of calcium-binding proteins expressed in the chick pineal gland. Journal of Pineal Research 34, 208-216.

Silva, E., Castro, A., Lopes, A., Rodrigues, A., Dias, C., Conceicao, A., Alonso, J., da Costa, J.M.C., Bastos, M., Parra, F., Moradas-Ferreira, P. and Silva, M. (2004) A recombinant antigen recognized by *Fasciola hepatica*-infected hosts. Journal of Parasitology 90, 746-751.

Smith, D.B. and Johnson, K.S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene 67, 31-40.

Smyth, D.R., Mrozkiewicz, M.K., McGrath, W.J., Listwan, P. and Kobe, B. (2003) Crystal structures of fusion proteins with large-affinity tags. Protein Science 12, 1313-1322.

Sommer, B., Friehs, K., Flaschel, E., Reck, M., Stahl, F. and Scheper, T. (2009) Extracellular production and affinity purification of recombinant proteins with *Escherichia coli* using the versatility of the maltose binding protein. Journal of Biotechnology 140, 194-202.

Song, J.A., Lee, D.S., Park, J.S., Han, K.Y. and Lee, J. (2011) A novel *Escherichia coli* solubility enhancer protein for fusion expression of aggregation-prone heterologous proteins. Enzyme and Microbial Technology 49, 124-130.

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

Sorensen, H.P. and Mortensen, K.K. (2005b) Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. Microbial Cell Factories 4.

Sorensen, H.P., Sperling-Petersen, H.U. and Mortensen, K.K. (2003a) A favorable solubility partner for the recombinant expression of streptavidin. Protein Expression and Purification 32, 252-259.

Sorensen, H.P., Sperling-Petersen, H.U. and Mortensen, K.K. (2003b) Production of recombinant thermostable proteins expressed in *Escherichia coli*: completion of protein synthesis is the bottleneck. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences 786, 207-214.

Stahl, S. and Nygren, P.A. (1997) The use of gene fusions to protein A and protein G in immunology and biotechnology. Pathologie Biologie 45, 66-76.

Stewart, E.J., Aslund, F. and Beckwith, J. (1998) Disulfide bond formation in the *Escherichia coli* cytoplasm: an in vivo role reversal for the thioredoxins. Embo Journal 17, 5543-5550.

Su, Y., Zou, Z.R., Feng, S.Y., Zhou, P. and Cao, L.J. (2007) The acidity of protein fusion partners predominantly determines the efficacy to improve the solubility of the target proteins expressed in *Escherichia coli*. Journal of Biotechnology 129, 373-382.

Takakura, Y., Oka, N., Kajiwara, H., Tsunashima, M., Usami, S., Tsukamoto, H., Ishida, Y. and Yamamoto, T. (2010) Tamavidin, a versatile affinity tag for protein purification and immobilization. Journal of Biotechnology 145, 317-322.

Tanaka, N. and Fersht, A.R. (1999) Identification of substrate binding site of GroEL minichaperone in solution. Journal of Molecular Biology 292, 173-180.

Terpe, K. (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Applied Microbiology and Biotechnology 60, 523-533.

Terpe, K. (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Applied Microbiology and Biotechnology 72, 211-222.

Thomas, J.G., Ayling, A. and Baneyx, F. (1997) Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from *E. coli*. To fold or to refold. Applied Biochemistry and Biotechnology 66, 197-238.

Tomme, P., Boraston, A., McLean, B., Kormos, J., Creagh, A.L., Sturch, K., Gilkes, N.R., Haynes, C.A., Warren, R.A.J. and Kilburn, D.G. (1998) Characterization and affinity applications of cellulose-binding domains. Journal of Chromatography B 715, 283-296.

Turner, P., Holst, O. and Karlsson, E.N. (2005) Optimized expression of soluble cyclomaltodextrinase of thermophilic origin in *Escherichia coli* by using a soluble

fusion-tag and by tuning of inducer concentration. Protein Expression and Purification 39, 54-60.

Vaillancourt, P., Zheng, C.F., Hoang, D.Q. and Briester, L. (2000) Affinity purification of recombinant proteins fused to calmodulin or to calmodulin-binding peptides. Applications of Chimeric Genes and Hybrid Proteins, Pt A 326, 340-362.

Vergis, J.M. and Wiener, M.C. (2011) The variable detergent sensitivity of proteases that are utilized for recombinant protein affinity tag removal. Protein Expression and Purification 78, 139-142.

Viljanen, J., Larsson, J. and Broo, K.S. (2008) Orthogonal protein purification - Expanding the repertoire of GST fusion systems. Protein Expression and Purification 57, 17-26.

Vinckier, N.K., Chworos, A. and Parsons, S.M. (2011) Improved isolation of proteins tagged with glutathione S-transferase. Protein Expression and Purification 75, 161-164.

Walsh, G. (2010) Biopharmaceutical benchmarks 2010. Nature Biotechnology 28, 917-924.

Wang, H.Y., Xiao, Y.C., Fu, L.J., Zhao, H.X., Zhang, Y.F., Wan, X.S., Qin, Y.X., Huang, Y.D., Gao, H.C. and Li, X.K. (2010) High-level expression and purification of soluble recombinant FGF21 protein by SUMO fusion in *Escherichia coli*. Bmc Biotechnology 10.

Wang, Z.Y., Li, N., Wang, Y.Y., Wu, Y.P., Mu, T.Y., Zheng, Y., Huang, L.Q. and Fang, X.X. (2012) Ubiquitin-intein and SUMO2-intein fusion systems for enhanced protein production and purification. Protein Expression and Purification 82, 174-178.

Waugh, D.S. (2005) Making the most of affinity tags. TRENDS in Biotechnology 23.

Waugh, D.S. (2011) An overview of enzymatic reagents for the removal of affinity tags. Protein Expression and Purification 80, 283-293.

Wilson, M.J., Haggart, C.L., Gallagher, S.P. and Walsh, D. (2001) Removal of tightly bound endotoxin from biological products. Journal of Biotechnology 88, 67-75.

Yasukawa, T., Kanei-Ishii, C., Maekawa, T., Fujimoto, J., Yamamoto, T. and Ishii, S. (1995) Increase of solubility of foreign proteins in *Escherichia coli* by coproduction of the bacterial thioredoxin. The Journal of Biological Chemistry 270, 25328-31.

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

Zhang, Y.J. and Cremer, P.S. (2006) Interactions between macromolecules and ions: the Hofmeister series. Current Opinion in Chemical Biology 10, 658-663.

Zhang, Y.B., Howitt, J., McCorkle, S., Lawrence, P., Springer, K. and Freimuth, P. (2004) Protein aggregation during overexpression limited by peptide extensions with large net negative charge. Protein Expression and Purification 36, 207-216.

Zhou, P., Lugovskoy, A.A. and Wagner, G. (2001) A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. Journal of Biomolecular Nmr 20, 11-14.

Zhou, Y., Yang, W., Kirberger, M., Lee, H.W., Ayalasomayajula, G. and Yang, J.J. (2006) Prediction of EF-hand calcium-binding proteins and analysis of bacterial EF-hand proteins. Proteins-Structure Function and Bioinformatics 65, 643-655.

Zou, Z.R., Cao, L.J., Zhou, P., Su, Y., Sun, Y.H. and Li, W.J. (2008) Hyper-acidic protein fusion partners improve solubility and assist correct folding of recombinant proteins expressed in *Escherichia coli*. Journal of Biotechnology 135, 333-339.

1.6. Appendices

1.6.1. Patent WO 2010/082097: Fusion proteins, its preparation process and its application on recombinant protein expression systems

Please, open here:

https://dl.dropbox.com/u/10833879/Fh8tag_Patents/WO2010082097.pdf

1.6.2. The Hofmeister series

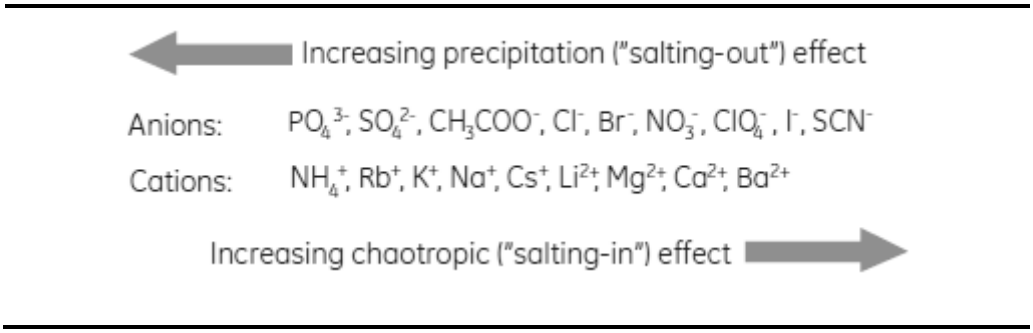


Figure A1.1. The Hofmeister series (adapted from GE Healthcare, 2006).

Chapter 2

The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology

Adapted from *Applied Microbiology and Biotechnology*, *In press*.

Abstract

The *Escherichia coli* host system is an advantageous choice for simple and inexpensive recombinant protein production but it still presents bottlenecks at expressing soluble proteins from other organisms. Several efforts have been taken to overcome *E. coli* limitations, including the use of fusion partners that improve protein expression and solubility. New fusion technologies are emerging to complement the traditional solutions. This work evaluates two novel fusion partners, the Fh8 tag (8 kDa) and the H tag (1 kDa), as solubility enhancing tags in *E. coli* and their comparison to commonly used fusion partners.

A broad range comparison was conducted in a small-scale screening and subsequently scaled-up. Six difficult-to-express target proteins (RVS167, SPO14, YPK1, YPK2, Frutalin and CP12) were fused to eight fusion tags (His, Trx, GST, MBP, NusA, SUMO, H and Fh8). The resulting protein expression and solubility levels were evaluated by sodium dodecyl sulfate polyacrylamide gel electrophoresis before and after protein purification and after tag removal. The Fh8 partner improved protein expression and solubility as the well-known Trx, NusA or MBP fusion partners. The H partner did not function as a solubility tag. Cleaved proteins from Fh8 fusions were soluble and obtained in similar or higher amounts than proteins from the cleavage of other partners as Trx, NusA or MBP.

The Fh8 fusion tag therefore acts as an effective solubility enhancer, and its low molecular weight potentially gives it an advantage over larger solubility tags by offering a more reliable assessment of the target protein solubility when expressed as a fusion protein.

2.1. Introduction

The production of soluble and functional protein in *Escherichia coli* is still a major challenge in biotechnology research. In spite of its fast growth, low cost, high-productivity, and extensive genetic characterization, *E. coli* occasionally still suffers from low expression and/or low solubility of target proteins (Terpe, 2006; Peti and Page, 2007; Demain and Vaishnav, 2009; Makino et al., 2011).

Several efforts have been exploited to prevent recombinant protein aggregation and to improve its soluble production by the use of different promoters, expression strains and induction conditions, co-expression of chaperones, and soluble fusion partners (Sorensen and Mortensen, 2005; Studier, 2005; Berrow et al., 2006; Makino et al., 2011; Vernet et al., 2011; Pacheco et al., 2012). The fusion of a highly soluble carrier to recombinant proteins has been generally used to improve protein solubility and expression in *E. coli* (Terpe, 2003; Waugh, 2005; Esposito and Chatterjee, 2006; Ohana et al., 2009), although success is not yet always guaranteed.

The Trx (LaVallie et al., 2000), GST (Smith and Johnson, 1988; Smith, 2000), MBP (Kapust and Waugh, 1999; Sachdev and Chirgwin, 2000), and NusA (Davis et al., 1999; De Marco et al., 2004) fusion partners are commonly employed as solubility enhancing carriers, but when producing a recombinant protein for structural and functional applications, these fusion partners must often be removed. The removal of fusion partners is usually made by specific protease sites included between the fusion tag and the target protein. However, after cleavage of the soluble fusion partner, precipitation of the target proteins can occur. Here, a major bottleneck appears as the target protein solubility can dramatically change in the presence and absence of the fusion partner. Meanwhile, new fusion solutions are constantly emerging and complementing the other fusion partners, as for instance, SUMO fusion technology (Michael P. Malakhov and Butt, 2004; Marblestone et al., 2006).

A novel fusion system (Hitag®) is presented in this work: the Fh8 and H partners. The Fh8 fusion partner is an 8-kDa calcium-binding recombinant protein (GenBank ID AF213970) extracted from the parasite *Fasciola hepatica* and it has been previously used on the diagnosis of parasite infections (Silva et al., 2004). The recombinant Fh8 was also studied before by directed mutagenesis, in which its N-terminal sequence

revealed to be important for Fh8 stability and production in *E. coli* (not published). From this analysis, a new fusion partner was suggested: the H tag that corresponds to the first eleven aminoacids of the Fh8 N-terminus, resulting in a molecular weight of 1 kDa.

In this work, both Fh8 and H fusion partners are explored as solubility enhancing partners and compared to the commonly used fusion partners Trx, GST, MBP, NusA, and SUMO. The study conducted here does not only evaluate the novel fusion system effect on protein solubility but also the behavior of target proteins after Fh8 and H tags removal in comparison to the other fusion partners. Six difficult-to-express target proteins in *E. coli* were fused to eight fusion tags and the resulting solubility compared in a broad screening before and after tag removal.

2.2. Materials and Methods

2.2.1. General

In this work, all the cloning PCRs used the Phusion High-Fidelity DNA Polymerase (New England Biolabs) with an annealing temperature of 55 ºC, according to the manufacturer's instructions. The colony PCRs were conducted using the in-house DNA Taq Polymerase with an annealing temperature of 55 ºC and with the T7 forward and reverse universal primers. Plasmid DNA extractions were performed using the Qiagen kits for maxi- and minipreps and the QIAquick DNA gel extraction kit or QIAquick PCR purification kit (Qiagen) were used for DNA purification. The restriction enzymes used in this work were from New England Biolabs. All the DNA ligations were carried out with the Rapid DNA Ligation kit (Roche). For plasmid maintenance and protein expression, different antibiotics were used depending on the strain and plasmid requirements. Antibiotic stock solutions were prepared, filtered through 0.2 µm, and stored at -20 ºC in the following concentrations: kanamycin, 30 $mg.mL^{-1}$; carbenicillin, 100 $mg.mL^{-1}$; and chloramphenicol, 10 $mg.mL^{-1}$.

2.2.2. Construction of the pETMFh8 and pETMH vectors

The insertion of the Fh8 tag with the TEV cleavage site into the pETM10 vector (Table 2.1) was carried out by DNA ligation of the *Nco*I-*Kpn*I digested plasmid and *Pci*I-*Kpn*I digested *fh8* PCR product. The pETMH fusion vector was obtained from the pETM11 backbone (Table 2.1), performing three PCRs: The PCR-I inserted part of the H tag (the first 28 nucleotides) after the His_6 tag sequence from the pETM11 plasmid. The PCR-II inserted part of the H tag (the final 28 nucleotides) before the TEV cleavage site of the pETM11 plasmid. As the H tag sequence has only 33 basepairs (bp), specific primers were designed for the PCR I and II in order to have 23 nucleotides of the H tag sequence matching in both PCRs. The universal primers T7 forward and reverse were used in the PCR-III for amplification of the His_6 tag/H tag/TEV-site sequence to be cloned into the pETM11 plasmid. The purified PCR-III product and the pETM11 plasmid were digested with *Xba*I and *Xho*I restriction enzymes, and the final pETMH vector was obtained by ligation of the digested DNAs.

E. coli DH5α competent cells were transformed with the constructed pETMFh8 and pETMH plasmids, and the obtained clones were analyzed by colony PCR. The novel pETMFh8 and pETMH fusion vectors were confirmed by sequencing with both T7 forward and reverse universal primers.

Table 2.1. Construction of the pETMFh8 and pETMH expression vectors:

Backbone	Vector	PCR	Primers	Sequence (5'→3')	Comments
pETM 10	pETMFh8	I	Fh8-FWD	TCTATT**ACATGT**CCCCTAGTGTTCAAGAGGTTGAAAAAC	In bold is the *PciI* restriction enzyme sequence and underlined is the initial part of the Fh8 tag sequence.
			Fh8-RV	AATAGA**GGTACC**GGAAC**CCATGG**A***GCCCTGAAAATAAAGATTCTC***TGACAAAATCGAAACGAG	In bold is the *KpnI* restriction enzyme sequence and in bold underlined is the *NcoI* restriction enzyme sequence. In ***italic bold*** is the TEV recognition cleavage site and underlined is the final part of the Fh8 tag sequence.
pETM 11	pETMH	I	T7-FWD	TAATACGACTCACTATAGGG	Underlined are the first 28 nucleotides of the H tag sequence. In bold underlined are the 23 common nucleotides used in both I and II PCRs. In ***italic bold*** is the His_6 tag sequence of the pETM11 plasmid.
			Htag-RV	**GTTTTTCAACCTCTTGAACACTA**GGCAT***GTGATGGTGATGGTGATG***TTTC	
		II	Htag-FWD	**TAGTGTTCAAGAGGTTGAAAAAC**TCCTT***GAGAATCTTTATTTTCAGGGC***	Underlined are the final 28 nucleotides of the H tag sequence. In bold underlined are the 23 common nucleotides used in both I and II PCRs. In ***italic bold*** is the TEV recognition cleavage site of the pETM11 plasmid.
			T7-RV	GCTAGTTATTGCTCAGCGG	

2.2.3. Cloning of the target genes into pETM vectors

The target genes *ypk1_frag*, *rvs167*, *spo14_frag* and *ypk2* used in this work (Table 2.2) are synthetic genes previously optimized for *E. coli* expression. All these genes were inserted into the pETM vectors (Table 2.3) using the *Nco*I/*Bam*HI - *Xho*I restriction sites, presented at the beginning of each gene sequence and after the stop codon, respectively. The *frutalin* and *cp12* genes (Table 2.2) were modified by PCR to have the *Nco*I/*Bam*HI-*Xho*I restriction sites at the same positions as the other target genes. A similar primer designing was used to amplify both *frutalin* and *cp12* sequences as follows: primer forward 5´-TCTATT**CCATG*G***GATCC*-22 initial nt of the target gene - 3' with the *NcoI* restriction site in bold, the *BamHI* restriction site in *italic* and the first 22 nucleotides of each gene underlined; and primer reverse 5'-AATAGA**CTCGAG**-22 final nt of the target gene-3' with the *XhoI* restriction site in bold and the final 22 nucleotides of each gene underlined. After ligation of the digested PCR products and plasmids, *E. coli* DH5α competent cells were transformed, and the resulting clones were analyzed and confirmed as previously mentioned.

2.2.4. Expression strains and culture conditions

Four different expression strains were evaluated in this work: the *E. coli* BL21 (DE3) Codon Plus-RIL; the *E. coli* Rosetta (DE3) cells, the *E. coli* Tuner (DE3) strain, and the *E. coli* SoluBL21 strain. Competent cells of the four different strains were prepared and transformed with the constructed fusion vectors. The resulting clones were confirmed by colony PCR, and one positive clone was selected for the expression trials. All the cells were grown in LB media with the appropriate antibiotics diluted with a factor of 1000 (see Table 2.3 for selection and maintenance of pETM vectors). In this work, the degradation-resistant carbenicillin was used instead of ampicillin. Precultures were grown overnight (o/n) at 37 ºC and a dilution factor of 100 was used for inoculation of all cultures (usually corresponding to a starting OD_{600nm} of 0.02). Cultures were performed in parallel, using 10 mL of culture media in 24 deep-well plates (25 mL capacity per well - DIA Nielsen GmbH&Co. KG, Germany) for small-scale screenings and 500 mL of culture media in 2 L flasks for the scale-up. Cultures were grown at 37 ºC and 200 rpm to a final OD_{600nm} of 0.4-0.6 before induction. In the small-scale study, two plates with the same strain and fusion proteins were used to test different induction

conditions: isopropyl-β-D-1-thiogalactopyranoside at 0.5 mM, 28 ºC and 3 hours (first plate) or at 0.2 mM, 18 ºC and o/n (second plate). Each 10-mL culture was divided in two 5-mL cultures. All cells were harvested for 25 min, at 4 ºC and 4500 rpm. Cell pellets from 500 mL cultures were washed once with 1x phosphate-buffered saline, and collected again by centrifugation. Bacterial pellets were flash frozen in liquid nitrogen and stored at -20 ºC.

2.2.5. Cell lysis

In the small-scale screening, two sonication protocols were tested using an eight-microtip sonicator or a single-tip sonicator (G. Heinemann, Germany). For the eight-microtip lysis, cells from a 5-mL culture pellet were resuspended in 1 mL of lysis buffer [50 mM Tris pH 8.0, 250 mM NaCl, 20 mM imidazole buffer supplemented with 1x complete free EDTA protease inhibitor (Roche), 5 mM $MgCl_2$ (Sigma), 5 $\mu g.mL^{-1}$ DNAse (Sigma) and 1 $mg.mL^{-1}$ lysozyme (Sigma)] and transferred to a 96-deep-well plate (2 mL capacity per well) and incubated at room temperature for 10 min. The plate was placed on ice, and cells were further lysed by sonication. The lysis efficiency was improved by adding 400 µL of 0.5 mm beads to each well. The 96-deep-well plates were then centrifuged at 4000 rpm, 4 ºC for 45 min, and the supernatant fraction was collected to a new 96-deep-well plate. For the single-tip lysis, 5-mL cell pellets were resuspended in 1 mL of Lysis buffer and transferred to a 2-mL tube containing 0.5 mm beads. After 10 min incubation at room temperature, bacterial cells were lysed by sonication, and 2 mL-tubes were centrifuged at 13000 rpm at 4 ºC for 25 min. The supernatant fractions were transferred to new 2-mL tubes.

In the scale-up experiments, cell pellets were thawed and resuspended in 10 mL of lysis buffer. After incubation at room temperature for 10 min, cells were sonicated (Branson Sonifier 250, G. Heinemann, Germany), and the soluble fraction was removed from the insoluble cell debris by ultracentrifugation at 40000 rpm, 4 ºC for 30 min. In all the experiments, aliquots of total lysates and supernatant samples were taken and prepared to be analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

Table 2.2. Features and properties of target genes used in this study**:**

Synthetic genes	GenBank/ Uniprot	Organism	Localization	Function or application	Gene size (nt)	Protein size (aa)	MW (Da)	pI	Cysteines (%)	GRAVY
Frutalin	Jacalin-related lectin: L03795.1 **Q38720**	*Artocarpus incisa*	Plant seeds	Biomarker of cancer cells [27]	477	158	17194	8.73	0	-0.089
cp12fragment	XM_625821.1 **Q5CR31**	*Cryptosporidium parvum*	Oocyst surface	Involved in the host-control interaction [25]	216	71	8414.2	4.72	0	-1.104
ypk1fragment	NM_001179692.1 **P12688**	*Saccharomyces cerevisiae*	Cell membrane and cytoplasm	Proliferation of yeast cells; involved in signaling pathways (Uniprot)	964	312	35629	6.54	1.6	-0.224
rvs167	NM_001180696.1 **P39743**	*Saccharomyces cerevisiae*	Cytoplasm and cytoskeleton	Formation of endocytic vesicles and cytoskeletal reorganization (Uniprot)	1516	496	54134	5.77	0.2	-0.470
spo14fragment	NM_001179821.1 **P36126**	*Saccharomyces cerevisiae*	Endosome, nucleus and prospore membrane	Meiosis and spore formation (Uniprot)	1192	388	44828	8.56	0.5	-0.579
ypk2	NM_001182604.1 **P18961**	*Saccharomyces cerevisiae*	Nucleus and cytoplasm	Proliferation of yeast cells; involved in signaling pathways (Uniprot)	2104	692	78111	7.33	0.9	-0.427

nt – nucleotides; aa – aminoacids; MW – molecular weight; Da – Dalton; pI – Isoelectric point; GRAVY – Grand average of hydropathicity

Table 2.3. Features of expression vectors and properties of the tags used in this work:

Expression vectors (pETM)	Fusion protein	Tag size (aa)	pI	MW tag (Da)	GRAVY	Promoter	Selection	Protease cleavage site
11 (His)	MK-***His$_6$***-TEV(*ENLYFQ↓G)*A**M**GS-**Target**	29	6.24	3,419.7	-1.09	T7/lac	kan	TEV
20 (Trx)	M-***Trx***- -His$_6$-TEV(*ENLYFQ↓G)*A**M**GS-**Target**	136	5.66	14,673	-0.189	T7/lac	amp	TEV
30 (GST)	MK-His$_6$-***GST***-TEV(*ENLYFQ↓G)*A**M**GS-**Target**	254	6.31	29,264	-0.446	T7/lac	kan	TEV
41 (MBP)	MK-His$_6$- ***MBP***-TEV(*ENLYFQ↓G)*A**M**GS-**Target**	404	5.49	44,549	-0.462	T7/lac	kan	TEV
60 (NusA)	M-***NusA***- His$_6$-TEV(*ENLYFQ↓G)*A**M**GS-**Target**	519	4.63	57,383	-0.306	T7/lac	kan	TEV
SUMO	MK-His$_6$-TEV(ENLYFQG)-***SUMO***-TGGS-**Target**	108	6.02	12,453	-0.882	T7/lac	kan	SUMO
H	MK-His$_6$-***Htag***-TEV(*ENLYFQ↓G)*A**M**GS-**Target**	30	6.56	3553.0	-0.773	T7/lac	kan	TEV
Fh8	MK-His$_6$-***Fh8tag***-TEV(*ENLYFQ↓G)*S**M**GS-**Target**	90	6.43	10,146	-0.780	T7/lac	kan	TEV

aa – aminoacids; ; pI – Isoelectric point; MW – molecular weight; Da – Dalton; GRAVY – Grand average of hydropathicity

2.2.6. Protein purification and tag removal

The selected fusion proteins were purified by nickel affinity chromatography using a spin protocol with Ni-NTA slurry (Qiagen) for the small-scale screening or a semi-automated system, in which 100 µL prepacked Ni-NTA superflow columns (Robot Columns, Atoll GmbH, Germany) were set on a 96-well-plate matrix, for the scale-up experiments. Both purifications were conducted according to manufacturer's instructions, using 50 mM Tris pH 8.0, 250 mM NaCl with 20 mM imidazole as running and washing buffer, and with 300 mM imidazole for the elution buffer. The collected fractions were analyzed by SDS-PAGE, and the total protein content of robot eluted samples was also estimated by Bradford assay.

For tag removal, the selected and purified proteins were digested with the TEV protease. After digestion, aliquots were taken, and samples were centrifuged at 13000 rpm for 10 min at 4 ºC. The cleaved proteins were then purified from the fusion tags and proteases by nickel affinity chromatography using the same above-mentioned protocol. The collected samples were prepared to be analyzed by SDS-PAGE.

2.2.7. Protein expression and solubility evaluation

The expression and solubility evaluation was conducted using a score of 0, 1, 2, and 3, based on the expected soluble protein production yields obtained after the purification of 1-L cultures (Berrow et al., 2006; Bird, 2011). The score 0 corresponds to no expression/soluble protein; a score of 1 was given to the soluble expressions that are expected to yield less than 0.5 mg; the score of 2 indicates a soluble expression yield between 0.5 mg and 5 mg; and a score of 3 specifies the soluble expressions with an expected yield higher than 5 mg. The soluble expression results were estimated from the SDS-PAGE analysis of the eluted fractions from nickel column and also from the corresponding supernatant fractions, according to the following equations:

Equation 2.1:

$$Protein\ Soluble\ Expression = \frac{number\ of\ scores\ 1{,}2\ and\ 3\ for\ the\ tested\ protein}{number\ of\ fusion\ tags\ tested}$$

Equation 2.2:

$$Tag\ Soluble\ Expression\ = \frac{number\ of\ soluble\ proteins\ fused\ to\ the\ tag\ with\ scores\ 1{,}2\ and\ 3}{number\ of\ total\ proteins\ tested}$$

Equation 2.3:

$$Tag\ Soluble\ Score\ 2+3 = \frac{number\ of\ soluble\ proteins\ fused\ to\ the\ tag\ with\ scores\ 2\ and\ 3}{number\ of\ total\ proteins\ tested}$$

The total expression results were estimated from the SDS-PAGE analysis of the total lysate fractions using the same scores 0-3, but here for the expected expression yields, as follows:

Equation 4:

$$Tag\ Total\ Expression\ = \frac{number\ of\ expressed\ proteins\ \ with\ scores\ 1{,}2\ and\ 3}{number\ of\ total\ proteins\ tested}$$

2.3. Results

2.3.1. The novel pETMFh8 and pETMH fusion vectors

The fusion tags in this study, Fh8 and H tags, were inserted into the pETM vector series (Dummler et al., 2005) resulting into two novel expression vectors: the pETMFh8tag and pETMHtag (see Appendix 2.7.1, Figure A2.1). The pETMFh8tag vector was constructed using the pETM10 backbone and the pETMHtag vector used the pETM11 backbone (Table 2.1). Both new vectors share common features with the rest of the pETM vectors used in this work (Table 2.3): they have similar multiple cloning sites (MCS) suitable for direct subcloning of the target genes; two His_6 tag sequences, one placed at the N-terminal of the fusion partner and the other after the MCS; and a TEV protease cleavage site between the fusion partner and the MCS. The novel fusion plasmids are kanamycin resistant.

In this work, only the N-terminal His_6 tag was used for protein purification. In the pETM series, the Fh8 and H tags have only 10 and 3.6 kDa, respectively (Table 2.3). Both new fusion tags have lower molecular weights than the studied fusion tags, even compared to the low molecular weight tags such as SUMO (12.5 kDa) and Trx (14.7 kDa). The analysis of fusion tags features (Table 2.3) shows that Fh8 and H tags have identical isoelectric points, which are also similar to the GST or SUMO values.

The Grand average of hydropathicity (GRAVY) shows a predominantly hydrophilic nature of the fusion tags (all the GRAVY values are negative). The Fh8 and H tags along with SUMO and His tags present the highest hydrophilic character.

2.3.2. Cloning of target proteins into the pETM expression vectors

The synthetic genes used in this work (Table 2.2) were cloned in parallel into the pETM vectors by using the same restriction enzymes for all the constructions: all the pETM vectors received *Nco*I-*Xho*I digested target genes with the exception for pETMSUMO, which received *Bam*HI-*Xho*I digested inserts. This strategy allowed a rapid and easy cloning procedure and a successful cloning rate of 98%. From the 48 fusion genes to be constructed, only the cloning of the *ypk1_frag* into the pETMHtag vector was not successful, as confirmed by sequencing at the GATC (Germany).

The six target proteins used in this work were selected regarding their previous difficulty of expression in soluble form in *E. coli*. Four of these proteins (two full length, RVS167 and YPK2; and two truncated proteins, SPO14 and YPK1) are found in the yeast *Saccharomyces cerevisiae* and were chosen due to the little/absent soluble expression in *E. coli* obtained when formerly fused to human SUMO3 tag (data not shown). The other two target proteins [the truncated CP12 derived from the *Cryptosporidium parvum* parasite (Yao et al., 2007) and the lectin Frutalin from the plant seed *Artocarpus incisa* (Oliveira et al., 2009; Oliveira et al., 2011)] were previously expressed as full lengths in *E. coli* presenting poor solubility.

The selected target proteins have different locations and functions or applications, and they also differ in size: the smallest target protein used in this work (a truncated form of the CP12 protein lacking its transmembrane region) has only 71 amino acids, and the largest one, YPK2, has approximately 700 amino acids. All the target proteins have low

cysteine content, and they range from a slight to moderate hydrophilic nature (Table 2.2).

2.3.3. Small-scale screening: selection of the expression strain and culture conditions

Figure 2.1 presents the small-scale comparison of the soluble expression for the different strains, fusion tags, and target proteins. Figure 2.1a presents the SDS-PAGE analysis of supernatant samples from the selected strains, using the two induction conditions tested. Figure 2.1b result from the analysis of scores 1-3 in supernatant fractions obtained from the 8-microtip sonication protocol, using Eq. 2.1.

For all the four strains, the number of soluble expressed proteins using the Fh8 tag was similar to the number obtained with the MBP or NusA tags, and higher than the number obtained using the Trx and GST tags. Proteins fused to the H tag and His_6 tag presented the lowest soluble expression along with proteins fused to SUMO, whose insolubility was previously evaluated for four of the six targets. The solubility of SUMO-Frutalin fusion was difficult to evaluate, as observed by the small protein content in lanes 12 and 13 of Figure 2.1a. SUMO-CP12 and SUMO-YPK2 fusion proteins presented amounts of soluble protein similar to those of Fh8-CP12 and Fh8-YPK2, respectively, as observed in Figure 2.1a. CP12 and Frutalin were soluble expressed within the highest number of fusion tags, and this result was observed amongst all the four strains used in this work (Figure 2.1b). This small-scale screening resulted in the following rank of the fusion tags for protein solubility: MBP > NusA > Fh8 ~Trx > GST ~H > His.

According to the obtained results in Figure 2.1, the *E. coli* BL21 (DE3) Codon Plus-RIL strain was established for the expression of the fusions with CP12, RVS167, and YPK2 target proteins, and the *E. coli* Rosetta (DE3) strain was selected for the expression of the fusions with Frutalin and truncated SPO14. Finally, fusions with the truncated YPK1 protein were later expressed using the *E. coli* Tuner (DE3) strain. The *E. coli* SoluBL21 strain was not used further in this study.

Regarding the two different culture conditions tested, the overnight induction condition presented the highest level of soluble expression for five of the six target proteins. The 3-h induction condition was only the best option for the fusions with CP12 target protein.

All the soluble expressed fusion proteins presented a molecular weight identical to the expected size (see Appendix 2.7.2, Table A2.1). SUMO fusions migrated on SDS-PAGE with a molecular weight of 3 kDa higher than the expected, which is also known from previous publications (Marblestone et al., 2006).

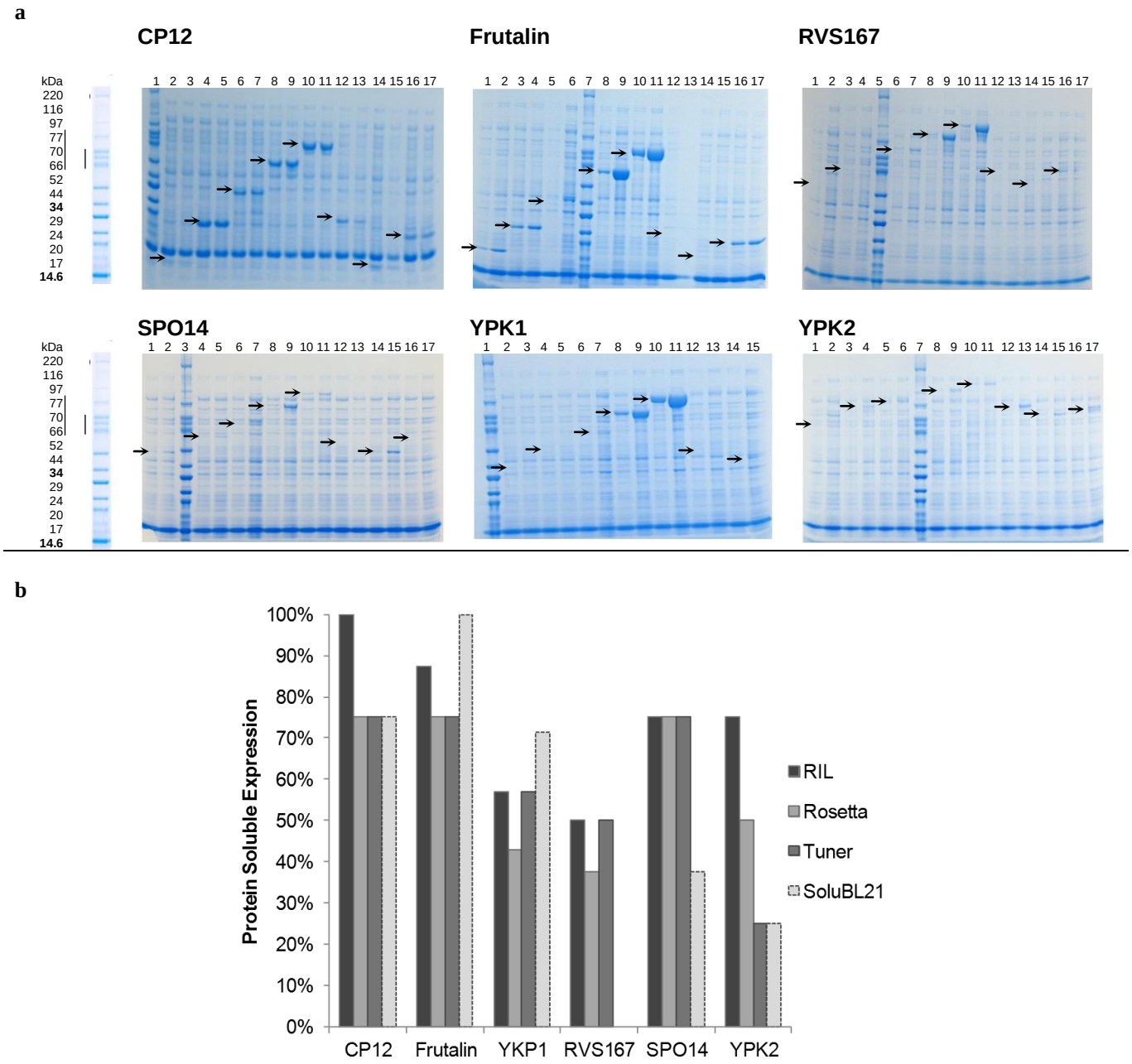


Figure 2.1. Comparison of fusion protein soluble expression using different *E. coli* strains: small-scale screening evaluation of the supernatant fractions from the cell lysis with the eight-tip sonicator. (**a**) SDS-PAGE of the supernatant samples from the two induction conditions tested. The 3-h induction samples were loaded aside with the o/n induction, according to the following fusion tags order: His – Trx – GST – MBP – NusA – SUMO – H – Fh8tag. The YPK1 gel follows the same order except for the H tag (which is not available with this target protein). The in-house protein marker was used for this analysis. *Arrows* indicate the expected molecular weights for all fusion proteins following the fusion tags order loaded on the gels. (**b**) Soluble expression (scores 1-3) per target protein with different strains. Plotted values were estimated by Equation 2.1.

2.3.4. Small-scale expression and solubility results: validation

For each target protein, four to six tags were selected to continue with the small-scale screening and to validate the initial solubility results. Table 2.4 summarizes the selected fusion proteins for the direct comparison of commonly used fusion tags and Fh8 tag.

Table 2.4. Selection of fusion proteins for small-scale screening – comparison groups:

Comparison group no.	Selected fusion tags	Target proteins
I	His, MBP, Fh8	Frutalin, RVS167, SPO14, YPK1, YPK2
II	His, NusA. Fh8	CP12, Frutalin, RVS167, YPK1
III	His, MBP, NusA, Fh8	Frutalin, RVS167, YPK1
IV	His, Htag, Fh8	CP12, Frutalin, YPK2
V	His, SUMO, Fh8	CP12, YPK2
VI	His, Trx, NusA, H, Fh8	CP12, Frutalin

Figure 2.2 presents the SDS-PAGE analysis of the total lysate and supernatant fractions of the selected fusion proteins that were obtained from the bacterial lysis with a single microtip sonication. These samples were also scored 0-3 (see Appendix 2.7.3, Figure A2.2). As observed, the MBP, NusA, and Trx fusions resulted in higher total expression than the Fh8 fusions. The Fh8 fusions resulted in a total expression identical to the H and His fusions.

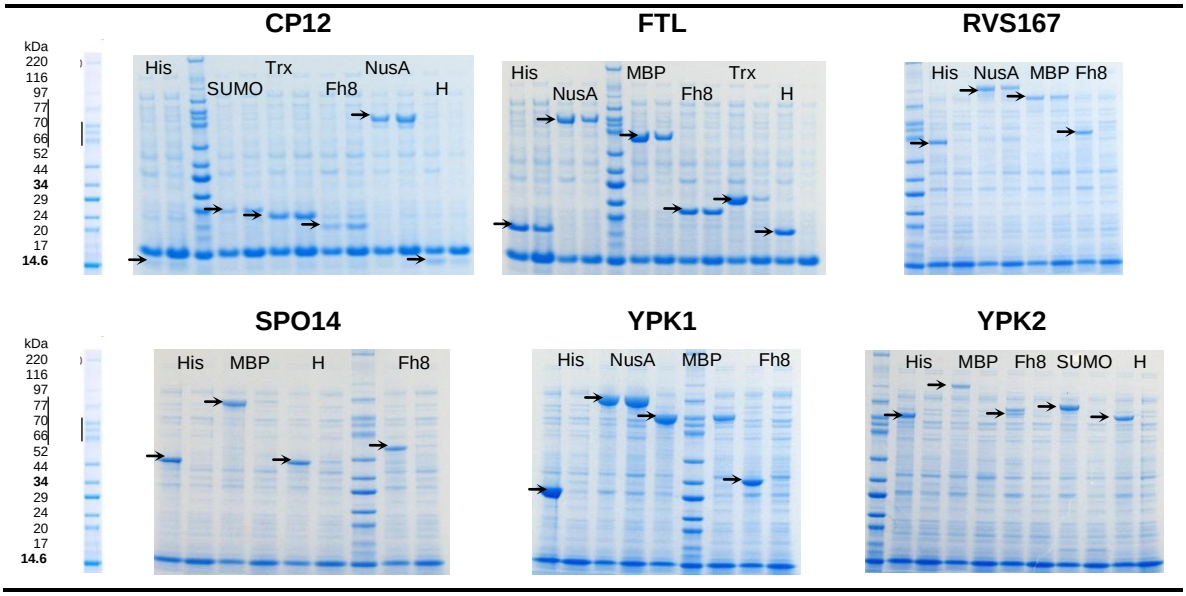


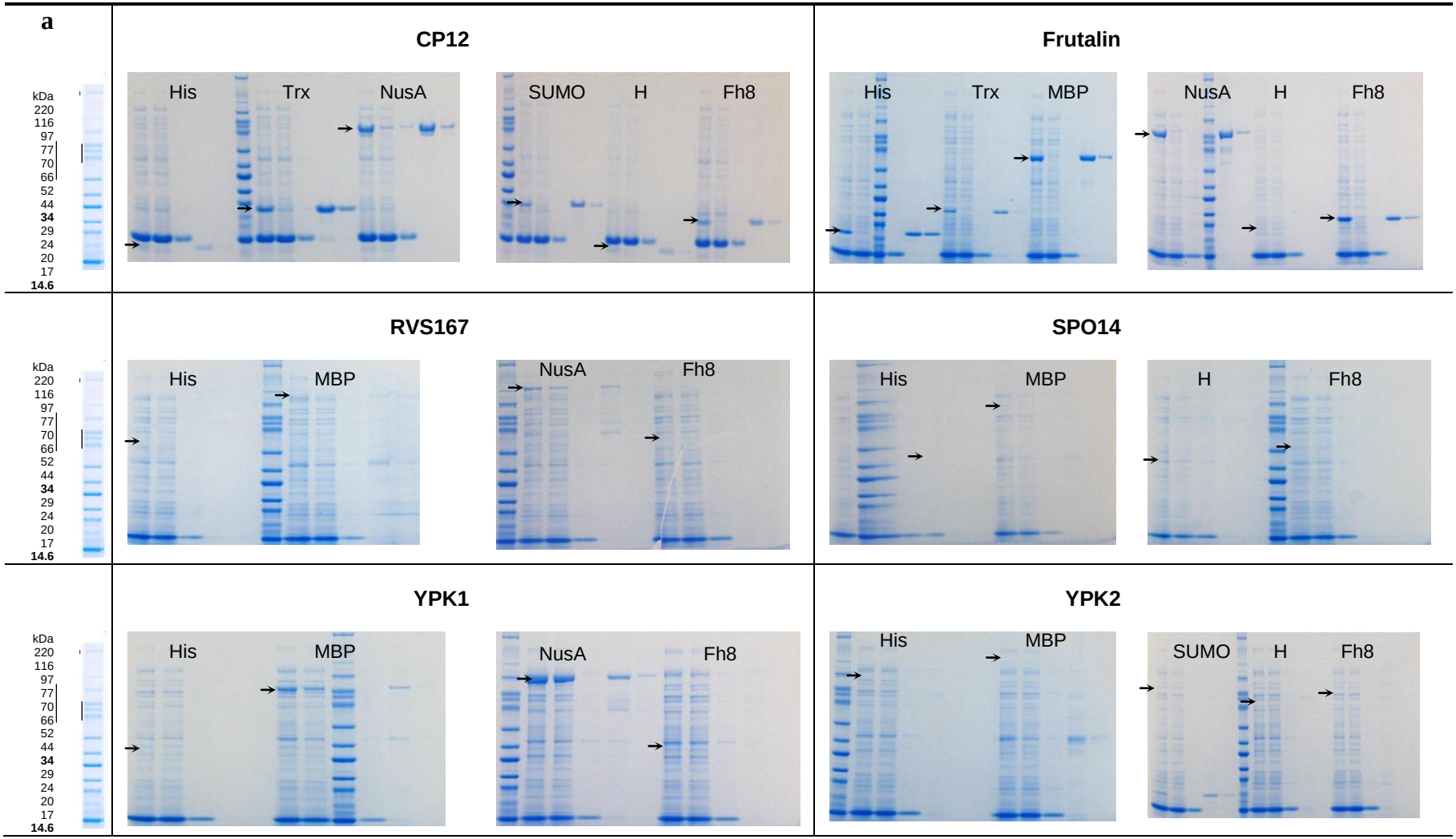
Figure 2.2. SDS-PAGE analysis of the total lysate and supernatant samples from the selected fusion proteins in the second small-scale screening. Samples were obtained after cell lysis with a single tip sonication. Gels were loaded with each total lysate sample aside with the corresponding supernatant (soluble) sample. *Arrows* indicate the expected molecular weights for all fusion proteins following the fusion tags order loaded on the gels.

To further validate the solubility screening, the supernatant samples presented in Figure 2.2 were processed by nickel affinity small-scale purification. Results from this analysis were evaluated by SDS-PAGE and also scored 0-3 (Figures 2.3a and 3b, respectively).

The MBP and NusA fusion proteins improved the solubility of a higher number of proteins than the Fh8 tag (groups I, II and III; Table 2.4). Proteins fused to MBP and NusA tags were, however, obtained in lower soluble amounts than those estimated before purification (see Appendix 2.7.3, Figure A2.2-e). For the Fh8 fusion proteins, no differences were observed before (see Appendix 2.7.3, Figure A2.2) and after purification. The same observation is valid for the fusions of the comparison groups IV, V, and VI of Table 2.4 (see Appendix 2.7.3, Figure A2.2, for the scoring results before purification). SUMO and Fh8 tags presented identical soluble amounts for CP12 and YPK2 proteins (Table 2.4, group V). The Fh8 fusions presented higher solubility than the H fusions. Interestingly, when directly compared using the CP12 and Frutalin

proteins (Table 2.4, group VI), Fh8 fusion proteins achieved the highest solubility, identical to proteins fused with NusA tag.

In general, the rank of the fusion tags obtained in this analysis for protein solubility after purification was identical to the previously observed, with the NusA and the MBP on the top of solubility tags, followed by the Fh8 and Trx tags. The H tag increased the *E. coli* protein expression identical to the Fh8 tag but did not improve protein solubility. The target protein SPO14 was not used further in this work as no soluble protein was detected in the small-scale expression and purification screenings.


a
CP12
His
Trx
NusA
SUMO
H
Fh8
Frutalin
His
Trx
MBP
NusA
H
Fh8
RVS167
His
MBP
NusA
Fh8
SPO14
His
MBP
H
Fh8
YPK1
His
MBP
NusA
Fh8
YPK2
His
MBP
SUMO
H
Fh8
kDa
220
116
97
77
70
66
52
44
34
29
24
20
17
14.6

b

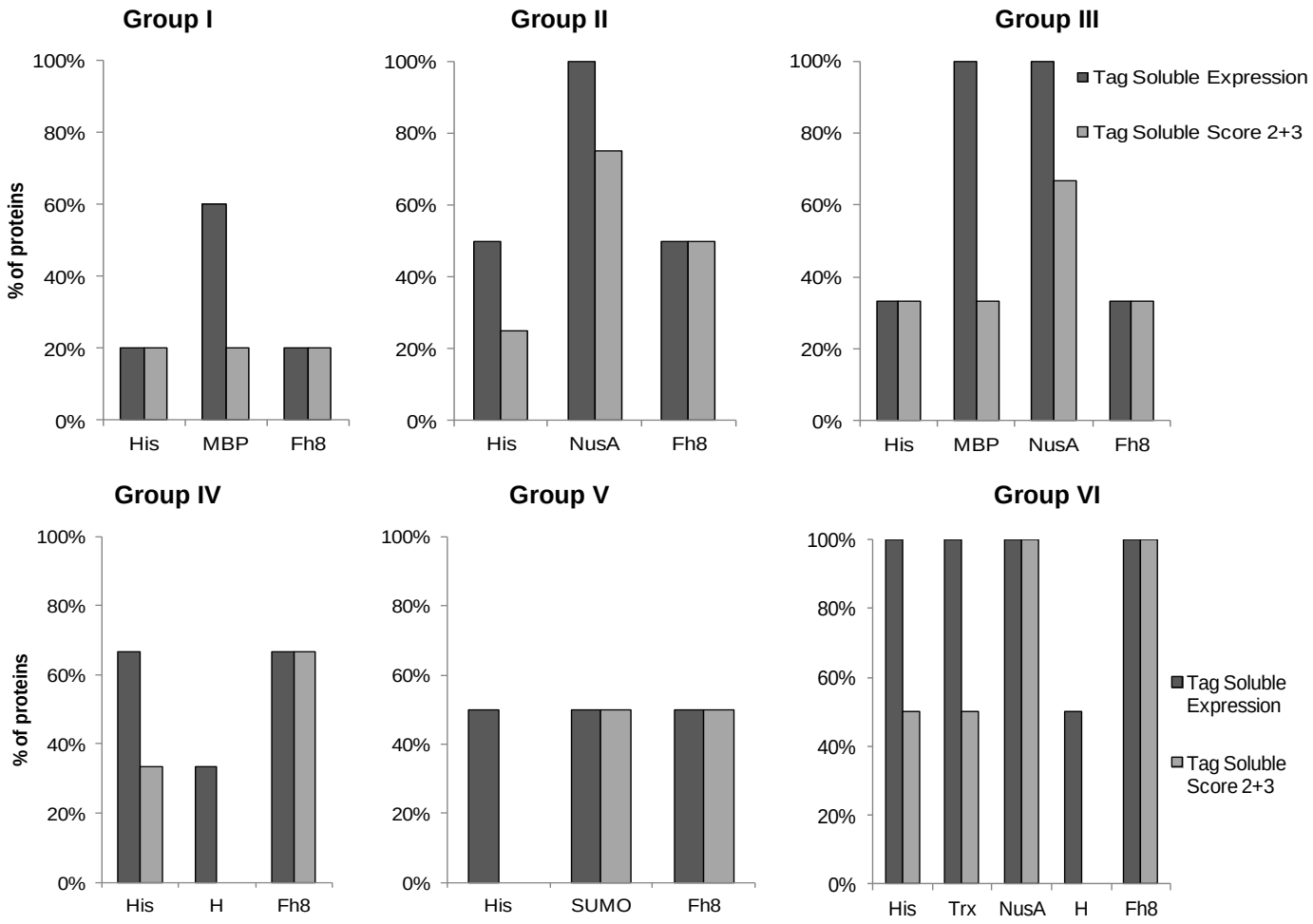


Figure 2.3. Fusion proteins nickel-affinity purification: small-scale processing. (**a**) SDS-PAGE analysis of nickel affinity purifications. Gels were loaded following this sequence: supernatant fraction – flow through – washing step – elution 1 – elution 2. *Arrows* indicate the location of fusion proteins at the observed/expected molecular weight. (**b**) Soluble expression comparison per fusion tag of the different groups of target proteins. The "tag soluble expression" refers to the percentage of proteins with scores 1-3, estimated by Eq. 2.2. The "tag soluble score 2+3" refers to the percentage of proteins with scores 2 and 3, estimated by Eq. 2.3.

2.3.5. Scale-up protein processing: evaluation of the fusion proteins solubility before and after tag removal

Figure 2.4 presents the SDS-PAGE results obtained for the scale-up screening. All the fusion proteins were successfully expressed in soluble form at the expected molecular weights (see Appendix 2.7.2, Table A2.1) in E. coli 500 mL cultures. Bradford estimations were also conducted to support the SDS-PAGE analysis (see Appendices 2.7.4 and 2.7.5, and corresponding Figures A2.3 and A2.4).

Making a wide comparison based on the group VI (Table 2.4), the Fh8 fusions resulted in similar soluble amounts as the Trx and NusA fusions, corroborating the small-scale screening results. After tag removal, the Fh8 fusions performed better than the NusA or MBP fusions, resulting in higher soluble amounts of the cleaved proteins. A case by case comparison established that NusA, Trx, and Fh8 fusions improved the Frutalin soluble expression following this order: NusA ~ Fh8 > Trx > His > Htag (Figure 2.4a).

Interestingly, after tag removal, the cleaved and purified Frutalin (17 kDa) from the Fh8 and Trx fusions presented higher protein amounts than the Frutalin cleaved from the NusA fusion protein. A similar result was obtained for the expression of CP12 target protein with the different fusion tags (Figure 2.4b). The Fh8-CP12 fusion protein achieved similar, but slightly lower, soluble amounts than the Trx and NusA fusions. After tag removal, the cleaved CP12 protein (8 kDa) from Fh8 fusion presented higher soluble amounts than the NusA-containing cleaved protein.

The RVS167 (Figure 2.4a) and YPK1 (Figure 2.4c) target proteins were obtained in higher soluble amounts when fused with the NusA tag than with the Fh8 tag. After tag removal, the cleaved YPK1 protein (35 kDa) from the NusA fusion yielded higher amounts than the one from the Fh8 fusion (Figure 2.4c). For the RVS167 tag removal experiment, only the NusA fusion was tested. The SDS-PAGE bands resulting from this digestion (Figure 2.4a) were difficult to distinguish as the molecular weight expected for the cleaved RVS167 (54 kDa) was close to the one (57 kDa) expected for the NusA.

The YPK2 target protein, of which the soluble expression was not easy to assess in the small-scale screening, presented interesting results in the scale-up experiments. Upon scale-up, this target protein revealed to be highly soluble expressed when fused to the Fh8 tag, out-performing the MBP fusion protein. The final cleaved YPK2 (78 kDa)

from the Fh8 fusion was obtained in 2.3-foldhigher amount than the cleaved protein from the MBP fusion.

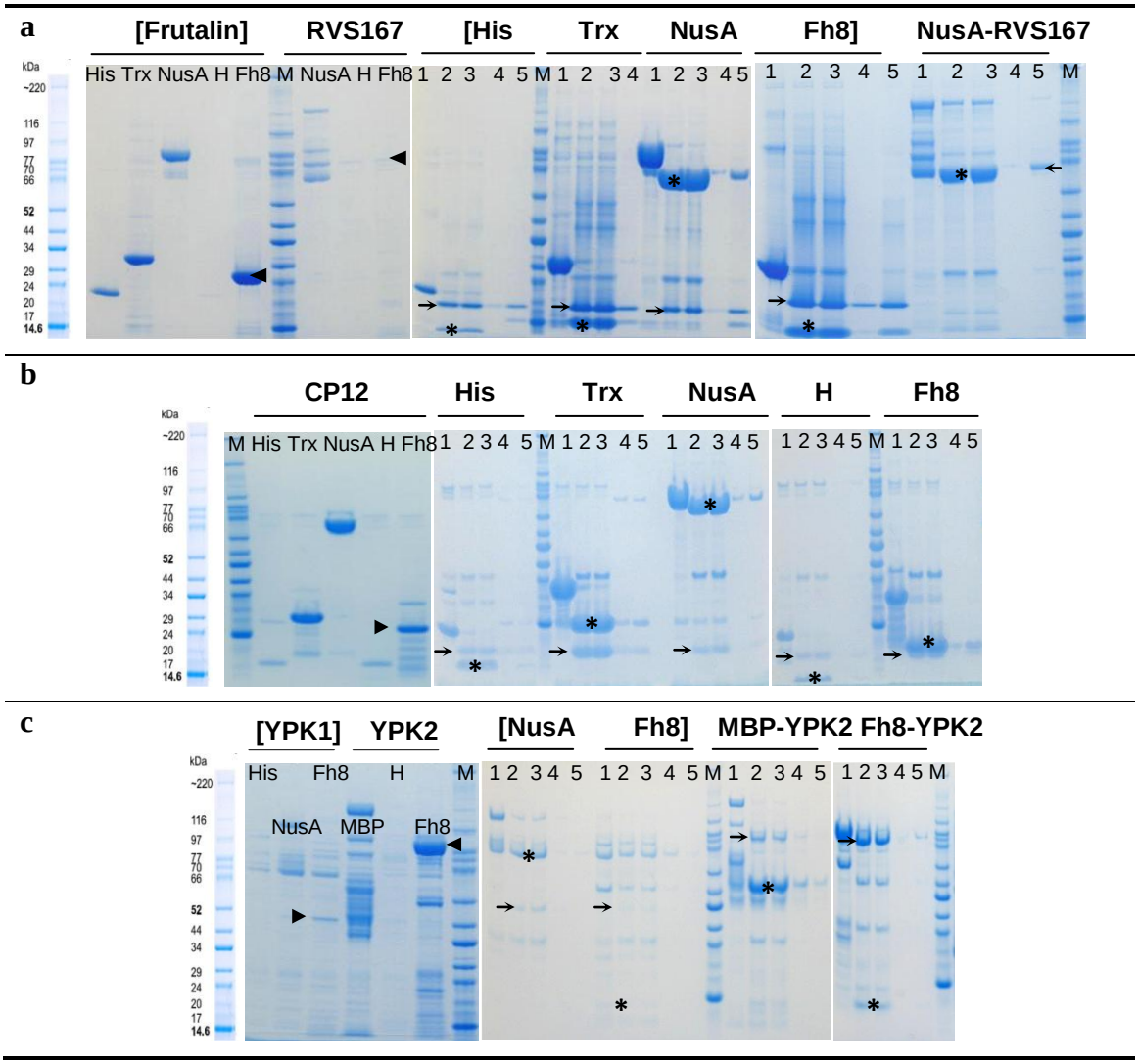


Figure 2.4. Fusion proteins nickel-affinity purification and tag removal: semi-automated processing. (**a**) Frutalin (*in brackets*) and RVS167 5th eluted fractions and corresponding samples before/after TEV digestion and after cleaved protein purification. (**b**) CP12 5th eluted fractions and corresponding samples before/after TEV digestion and after cleaved protein purification. (**c**) YPK1 (*in brackets*) and YPK2 4th eluted fractions and corresponding samples before/after TEV digestion and after cleaved protein purification. *M* In house protein marker. *1-3* Samples before TEV, after TEV, and after TEV plus centrifugation. *4 and 5* Flow through and washing samples from the purification of the cleaved proteins. The *right arrowhead marks* show the Fh8-fusion proteins on the gel. Arrows (→) indicate the expected molecular weight of cleaved proteins. *Asterisk* position the fusion tags after TEV digestion.

2.4. Discussion

In this work, a novel fusion system for soluble protein overexpression in *E. coli* – the Fh8 tag (8 kDa) and H tag (1 kDa) – is presented and compared to the traditional fusion partners using a screening methodology. For the soluble protein expression screening, both Fh8 and H partners were inserted into the pETM vector series (Dummler et al., 2005). Using the same backbone, promoter, and cloning procedures, it was possible to achieve a systematic and consistent comparative analysis of the different solubility tags.

All the selected pETM vectors are identical, differing only at the N-terminal fusion partners. Here, any differences found in target protein expression levels will probably be caused by the fusion partners sequence specific properties. The N-terminal position of the fusion partners seems to be a good option for optimal protein expression compared to the C-terminal position (not tested in this study) as it will allow the fusion partner to be translated first, providing time for the correct folding of the target proteins (Dyson et al., 2004).

The Trx, GST, MBP, and NusA fusion partners were used in the screening comparison due to their known expression and solubility enhancing features and also because they are the most widely used fusion partners for recombinant protein production. The SUMO partner (Michael P. Malakhov and Butt, 2004; Marblestone et al., 2006) was determinant for the selection of four target proteins, which were previously insoluble with this tag. Nevertheless, it was also included in this study to evaluate its solubility effect among the other fusion tags in the Frutalin and CP12 target proteins. The solubility of SUMO-Frutalin fusion was, however, not well evaluated probably due to a technical problem during the eight-tip sonication lysis, in which the ultrasound may not be equally distributed among the eight tips, resulting in the lower protein content observed in the small-scale screening of this fusion protein.

The first expression and solubility comparisons between proteins fused to the Fh8 and H fusion partners and to the other tags were conducted in a small-scale screening, using qualitative SDS-PAGE and Bradford analyses. The small-scale screening strategy already proved to be a reliable tool for the comparison and selection of soluble proteins among different constructs, and it is a useful indicator of the expected protein production amounts upon scale-up expressions (Dummler et al., 2005; Berrow et al.,

2006). Gel bands of fusion proteins were scored according to the 0-3 scale used by Berrow et al. and Bird (Berrow et al., 2006; Bird, 2011). In this study, scores 1-3 were given to an increased protein solubility level (for purified and supernatant samples) and were also used to estimate the total expression levels of fusion proteins. At a small-scale screening, scores 2 and 3 proved to be more consistent solubility predictors than the score 1, which may not represent a proper soluble expression in scale-up cultures (Bird, 2011). Thus, three different evaluations were conducted here: the comparison of total expression (scores 1-3), the comparison of soluble expression (scores 1-3) and the comparison of soluble proteins with scores 2 and 3.

The small-scale comparisons showed that the Fh8 fusion partner stands among the well-described best fusion partners, MBP, NusA, and Trx, for soluble protein expression (Nallamsetty and Waugh, 2006; Kohl et al., 2008). Interestingly, the Fh8 fusions presented lower total expression levels than these known tags. This difference in the expression levels may explain the better solubility results of some proteins when fused to the Fh8 tag. In spite of presenting higher total expression levels than the His tag fusions and identical to the Fh8 fusions, the H tag performed poorly in the protein solubility analysis, not working as solubility enhancing tag in *E. coli*. The GST fusion partner was previously evaluated as a relatively poor tag in several comparison studies (Hammarstrom, 2006; Ohana et al., 2009; Bird, 2011), which was confirmed also in our study.

In addition to the fusion partners comparison, four different expression strains and two induction conditions were also tested in the small-scale screening. The best induction condition revealed to be correlated with the molecular weight of target proteins. Proteins with high molecular weights performed better at lower induction temperatures, in contrast to the CP12 protein, the smallest molecular weight target studied, which performed better with an induction temperature of 30 ºC for 3 hours. In fact, a lower induction temperature will slow down the protein expression rate, promoting a less stressful environment to the cell for protein production. A slow translation rate may improve the correct folding of higher molecular weight proteins and consequently their solubility (Sorensen and Mortensen, 2005; Berrow et al., 2006; Terpe, 2006; Pacheco et al., 2012). Most of the fusion proteins presented higher solubility when expressed in *E. coli* strains engineered with extra copies of rare codons [BL21 (DE3) Codon Plus-RIL

and Rosetta (DE3)], thus confirming the importance of protein expression optimization via its host cell (Makino et al., 2011; Vernet et al., 2011; Pacheco et al., 2012).

The small-scale screening was reproducible as shown by the comparison screenings before and after protein purification, which is essential to assess the real protein solubility among the different fusion tags (Dummler et al., 2005). In general, results from the scale-up analysis were consistent with the small-scale screening: the Fh8 fusion partner is among the best expression and solubility enhancing tags (as NusA and Trx). The YPK2 target (not successfully expressed in small-scale screening) was an exception, resulting in an improved solubility when fused to the Fh8 tag at the scale-up expression. Thus, when conducting a small-scale high throughput analysis, one must be aware of the balance between the loss of some target proteins for scale-up and the gain in increased number of parallel evaluations (Berrow et al., 2006). The YPK1 and RVS167 target proteins were, effectively, difficult to express as soluble proteins even using the NusA or Fh8 tags. In both screening methodologies, the position of the His_6 tag did not interfere with the expression and purification results since fusion constructs with the affinity tag in the middle of the fusion partner and the TEV cleavage site (NusA and Trx) performed as well as the fusion constructs with the affinity tag in the N-terminus (Trx, MBP and Fh8).

In both small-scale and scale-up experiments, fusion partners were removed using the TEV protease, leaving a glycine residue in the N-terminal of target proteins. The Fh8-derived target proteins presented similar final solubility as the Trx-derived ones and, curiously, performed better than the target proteins cleaved from NusA and MBP fusions. In fact, larger fusion partners are good solubilizers, but their large size can lead to an overoptimistic evaluation of protein solubility and yield (Kapust and Waugh, 1999; Shih et al., 2002; Dyson et al., 2004; Hammarstrom, 2006), as shown by the different solubility results before and upon tag removal. Looking among the solubility fusion partners used in this work, the Fh8 tag has the lowest contribution on the final size of the fusion protein (see Appendix 6, Table A2.2), which can explain the apparent lower solubility of Fh8 fusions in comparison to NusA or MBP fusions. In this context, the YPK2 target protein was again a particular case, in which the solubility effect of the Fh8 tag over the MBP tag was noticed in both “before” and “after” tag removal experiments.

The SDS-PAGE and Bradford screening methodology used for the assessment of soluble expression before and after protein purification as well as upon tag removal is protein dependent, requiring a calibration for each protein (Hammarstrom, 2006). Even so, a good agreement between the two methodologies was obtained, indicating that it can be used to predict and compare the protein soluble expression levels among different constructs.

2.5. Conclusions

The novel Fh8 fusion partner presented in this work revealed to be an effective tool for the improvement of protein solubility in *E. coli*.

The conducted study pointed that (1) there is no "the best tag" for protein soluble expression, so multiple tags need to be tested with different proteins; (2) larger tags usually result in higher production yields, but these can lead to overestimation of the amount of soluble protein; (3) the smaller the size of the fusion tag, the easier it is to assess the solubility of the target protein. Thus, the Fh8 tag is an excellent candidate for testing expression and solubility next to the other well-known fusion tags. Its low molecular weight and its solubility enhancing effect make Fh8 an advantageous option compared to larger fusion tags for soluble protein production in *E. coli*.

Acknowledgments

The financial support of the EMBL Heidelberg, Germany, is acknowledged. We wish to acknowledge Anne-Claude Gavin for providing four of the constructs for this study (RVS167, SPO14, YPK1, and YPK2) and Emmanuel Poilpré for the experimental help (both from the EMBL Heidelberg, Germany).

2.6. References

Berrow, N.S., Bussow, K., Coutard, B., Diprose, J., Ekberg, M., Folkers, G.E., Levy, N., Lieu, V., Owens, R.J., Peleg, Y., Pinaglia, C., Quevillon-Cheruel, S., Salim, L., Scheich, C., Vincentelli, R. and Busso, D. (2006) Recombinant protein expression and solubility screening in *Escherichia coli*: a comparative study. Acta Crystallographica Section D-Biological Crystallography 62, 1218-1226.

Bird, L.E. (2011) High throughput construction and small scale expression screening of multi-tag vectors in *Escherichia coli*. Methods 55, 29-37.

Davis, G.D., Elisee, C., Newham, D.M. and Harrison, R.G. (1999) New fusion protein systems designed to give soluble expression in *Escherichia coli*. Biotechnology and Bioengineering 65, 382-388.

De Marco, V., Stier, G., Blandin, S. and de Marco, A. (2004) The solubility and stability of recombinant proteins are increased by their fusion to NusA. Biochemical and Biophysical Research Communications 322, 766-771.

Demain, A.L. and Vaishnav, P. (2009) Production of recombinant proteins by microbes and higher organisms. Biotechnology Advances 27, 297-306.

Dummler, A., Lawrence, A.M. and de Marco, A. (2005) Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. Microbial Cell Factories 4.

Dyson, M.R., Shadbolt, S.P., Vincent, K.J., Perera, R.L. and McCafferty, J. (2004) Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. Bmc Biotechnology 4.

Esposito, D. and Chatterjee, D.K. (2006) Enhancement of soluble protein expression through the use of fusion tags. Current Opinion in Biotechnology 17, 353-358.

Hammarstrom, M. (2006) Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein. Journal of Structural and Functional Genomics 7, 1-14.

Kapust, R.B. and Waugh, D.S. (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Science 8, 1668-1674.

Kohl, T., Schmidt, C., Wiemann, S., Poustka, A. and Korf, U. (2008) Automated production of recombinant human proteins as resource for proteome research. Proteome Science 6.

LaVallie, E.R., Lu, Z.J., Diblasio-Smith, E.A., Collins-Racie, L.A. and McCoy, J.M. (2000) Thioredoxin as a fusion partner for production of soluble recombinant proteins in *Escherichia coli*. Applications of Chimeric Genes and Hybrid Proteins, Pt A 326, 322-340.

Makino, T., Skretas, G. and Georgiou, G. (2011) Strain engineering for improved expression of recombinant proteins in bacteria. Microbial Cell Factories 10.

Marblestone, J.G., Edavettal, S.C., Lim, Y., Lim, P., Zuo, X. and Butt, T.R. (2006) Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. Protein Science 15, 182-189.

Michael P. Malakhov, M.R.M., Oxana A. Malakhova, Mark Drinker, Stephen D. and Butt, W.T.R. (2004) SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. Journal of Structural and Functional Genomics 5, 75–86.

Nallamsetty, S. and Waugh, D.S. (2006) Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. Protein Expression and Purification 45, 175-182.

Ohana, R.F., Encell, L.P., Zhao, K., Simpson, D., Slater, M.R., Urh, M. and Wood, K.V. (2009) HaloTag7: A genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification. Protein Expression and Purification 68, 110-120.

Oliveira, C., Costa, S., Teixeira, J.A. and Domingues, L. (2009) cDNA Cloning and Functional Expression of the alpha-D-Galactose-Binding Lectin Frutalin in *Escherichia coli*. Molecular Biotechnology 43, 212-220.

Oliveira, C., Nicolau, A., Teixeira, J.A. and Domingues, L. (2011) Cytotoxic Effects of Native and Recombinant Frutalin, a Plant Galactose-Binding Lectin, on HeLa Cervical Cancer Cells. Journal of Biomedicine and Biotechnology.

Pacheco, B., Crombet, L., Loppnau, P. and Cossar, D. (2012) A screening strategy for heterologous protein expression in *Escherichia coli* with the highest return of investment. Protein Expression and Purification 81, 33-41.

Peti, W. and Page, R. (2007) Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. Protein Expression and Purification 51, 1-10.

Sachdev, D. and Chirgwin, J.M. (2000) Fusions to maltose-binding protein: Control of folding and solubility in protein purification. In: Applications of Chimeric Genes and Hybrid Proteins, Pt A, Vol. 326, p. 312-321.

Shih, Y.P., Kung, W.M., Chen, J.C., Yeh, C.H., Wang, A.H.J. and Wang, T.F. (2002) High-throughput screening of soluble recombinant proteins. Protein Science 11, 1714-1719.

Silva, E., Castro, A., Lopes, A., Rodrigues, A., Dias, C., Conceicao, A., Alonso, J., da Costa, J.M.C., Bastos, M., Parra, F., Moradas-Ferreira, P. and Silva, M. (2004) A recombinant antigen recognized by *Fasciola hepatica*-infected hosts. Journal of Parasitology 90, 746-751.

Smith, D.B. (2000) Generating fusions to glutathione S-transferase for protein studies. Applications of Chimeric Genes and Hybrid Proteins, Pt A 326, 254-270.

Smith, D.B. and Johnson, K.S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene 67, 31-40.

Sorensen, H.P. and Mortensen, K.K. (2005) Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. Microbial Cell Factories 4.

Studier, F.W. (2005) Protein production by auto-induction in high-density shaking cultures. Protein Expression and Purification 41, 207-234.

Terpe, K. (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Applied Microbiology and Biotechnology 60, 523-533.

Terpe, K. (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Applied Microbiology and Biotechnology 72, 211-222.

Vernet, E., Kotzsch, A., Voldborg, B. and Sundstrom, M. (2011) Screening of genetic parameters for soluble protein expression in *Escherichia coli*. Protein Expression and Purification 77, 104-111.

Waugh, D.S. (2005) Making the most of affinity tags. TRENDS in Biotechnology 23.

Yao, L., Yin, J., Zhang, X., Liu, Q., Li, J., Chen, L., Zhao, Y., Gong, P. and Liu, C. (2007) *Cryptosporidium parvum*: identification of a new surface adhesion protein on sporozoite and oocyst by screening of a phage-display cDNA library. Exp Parasitol 115, 333-8.

2.7. Appendices

2.7.1. Schematic representation of the pETMFh8 and pETMH vectors

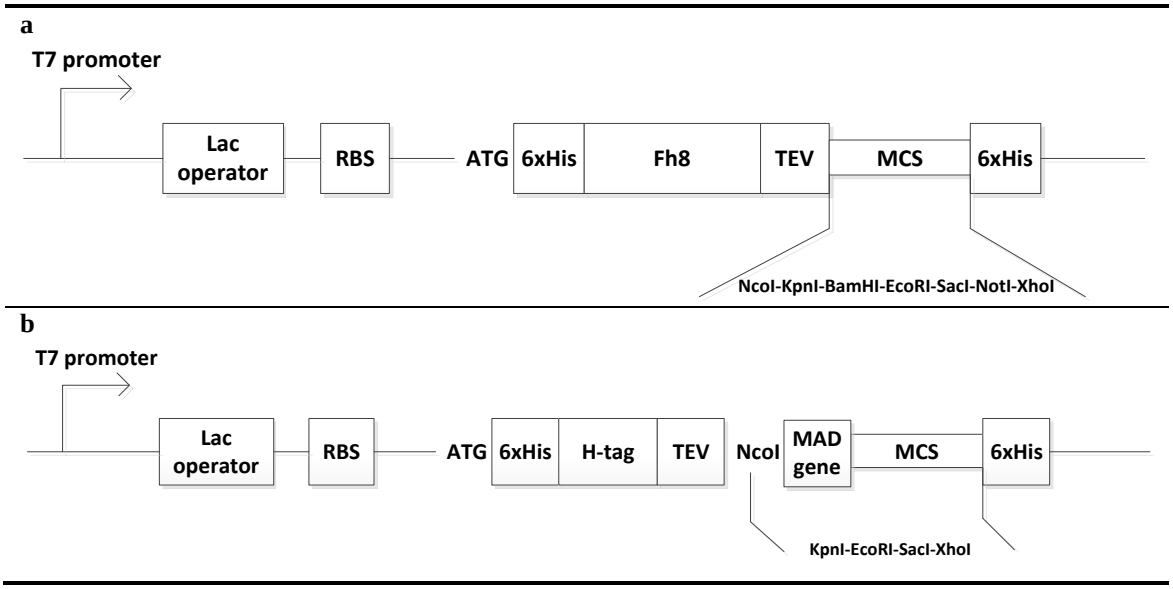


Figure A2.1. Schematic representation of the pETMFh8 vector (**a**) and pETMH vector (**b**). Both Fh8 and H tags are cloned between the His_6 tag and TEV cleavage site. The novel fusion vectors have similar multiple cloning sites (MCS). An extra His_6 tag is also available at the end of the MCS. The pETMH vector has a stuffer gene for cloning control.

2.7.2. Molecular weights of fusion proteins

Table A2.1. Molecular weights (kDa) of fusion proteins estimated by the ProtParam tool (Expasy.org):

Fusion Tags	Targets					
	RVS167	**SPO14**	**YPK1**	**FTL**	**CP12**	**YPK2**
His	57.2	48.0	38.8	20.5	11.8	81.8
Trx	68.5	59.2	50.0	31.7	23.1	92.6
GST	83.1	73.8	64.6	46.3	37.7	107
MBP	98.4	89.1	79.9	61.6	52.9	122
NusA	111	102	92.7	74.4	65.8	135
SUMO	66.3	57.0	47.8	29.5	20.8	90.3
H	57.4	48.1	-	20.6	11.9	81.4
Fh8	64.0	54.7	45.5	27.2	18.5	88.0

2.7.3. Small-scale screening evaluation before protein purification: expression and solubility

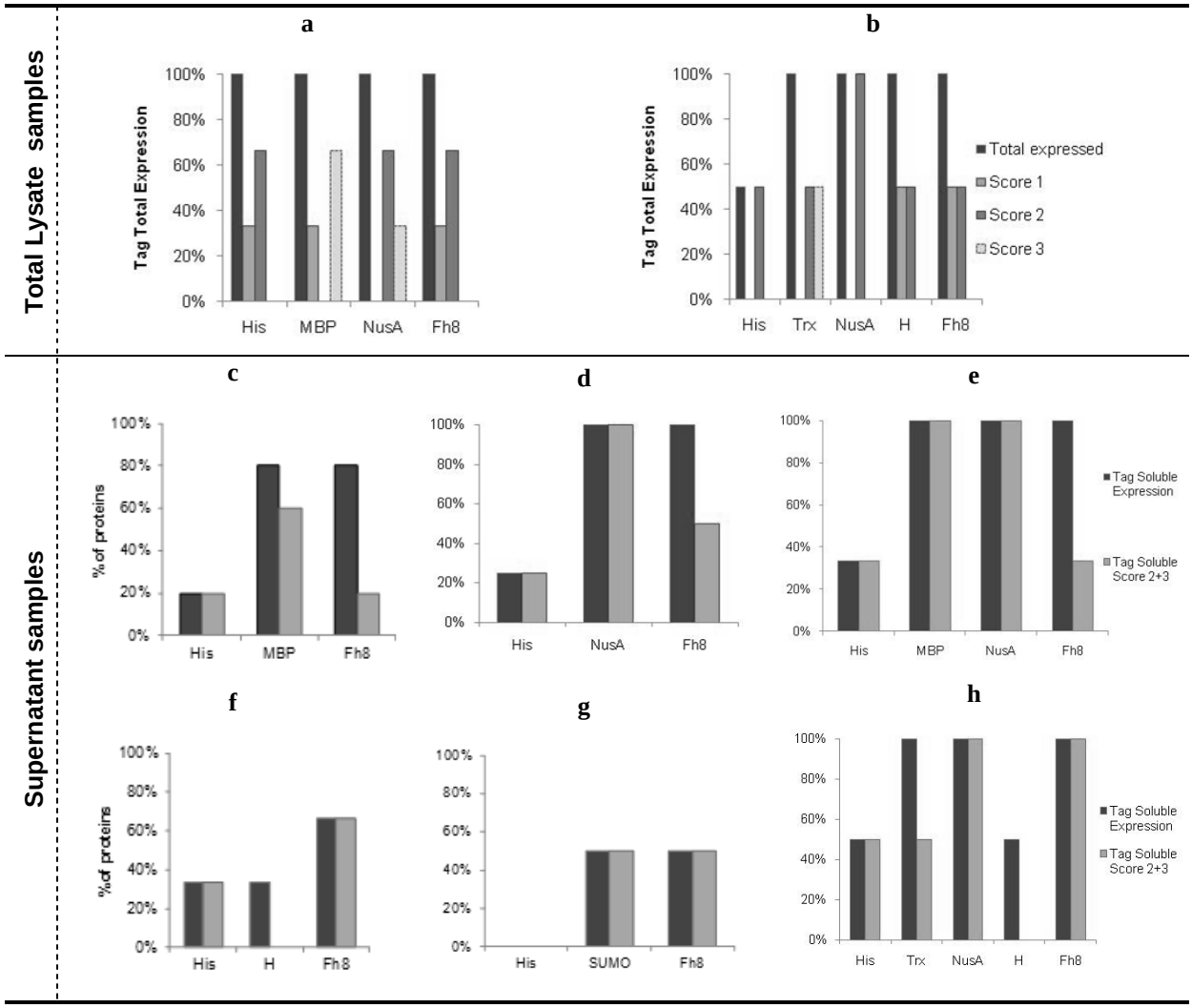


Figure A2.2. Individual comparisons of fusion proteins total expression and solubility: small-scale screening evaluation before protein purification. Fusion proteins were scored 0-3 according to their expression and solubility level. (**a**) Total expression comparison per fusion tag of group III – Table 2.4. (**b**) Total expression comparison per fusion tag of group VI – Table 2.4. In both (a) and (b), "Total expressed" refers to the % of proteins with scores 1+2+3, estimated by Equation 2.4. (**c** to **h**) Soluble expression comparison per fusion tag. From (c) to (h), the "Tag Soluble Expression" refers to the % of proteins with scores 1, 2 and 3, estimated by Eq. 2.2. The "Tag Soluble Score 2+3" refers to the % of proteins with scores 2 and 3, estimated by Eq. 2.3.

2.7.4. Estimated protein production yields

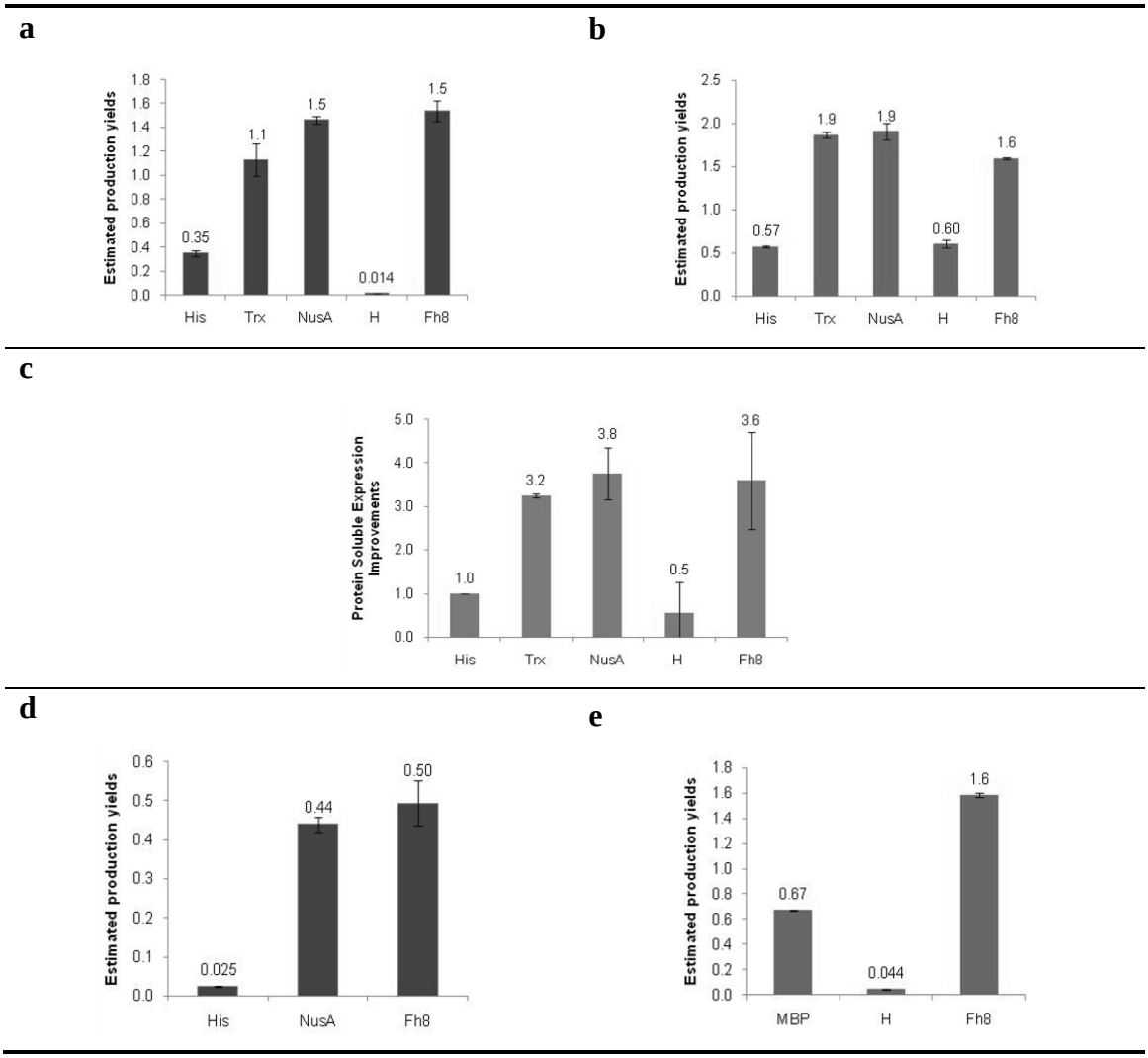


Figure A2.3. Fusion protein estimated production amounts (mg) after semi-automated purification from 15% of 500 mL *E. coli* cultures. (**a**) Fusions with Frutalin. (**b**) Fusions with CP12. (**c**) Average improvements in soluble protein expression for fusions with Frutalin and CP12 (Group VI – Table 2.4). (**d**) Fusions with YPK1. (**e**) Fusions with YPK2.

2.7.5. Estimated protein production yields after tag removal

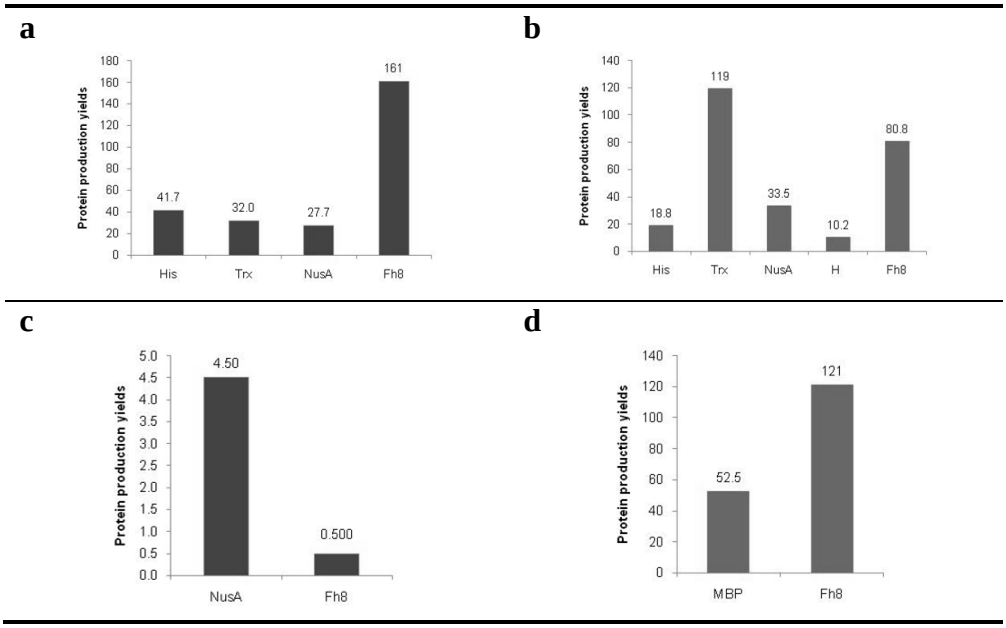


Figure A2.4. Estimated protein production amounts (µg) after tag removal. (**a**) Frutalin cleaved proteins. (**b**) CP12 cleaved proteins. (**c**) YPK1 cleaved proteins. (**d**) YPK2 cleaved proteins.

2.7.6. Percentage of the fusion tag in the molecular weight of fusion proteins

Table A2. 2. Molecular weight ratio (%) between fusion tags and total fusion proteins:

Fusion Tags	Targets					
	RVS167	**SPO14**	**YPK1**	**FTL**	**CP12**	**YPK2**
His	6%	7%	9%	17%	29%	4%
Trx	21%	25%	29%	46%	64%	16%
GST	35%	40%	45%	63%	78%	27%
MBP	45%	50%	56%	72%	84%	36%
NusA	52%	56%	62%	77%	87%	42%
SUMO	19%	22%	26%	42%	60%	14%
H	6%	7%	-	17%	30%	4%
Fh8	16%	19%	22%	37%	55%	12%

Chapter 3

The novel adjuvant-free H fusion partner for the production of recombinant immunogens in *Escherichia coli*: its application to a 12-kDa antigen from *Cryptosporidium parvum*

Abstract

The production of recombinant antigens and specific polyclonal antibodies for diagnosis and therapy is still a challenge for world-wide researchers. Several different strategies have been explored to improve both antigen and antibody production, all of them depending on a successful correlation between its expression and immunogenicity. *Escherichia coli* host cell is widely used for recombinant antigen production. This host cell has, however, its limitations at producing recombinant antigens and these are often poor immunogens, being unable to elicit itself an adequate immune response in the organism of interest. Gene fusion technology attempts to overcome these problems: fusion partners have been applied to optimise recombinant antigen production in *E. coli*, and to increase protein immunogenicity.

This work presents the effects of the novel H fusion partner in the *E. coli* expression of a 12-kDa surface adhesion antigen from *Cryptosporidium parvum* (CP12), and its subsequent adjuvant-free immunisation. The solubility enhancer partner Fh8 is also here used for the production of CP12. Both fusion partners increased the recombinant CP12 antigen availability via its production in *E. coli*.

The H tag efficiently triggered a CP12-specific immune response, and it also improved the immunisation procedure without requiring co-administration of adjuvants. Polyclonal antibodies raised against the HCP12 fusion antigen detected native antigen structures displayed on the surface of *C. parvum* oocysts, making this novel fusion system a promising tool for the diagnosis and therapy of *C. parvum* infections in animals and humans.

Taking CP12 by example, the H fusion partner appears as an attractive option for the development of recombinant immunogens and its adjuvant-free immunisation.

3.1. Introduction

Antibodies are important tools in biomedical research. They allow the identification of new genes, the purification of proteins, and the study of protein properties such as structure, function and localisation (Lowenadler et al., 1986; Chowdhury et al., 2001). Most of these applications use polyclonal antibodies, which are produced in response to multiple epitopes of the same target protein (antigen). Polyclonal antibodies are usually raised against a specific protein by immunising an animal with the target protein in its purified form (Chowdhury et al., 2001).

The most effective way of obtaining the high quantities of antigenic protein required for efficient immunisation is by heterologous expression in a host cell. The bacteria *Escherichia coli* is one of the most widely used organisms for this purpose, as it is easy to manipulate, has a fast growth rate and is relatively cheap to use (Jana and Deb, 2005; Sorensen and Mortensen, 2005b; Sorensen and Mortensen, 2005a; Terpe, 2006; Peti and Page, 2007; Demain and Vaishnav, 2009). However, it also has its limitations; the recombinant protein it produces is not always correctly folded or sufficiently soluble for use in immunisation. The development of alternative strategies for protein production that overcome these drawbacks is therefore highly desirable (Villaverde and Carrio, 2003; Vallejo and Rinas, 2004; Peti and Page, 2007). One such strategy is gene fusion technology, whereby the gene coding for the protein of interest is fused to a polypeptide chain, known as a fusion partner.

Fusion partners can simplify protein purification, improve protein production yield, reduce susceptibility to proteolysis and increase protein solubility (Makrides, 1996; Nilsson et al., 1997; Sorensen and Mortensen, 2005a; Waugh, 2005; Esposito and Chatterjee, 2006). They have also been reported to increase protein immunogenicity (Kaslow and Shiloach, 1994; Larsson et al., 1996; Sjolander et al., 1997; Kink and Williams, 1998; Libon et al., 1999; Chowdhury et al., 2001; Chuang et al., 2009).

Fusion partners, such as SpA (Lowenadler et al., 1986), GST (Lopez-Monteon et al., 2003), BB-SpG (Sjolander et al., 1997), MBP (Kink and Williams, 1998) and Trx (Barrell et al., 2004), have been used to improve both antigen expression and antibody production. However, these fusion partners have also been shown to have drawbacks. In some cases, the resulting recombinant antigen is not sufficiently soluble or pure

(Chowdhury et al., 2001; Lopez-Monteon et al., 2003). In others, the immune response obtained against fusion antigens is often triggered predominantly by the fusion partner itself, rather than by the target antigen (Lowenadler et al., 1986; Knuth et al., 2000; Lopez-Monteon et al., 2003). And in others still, the fusion partners have shown inadequate immunopotentiating properties to elicit the production of sufficient quantities of the antibodies of interest. In such cases, during immunisation, it is necessary to co-administer an adjuvant. Despite being widely used for routine antibody production in animals, adjuvants are undesirable because they may trigger non-specific immune responses and cause several side effects and lesions at the injection site (Lowenadler et al., 1986; Sjolander et al., 1997; Andersson et al., 2000; Knuth et al., 2000). Evidently, therefore, the use of recombinant fusion partners for protein and antibody production has much scope for improvement.

In this work, we show how the novel H fusion partner can be used for the expression in *E. coli* and subsequent adjuvant-free immunisation of the 12-kDa recombinant protein, CP12 (GenBank ID: XM625821.1), belonging to the parasite *Cryptosporidium parvum (C. parvum)*. The solubility enhancer partner Fh8 (Chapter 2) (Costa et al., 2012) is here also used for the production of Fh8CP12 fusion protein, which will help to clarify the specificity of the immune response in study.

As a surface adhesion protein, CP12 plays a major role in the diagnosis of *C. parvum* infections in various mammals, including humans (Yao et al., 2007). A higher availability of this antigen and its specific polyclonal antibodies is therefore important for cryptosporidiosis prevention and therapy. The novel fusion system presented here offers a strategy for obtaining CP12 and anti-CP12 antibodies in quantities that make them fit for this purpose.

The H partner consists of the N-terminal part of the Fh8 sequence. Fh8 (GenBank ID: AF213970.1) is a calcium-binding protein excreted and secreted by the adult worm of *Fasciola hepatica*. Fh8 is highly soluble and stable in *E. coli*, can be easily purified by nickel affinity chromatography, and it was previously used for the detection of *F. hepatica* infections (Silva E, 2004). Previous structural studies carried out on Fh8 protein (unpublished work) revealed that the H fusion partner region may be critical for the stability and production of the entire protein.

The low molecular weight of the H fusion partner (1 kDa) makes it a particularly attractive option for use in antibody production.

3.2. Material and Methods

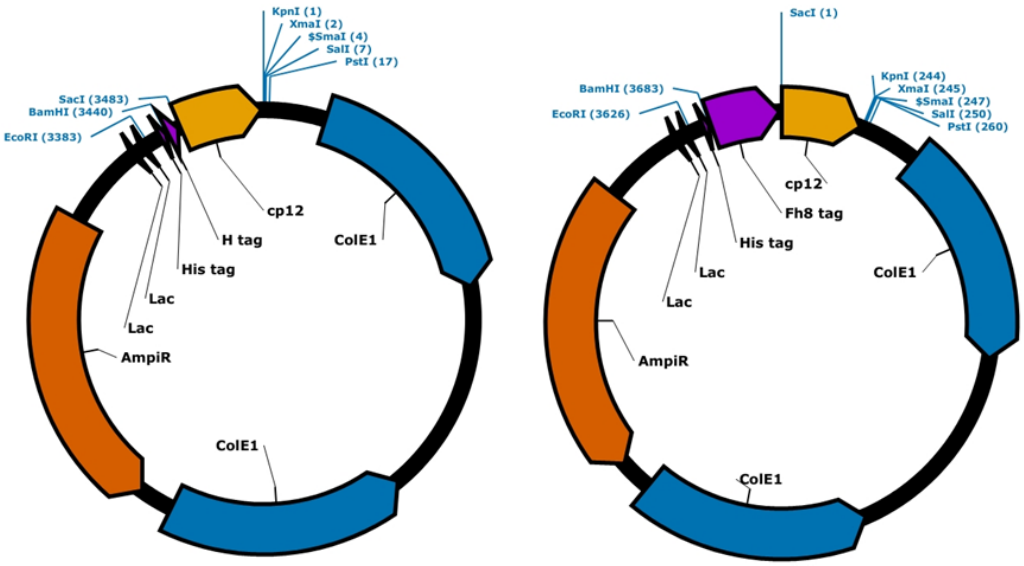
3.2.1. Plasmids, strains and media

Plasmids pGEM-T easy (Promega) and pQE-30 (Qiagen) were used to sub-clone the *cp12* gene obtained as described below. *Escherichia coli* XL1 and M15 [pREP4] cells were chosen for cloning and expression of the CP12 protein. Bacteria cultures were grown in Luria Broth medium with proper antibiotics (see subsection 3.2.4).

3.2.2. Fusion vectors

Fh8 and H fusion partners containing the restriction sites for *Bam*HI and *Sac*I enzymes were introduced into pQE-30 vectors between the restriction sites for the same enzymes (Figure 3.1). The resulting fusion vectors are pQE-30 modified vectors with a Fh8 tag sequence (GenBank ID: AF213970.1), designated by pQEFh8, or a H tag sequence site, designated by pQEH, after N-terminal polyhistidine. Both Fh8 and H tags are low molecular weight peptides with 8 kDa and 1 kDa, respectively.

a)



b)

```
M   R   G   S   H   H   H   H   H   H   G   S   M   P   S   V
ATG AGA GGA TCG CAT CAC CAT CAC CAT CAC GGA TCC ATG CCT AGT GTT
                 His-tag           BamH I

 Q   E   V   E   K   L   L   E   L   E   G   V   L   W   H   N
CAA GAG GTT GAA AAA CTC CTT GAG CTC GAA GGA GTA CTT TGG CAC AAC
      H-tag               Sac I

 V   F   E   Q   D   R   L   Q   W   Q   P   E   R   N   D   A
GTA TTC GAA CAG GAT AGG CTC CAG TGG CAA CCT GAA AGA AAC GAT GCG

 Q   N   F   T   N   G   N   Q   Y   N   Y   I   Q   V   P   T
CAA AAT TTT ACC AAT GGT AAT CAA TAT AAC TAT ATT CAA GTT CCA ACC
                          cp12

 D   F   N   S   V   M   G   G   L   Q   S   P   S   E   M   A
GAT TTT AAT AGT GTG ATG GGA GGG CTA CAG TCG CCT TCA GAA ATG GCA

 R   T   I   E   R   N   I   E   K   K   Q   M   N   E   Q   I
AGA ACA ATT GAA AGG AAC ATA GAA AAG AAA CAG ATG AAT GAA CAA ATA

 *
TAG TTA GTA TTC TTG ATA ATG AAA GGT ACC
                                Kpn I
```

Figure 3.1. Fusion vectors. (**a**) Schematic maps of Fh8 and H fusion vectors containing *cp12* gene. His tag – six-histidine tag sequence. Fh8/H – novel fusion tags sequence. *cp12* – *cp12* gene sequence. (**b**) Amino acid and nucleotide sequences of the HCP12 codifying gene.

3.2.3. Construction of expression vectors

The *cp12* gene used in this work (Figure 3.1) corresponds to a truncated form of the original *cp12* gene: it lacks the original N-terminal peptide signal and transmembrane region. This truncated gene (*cp12* gene) was obtained from genomic DNA of *C. parvum* by a PCR using the forward primer: 5' – CATACTGGTAT**GAGCTC**GAAGGAGTAC - 3' and the reverse primer: 5' – CATTAAAA**GGTACC**TTTCATTATCAAG - 3'. The forward primer introduced a restriction site for *SacI* enzyme (bold) between the *C. parvum* genomic DNA sequence and the initial part of *cp12* gene (underlined). The reverse primer introduced a restriction site for *KpnI* enzyme (bold) between the final part of *cp12* gene (underlined) and the *C. parvum* genomic DNA sequence. The PCR mix was first preheated at 95 ºC for 4 min, followed by 30 amplification cycles: a 95 ºC denaturation for 30 sec, a 50 ºC annealing for 1 min and a 72 ºC extension. A final elongation step was carried out at 72 ºC for 7 min.

The PCR product was cloned into pGEM T-easy vector (Promega), according to manufacturer instructions, and used to transform *E. coli* XL1 strain. Positive clones were selected and sequenced using pQE universal primers.

pGEM plasmid containing the *cp12* gene was digested with *SacI* and *KpnI* restriction enzymes (Promega) and the *cp12* gene was then cloned into the same restriction sites of pQE-30, pQEH and pQEFh8 vectors, using T4 DNA ligase (Promega). Both H and Fh8 fusion vectors contain a MCS similar to pQE-30 vector. The H partner sequence corresponds to the initial N-terminal sequence of Fh8 partner and both are preceded by an N-terminal hexahistidine tag from pQE-30 vector. Codifying gene sequences are allowed to be cloned immediately after fusion tags using the specific restriction site for *SacI* enzyme.

Following this strategy, *E. coli* M15 [pREP4] cells were transformed with pQE-30 (Qiagen), pQEH and pQEFh8 expression vectors containing the 237 bp *cp12* gene. Positive clones were confirmed by DNA sequencing (as mentioned above) and the resulting sequence was aligned to *cp12* (GenBank ID: XM625821.1).

LB medium plates supplemented with ampicilin (100 $\mu g.mL^{-1}$) and kanamicin (50 $\mu g.mL^{-1}$) were used to select and maintain *E. coli* M15 [pREP4] transformants.

3.2.4. Expression and purification of fusion proteins in *E. coli*

E. coli M15 [pREP4] cells were transformed with fusion vectors, pQEH and pQEFh8, containing *cp12* gene, and also with pQE-30 harbouring the same *cp12* gene. Cells were grown at 37 ºC in LB medium supplemented with ampicilin and kanamicin. The *E. coli* culture was induced with 1 mM of isopropyl-β-D-1-thiogalactopyranoside (IPTG) for 5 hours at 37 ºC. After induction, cells were harvested by centrifugation at 4000 g during 15 minutes at 4 ºC, and the resultant pellet was lysed with 8 M urea, pH 8.0, overnight at room temperature, 150 rpm. An aliquot of *E. coli* total lysates was taken and then, cell extracts were centrifuged at 10000 g, 20 minutes. The resulting supernatants and pellets were collected for further analyses (SDS-PAGE and protein purification).

The supernatant fraction was applied onto a Ni-NTA column (GE Healthcare), pre-equilibrated with 8 M urea, pH 8.0. Ni-NTA column was washed with 5 column volumes (CVs) of 8 M urea, pH 8.0 and 5 CVs of 8 M urea, pH 6.5. Fusion proteins and non-fused protein were eluted with 8 M urea, pH 4.5.

All CP12, HCP12 and Fh8CP12 production procedures were repeated in three different periods of time to obtain biological independent replicates. Recombinant antigens (HCP12 and CP12) to be administrated to CD1 mice were first dialysed with phosphate buffer saline 1x (PBS) prepared in apyrogenic water (B Braun), pH 7.2, and then, concentrated using Centricon 3 (Millipore). Pyrogens from concentrated recombinant antigen suspensions were removed with Detoxi-Gel Endotoxin Removing Gel (Pierce) according to the manufacturer's instructions. Recombinant antigen samples were recovered in pyrogen-free water (Braun) and stored at –20ºC until use.

3.2.5. Protein electrophoresis and protein determination

The purity of collected fractions from Ni-NTA chromatography as well as the *E. coli* total lysates and supernatant samples were analysed by 17%- or 12%-4% SDS-PAGE gels (Laemmli, 1970) stained with Coomassie Brilliant Blue dye. 15%-4% tris-tricine gels (Schagger and von Jagow, 1987) were also used for Western blot analyses, as described below. Protein content of each collected fraction and final production yields were determined by Bradford assay (Bradford, 1976).

3.2.6. Immunisations

Polyclonal antibodies were produced in CD1 female mice with 10 to 12 weeks old purchased from Charles River SA, Barcelona. The animals were housed and maintained with food and drink *ad libitum*.

For immunisation, groups of 3 mice were injected intraperitoneally (IP) at 2 week intervals with 20 µg of antigen (CP12 or HCP12) in 200 µl of inocula, without using Freund's adjuvant. Purified recombinant antigens were diluted in pyrogen-free physiological serum (Braun) to a concentration of 0.5 $mg.ml^{-1}$, filtered by 0.2 µm membrane (Sarstedt) in sterile conditions, prior to administration**.** Control animals harbouring the same age and characteristics did not receive either protein or inocula.

Blood collection was carried out for five times at 3^{rd} IP, 5^{th} IP, 7^{th} IP, 8^{th} IP at the tail vein and mice were sacrificed 1 week later. After blood collection, mice sera were separated by centrifugation at 2500 rpm for 10 minutes and then stored at -20 ºC.

All animal experiments were carried out in accordance to the European Communities Council Directive of 24 November 1986 (86/609/EEC).

3.2.7. ELISA

The antibody titer was determined by ELISA. Briefly, 96-well microtiter plates were coated with 100 µL per well of HCP12 fusion antigen or CP12 antigen (both at 10 $\mu g.mL^{-1}$) by incubating overnight at 4 ºC. The plates were washed twice with PBS 1x containing 0.3% (v/v) of Tween 20 (PBST), and the remaining binding sites were blocked with 3% (w/v) of Bovine serum albumin (BSA) in PBST, for 1h at 37 ºC. After washing plates three times with PBST, 100 µL of diluted (1:400 in PBST) anti-HCP12, anti-CP12 or negative serum (from the non-injected group) was dispensed into each well and plates were incubated at 4 ºC overnight. Plates were then washed three times with PBST, and 100 µL per well of diluted Protein A-Horseradish Peroxidase mixed with Protein G- Horseradish Peroxidase (1:4000 in PBST) was added. Plates were incubated for 2h at 37 ºC. After several washes, 100 µL per well of *o*-phenylenediamine in 0.2 M phosphate buffer (pH 5.6) were added, and plates were incubated at room temperature (RT). The reaction was stopped by adding 100 µL per well of 3 M HCl solution.

The optical density (O.D.) was measured at 492 nm using an ELISA reader (Biorad). ELISA measurements were submitted to a Mann-Whitney U test with 95% confidence level.

3.2.8. H tag specific humoral response assay

For testing the humoral response against H tag, several ELISAs were performed using two different proteins as coating antigens, which contain the H-tag sequence, Fh8 and HTgOWP - a *Toxoplasma gondii* oocyst wall protein, with anti-CP12, anti-HCP12 and negative sera. 96-well microtiter plates were coated with 100 μL per well of Fh8 or HTgOWP antigens (both at 10 $\mu g.mL^{-1}$), and were subsequently treated following the above-mentioned ELISA protocol.

3.2.9. Western blot

Tris-Tricine gel containing recombinant Fh8CP12 or Fh8 antigens were transferred to nitrocellulose membranes using a sandwich system. Membranes were saturated with a 5% (w/v) PBS-Milk solution for 1h at room temperature (RT). Diluted anti-HCP12, anti-CP12 or negative group sera (1:1000 in PBST) were added to each membrane and incubated overnight at 4 ºC. Membranes were washed three times with 0.3% (v/v) PBST and incubated with diluted Protein G-peroxidase (1:1000 in PBST) for 2h at RT. Protein detection was carried out using 4-chloro-1-naphthol in cold methanol, phosphate buffer saline and hydrogen peroxide.

3.2.10. Immunofluorescence assay (IFA)

C. parvum oocysts were used for IFA assays. *C. parvum* oocysts were isolated from positive fecal samples of animals by using cesium chloride density gradients (Arrowood and Donaldson, 1996). Isolated oocysts were counted using a Newbauer chamber, and added to the slides for testing. Oocysts were air-dried onto slides, and then fixed with acetone, and incubated at 37 ºC in a humid atmosphere with either anti-HCP12 or negative control sera, both diluted 1:20. After being washed twice with PBS 1x, oocysts were incubated with anti-mice IgG conjugated with FITC (Sigma) and diluted 1:80,

according to the manufacturer's instructions. This conjugate was used to reveal the bound antibodies, which were then observed using a fluorescence microscope.

3.3. Results

3.3.1. Expression and purification of *cp12* codifying gene in *E. coli* using H and Fh8 fusion partners

E. coli M15 [pREP4] cells were transformed with both H and Fh8 fusion vectors and pQE-30 vector containing the *cp12* codifying sequence (Figure 3.1). This sequence matches 100% identity with the original *cp12* sequence except that it lacks the original N-terminal peptide signal and transmembrane region.

SDS-PAGE analysis of *E. coli* cultures (Figure 3.2.a) showed that both fusion proteins (HCP12 and Fh8CP12) were expressed at higher amounts than the non-fused CP12 protein. This improvement in protein expression was most notable for Fh8CP12 fusion protein, as demonstrated by its total lysate and supernatant samples.

SDS-PAGE analysis also confirmed that CP12 and HCP12 proteins were expressed at the predicted 10 kDa and 11 kDa, respectively. Fh8CP12 protein appeared at 18 kDa as expected, and also at 36 kDa (Figure 3.2.b), suggesting dimer forms of this recombinant fusion protein.

After being purified, CP12 non-fused protein achieved a production yield of 0.40 ± 0.050 mg per litre of *E. coli* culture while HCP12 fusion protein achieved a production yield of 1.5 ± 0.30 mg per litre of *E. coli* culture. Further improvements on protein production levels were obtained with the Fh8 partner (4.7 ± 0.81 mg per litre of *E. coli* culture).

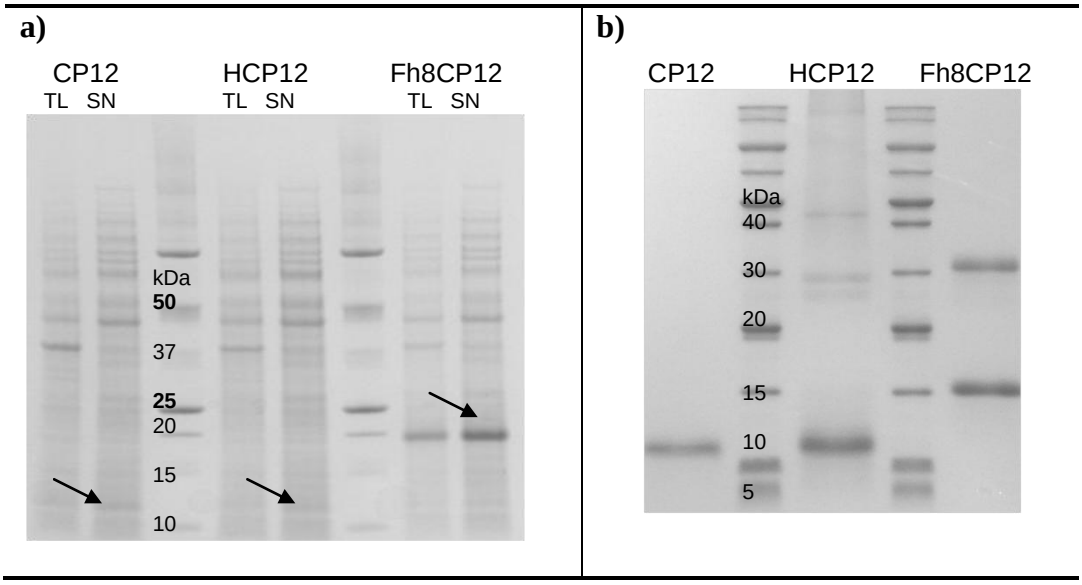


Figure 3.2. CP12, HCP12 and Fh8CP12 proteins production in *E. coli*. (**a**) SDS-PAGE analysis of TL – *E. coli* total lysate and S – supernatant samples obtained from total lysate, as described in the Materials and Methods section, of CP12, HCP12 and Fh8CP12 cultures. (**b**) SDS-PAGE analysis of purified proteins after nickel affinity chromatography (5 µg of protein).

3.3.2. Production of polyclonal antibodies anti-CP12 using the H tag

ELISA assays were performed with sera from three different groups of mice: HCP12-injected mice, non-fused CP12-injected mice and non-injected mice, and plates were coated with recombinant CP12 antigen and with HCP12 fusion antigen.

Similar results were achieved using both CP12 and HCP12 coatings: an increased humoral immune response was observed in CP12 and HCP12-injected mice. CP12-injected mice produced polyclonal antibodies anti-CP12 statistically different from the non-injected group at 49 days post injection (CP12-coated plate, Figure 3.3.a), and at 42 days post injection (HCP12-coated plate, Figure 3.3.b). This result shows that CP12 antigen is itself an immunogenic protein. The IgG levels for HCP12-injected mice started to increase after the 14th day post injection (Figure 3.3.a), which reveals an earlier immune response than the obtained in CP12-injected mice.

After 28 days post injection, the production of polyclonal anti-HCP12 antibodies was higher than the production of polyclonal anti-CP12 antibodies, but these titers were only statistically different at 49 days post injection (CP12 and HCP12-coated plates, Figures 3.3.a and 3.3.b).

Results from ELISA showed that, when fused to the H partner, CP12 antigen was able to trigger an earlier immune response than the non-fused antigen. Furthermore, HCP12 mice group achieved higher polyclonal antibody titers than those obtained with CP12-injected group.

3.3.3. Humoral response against H partner

The serum collected from HCP12-injected mice was also analysed by ELISA to determine if the H partner elicited itself a humoral response. 96-well plates were coated with Fh8 and HTgOWP antigens and polyclonal antibodies anti-CP12, anti-HCP12 and negative sera were added to the assay. Both Fh8 and HTgOWP antigens have the H partner sequence at its N-terminal, which make them suitable for studying the specific humoral response against this fusion partner. For all sera analysed, no significant IgG levels were observed in plates coated with Fh8 or HTgOWP antigens. Neither anti-CP12 nor anti-HCP12 polyclonal antibodies reacted with Fh8 (Figure 3.3.c) and HTgOWP (Figure 3.3.d) antigens.

These assays demonstrated that the H partner did not triggered itself a humoral response since polyclonal antibodies anti-HCP12 were not capable to interact with H-containing antigens.

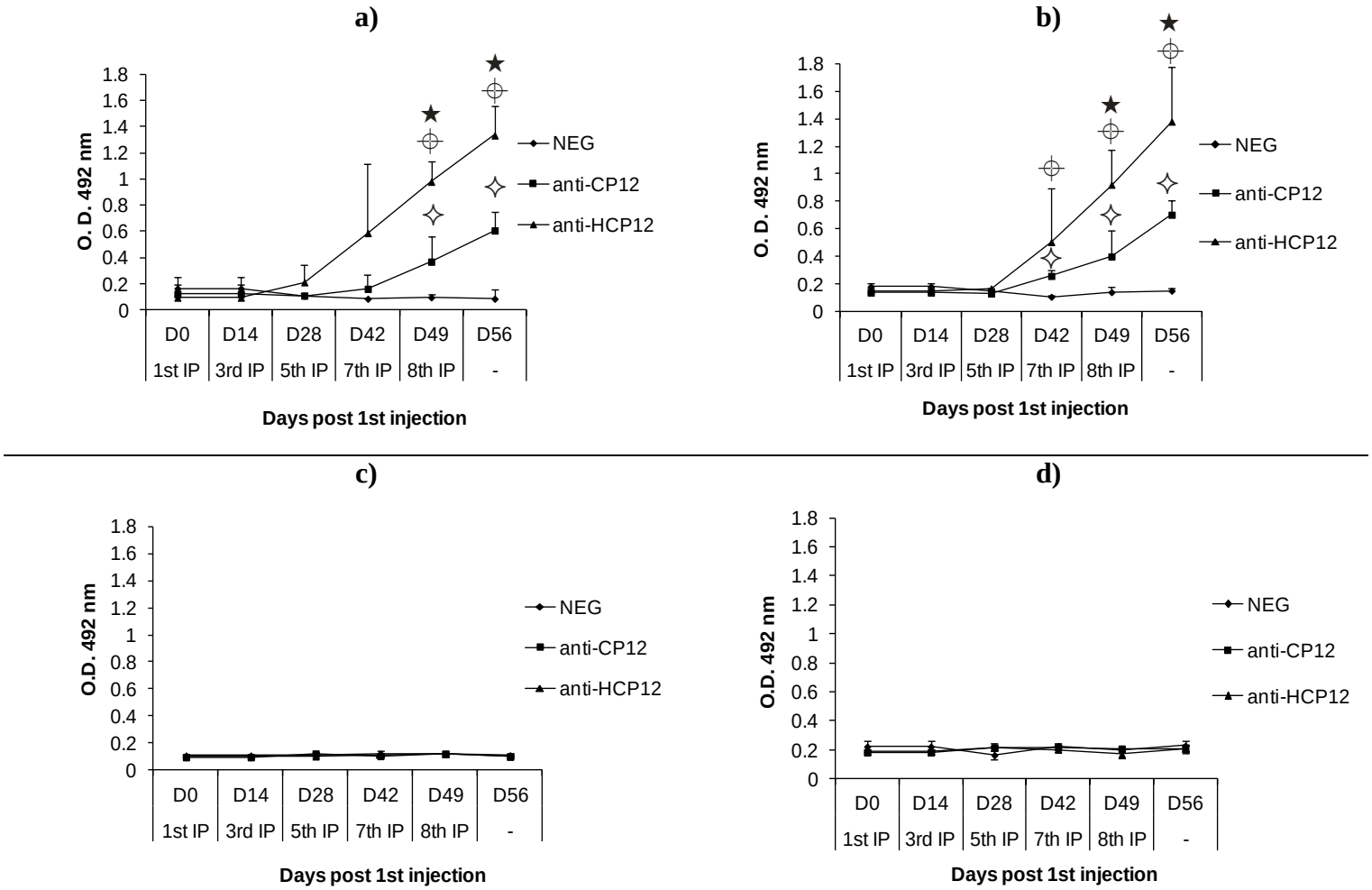


Figure 3.3. Evaluation of IgG production against CP12 and HCP12 antigens during the course of immunisation. ELISAs were performed with sera collected periodically from CD1 mice experimentally injected with HCP12 (here presented as *anti-HCP12*), with CP12 (here presented as *anti-CP12*), and from non-injected mice (here presented as *NEG*). $1^{st}/3^{rd}/5^{th}/7^{th}/8^{th}$ IP – Blood collection before $1^{st}/3^{rd}/5^{th}/7^{th}/8^{th}$ intraperitoneal injections. D56 – Blood collection at seven days after 8^{th} IP. (⌖) and (✧) – O. D. values statistically different from negative control. (★) - O. D. values statistically different from CP12-injected mice. (**a**) Plates coated with 10 μg.mL^{-1} of CP12. (**b**) Plates coated with 10 μg.mL^{-1} of HCP12. (**c**) Plates coated with 10 μg.mL^{-1} of HTgOWP. (**d**) Plates coated with 10 μg.mL^{-1} of Fh8.

3.3.4. CP12 specific polyclonal antibodies

Western blot assays were also performed to evaluate the CP12 specificity of polyclonal antibodies produced by HCP12-injected mice. In this analysis we used Fh8CP12 fusion

antigen to test if there were cross reactions with the H partner, which DNA sequence is part of Fh8 sequence.

Western analyses with sera of CP12 and HCP12-injected mice (collected 7 days after 8th IP) revealed the formation of blots at a molecular weight that corresponds to dimeric Fh8CP12 antigen (36 KDa). Arrows presented in Figure 3.4.a, lanes 1 and 2, focus this observation. The Western analysis performed with the mice control group sera did not reveal blot formation, as expected (Figure 3.4.a, lane 3).

An immunoblotting analysis using Fh8 recombinant antigen was also carried out with mice sera of the three groups (CP12, HCP12 and negative), in the same conditions as above, but no blots were observed (Figure 3.4.b).

The Western blot analyses showed also that polyclonal antibodies produced by HCP12-injected mice group specifically recognized the CP12 antigen and do not cross react with the H partner.

3.3.5. Recognition of native CP12 epitopes

Immunofluorescence assays were conducted to evaluate the potential of polyclonal antibodies produced by HCP12-injected mice to recognize native CP12 epitopes presented on the surface of the parasite *C. parvum*.

The fluorescence observed in Figure 3.4.c highlights the surface of *C. parvum* oocysts, thus confirming the ability of these polyclonal antibodies to detect the native CP12 epitopes displayed on the surface of parasite oocysts.

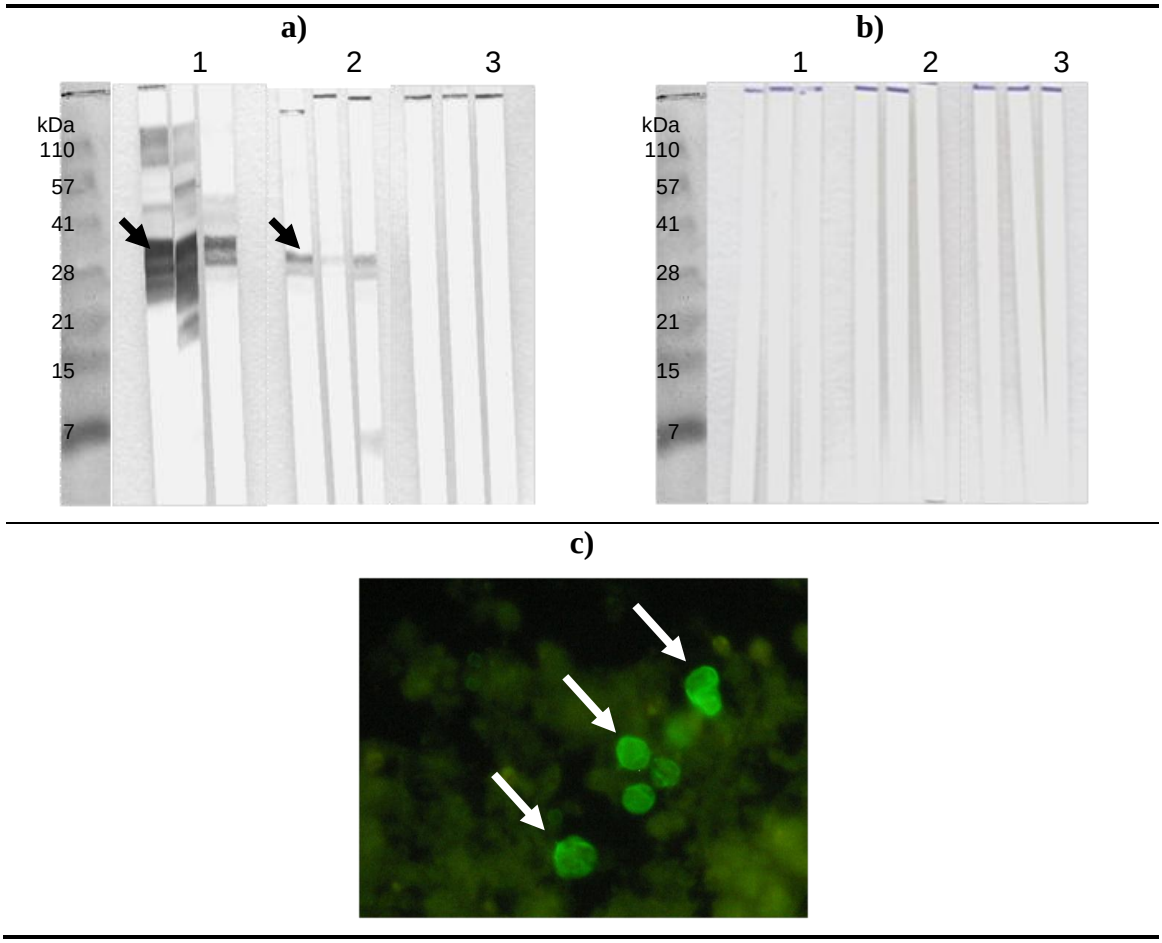


Figure 3.4. Evaluation of polyclonal antibodies specificity against CP12 antigen. (**a**) Western blot analysis of Fh8CP12 antigen. (**b**) Western blot analysis of Fh8 antigen. In all images, 1 – Polyclonal antibodies produced by HCP12-injected mice (presented in Figure 3.3 as *anti-HCP12*); 2 – Polyclonal antibodies produced by CP12-injected mice (presented in Figure 3.3 as *anti-CP12*); 3 - Negative sera (presented in Figure 3.3 as *NEG*). (**c**) Immunofluorescence assay using anti-HCP12 polyclonal antibodies: detection of *C. parvum* oocysts surface.

3.4. Discussion

A novel approach to produce CP12 specific polyclonal antibodies is presented in this work, in which the H partner plays an important role for adjuvant-free immunisation.

CP12 is an adhesion protein that was previously identified on *C. parvum* sporozoite and oocyst surface (Yao et al., 2007). Due to its surface localization and its contribution in host-parasite interactions, CP12 is a potential target for cryptosporidiosis prevention and therapy (Yao et al., 2007). Taking into account its diagnostic and prophylactic relevance in the control of *C. parvum* infections in various organisms, novel strategies to improve CP12 antigen production and its specific polyclonal antibodies availability are desired.

Generally, to raise a specific antibody against a certain target antigen it is first required to obtain substantial amounts of the purified and soluble antigen and second, the target antigen has to trigger an immune response in the organism (to be immunogenic) (Chowdhury et al., 2001). Most of the times, however, target proteins do not meet these criteria, being inefficiently expressed, purified and/or presenting a misfolded structure in *E. coli*, and acting as poor immunogens.

Following a production strategy from gene to antibody, the study conducted here demonstrates that both Fh8 and H fusion partners increase recombinant protein production of CP12 in *E. coli*, being a versatile tool for the generation of antigenic proteins. Furthermore, the H fusion partner improves the immunogenic properties of CP12, without using adjuvants or additional immunisation procedures. The immunopotentiating properties of the H fusion partner make it an innovative alternative to the traditional immunisation methods.

In the cloning strategy, we used a truncated *cp12* gene, which lacks the original N-terminal transmembrane region of *cp12* because it may hamper soluble protein expression. This truncated *cp12* gene was inserted into pQE modified vectors containing Fh8 and H fusion partners and also into pQE-30 vector, which was used for expression and immunisation control.

Both fusion partners increased CP12 production yields, allowing also a rapid and easy purification of this antigen by the use of the same hexahistidine tag from pQE-30 plasmid. HCP12 and Fh8CP12 fusion proteins achieved a production yield four-fold and twelve-fold higher than the non-fused protein in *E. coli*, respectively. These

production yields may, however, be overestimated because contaminant proteins were also quantified by Bradford.

Findings from this work reveal several advantages of using H partner to produce polyclonal antibodies. The H partner improves specific antibody production against CP12 antigen, resulting in an earlier immune response and in higher polyclonal antibody titers when compared to the non-fused CP12 immunisation. Moreover, polyclonal antibodies anti-HCP12 reacted only with CP12 antigen but not with the H moiety contained in the N-terminal part of Fh8 partner as well as in the HTgOWP antigen. This CP12 specific immune response resulted from immunisation of HCP12 antigen with no addition of adjuvants, whereby the immunogenicity rise may be exclusively due to the fusion of CP12 recombinant antigen with H partner. Further studies are actually being conducted to clarify the mechanism behind this effect, in which the ability of H partner to function as an immunomodulator or adjuvant is tested *in vitro*.

CP12 antigen has a low molecular weight that can hinder the production of polyclonal antibodies. When producing such low molecular weight antigens, fusion partners like GST (Lopez-Monteon et al., 2003), Trx (Barrell et al., 2004), BB-SpG (Sjolander et al., 1997), MBP (Kink and Williams, 1998) and SpA (Lowenadler et al., 1986) are usually used to increase the molecular weight of target proteins, making them detectable by the immune system. However, these fusion partners are often not capable to elicit a satisfactory immune response, requiring therefore the use of adjuvants, such as Freund's complete/incomplete adjuvants. Fusion partners as SSNAP (Knuth et al., 2000) and CTB (Holmgren et al., 1994; Liljeqvist et al., 1997) can also act themselves as adjuvants by expressing fusion proteins in inclusion bodies or by forming multimeric structures. Here, a local immune response is stimulated at the injection site and a sustained release of small antigen quantities occurs over a long period of time.

In fact, all of these adjuvant stimulations are not desirable as they can be difficult to prepare or may cause several side effects, namely, chronic inflammation response at the injection site. In addition, the immune responses obtained with these fusion partners are frequently not specific for the target antigen, resulting in polyclonal antibodies production against both fusion moiety and target antigen. For most antibody applications, extra purification steps need to be conducted, which makes this a time-consuming and expensive methodology.

Taking into account all these aspects, the ability of H-partner to effectively improve the specific immunogenicity of CP12 protein without adjuvants makes this fusion partner an attractive option to the existent and commonly used fusion partners. Furthermore, H partner has only 1 kDa, which is, to our knowledge, the lowest molecular weight partner used for protein and antibody production.

The immunofluorescence assay conducted in this work revealed also that polyclonal antibodies raised against HCP12 antigen were effective on detecting the native CP12 antigen structure presented on the surface of *C. parvum* oocysts. Besides allowing an adjuvant-free immunisation procedure and a faster humoral response than usual, the fusion of the H partner to the CP12 antigen produced polyclonal antibodies that can be used as a diagnostic tool for immunodetection of *C. parvum* infections in humans or animals. A small fusion partner like the H tag may not interfere negatively with target antigen conformation and, hence, polyclonal antibodies raised against antigens fused to the H partner are able to recognize native antigen structures.

The low molecular weight of H partner may also clarify the reason why this moiety does not trigger an immune response against itself when fused to a target antigen.

Apart from CP12, the H partner was already successfully applied to the production of other antigens (Conceição et al., 2010) and their corresponding polyclonal antibodies, such as, the human interleukin-5 (IL-5), the cyst wall protein-1 from *Toxoplasma gondii* (TgOWP) that was used above as a control for cross reactions with the H partner, the cyst wall protein from *Giardia lamblia* cysts (CWG), the β-giardin cytoskeletal protein of the ventral disk from the *Giardia lamblia* trophozoite (βG), the cyst wall specific-glycoprotein Jacob from *Entamoeba histolytica* (Ent), and the falcipain-1 trophozoite cysteine proteinase from *Plasmodium falciparum* (Pfsp), among others (Conceição et al., 2011).

3.5. Conclusion

The novel fusion system studied in this work has overcome problems related to the production of CP12 antigen and its polyclonal antibodies such as low antigen quantity, poor immunogenicity, adverse effects of adjuvants and unspecific antibody production.

When fused to CP12, the H partner improves its immunogenic properties without being removed from the fusion antigen and without co-administration of adjuvants, resulting in a more effective and earlier immune response.

We suggest this novel fusion system as an improved strategy and promising technology for immunodiagnostic and immunoprophylactic purposes in the control of infections, allowing a rapid and easy recombinant procedure (see Patents WO 2010/082097 and WO 2011/071404 in Appendices 1.6.1 and 3.7.1, respectively).

Acknowledgements

We thank Lurdes Delgado and Sónia Soares for the *Cryptosporidium* oocysts isolation from fecal samples.

3.6. References

Andersson, C., Sandberg, L., Wernerus, H., Johansson, M., Lovgren-Bengtsson, K. and Stahl, S. (2000) Improved systems for hydrophobic tagging of recombinant immunogens for efficient iscom incorporation. J Immunol Methods 238, 181-93.

Arrowood, M.J. and Donaldson, K. (1996) Improved purification methods for calf-derived *Cryptosporidium parvum* oocysts using discontinuous sucrose and cesium chloride gradients. Journal of Eukaryotic Microbiology 43, S89-S89.

Barrell, P.J., Liew, O.W. and Conner, A.J. (2004) Expressing an antibacterial protein in bacteria for raising antibodies. Protein Expr Purif 33, 153-9.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72, 248-54.

Chowdhury, P.S., Gallo, M. and Pastan, I. (2001) Generation of high titer antisera in rabbits by DNA immunization. Journal of Immunological Methods 249, 147-154.

Chuang, C.K., Su, Y.S., Fan, C.T., Lee, W.C. and Chen, M.Y. (2009) A dual-functional *E. coli* vector for expressing recombinant protein with high solubility and antigen presentation ability. Protein Expression and Purification 65, 51-56.

Conceição, M., Costa, S., Castro, A. and Almeida, A. (2010) Fusion proteins, its preparation process and its application on recombinant protein expression systems. In: European Patent Office. WO 2010/082097.

Conceição, M., Costa, S., Castro, A. and Almeida, A. (2011) Immunogens, process for preparation and use in systems of polyclonal antibodies production. In: European Patent Office. WO 2011/071404.

Costa, S.J., Almeida, A., Castro, A., Domingues, L. and Besir, H. (2012) The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology. Applied Microbiology and Biotechnology.

Demain, A.L. and Vaishnav, P. (2009) Production of recombinant proteins by microbes and higher organisms. Biotechnol Adv 27, 297-306.

Esposito, D. and Chatterjee, D.K. (2006) Enhancement of soluble protein expression through the use of fusion tags. Curr Opin Biotechnol 17, 353-8.

Holmgren, J., Czerkinsky, C., Lycke, N. and Svennerholm, A.M. (1994) Strategies for the Induction of Immune-Responses at Mucosal Surfaces Making Use of Cholera-Toxin-B Subunit as Immunogen, Carrier, and Adjuvant. American Journal of Tropical Medicine and Hygiene 50, 42-54.

Jana, S. and Deb, J.K. (2005) Strategies for efficient production of heterologous proteins in *Escherichia coli*. Appl Microbiol Biotechnol 67, 289-98.

Kaslow, D.C. and Shiloach, J. (1994) Production, purification and immunogenicity of a malaria transmission-blocking vaccine candidate: TBV25H expressed in yeast and purified using nickel-NTA agarose. Biotechnology (N Y) 12, 494-9.

Kink, J.A. and Williams, J.A. (1998) Antibodies to recombinant *Clostridium difficile* toxins A and B are an effective treatment and prevent relapse of C-difficile-associated disease in a hamster model of infection. Infection and Immunity 66, 2018-2025.

Knuth, M.W., Okragly, A.J., Lesley, S.A. and Haak-Frendscho, M. (2000) Facile generation and use of immunogenic polypeptide fusions to a sparingly soluble non-antigenic protein carrier. Journal of Immunological Methods 236, 53-69.

Laemmli, U.K. (1970) Cleavage of Structural Proteins during Assembly of Head of Bacteriophage-T4. Nature 227, 680-685.

Larsson, M., Brundell, E., Nordfors, L., Hoog, C., Uhlen, M. and Stahl, S. (1996) A general bacterial expression system for functional analysis of cDNA-encoded proteins. Protein Expr Purif 7, 447-57.

Libon, C., Corvaia, N., Haeuw, J.F., Nguyen, T.N., Stahl, S., Bonnefoy, J.Y. and Andreoni, C. (1999) The serum albumin-binding region of streptococcal protein G (BB) potentiates the immunogenicity of the G130-230 RSV-A protein. Vaccine 17, 406-14.

Liljeqvist, S., Stahl, S., Andreoni, C., Binz, H., Uhlen, M. and Murby, M. (1997) Fusions to the cholera toxin B subunit: influence on pentamerization and GM1 binding. Journal of Immunological Methods 210, 125-135.

Lopez-Monteon, A., Ramos-Ligonio, A., Perez-Castillo, L., Talamas-Rohana, P. and Rosales-Encina, J.L. (2003) Specific antibody immune response against the parasitic portion of a glutathione-S-transferase fusion protein. FASEB J 17, 621-7.

Lowenadler, B., Nilsson, B., Abrahmsen, L., Moks, T., Ljungqvist, L., Holmgren, E., Paleus, S., Josephson, S., Philipson, L. and Uhlen, M. (1986) Production of specific antibodies against protein A fusion proteins. EMBO J 5, 2393-8.

Makrides, S.C. (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. Microbiol Rev 60, 512-38.

Nilsson, J., Stahl, S., Lundeberg, J., Uhlen, M. and Nygren, P.A. (1997) Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. Protein Expr Purif 11, 1-16.

Peti, W. and Page, R. (2007) Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. Protein Expr Purif 51, 1-10.

Schagger, H. and von Jagow, G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem 166, 368-79.

Silva E, C.A., Lopes A, Rodrigues A, Dias C, Conceição A, Alonso J, Correia da Costa J. M. , Bastos M, Parra F, Moradas-Ferreira P and Silva M. (2004) A recombinant antigen recognized by *Fasciola hepatica*-infected hosts. J Parasitology 90, 746-751.

Sjolander, A., Nygren, P.A., Stahl, S., Berzins, K., Uhlen, M., Perlmann, P. and Andersson, R. (1997) The serum albumin-binding region of streptococcal protein G: a bacterial fusion partner with carrier-related properties. J Immunol Methods 201, 115-23.

Sorensen, H.P. and Mortensen, K.K. (2005a) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J Biotechnol 115, 113-28.

Sorensen, H.P. and Mortensen, K.K. (2005b) Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. Microb Cell Fact 4, 1.

Terpe, K. (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 72, 211-22.

Vallejo, L.F. and Rinas, U. (2004) Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. Microb Cell Fact 3, 11.

Villaverde, A. and Carrio, M.M. (2003) Protein aggregation in recombinant bacteria: biological role of inclusion bodies. Biotechnol Lett 25, 1385-95.

Waugh, D.S. (2005) Making the most of affinity tags. Trends Biotechnol 23, 316-20.

Yao, L., Yin, J., Zhang, X., Liu, Q., Li, J., Chen, L., Zhao, Y., Gong, P. and Liu, C. (2007) *Cryptospodium parvum*: identification of a new surface adhesion protein on sporozoite and oocyst by screening of a phage-display cDNA library. Exp Parasitol 115, 333-8.

3.7. Appendices

3.7.1. Patent WO 2011/071404: Immunogens, process for preparation and use in systems of polyclonal antibodies production

Please, open here:

https://dl.dropbox.com/u/10833879/Fh8tag_Patents/WO2011071404.pdf

Chapter 4

The Fh8 tag: a novel fusion partner for simple and inexpensive protein purification in *Escherichia coli*

Abstract

Downstream processing is still a major bottleneck in recombinant protein production representing most of its costs. Hence, there is a constant demand of novel and cost-effective purification processes aiming at the recovery of pure and active target protein.

In this work, a novel purification methodology is presented, using a calcium binding protein as fusion handle: the Fh8 tag. The binding properties of Fh8 tag to a hydrophobic matrix were first studied via hydrophobic interaction chromatography (HIC). The Fh8 tag was then evaluated as a purification handle by its fusion to green fluorescent protein and superoxide dismutase. The purification efficiency of the Fh8-HIC strategy was compared to the immobilized metal ion affinity chromatography (IMAC) using the His_6 tag.

Results showed that the Fh8-HIC binding mechanism is calcium dependent in a low salt medium, making the purification process highly selective. Both target proteins were biologically active and successfully purified by HIC when fused to Fh8, achieving efficiencies identical to those of IMAC. Thus, Fh8 can be used as an affinity tag, allowing the design of inexpensive and effective purification processes, complementing its ability to promote the soluble expression of recombinant proteins in *E. coli*.

4.1. Introduction

The continuing growth of biotechnology industry demands new strategies for the rapid and economical recombinant protein production through the use of a variety of host organisms and solubility as well as affinity partners (Hearn and Acosta, 2001; Waugh, 2005; Esposito and Chatterjee, 2006; Malhotra, 2009; Vernet et al., 2011).

The purification of a protein of interest from biological mixtures using rapid, robust and cost-effective methodologies is still a current challenge for a good number of scientists in academia and industry. Taking into account that the downstream processing comprises up to 80% of the production costs (Hearn and Acosta, 2001), novel solutions that simplify the protein purification process are essential for the biotechnology's progress. This purification bottleneck has been somewhat overcome by the use of affinity fusion partners together with DNA recombinant techniques that allow to clone in frame the peptide or protein affinity tag at the N- or C-terminal end of the target construct (Hearn and Acosta, 2001). These fusion partners or tags diverge in molecular size and complexity and can also be used to improve soluble protein production, besides facilitating its purification by specific interaction with a known ligand/adsorption matrix (Malhotra, 2009).

Several affinity tags are commercially available for research or large scale protein production as, for instance, the Glutathione S-Transferase (GST) tag (Smith, 1988), Maltose Binding Protein (MBP) tag (Di Guan et al., 1988) and Hexahistidine tag (Hochuli et al., 1987) that have affinity to bind immobilized glutathione, maltose and nickel, respectively. Although being widely used, these fusion tags often yield low levels of protein purity due to unspecific and/or weak interactions with the matching matrices (Arnau et al., 2006). New purification tags are constantly emerging, outperforming the existing techniques and advancing the affinity concept or protein detection, as for instance, the Si-tag (Ikeda et al., 2010), Tamavidin tag (Takakura et al., 2010), Tab2 tag (Crusius et al., 2006), intein-mediated purifications (Wang et al., 2010), Heme tag (Asher and Bren, 2010), Z-basic tag (Hedhammar and Hober, 2007), Dock tag (Kamezaki et al., 2010) and the HiCaM tag (McCluskey et al., 2007).

The ideal purification process should: allow efficient and high yield protein recovery from a biological mixture; be universally applied to any protein without disturbing its

function; use a small fusion partner; be compatible with native conditions; and it should offer great control over selectivity (binding and elution of the protein of interest) using inexpensive and high capacity resins (Lichty et al., 2005; Waugh, 2005).

In this work, we investigate the Fh8 tag as a novel fusion purification strategy that meets these criteria by combining the calcium-binding intrinsic property of the Fh8 molecule with an inexpensive hydrophobic resin (phenyl-Sepharose). The Fh8 (GenBank ID: AF213970) was first isolated from the excreted/secreted proteins of the *Fasciola hepatica* parasite and recombinantly produced in *Escherichia coli* for diagnostic purposes, presenting a molecular weight of 8 kDa (Silva et al., 2004). This recombinant protein has been successfully used in *E. coli* as a solubility enhancer tag (Chapter 2) (Costa et al., 2012) and it was previously characterized as a calcium sensor protein that changes its structure upon calcium binding, exposing its hydrophobic residues to interact with its targets or other molecules (Fraga et al., 2010).

We demonstrate here that the Fh8 tag interacts with the phenyl-Sepharose hydrophobic resin with an identical mechanism as other calcium-binding proteins (Rozanas, 1998; Shimizu et al., 2003). Furthermore, using the green fluorescent protein (GFP) and the superoxide dismutase (SOD) as target model proteins, we also reveal that the Fh8 tag can be successfully applied as a purification handle for simple, rapid, and low cost recover of biologically active proteins.

4.2. Materials and Methods

4.2.1. Cloning of *sod* and *gfp* genes into expression vectors

Both *gfp* and *sod* genes were first modified and amplified by PCR to be later sub-cloned into the pETM11 and pETMFh8 expression vectors. The PCRs were conducted using minipreps (GenElute™ Plasmid Miniprep Kit - Sigma) of *gfp* and *sod* harboring plasmids as templates and the Phusion High-Fidelity DNA Polymerase (New England Biolabs - NEB), according to the manufacturer's instructions. Specific primers were design as follows: Forward primer: 5' – TCTATTCC*ATG*GGATCC*+18 nt gfp/sod* – 3' and Reverse primer: 5' – AATAGACTCGAG**TTA***+21 nt gfp/sod* – 3', to introduce the *Nco*I/*BamH*I restriction sites (underlined) at the N-terminal of both genes and the *Xho*I restriction site (underlined) after the stop codon (bold) at the C-terminal of both genes.

After DNA purification (QIAquick DNA gel extraction kit - Qiagen), the PCR products and the expression vectors were digested using the *Nco*I and *Xho*I restriction enzymes (New England Biolabs) and DNA ligations were carried out using the Rapid DNA Ligation kit (Roche). *E. coli* TOP10 competent cells were transformed with the obtained vectors and the resulting clones were verified by colony PCR using the T7 forward and reverse universal primers.

The correct insertion of *gfp* and *sod* genes into the pETM11 and pETMFh8 expression vectors was additionally confirmed by sequencing with both T7 forward and reverse universal primers.

4.2.2. Expression of Fh8 tag and HisGFP/SOD and Fh8GFP/SOD fusion proteins in *E. coli*

The expression strains and induction conditions used in this work were selected from a small-scale screening using 10 mL cultures (see Appendix 4.7.1, Figure A4.1). GFP recombinant proteins and the Fh8 tag were expressed in 2 L cultures using the *E. coli* BL21 Codon Plus-RIL strain. SOD recombinant proteins were expressed in 2 L cultures of *E. coli* Rosetta strain. Recombinant proteins were expressed as follows: pre-cultures were grown overnight (o/n) at 37 ºC and a dilution factor of 50 was used for inoculation of all cultures. Eight flasks of 250 mL of culture media (total culture volume of 2 L) were grown in 1 L flasks at 37 ºC and 200 rpm to a final $O.D._{600nm}$ of 0.4-0.6. *E. coli* cultures were induced with isopropyl-β-D-1-thiogalactopyranoside (IPTG) 0.2 mM, 18 ºC, o/n (for SOD and GFP expressions) or 1 mM, 30 ºC, 4 hours (for Fh8 tag expression).

After induction, three 5 mL samples were taken from each culture for dry weight estimation and the remaining cells were harvested for 25 minutes, at 4 ºC and 4000 rpm. Cell pellets were washed once with phosphate buffer saline 1x and collected again by centrifugation. Bacterial pellets were then stored at -20 ºC.

Stock solutions of antibiotics for plasmid maintenance and protein expression were prepared, filtered through 0.2 µm and stored at -20 ºC to be used in culture media with the following concentrations: kanamycin 50 $\mu g.mL^{-1}$ and chloramphenicol 10 $\mu g.mL^{-1}$.

4.2.3. Fh8 purification by HIC

Three different HIC purifications were tested to evaluate the interaction of Fh8 with the hydrophobic resin in the presence and absence of calcium, following the strategy presented in Appendix 4.7.2, Figure A4.2.

Cell pellets of 1.5 L culture of Fh8 tag were thawed and resuspended in a total volume of 3x25 mL of lysis buffer (50 mM Tris pH 7.6 and 150 mM NaCl, supplemented with 1x complete free EDTA protease inhibitor (Roche), 5 $\mu g.mL^{-1}$ DNAse (Sigma) and 1 $mg.mL^{-1}$ lysozyme (Sigma)). The lysis buffer was also supplemented with 5 mM $CaCl_2$, accordingly. After ressuspension, cells were incubated at room temperature for 10 minutes and then lysed by sonication (Branson) for six cycles of 30 seconds each, with 30 seconds intervals. Aliquots of 100 µL of total lysates were taken and stored at 4 ºC. Supernatant fractions were collected from the insoluble debris by centrifugation at 10000 rpm, 30 minutes, 4 ºC and aliquots of 100 µL were stored at 4 ºC for Bradford estimation of the total protein content and for *s*odium dodecyl sulfate polyacrylamide gel electrophoresis *(*SDS-PAGE) analysis.

The Fh8 tag purifications were conducted in the Biologic Duoflow FPLC system (Bio Rad) using a 5 mL pre-packed Phenyl Sepharose 6 Fast Flow High Sub column (GE Healthcare). Supernatant samples were loaded onto the HIC column, using the following buffers: in the HIC-1, the Fh8 tag was purified by HIC using a Tris NaCl buffer without calcium addition (50 mM Tris pH 7.6 and 150 mM NaCl*).* In the HIC-2, the Fh8 tag was purified by HIC using the Tris NaCl buffer supplemented with 5 mM $CaCl_2$. For both purifications, the Elution buffer was used in the same concentration as indicated in Table 4.1. In the HIC-3, the Fh8 tag was purified by HIC using the Tris NaCl buffer supplemented with 5 mM $CaCl_2$ but with two different Elution buffers: a first Elution buffer with EDTA (50 mM Tris pH 7.6, 150 mM NaCl and 5 mM EDTA) and a second Elution buffer identical to the described at Table 4.1 (50 mM Tris pH 10). Aliquots of all supernatant and flowthrough samples, washing steps and eluted samples were stored at 4 ºC to be further analysed.

Table 4.1. Composition of purification buffers:

Purification technique	Buffer	Composition
IMAC	Binding	50 mM Tris pH 7.6 150 mM NaCl 20 mM Imidazole
	Washing	50 mM Tris pH 7.6 150 mM NaCl 50 mM Imidazole
	Elution	50 mM Tris pH 7.6 150 mM NaCl 300 mM Imidazole
HIC	Binding	50 mM Tris pH 7.6 150 mM NaCl 5 mM CaCl2
	Washing = Binding 1:2	25 mM Tris pH 7.6 75 mM NaCl 2.5 mM CaCl2
	Elution	50 mM Tris pH 10

4.2.4. Protein purification by HIC and by IMAC

Cell pellets of HisSOD, Fh8SOD, HisGFP and Fh8GFP proteins (one pellet per protein, corresponding to 1 L *E. coli* culture) were thawed and resuspended in a total volume of 2x20 mL of lysis buffer (50 mM Tris pH 7.6 and 150 mM NaCl, supplemented with 1x complete free EDTA protease inhibitor (Roche), 5 $\mu g.mL^{-1}$ DNAse (Sigma) and 1 $mg.mL^{-1}$ lysozyme (Sigma)) with the addition of 20 mM imidazole, for IMAC purifications (1x20 mL), or 5 mM $CaCl_2$, for HIC purifications (1x20 mL). Cells were lysed as mentioned in the Fh8 purification (section 4.2.3) and aliquots of total lysates and supernatant samples were stored at 4 ºC for Bradford estimation of the total protein content and for SDS-PAGE analysis.

SOD and GFP target proteins were purified in parallel by HIC, using the same column as for the Fh8 purification, and by IMAC, using a 5 mL Histrap pre-packed column (GE Healthcare). All the proteins were purified following an identical strategy (see

Appendix 4.7.2, Figure A4.2). The composition of specific buffers used for each purification methodology is described in Table 4.1.

The obtained purified SOD and GFP fusion proteins were dialysed in 50 mM Tris pH 7.6, 150 mM NaCl buffer, filtered through 0.2 µm and stored at 4 ºC and -20 ºC until used. Columns regeneration and storage was performed according to the manufacturer's instructions.

4.2.5. Dual protein purification using HIC/IMAC and IMAC/HIC

Fh8GFP and Fh8SOD fusion proteins were purified by HIC followed by IMAC using the same protocols mentioned in the previous section 4.2.4, with the following modifications: after HIC purification, eluted samples were dialysed in 50 mM Tris pH 7.6, 150 mM NaCl buffer supplemented with 20 mM imidazole.

The IMAC/HIC purification was only conducted for the Fh8SOD protein. Eluted samples from the IMAC purification were dialysed in 50 mM Tris pH 7.6, 150 mM NaCl buffer supplemented with 5 mM $CaCl_2$, following an identical protocol as above-mentioned.

At the end of the HIC/IMAC or IMAC/HIC purifications, samples were dialysed in 50 mM Tris pH 7.6, 150 mM NaCl buffer and stored at 4 ºC and -20 ºC until needed. Aliquots of all purification samples were stored at 4 ºC to be further analysed.

4.2.6. Protein electrophoresis and protein quantification

SDS-PAGE of Fh8 and His fusion proteins was conducted according to the Laemmli method (Laemmli, 1970) using 12%-4% gels. SDS-PAGE of Fh8 tag expressed samples was conducted according to the Schagger and Jagow method (Schagger and von Jagow, 1987), using 15%-4% gels. Gels were loaded with the PageRuler Unstained Broad Range Protein Ladder (Thermo Scientific).

Protein purity and correspondent molecular weights in the SDS-PAGE and Tris-Tricine gels were estimated by densitometry, conducting three independent readings in the Image Lab 2.0 software (Bio Rad), using the Molecular Imager Chemidoc XRS+ system (Bio Rad).

The total protein content of supernatant samples and purification samples was estimated by Bradford method (Bradford, 1976), using the Bio Rad protein assay dye reagent and bovine serum albumin as standard. Protein quantifications were also conducted by reading the absorbance of eluted samples at 280 nm. All protein quantifications were conducted in triplicates.

Purification efficiencies were estimated by the ratio between the protein amount in eluted samples and the protein amount loaded onto the purification column.

4.2.7. GFP fluorescence measurements and SOD activity evaluation

The GFP target protein was considered to be active by emitting green fluorescence. The fluorescence intensity of GFP eluted samples was measured in triplicates using a spectrofluorometer with an excitation filter of 475 nm and an emission filter of 505 nm. The resulting fluorescence intensities were normalized by the protein amount (in milligrams).

The SOD activity was evaluated according to the method of Marklund and Marklund (Marklund and Marklund, 1974), in three independent assays, estimating the inhibition of pyrogallol autoxidation promoted by the eluted samples, using a spectrophotometer at 420 nm.

4.3. Results

4.3.1. Cloning of SOD and GFP fusion proteins

The insertion of *gfp* and *sod* genes into the pETM11 and pETMFh8 vectors was successfully confirmed by sequencing and further analysed by BLASTN and BLASTP tools at the NCBI website. The obtained *gfp* gene presented 100% identity with the *Aequorea coerulescens*-derived *gfp* gene cloned into vector pT7XbG2-AcGFP1 (GenBank: AB255038.1) and 76% identity with the original *gfp* gene from *Aequorea coerulescens* (GenBank: AY151052.1). The obtained *gfp* gene presented also 89% identity with the *enhanced gfp* gene cloned into vector pEGFP-1 (GenBank: U55761.1).

At the amino acid level, the obtained GFP protein presented 91% homology with the enhanced GFP (GenBank: AAB02572.1) and 98% homology with the AcGFP (GFP from *Aequorea coerulescens*, GenBank: AAN41637.1). The obtained *sod* gene presented 100% identity with *Saccharomyces cerevisiae* copper-zinc superoxide dismutase gene (GenBank: AY690619.1). The constructed pETMFh8SOD and pETMFh8GFP vectors presented 100% identity with the mRNA from *Fasciola hepatica* putative calcium-binding protein (GenBank: AF213970.1).

4.3.2. The Fh8 tag interaction with the hydrophobic resin

The SDS-PAGE analysis of the three HIC experimental protocols tested for the Fh8 tag is presented in Figure 4.1 and the densitometry results can be consulted in the Appendix 4.7.3, Table A4.1.

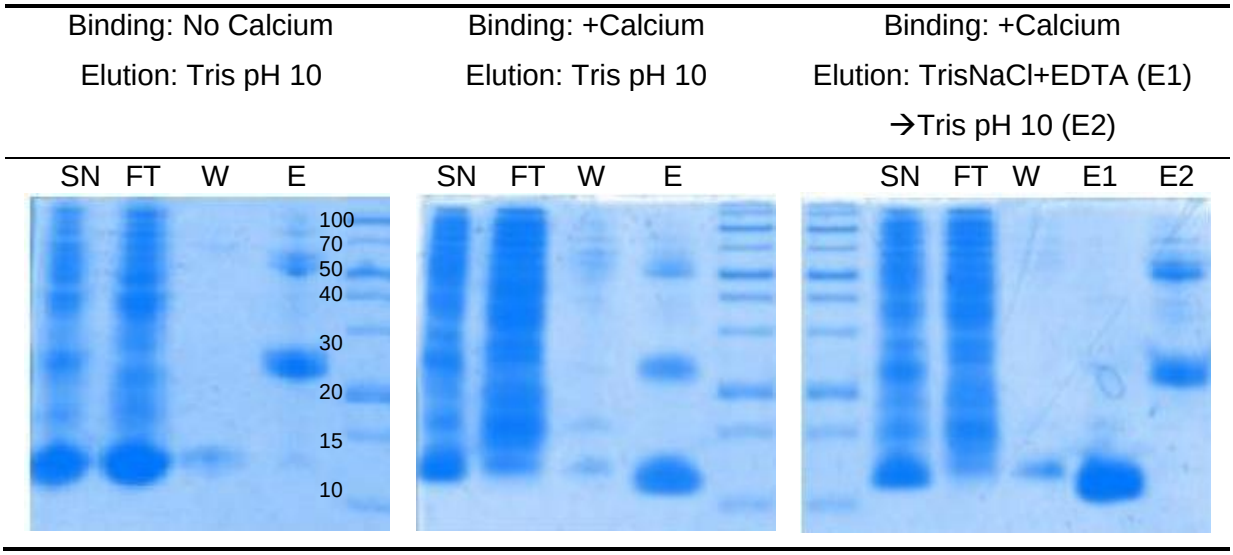


Figure 4.1. SDS-PAGE analyses of the three HIC purifications of the Fh8 tag, using TrisNaCl buffers from Table 4.1 with or without 5 mM $CaCl_2$. SN – supernatant sample loaded onto the columns; FT – flow-through sample; W – washing sample; E – eluted sample; E1 - eluted sample using the buffer: 50 mM Tris pH 7.6, 150 mM NaCl, 5 mM EDTA; E2 – eluted sample using the buffer: 50 mM Tris pH 10 (Table 4.1).

In the HIC purification without calcium supplementation in the binding step, most of the loaded Fh8 (gel band of 12 kDa) is observed in the flow-through (FT lane, gel band of 12 kDa) and washing (W lane, gel band of 12 kDa) samples. The eluted sample is

majorly composed by two gel bands, one of 23 and other of 54 kDa, as estimated by densitometry (see Appendix 4.7.3, Table A4.1).

In the HIC purifications using buffers supplemented with calcium, the Fh8 is mostly visible in the eluted samples (E lanes, gel band of 12 kDa) and small leakages are observed in the washing steps. When performing the elution with 50 mM Tris pH 10, it is possible to observe the recovery of Fh8 together with the other two gel bands of 23 and 54 kDa. On the other hand, when a first elution step with EDTA supplementation was performed, only the 12 kDa gel band is observed (lane E1). The other high molecular weight gel bands elute in 50 mM Tris pH 10 buffer (lane E2).

Table 4.2 presents the total protein content of the collected samples from all the HIC purifications as well as the corresponding purification efficiency. As shown in this table, for the assay with no calcium addition, there was no eluted Fh8 from the total protein loaded (0% purification efficiency). In the assays with calcium addition, the purification efficiency using the two elution strategies was similar (82±6.2% and 86±4.3%, respectively). These results are in good agreement to the SDS-PAGE analysis, showing that a higher recovery of the Fh8 tag is obtained when using the TrisNaCl buffers supplemented with calcium.

Table 4.2. Fh8 tag purification results using three different HIC protocols:

	HIC-1	**HIC-2**	**HIC-3**
	Binding: No Calcium Elution: Tris pH 10	Binding: + 5mM $CaCl_2$ Elution: Tris pH 10	Binding: + 5mM $CaCl_2$ Elution: TrisNaCl+EDTA→Tris pH 10
Loaded (mg)[a]	37 ± 1.4	22 ± 0.7	35 ± 0.9
Eluted (mg)[b]	0 ± 0	18 ± 1.3	31 ± 1.5
Purification Efficiency (%) [c]	0 ± 0	82 ± 6.2	86 ± 4.3

[a, b] Values were determined by taking into account the SDS-PAGE densitometric analysis of each target protein and the total protein amount presented in the loading ([a]) or elution step ([b]).

[c] Efficiency is the ratio between the target protein amount in the elution step and the initially loaded amount of each target protein.

4.3.3. SOD and GFP purification by HIC and by IMAC

The Fh8 tag was evaluated as a purification handle using the HIC in the presence of calcium. Fh8GFP and Fh8SOD recombinant fusion proteins were purified by HIC in parallel with IMAC, and the equivalent His-fused proteins were also used as a reference control for both HIC and IMAC purifications. Figures 4.2 and 4.3 present the main comparison results obtained for HIC and IMAC purifications: Figure 4.2 shows the SDS-PAGE of the comparison analysis and Figure 4.3 shows the resulting purification efficiency of each target protein that was estimated from the ratio between the amount of eluted proteins and the corresponding amount of loaded proteins (see Table 4.3).

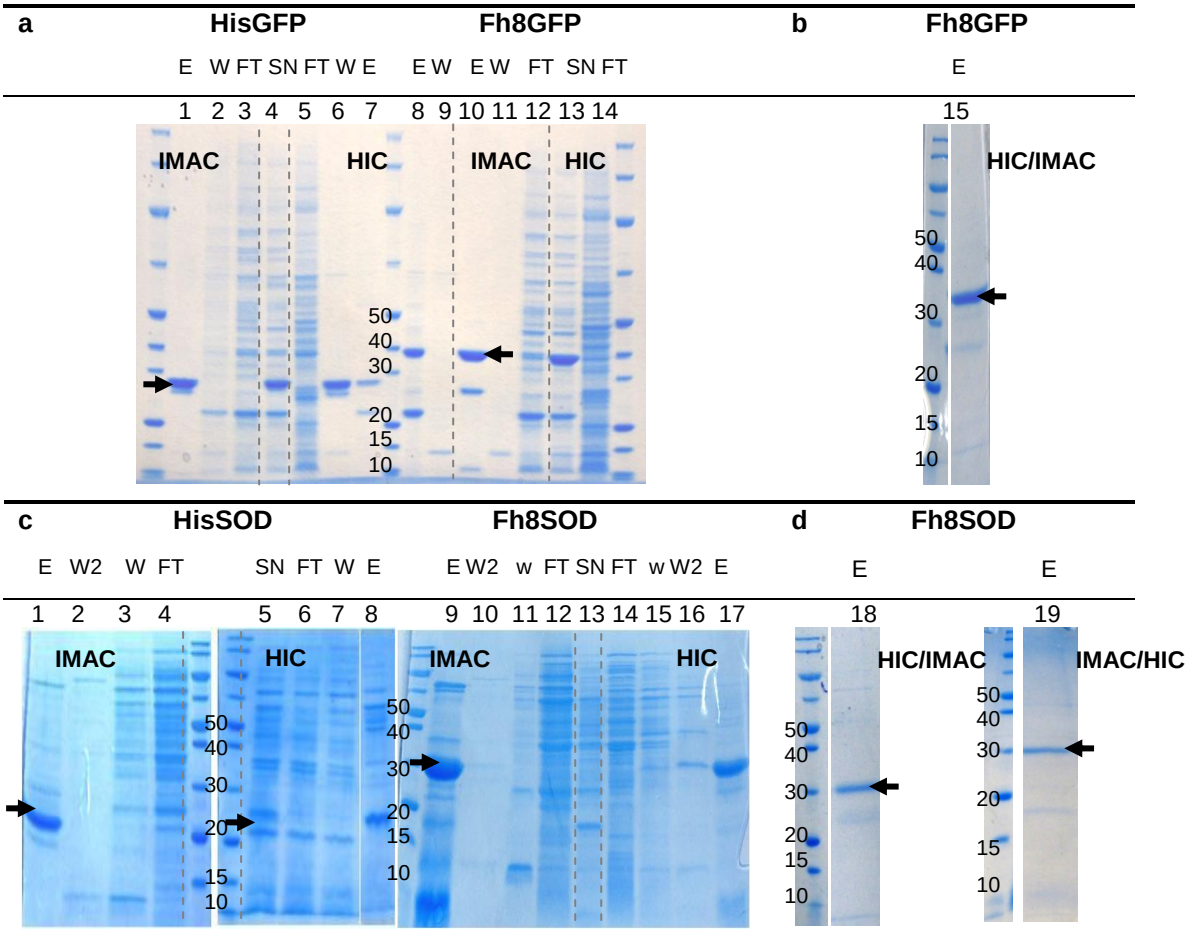


Figure 4.2. SDS-PAGE of HIC and IMAC purifications of: (**a**) HisGFP and Fh8GFP, (**b**) Fh8GFP by HIC/IMAC, (**c**) HisSOD and Fh8SOD, (**d**) Fh8SOD by HIC/IMAC and IMAC/HIC. Aliquots of all samples were prepared and resolved by SDS-PAGE as follows: supernatant (SN) and flow-through (FT) aliquots contain 10 µg of total protein; washing samples (W, w and W2) contain 2 µg of total protein; eluates (E) contain 5 µg of total protein. Proteins were purified by HIC or IMAC using the buffers presented at Table 4.1. *Arrows* indicate the position of each recombinant protein in SDS-PAGE.

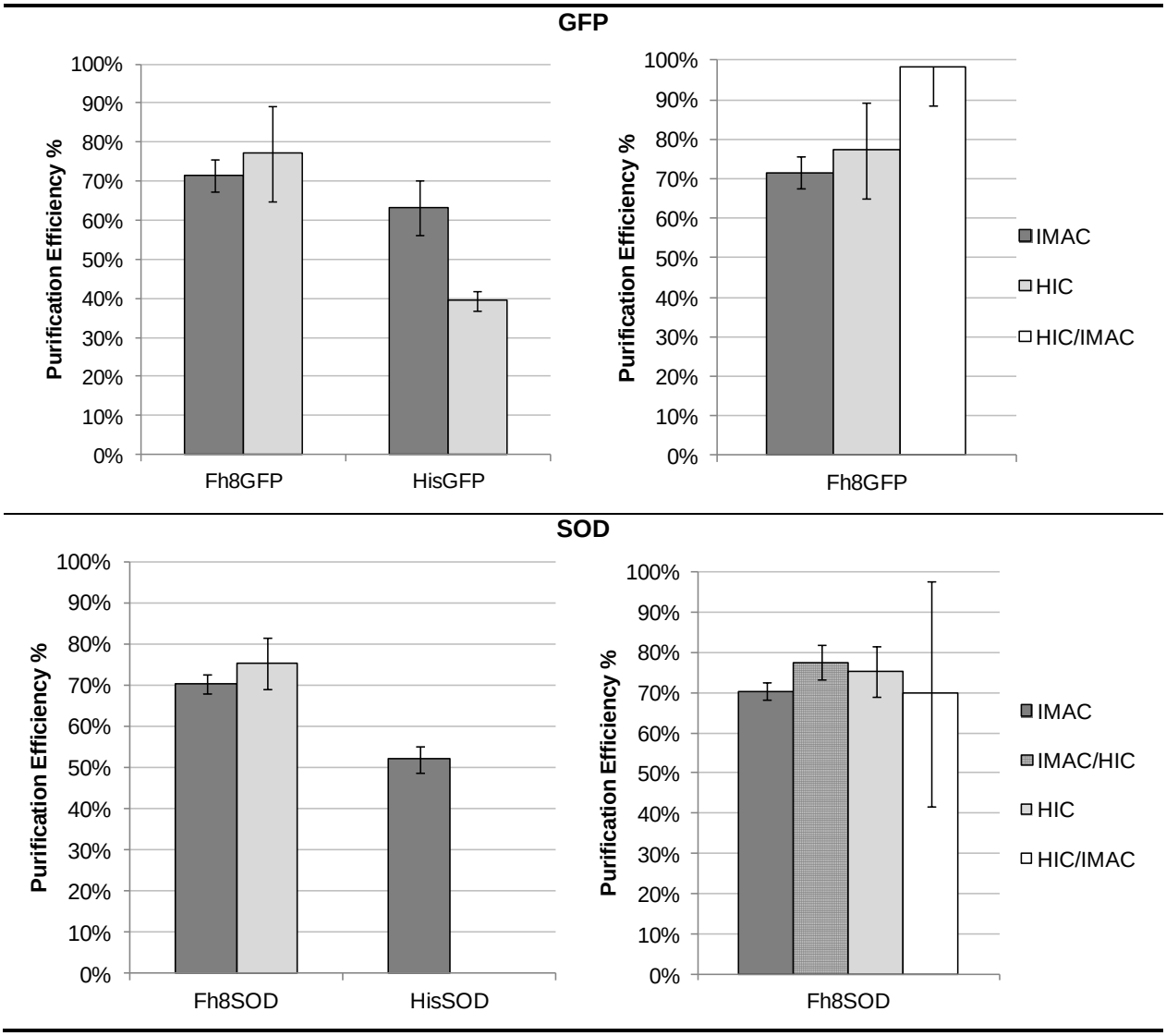


Figure 4.3. GFP and SOD protein purification efficiency by HIC in comparison to IMAC and/or IMAC/HIC and/or HIC/IMAC: values are the ratio between the target protein amount (average±standard deviation) in the elution step (Eluted yield - Table 4.3) and the initially loaded amount (average±standard deviation) of each target protein (Loaded – Table 4.3).

In Figure 4.2.a, the HisGFP samples (30 kDa – see Appendix 4.7.3, Table A4.1) were loaded at lanes 1-7 and the Fh8GFP samples (37 kDa – see Appendix 4.7.3, Table A4.1) at lanes 8-14. As observed in this figure, the Fh8GFP purification by the Fh8-HIC methodology resulted in an identical purification profile as the Fh8GFP IMAC purification, with low protein amount in the flow-through samples (lanes 12 and 14) and with most of the protein in the elution samples (lanes 8 and 10). In both HIC and IMAC

elution samples, it is also possible to observe a second gel band (at 22 kDa for HIC and at 28 kDa for IMAC) that affected the purity of Fh8GFP samples. The Fh8GFP purification efficiencies by Fh8-HIC (77%±12%) and by IMAC (72%±4%) were identical (Figure 4.3) and these results were in good agreement to what it was observed in SDS-PAGE analysis. The HisGFP was successfully purified by IMAC (lane 1) but not so well by HIC, in which most of the protein was found in the washing sample (lane 6) rather than in the elution sample (lane 7). The HisGFP purification by IMAC yielded a similar efficiency (63%±7%) as the Fh8GFP purifications by IMAC or Fh8-HIC (Figure 4.3).

Figure 4.2.c presents the purification results obtained for SOD recombinant proteins. Fh8SOD (26 kDa) and HisSOD (19 kDa) proteins migrated in SDS-PAGE with a molecular weight 3 kDa higher than expected (see Appendix 4.7.3, Table A4.1). The Fh8SOD purification profile by Fh8-HIC (lanes 14-17) was identical to the purification profile by IMAC (lanes 9-12). In both chromatographic techniques, the flow-through and washing samples presented low protein amount and the Fh8SOD protein was mostly recovered in the elution step. The eluted samples were not completely pure, presenting other protein bands of different molecular weights.

As observed in Figure 4.3, the Fh8SOD recovered from Fh8-HIC strategy (75%±6%) resulted in a similar purification efficiency as the Fh8SOD recovered from IMAC (70%±2%).

The purification efficiency of Fh8SOD fusion protein was higher than the one of HisSOD fusion protein. The HisSOD purifications showed a similar pattern as the HisGFP purifications: HisSOD was purified by IMAC (Figure 4.2.c, lane 1) but, without the presence of Fh8 tag, this target protein was not purified by HIC, coming out of the column in the flow-through and washing samples (Figure 4.2.c, lanes 6 and 7). In the HIC purification, other gel band from the soluble extract of *E. coli* expressing the HisSOD protein was observed nearby the HisSOD protein gel band. We could confirm that this gel band did not correspond to the HisSOD protein as it was not purified by IMAC (Figure 4.2.c, lanes 3 and 4) and its densitometric analysis revealed a molecular weight of 3 kDa higher than the observed for HisSOD (21 kDa – see Appendix 4.7.3, Table A4.1). Also a protein loss of 96%±8% was estimated for the HisSOD purification by HIC, taking into account the total protein content and corresponding densitometric analysis in the flowthrough and washing samples.

4.3.4. Dual purification system: HIC/IMAC and IMAC/HIC

The purity of eluted Fh8GFP and Fh8SOD proteins from the Fh8-HIC purification was further evaluated by a second purification with IMAC. A second HIC purification was also conducted with the Fh8SOD protein after IMAC purification.

As observed in Figures 4.2.b and 4.2d, the HIC/IMAC or IMAC/HIC purifications improved both Fh8GFP and Fh8SOD proteins purity, achieving purity levels between 85% and 92% of the target protein, evaluated by SDS-PAGE densitometry analysis. In Figure 4.3, the efficiency of the two-step purifications was compared to the efficiency of single purifications. The Fh8GFP protein recovered after HIC/IMAC presented a similar purification yield as the Fh8GFP from Fh8-HIC purification. Fh8SOD proteins after HIC/IMAC or IMAC/HIC yielded similar purification efficiencies as in the corresponding single purifications.

Table 4.3. Summary of Fh8 and His fusion proteins purification results by HIC in comparison to IMAC:

	Fh8-GFP		**His-GFP**		**Fh8-SOD**		**His-SOD**	
	IMAC	HIC	IMAC	HIC	IMAC	HIC	IMAC	HIC
Loaded (mg)[a]	55±2.1	69±5.2	48±2.5	58±4.0	7.9±0.014	19±1.2	9.7±0.56	8.8±0.37
Eluted (mg)[a]	40±2.2	54±8.4	30±3.3	23±1.4	5.5±0.18	14±1.2	5.0±0.30	0
Yield (mg) per g of *E. coli* dry weight[b]	61±3.4	82±13	46±5.1	35±2.1	14±0.45	36±3.0	19±1.1	0

[a] Values (mg) per litre of *E. coli* culture. These values were determined by taking into account the SDS-PAGE densitometric analysis of each target protein and the total protein amount presented in the loading or elution step, respectively.

[b] Values (mg per g of *E. coli* dry weight) were obtained by the ratio between the eluted yield of each target protein (in mg per litre of *E. coli* culture) and the dry weight of the corresponding *E. coli* culture (in g/L). These values are calculated by the SDS-PAGE densitometric analysis for each target protein.

4.3.5. Functional assays of purified fusion proteins

The Fh8-HIC purification methodology was further compared to the IMAC purification regarding the biological activity of the purified protein. Figure 4.4 presents the activity results per mg of target protein.

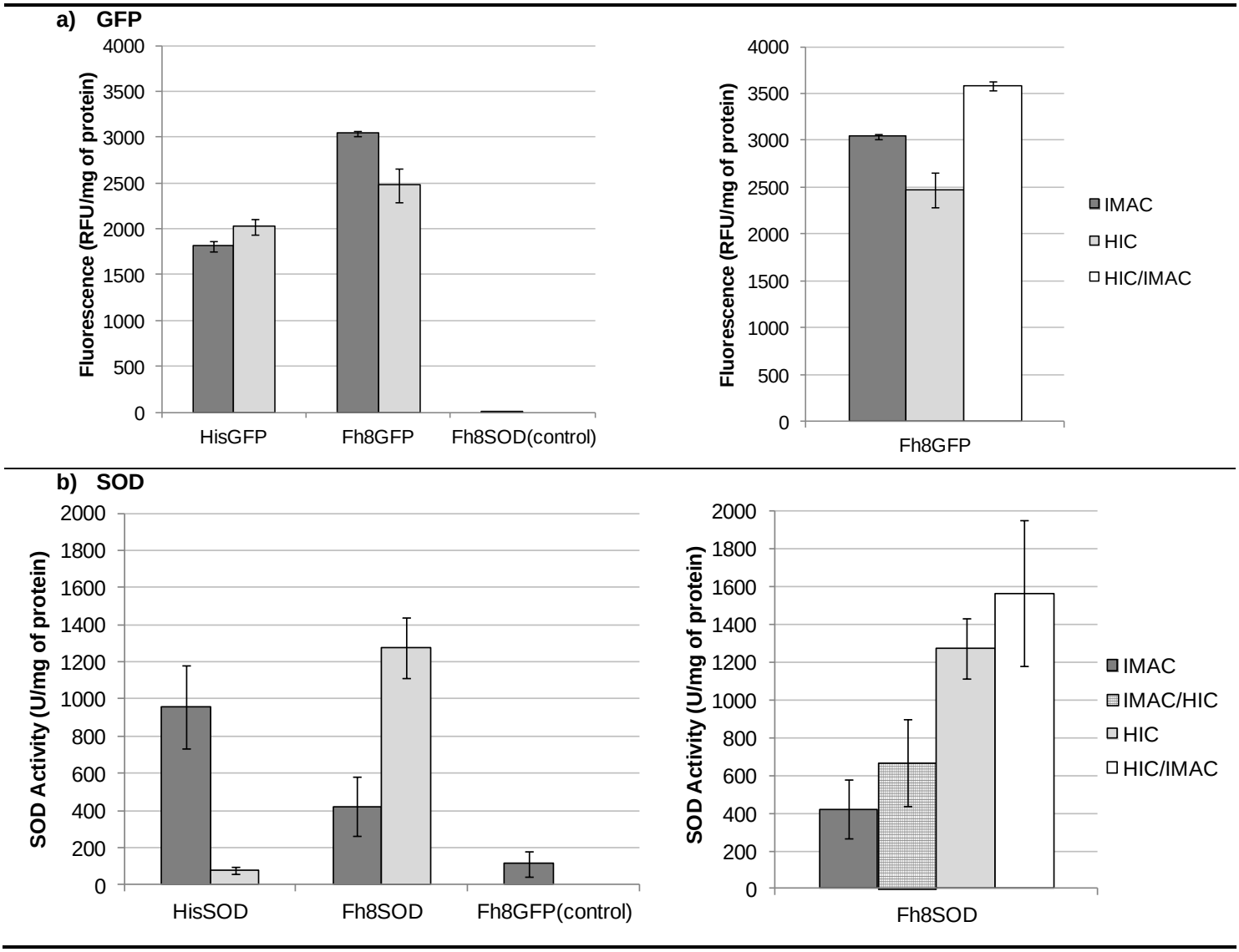


Figure 4.4. GFP (**a**) and SOD (**b**) activity measurements: comparison of protein activity after HIC, IMAC, HIC/IMAC and/or IMAC/HIC purifications. The presented values for GFP or SOD activity are the ratio between the obtained results of three activity measurements per target protein (average±standard deviation) and the target protein amount (mg). GFP fluorescence was measured with an excitation filter of 475 nm and an emission filter of 505 nm. One SOD activity unit is defined as the amount of SOD that inhibits the rate of pyrogallol autoxidation by half at pH 8.2 and 25 ºC (Marklund and Marklund, 1974).

In general, the GFP proteins purified by IMAC resulted in higher relative fluorescence units (RFU) per mg of protein than the GFP proteins purified by HIC. The Fh8 tag did not interfere with the GFP fluorescence as shown by the higher fluorescence readings of the Fh8-HIC and IMAC eluted Fh8GFPs compared to the eluted HisGFPs. Even so, Fh8GFP and HisGFP proteins yielded high fluorescence measurements (>1500 RFU per mg of protein), indicating that their natural ability to fluoresce was not affected by the purification steps. The Fh8GFP purified protein from HIC/IMAC presented also high RFU values. The several GFP target proteins did also exhibit green light under natural daylight. Photos of all purified GFP proteins are available at the Appendix 4.7.4, Figure A4.3.

The Fh8SOD fusion protein purified by the Fh8-HIC strategy presented similar biological activity as the HisSOD protein purified by IMAC, and higher biological activity than the Fh8SOD purified by IMAC. In the eluted sample of HisSOD purified by HIC, no considerable SOD activity was detected, thus confirming the SDS-PAGE results of Figure 4.2. In good agreement with the results observed for the GFP activity, the Fh8 tag did not affect the SOD activity as fusion protein.

4.4. Discussion

In this study, a novel methodology for protein purification using the Fh8 tag was investigated. The chromatographic properties of the Fh8 tag were firstly demonstrated through simple proof-of-principle experiments and the usefulness of the Fh8 as purification tag was then evaluated by its fusion to two different model proteins: GFP and SOD. The purification efficiency of Fh8-fused proteins by HIC was also compared to the His tag technology.

The Fh8 tag has been used as solubility tag for protein overexpression in *E. coli* (Chapter 2) (Costa et al., 2012) and it was previously characterized as a calcium binding sensor protein from the Calmodulin-like protein family. In the presence of calcium, the Fh8 molecule undergoes conformational changes, exposing its hydrophobic region for interaction with its targets or other molecules (Ikura, 1996; Fraga et al., 2010; Russell et al., 2012). Taking into account this calcium-binding property, the Fh8 was explored as a purification tag using the hydrophobic interaction chromatography. Experiments using the Fh8 tag alone were performed in order to demonstrate the specificity of the binding mechanism and the consequent behavior of Fh8 as a calcium binding protein in terms of hydrophobicity and affinity to the phenyl-Sepharose hydrophobic resin.

Our results showed that the Fh8 could only bind the hydrophobic matrix in the presence of calcium in the mobile phase. Without calcium, most of the Fh8 was collected in the flow through, revealing low affinity for the matrix. The Fh8 tag was also able to bind to the phenyl-Sepharose resin under low salt concentration in the mobile phase, corroborating the results obtained for other calcium binding proteins (Rozanas, 1998; Shimizu et al., 2003). The salt concentration in the mobile phase has significant contribution for the HIC performance: when using anti-chaotropic salts as sodium chloride, in a medium of high salt concentration, the bound form of the protein is thermodynamically more favorable than the unbound state. These salts bind water molecules strongly, excluding them from the protein and ligand surfaces, which start to interact hydrophobically (salting-out effect) (Queiroz et al., 2001; Lienqueo et al., 2007; Tsumoto et al., 2007). Therefore, when using low salt concentrations in the mobile phase, the binding of proteins to the HIC matrix is not favored. Considering that the purification buffers used in this work have low salt concentrations, most of the

contaminant proteins in the *E. coli* extract are excluded in the binding step, promoting selectivity towards the purification of the desired fusion protein.

The Fh8 tag eluted from the phenyl-Sepharose column either with EDTA or with pH 10. The use of a calcium chelating agent was already demonstrated to be effective for the elution of other calcium binding proteins (Rozanas, 1998; Shimizu et al., 2003; McCluskey et al., 2007). This elution mechanism proved to be highly selective towards Fh8 tag since other proteins of superior molecular weight (between 20-25 kDa and 50-75 kDa) were only observed in the elution with pH 10, after the first elution with EDTA. These other proteins could, however, correspond to oligomer forms of the Fh8 tag, as suggested by their molecular weight. In fact, the Fh8 molecule was previously described to form dimmers upon calcium binding that can resist to the denaturating conditions used in the SDS-PAGE analysis (Fraga et al., 2010).

Overall, the Fh8 tag interacts with the HIC resin as a calcium binding protein and it has the properties required for protein purification by HIC, offering the possibility to control the binding, selectivity and elution steps with the exclusion of major *E. coli* contaminants.

In order to investigate if the chromatographic properties of Fh8 were preserved after fusion to target proteins, we selected two proteins with different characteristics, GFP and SOD. These proteins were fused to the Fh8 tag and to the His_6 tag, and a Fh8-HIC purification protocol with mild conditions that do not interfere with target biological activity of target proteins was developed.

Results from this work demonstrated that the purification efficiency and biological activity obtained for the Fh8-fused proteins after Fh8-HIC purification is comparable to those of Fh8-fused and His-fused proteins after IMAC purification.

The higher biological activity observed for the Fh8SOD fusion protein purified by HIC or by HIC/IMAC may be correlated to the calcium addition during the HIC purification protocol and its effect on SOD activity itself. Actually, the addition of calcium has a positive effect on SOD activity (Bakardjieva et al., 2000). In order to corroborate this effect, we also compared the biological activity of Fh8SOD protein purified by IMAC with or without the addition of $CaCl_2$ 5 mM, and an increase in SOD activity per mg of protein was observed (data not shown).

Results from this work also established that the HIC purification of both GFP and SOD proteins was only possible when fused to Fh8 tag as shown by the weak interaction of the His_6-tagged proteins with the HIC matrix (most of the proteins were found in the flow-through samples). This unfavorable interaction to the HIC matrix can be explained by the previously-mentioned low salt concentration used in the binding buffer. The difference between the pH value used in both binding and washing buffers (pH 7.6) and the p*I* values of target proteins (5.8 for HisGFP and 5.9 for HisSOD) is probably causing a net charge shield around these proteins that will also decrease the chance for hydrophobic interactions (Xia et al., 2005; Tsumoto et al., 2007). In the HisGFP purification by HIC, the HisGFP protein was mostly detected in the washing sample but it was also identified in its active form in the elution sample. However, the fact of finding active GFP in the elution sample can be mostly caused by a non-optimized washing step rather than a specific hydrophobic interaction due to the above-mentioned conditions (low salt concentration and different pH and p*I* values).

The use of dual tags for expression and purification of recombinant proteins has become an increasingly popular method that simplifies purification and yields homogeneous preparations of the protein of interest (Terpe, 2003). Our results showed that the Fh8-HIC and IMAC purification strategies can be used in a sequential step, complementing each other, to obtain an active and more purified protein. The use of two consecutive purification steps and the distinct nature of HIC and IMAC methodologies allows for the efficient removal of contaminating proteins (McCluskey et al., 2007).

The data presented here have proven the feasibility of the Fh8-HIC purification strategy as a rapid, easy and low cost methodology for protein recover from *E. coli* extracts, even without an optimized purification protocol. Proteins purified by the Fh8-HIC strategy have the extra feature of being free of *E. coli* endotoxins, since the HIC is itself one of the strategies used for the removal of endotoxins (Wilson et al., 2001; Magalhães et al., 2007; Ongkudon et al., 2012).

The Fh8-HIC purification strategy is calcium-dependent and, consequently, chelating agents must be avoided during protein binding and washing steps. This could be considered a limitation of the process but it is important to mention that this limitation can also be found in the IMAC technology. Due to the calcium-dependent mechanism for protein purification using the Fh8 tag, this Fh8-HIC strategy can be potentially

applied for several different target proteins without requiring the development and optimization of a new system for each new protein of interest.

4.5. Conclusion

By taking part of a one-step purification process, we successfully established a novel purification tag – Fh8 tag – that offers several benefits: a low molecular weight (8 kDa) that may not disturb the biological activity of target proteins, highly soluble and easy protein production in *E. coli*, besides simplicity and economy of the purification process (it does not require specialized buffers and substrates for elution and it makes use of inexpensive and high-capacity matrices).

The novel Fh8 purification tag can be of most utility for the inexpensive large scale production and purification of several proteins. In addition, it can also be used in a two-step purification procedure together with IMAC methodology, as well as other purification strategy, to improve the protein purity level.

Acknowledgements

We thank Dr. Hüseyin Besir, Protein Expression and Purification Facility Core, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany, for supplying the pETM11 expression vector, and also Dr. Vítor Costa for kindly providing the SOD target gene used in this work.

4.6. References

Arnau, J., Lauritzen, C., Petersen, G.E. and Pedersen, J. (2006) Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expression and Purification 48, 1-13.

Asher, W.B. and Bren, K.L. (2010) A heme fusion tag for protein affinity purification and quantification. Protein Science 19, 1830-1839.

Bakardjieva, N.T., Christov, K.N. and Christova, N.V. (2000) Effect of calcium and zinc on the activity and thermostability of superoxide dismutase. Biologia Plantarum 43, 73-78.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72, 248-54.

Costa, S.J., Almeida, A., Castro, A., Domingues, L. and Besir, H. (2012) The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology. Applied Microbiology and Biotechnology.

Crusius, K., Finster, S., McClary, J., Xia, W., Larsen, B., Schneider, D., Lu, H.T., Biancalana, S., Xuan, J.A., Newton, A., Allen, D., Bringmann, P. and Cobb, R.R. (2006) Tab2, a novel recombinant polypeptide tag offering sensitive and specific protein detection and reliable affinity purification. Gene 380, 111-119.

Di Guan, C., Li, P., Riggs, P.D. and Inouye, H. (1988) Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. Gene 15;67, 21-30.

Esposito, D. and Chatterjee, D.K. (2006) Enhancement of soluble protein expression through the use of fusion tags. Current Opinion in Biotechnology 17, 353-358.

Fraga, H., Faria, T.Q., Pinto, F., Almeida, A., Brito, R.M.M. and Damas, A.M. (2010) FH8-a small EF-hand protein from *Fasciola hepatica*. Febs Journal 277, 5072-5085.

Hearn, M.T.W. and Acosta, D. (2001) Applications of novel affinity cassette methods: use of peptide fusion handles for the purification of recombinant proteins. Journal of Molecular Recognition 14, 323-369.

Hedhammar, M. and Hober, S. (2007) Z(basic) - A novel purification tag for efficient protein recovery. Journal of Chromatography A 1161, 22-28.

Hochuli, E., Dobeli, H. and Schacher, A. (1987) New Metal Chelate Adsorbent Selective for Proteins and Peptides Containing Neighboring Histidine-Residues. Journal of Chromatography 411, 177-184.

Ikeda, T., Ninomiya, K., Hirota, R. and Kuroda, A. (2010) Single-step affinity purification of recombinant proteins using the silica-binding Si-tag as a fusion partner. Protein Expression and Purification 71, 91-95.

Ikura, M. (1996) Calcium binding and conformational response in EF-hand proteins. Trends in Biochemical Sciences 21, 14-17.

Kamezaki, Y., Enomoto, C., Ishikawa, Y., Koyama, T., Naya, S., Suzuki, T. and Sakka, K. (2010) The Dock tag, an affinity tool for the purification of recombinant proteins, based on the interaction between dockerin and cohesin domains from *Clostridium josui* cellulosome. Protein Expression and Purification 70, 23-31.

Laemmli, U.K. (1970) Cleavage of Structural Proteins during Assembly of Head of Bacteriophage-T4. Nature 227, 680-685.

Lichty, J.J., Malecki, J.L., Agnew, H.D., Michelson-Horowitz, D.J. and Tan, S. (2005) Comparison of affinity tags for protein purification. Protein Expression and Purification 41, 98-105.

Lienqueo, M.E., Mahn, A., Salgado, J.C. and Asenjo, J.A. (2007) Current insights on protein behaviour in hydrophobic interaction chromatography. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences 849, 53-68.

Magalhães, P.O., Lopes, A.M., Mazzola, P.G., Rangel-Yagui, C., Penna, T.C.V. and Pessoa Jr., A. (2007) Methods of endotoxin removal from biological preparations: a review. Journal of Pharmacy & Pharmaceutical Sciences 10, 388-404.

Malhotra, A. (2009) Tagging for Protein Expression. In: Methods in Enzymology, Vol. 463. Elsevier Inc.

Marklund, S. and Marklund, G. (1974) Involvement of Superoxide Anion Radical in Autoxidation of Pyrogallol and a Convenient Assay for Superoxide-Dismutase. European Journal of Biochemistry 47, 469-474.

McCluskey, A.J., Poon, G.M.K. and Gariepy, J. (2007) A rapid and universal tandem-purification strategy for recombinant proteins. Protein Science 16, 2726-2732.

Ongkudon, C.M., Chew, J.H., Liu, B. and Danquah, M.K. (2012) Chromatographic removal of endotoxins: a bioprocess engineer's perspective. International Scholarly Research Network - ISRN Chromatography 649746.

Queiroz, J.A., Tomaz, C.T. and Cabral, J.M.S. (2001) Hydrophobic interaction chromatography of proteins. Journal of Biotechnology 87, 143-159.

Rozanas, C. (1998) Purification of calcium-binding proteins using hydrophobic interaction chromatography. Life Science News I.

Russell, S.L., McFerran, N.V., Moore, C.M., Tsang, Y., Glass, P., Hoey, E.M., Trudgett, A. and Timson, D.J. (2012) A novel calmodulin-like protein from the liver fluke, *Fasciola hepatica*. Biochimie 94, 2398–2406.

Schagger, H. and von Jagow, G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem 166, 368-79.

Shimizu, F., Sanada, K. and Fukada, Y. (2003) Purification and immunohistochemical analysis of calcium-binding proteins expressed in the chick pineal gland. Journal of Pineal Research 34, 208-216.

Silva, E., Castro, A., Lopes, A., Rodrigues, A., Dias, C., Conceicao, A., Alonso, J., da Costa, J.M.C., Bastos, M., Parra, F., Moradas-Ferreira, P. and Silva, M. (2004) A recombinant antigen recognized by *Fasciola hepatica*-infected hosts. Journal of Parasitology 90, 746-751.

Smith, D.B.a.K.S.J. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene 67, 31-40.

Takakura, Y., Oka, N., Kajiwara, H., Tsunashima, M., Usami, S., Tsukamoto, H., Ishida, Y. and Yamamoto, T. (2010) Tamavidin, a versatile affinity tag for protein purification and immobilization. Journal of Biotechnology 145, 317-322.

Terpe, K. (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Applied Microbiology and Biotechnology 60, 523-533.

Tsumoto, K., Ejima, D., Senczuk, A.M., Kita, Y. and Arakawa, T. (2007) Effects of salts on protein-surface interactions: Applications for column chromatography. Journal of Pharmaceutical Sciences 96, 1677-1690.

Vernet, E., Kotzsch, A., Voldborg, B. and Sundstrom, M. (2011) Screening of genetic parameters for soluble protein expression in *Escherichia coli*. Protein Expression and Purification 77, 104-111.

Wang, L., Kang, J.H., Kim, K.H. and Lee, E.K. (2010) Expression of intein-tagged fusion protein and its applications in downstream processing. Journal of Chemical Technology and Biotechnology 85, 11-18.

Waugh, D.S. (2005) Making the most of affinity tags. TRENDS in Biotechnology 23.

Wilson, M.J., Haggart, C.L., Gallagher, S.P. and Walsh, D. (2001) Removal of tightly bound endotoxin from biological products. Journal of Biotechnology 88, 67-75.

Xia, F., Nagrath, D. and Cramer, S.M. (2005) Effect of pH changes on water release values in hydrophobic interaction chromatographic systems. Journal of Chromatography A 1079, 229-235.

4.7. Appendices

4.7.1. Solubility small-scale screening of SOD and GFP expression in *E. coli*

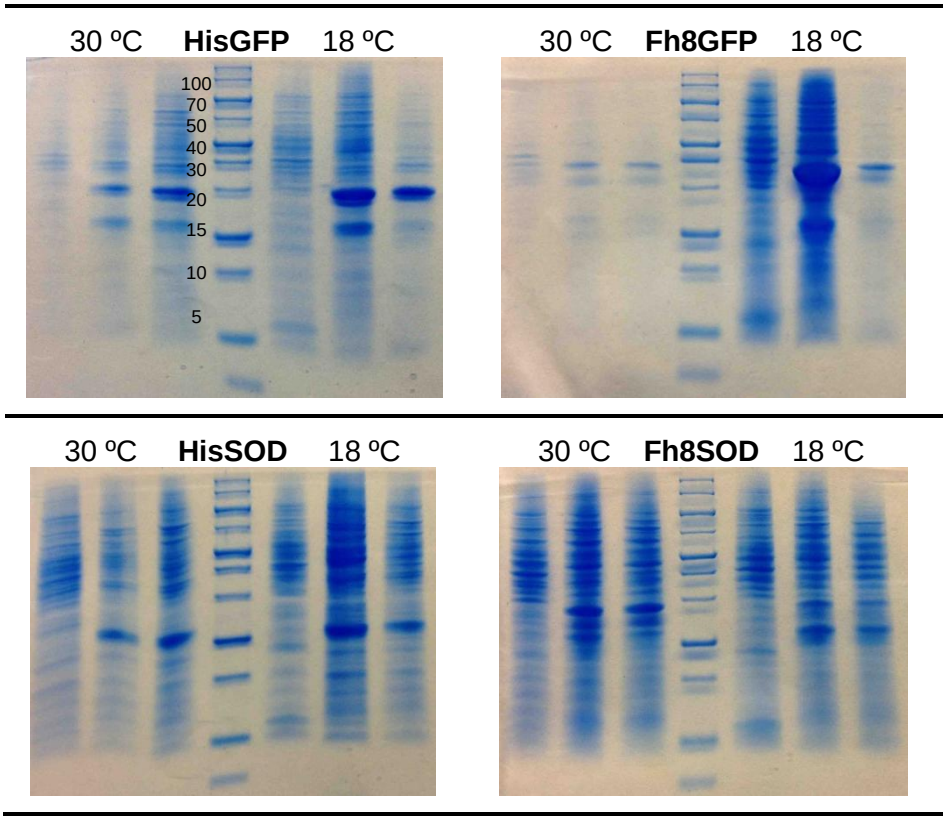


Figure A4.1. Solubility small scale screening evaluation in different *E. coli* strains by SDS-PAGE Gels were loaded as follows: supernatant samples expressed in *E. coli* Tuner, in BL21 Codon Plus-RIL and in Rosetta strains. The left side of the protein marker corresponds to proteins from the induction at 30 ºC, 0.5 mM IPTG for 3 hours. The right side of the protein marker presents proteins from the overnight induction at 18 ºC and 0.2 mM IPTG.

4.7.2. Purification strategy

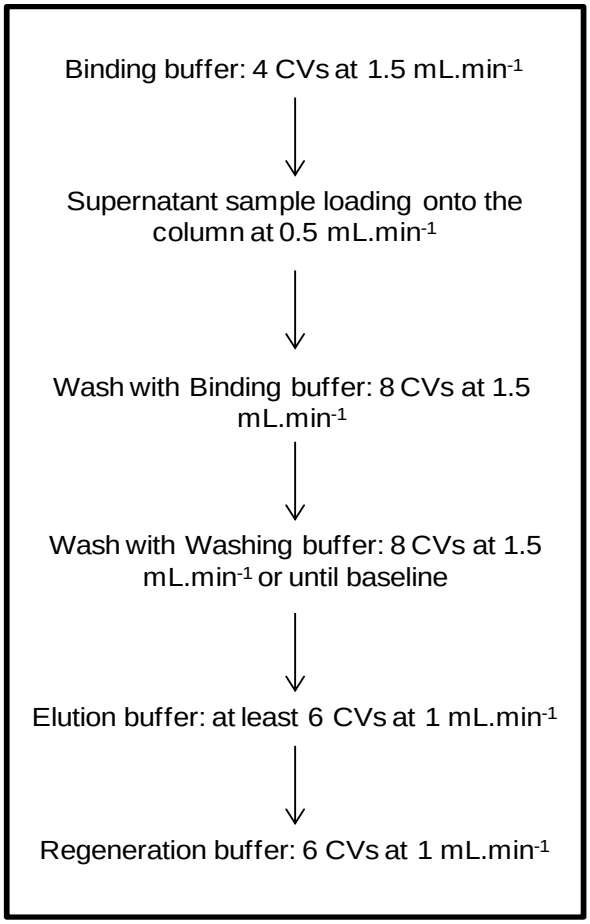


Figure A4.2. Purification strategy conducted in this work.

4.7.3. Protein molecular weights

Table A4.1. Protein molecular weights estimated by the Expasy ProtParam tool and by densitometry of SDS-PAGE (values are "average±standard deviation" from three independent analyses of the same gel):

Protein	MW by ProtParam (kDa)	Densitometry (kDa)		
		HIC	IMAC	all
Fh8 tag	12	12 ± 0.61 Ea: 23 ± 0.56 Eb: 54 ± 2.1	-	-
Fh8GFP	37	38 ± 0.90	40 ± 1.3	39 ± 1.4
HisGFP	30	29 ± 0.57	30 ± 1.6	30 ± 1.3
Fh8SOD	26	30 ± 0.59	28 ± 1.5	29 ± 1.4
HisSOD	19	21 ± 0.42	21 ± 0.56	21 ± 0.50

Ea: 2nd gel band (of three) observed in the elution; Eb: 3rd gel band (of three) observed in the elution.

4.7.4. GFP purification samples

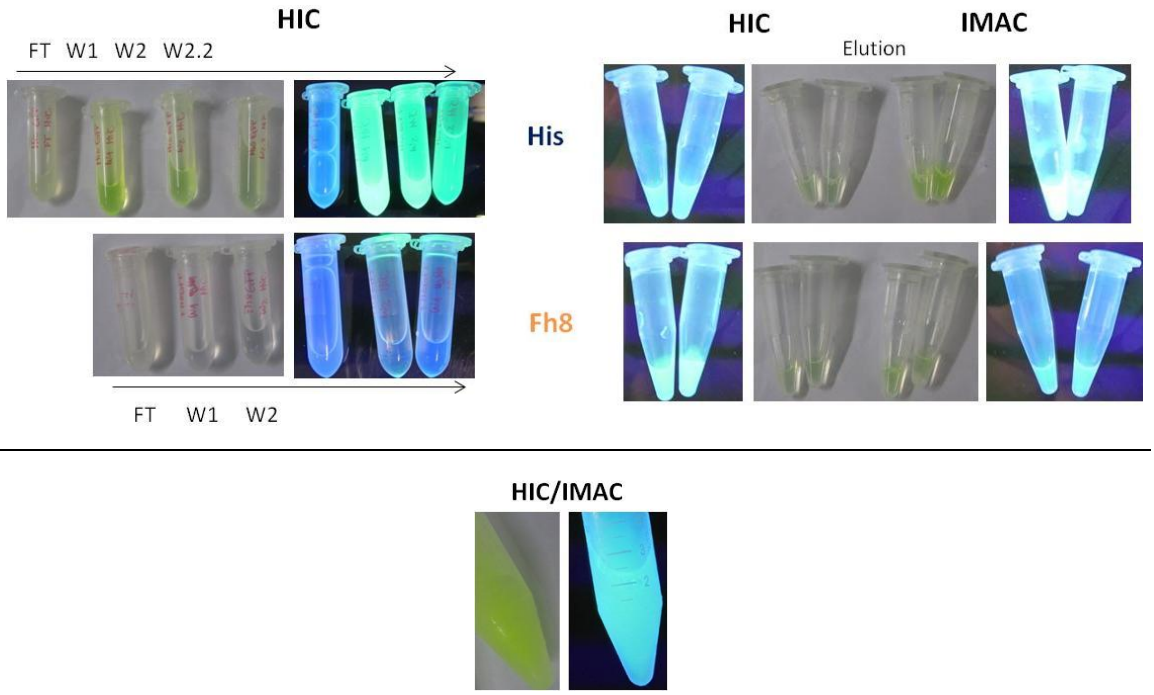


Figure A4.3. Photos of GFP purification samples at UV light.

Chapter 5

Soluble expression in *Escherichia coli* of difficult-to-express recombinant bone morphogenetic protein-2 and interleukin-10 proteins by fusion to the Fh8 solubility partner

Abstract

Bone morphogenetic proteins and interleukins are important components of the immune system, being crucial for bone formation and repair, and in host defence, respectively. Among these proteins, the bone morphogenetic protein-2 (BMP-2) and the interleukin-10 (IL-10) have been intensively studied, being usually produced in recombinant hosts such as *Escherichia coli*. Both BMP-2 and IL-10 are produced in this bacterial cell as insoluble inclusion bodies, requiring several time-consuming solubilisation and refolding steps to acquire their biologically active structure.

In this work, a novel strategy for BMP-2 and IL-10 soluble *E. coli* expression and purification is presented by using the new Fh8 tag. Both proteins are directly expressed in the soluble form as Fh8BMP-2 and Fh8IL-10 fusion proteins in *E. coli*, in contrast to the His_6-tagged proteins that are insoluble in this host cell. The novel purification strategy via Fh8 hydrophobic interaction allowed the successful recovery of both fusion proteins, while some steric hindrance was observed during their nickel affinity purification. Moreover, Fh8BMP-2 and Fh8IL-10 purified fusion proteins presented secondary structures similar to those reported for BMP-2 and IL-10 proteins, and were obtained in dimeric and oligomeric forms, which are essential for their biological activity.

In spite of their soluble production, some issues were found in the immunodetection of purified proteins that, together with the biological inactivity of Fh8BMP-2 and difficulties in the tag removal, suggest a structural complexity of the soluble Fh8-fused proteins, which must be circumvented.

5.1. Introduction

Bone morphogenetic proteins and interleukins play important and mutual roles in the immune system, and their potential application in diverse medical areas has been growing over the years (Asadullah et al., 2003; Bessa et al., 2008b; Kwan, 2011).

Bone morphogenetic proteins (BMPs) belong to the transforming growth factor-β (TGF-β) superfamily, and are identified by their ability to induce bone and cartilage formation (Urist, 1965; Szpalski and Gunzburg, 2005; Bessa et al., 2008a; Kwan, 2011). Biologically active BMPs are homo or heterodimers, which chains are connected via disulfide bonds (Israel et al., 1996; Scheufler et al., 1999).

To date, at least fifteen BMPs have been identified and characterized, and among them, the bone morphogenetic protein-2 (BMP-2) protein has received a lot of attention due to its osteoinduction properties during skeletal development and repair. BMP-2 is localized in bone tissue and is released in response to bone damage, stimulating differentiation, via the Smad pathway, of pluripotent mesenchymal stem cells into multiple cell lineages, such as bone, cartilage, muscle, or fat cells (Shea et al., 2003; Yu et al., 2010; Kwan, 2011). BMP-2 is one of the two bone morphogenetic proteins approved by the Food and Drug Administration (FDA) as biological method to stimulate bone repair (Sharapova et al., 2010; von Einem et al., 2010; Oliveira et al., 2011), and it has been studied as an alternative to autologous bone grafting in many clinical situations, including spinal fusion, osteoporosis, treatment of bone defects, non-union fractures and root canal surgery (Azari et al., 2001; Bessa et al., 2008a). In terms of structure, human BMP-2 consists of a long precursor protein of 396 amino acids, which is glycosylated, proteolytically cleaved and dimerized to form the mature homodimeric protein consisting of two 114 residue subunits (Scheufler et al., 1999; Hillger et al., 2005).

The biological activity of BMP-2 has been reported by the ability to promote the differentiation of mesenchymal stem cells into bone, involving different strategies, such as, the study of phenotypic characteristics of skeletal cells, the evaluation of alkaline phosphatase activity and extracellular matrix mineralization, the expression of various extracellular matrix proteins (collagen type II or osteocalcin), and the terminal phenotypic markers of the differentiated state of these cells (Shea et al., 2003; Vallejo

and Rinas, 2004a; Bessa et al., 2008c; Ihm et al., 2008; Sharapova et al., 2010; von Einem et al., 2010; Yu et al., 2010; Zhang et al., 2011).

Interleukin-10 (IL-10) is a pleiotropic cytokine that exerts potent effects on numerous cell populations, in particular, circulating and resident immune cells as well as epithelial cells, by its specific cell surface receptor complex (IL-10R). IL-10 is, thus, an important broad effector molecule in immunoregulation and host defence (Moore et al., 2001; Pestka et al., 2004). Taking into account its anti-inflammatory and immunostimulatory properties, IL-10 has been proposed to be used in several clinical applications, such as vaccination (Berzofsky et al., 2001) or treatment of allergies (Pullerits, 2002), infectious diseases (Hubel et al., 2002), acute and chronic inflammatory diseases (Asadullah et al., 2003). Human IL-10 is a homodimer with a molecular mass of 37 kDa and each monomer consists of 160 amino acids with a molecular mass of 18.5 kDa. Murine IL-10 has about 80% homology with human IL-10 (Asadullah et al., 2003) and it consists of a polypeptide chain of 157 amino acids.

Both BMP-2 and IL-10 proteins can be obtained from their corresponding natural sources, and several commercial solutions are now available, but the high cost, time-consuming, laborious protein purification, and low yields limit their therapeutic use. Therefore, several strategies have been applied to produce BMP-2 and IL-10 recombinantly, including the use of mammalian cell cultures and insect cells, but prokaryotic expression systems are still the best choice due to its advantageous high yield, low cost cultivation/production and high bio-safety (Bessho et al., 1999; Asadullah et al., 2003; Bessa et al., 2008b; Klompus et al., 2008; Carvalho et al., 2010).

BMP-2 and IL-10 have been produced in *Escherichia coli* as insoluble inclusion bodies (Long et al., 2006; Zhang et al., 2006; Bessa et al., 2008c; Klompus et al., 2008; Carvalho et al., 2010; Sharapova et al., 2010; Zhang et al., 2010). Despite all the optimizations conducted so far, this production methodology is disadvantageous as it requires solubilisation and refolding steps that are time-consuming and limit their production at large scale (Makrides, 1996; Terpe, 2006; Demain and Vaishnav, 2009).

Taking into account the limitations found in the production of BMP-2 and IL-10, a novel strategy for their soluble *E. coli* expression and purification is presented in this work by using the new Fh8 tag (Chapters 2 and 4) (Costa et al., 2012). Both BMP-2 and IL-10 are fused to the Fh8 tag, and their soluble expression is compared to that of His_6-fused proteins. The secondary structure of Fh8BMP-2 and Fh8IL-10 is also reported. In

addition, the *in vitro* bioactivity of soluble Fh8BMP-2 fusion protein is further explored using C2C12 myoblast cell cultures.

5.2. Materials and Methods

5.2.1. Materials

In this work, all the cloning PCRs used the Phusion High-Fidelity DNA Polymerase (Thermo Scientific), according to manufacturer's instructions. The colony PCRs were conducted using the Taq Polymerase (Nzytech) with an annealing temperature of 55 ºC and T7 forward and reverse universal primers. Plasmid DNA extractions were performed using the GenElute™ plasmid miniprep kit (Sigma) for minipreps, and the QIAquick DNA gel extraction kit or QIAquick PCR purification kit (Qiagen) were used for DNA purification. The restriction enzymes used in this work were from New England Biolabs. The DNA ligations were carried out with the T4 DNA ligase (Promega) or with the Rapid DNA Ligation kit (Roche). Kanamycin and chloramphenicol antibiotics were used for plasmid maintenance and protein expression at a final concentration of 50 $\mu g.mL^{-1}$ and 10 $\mu g.mL^{-1}$, respectively. Antibiotic stock solutions were prepared at 1000x, filtered through 0.2 µm and store at -20 ºC.

5.2.2. Cloning of *bmp-2* and *il-10* genes into the pETM11 and pETMFh8 vectors

The *bmp-2* (Swiss Prot ID: P12643.1, Bioclone Inc cat# RPA0305) and *il-10* genes (Carvalho, 2010) used in this work are synthetic genes previously optimized for *E. coli* expression. The cloning of these genes into pETM11 (EMBL) and pETMFh8 plasmids (Chapter 2) (Costa et al., 2012) followed an identical strategy, in which both sequences were modified and amplified by PCR in order to have the *Nco*I/*Bam*HI cleavage site at the 5' position and the *Xho*I cleavage site at the 3' position. An annealing temperature of 68 ºC and 58 ºC was used to modify and amplify the *bmp-2* and *il-10* genes, respectively. Primers used in both PCRs are listed at Table 5.1.

Table 5.1. Primers used in *bmp-2* and *il-10* PCRs:

Primer	Sequence	Comments
BMP-2_*fwd*	5'-TCTATT**CCATGG***GATCC*ACTTTCGGC CACGATGGTAAAGG-3'	the *Nco*I restriction site is in bold, the *Bam*HI restriction site in italic and the first 23 nucleotides of the *bmp-2* gene underlined;
BMP-2_*rv*	5'-AATAGA**CTCGAG***CTA*GCGACAGCC ACAACCCTCCACAAC-3´	the *Xho*I restriction site in bold, the stop codon in italic and the final 24 nucleotides of the *bmp-2* gene underlined;
IL-10_*fwd*	5´-TCTATT**CCATGG***GATCC*TCTCGTGG CCAGTACTCTC-3´	the *Nco*I restriction site is in bold, the *Bam*HI restriction site in italic and the first 23 nucleotides of the *il-10* gene underlined;
IL-10_*rv*	5'-AATAGA**CTCGAG***CTA*GCTTTTCATTT TGATCATCATG-3'	the *Xho*I restriction site in bold, the stop codon in italic and the final 24 nucleotides of the *il-10* gene underlined.

fwd – forward; *rv* – reverse

The resulting PCR products were purified and digested with *Nco*I and *Xho*I restriction enzymes. DNA ligation was conducted at a molar ratio of 1:3 using the *Nco*I-*Xho*I digested PCR products and the *Nco*I-*Xho*I digested pETM11 and pETMFh8 vectors. *E. coli* TOP10 competent cells were transformed with the obtained vectors and the resulting clones were verified by colony PCR. The insertion of *bmp-2* and *il-10* genes into the pETM11 and pETMFh8 expression vectors was additionally confirmed by sequencing with both T7 forward and reverse universal primers.

5.2.3. Recombinant protein soluble expression and extraction in *E. coli*

The expression of recombinant Fh8BMP-2 and Fh8IL-10 proteins was conducted using the *E. coli* BL21 (DE3) Codon Plus-RIL strain that was selected from a small-scale screening as the best host for soluble expression of these proteins (data not shown). The same strain was also used for the expression of BMP-2 and IL-10 proteins without the Fh8 tag.

Each fusion protein was expressed using different culture volumes, as follows: BMP-2 proteins were expressed in 100-mL Erlenmeyers containing 20 mL cultures, and in several 1-L flasks containing 250 mL cultures; IL-10 proteins were expressed in 250-mL, and 1-L flasks containing 50 mL, and 250 mL cultures, respectively. All precultures were grown in LB media containing kanamycin and chloramphenicol antibiotics, at 37 ºC, overnight. Precultures were diluted 1:50 for inoculation of LB media and cells were grown at 37 ºC, 200 rpm, to a final $OD_{600\ nm}$ of 0.4-0.6. The induction of protein expression occurred at 18 ºC, o/n, with 0.2 mM isopropyl-β-D-1-thiogalactopyranoside *(*IPTG).

After induction, cells were harvested and cell pellets were stored at -20 ºC. For the protein extraction from *E. coli* cells, bacterial pellets were resuspended in lysis buffer [50 mM Tris pH 7.4, 250 mM (for BMP-2 proteins) or 150 mM (for IL-10 proteins) NaCl buffer with 1x complete free EDTA protease inhibitor (Roche), 5 mM $MgCl_2$ (Sigma), 5 µg.mL^{-1} DNAse (Sigma) and 1 mg.mL^{-1} lysozyme (Sigma)], and incubated at room temperature for 10 minutes prior sonication.

Total lysate samples were collected after cell lysis and the supernatant fractions were recovered by centrifugation at 10000 rpm, 30 minutes, 4 ºC. Aliquots of soluble and insoluble fractions as well as total lysates were prepared for sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot analyses.

5.2.4. IMAC protein purification

Supernatant samples were resuspended as mentioned above, supplemented with 20 mM imidazole, and centrifuged for 10 minutes at 10000 rpm, 4 ºC. Two purification protocols were then conducted using Ni-NTA beads (GE Healthcare) for small-scale spin purification, and a prepacked Histrap column (GE Healthcare) for purification of large sample volumes in a manual pump system. Supernatant fractions were filtered through 0.45 µm before each purification. Both protocols were performed according to manufacturer's instructions, using 50 mM Tris pH 7.4, 250/150 mM NaCl buffer in all purification steps, supplemented with 20 mM imidazole (for the binding step), with 40 mM imidazole (for the washing step), and with 300 mM imidazole (for the elution step). Aliquots of flow-through, washing and eluted samples were prepared and analysed by SDS-PAGE. Protein content from the eluted samples was estimated by Bradford assay.

5.2.5. Fh8/HIC protein purification

Fh8BMP-2 and Fh8IL-10 cell pellets were resuspended in lysis buffer supplemented with 5 mM $CaCl_2$ and incubated at room temperature for 10 minutes prior sonication. Total lysate samples were collected after cell lysis and the supernatant fractions were recovered by centrifugation as mentioned above. Supernatant fractions were filtered through 0.45 µm and applied onto a pre-equilibrated Phenyl Sepharose™ 6 Fast Flow (high substitution) column (GE Healthcare) to be purified using the Fh8 as purification tag via hydrophobic interaction chromatography (HIC), in a manual pump system. An identical protocol was performed for both fusion proteins using the 50 mM Tris pH 7.4, 250/150 mM NaCl, 5 mM $CaCl_2$ buffer as binding buffer. Two washing steps were conducted: the first with the binding buffer diluted 1:2, and the second with distilled water. Fh8BMP-2 and Fh8IL-10 fusion proteins were eluted from the HIC resin using 50 mM Tris at pH 10 as elution buffer. Aliquots of flow-through, washing and eluted samples were prepared and analyzed by SDS-PAGE. Protein content from the eluted samples was estimated by Bradford assay. Purified Fh8BMP-2 and Fh8IL-10 fusion proteins were finally dialyzed in phosphate buffer saline (PBS) 1x, pH 7.4, filtered through 0.2 µm, and stored at -20 ºC. Aliquots of purified samples were also used to estimate protein concentration by absorbance at 280 nm.

5.2.6. Protein electrophoresis and quantification

SDS-PAGE of Fh8 and His fusion proteins was conducted according to the Laemmli method (Laemmli, 1970) using 12%-4% gels. The total protein content of supernatant samples and purification samples was estimated by Bradford method (Bradford, 1976), using the Biorad protein assay dye reagent and bovine serum albumin as standard. Protein quantifications were also conducted by reading the absorbance of eluted samples at 280 nm.

5.2.7. Immunodetection of Fh8BMP-2 and Fh8IL-10 soluble fusion proteins

After SDS-PAGE, supernatant and purified samples of Fh8BMP-2 and Fh8IL-10 fusion proteins were transferred to nitrocellulose membranes using a sandwich system. The

immunodetection of BMP-2 was conducted with the in-house polyclonal mouse antibody anti-BMP-2 as primary antibody, and anti-mouse IgG HRP (Biorad) together with protein G HRP (Biorad) as conjugates. Anti-BMP-2 polyclonal antibodies were produced using the H partner as described in Chapter 3. The IL-10 analysis was carried out with the biotinylated anti-mouse IL-10 monoclonal antibody and streptavidin-horseradish peroxidase conjugate (BD Biosciences). The substrate 4-chloro-naftol (Sigma) was used for membrane revelation.

5.2.8. DLS and CD

Dynamic Light Scattering (DLS) measurements were conducted in the Zetasizer Nano ZS (Malvern Instruments), using a protein concentration between 0.5-1 $mg.mL^{-1}$. Protein samples were centrifuged and filtered before measurements to remove any aggregates.

The circular dichroism (CD) spectra of purified recombinant proteins (30 µg of Fh8BMP-2 or 60 µg of Fh8IL-10 fusion proteins in 10 mM sodium phosphate buffer, pH 8.0) were collected with a 1 mm path length cuvette, between 190 and 260 nm, set up to 1 nm band width, continuous scan mode at 200 nm/min. The presented spectra are the average of 3 scans with the average buffer control spectrum subtracted. Spectra were acquired in a J-815 circular dichroism spectropolarimeter (Jasco), at 20 ºC. The results are expressed in terms of molar ellipticity (mean residue ellipticity, θ) in $deg.cm^2.dmol^{-1}$, according to the equation: $(\theta) = \frac{\theta obs}{10 \times Cr \times l}$, where θ_{obs} is the observed ellipticity in mdeg, C_r is the mean residue molar concentration (mol/L), and l is the cuvette path length in cm. The mean residue molar concentration is obtained as follows: $Cr = \frac{1000 \times n \times C'}{MW}$, where n is the number of residues of the protein, C' is the protein concentration (g/mL), and MW is the protein molecular mass (g/mol).

5.2.9. Recombinant Fh8BMP-2 biological activity test

The biological activity of the purified Fh8BMP-2 was evaluated by the induction of alkaline phosphatase (ALP) activity in C2C12 cells, at 10^4 $cells.cm^{-2}$ in a 48-well plate, as described elsewhere (Abarrategi et al., 2012) with the following modifications: after cell seeding in Dulbecco's modified Eagle's medium supplemented with 10% fetal

bovin serum (FBS), recombinant Fh8BMP-2 was added at 0, 1, 2, 5, and 10 µg.mL^{-1} and incubated for four days. Culture medium was removed at the end of incubation period and wells were washed with PBS (400 µL per well). After that, lysis buffer (50 mM Tris pH 6.8, 0.1% Triton X-100, 2 mM $MgCl_2$) was added to each well and three cycles of temperature (-80 /37 ºC) were carried out for 30 minutes each.

ALP activity was determined using p-nitrophenyl phosphate dissolved in alkaline buffer solution (Sigma) as a substrate. The reaction was stopped with 0.5 M NaOH and the absorbance was measured at 405 nm on a Microplate Reader (Biotek FL-600). Protein content of cell lysate was estimated by Bradford assay, according to the manufacturer's instructions. ALP results are presented as Relative Fluorescence Units (RFU) per µg of Fh8BMP-2 protein.

5.2.10. Tag removal of Fh8BMP-2

The Fh8 tag removal from the Fh8BMP-2 fusion protein was conducted with the recombinant HisTEV protease (EMBL). Three different buffer incubation conditions were initially tested: (A) 50 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA buffer; (B) the same buffer supplemented with 0.1 % Triton X100; and (C) the same buffer supplemented with 0.1 % Tween 20. In this analysis, 150 µg of protein were used, and HisTEV (1 mg.mL^{-1}) was added at different concentrations (diluted 1:50, 1:20, and 1:10), followed by an o/n incubation at 4 ºC. After selecting the best buffer and protease concentration, 15 mg of Fh8BMP-2 protein were digested as referred before. The efficiency of the Fh8 tag removal from Fh8BMP-2 was assessed by SDS-PAGE.

The cleaved BMP-2 was purified further from the Fh8 tag and HisTEV protease by IMAC followed by ion exchange chromatography (IEX), performing both purifications according to the manufacturer's instructions. The IMAC purification was conducted in a 5-mL prepacked Histrap column (GE Healthcare), using the 50 mM Tris pH 7.6, 250 mM NaCl buffer supplemented with 20 mM imidazole for the binding step, supplemented with 40 mM imidazole for the washing step, and supplemented with 300 mM imidazole for the elution step. In IEX purification, a 1-mL prepacked Hitrap SP column (GE Healthcare) was used with the 50 mM Tris pH 7.0 as binding buffer. The same buffer supplemented with a gradient of NaCl 1 M was used for washing and elution of cleaved BMP-2. The final purity of cleaved BMP-2 was evaluated by SDS-PAGE and samples stored at -20 ºC.

5.3. Results

5.3.1. Expression of soluble BMP-2 and IL-10 as Fh8 fusion proteins

BMP-2 and IL-10 codifying genes were successfully cloned into the pETM11 (EMBL) and pETMFh8 (Chapter 2) (Costa et al., 2012) plasmids, as confirmed by sequencing. The obtained plasmids have the His_6 tag or the His_6-Fh8 tags at the N-terminal, and a Tobacco Etch Virus (TEV) protease recognition site is placed between fusion tags and BMP-2 and IL-10. At the nucleotide level, the obtained sequences presented 89% identity with the *Homo sapiens* BMP-2 (GenBank: GQ335530.1), and 92% identity with the *Mus musculus* IL-10 (GenBank: NM_010548.2). The obtained amino acid sequences presented 100% identity with the BMP-2 from *Homo sapiens* (GenBank: EAX10386.1), and with the IL-10 from *Mus musculus* (GenBank: AAI20613.1).

The soluble BMP-2 and IL-10 recombinant production in *E. coli* was evaluated using the Fh8 tag as solubility partner and the His_6 tag as purification partner. Figure 5.1 shows the small-scale solubility results obtained for these two proteins before and after purification by nickel affinity chromatography.

As observed in Figures 5.1.a and 5.1.b, recombinant BMP-2 was soluble expressed when fused to the Fh8 tag at a molecular weight of about 25 kDa, identical to the estimated by the ProtParam tool (Expasy.org). When only fused to the His_6 tag, the BMP-2 is highly expressed but not in the soluble form, as observed by the presence of a gel band at the estimated molecular weight of 18 kDa in lane 3 (total lysate sample) and the absence of a gel band of the same molecular weight in lane 4 (supernatant sample). The recombinant IL-10 was also soluble expressed as a Fh8-fused protein of approximately 29 kDa (Figures 5.1.c-d). Without the Fh8 tag, the recombinant HisIL-10 was highly expressed in the insoluble fraction with an estimated molecular weight of 22 kDa (lane 2 of Figure 5.1.c), and no soluble production was observed in supernatant and eluted samples from Ni-NTA purifications (lanes 3, 6 and 9 of Figures 5.1.c-d).

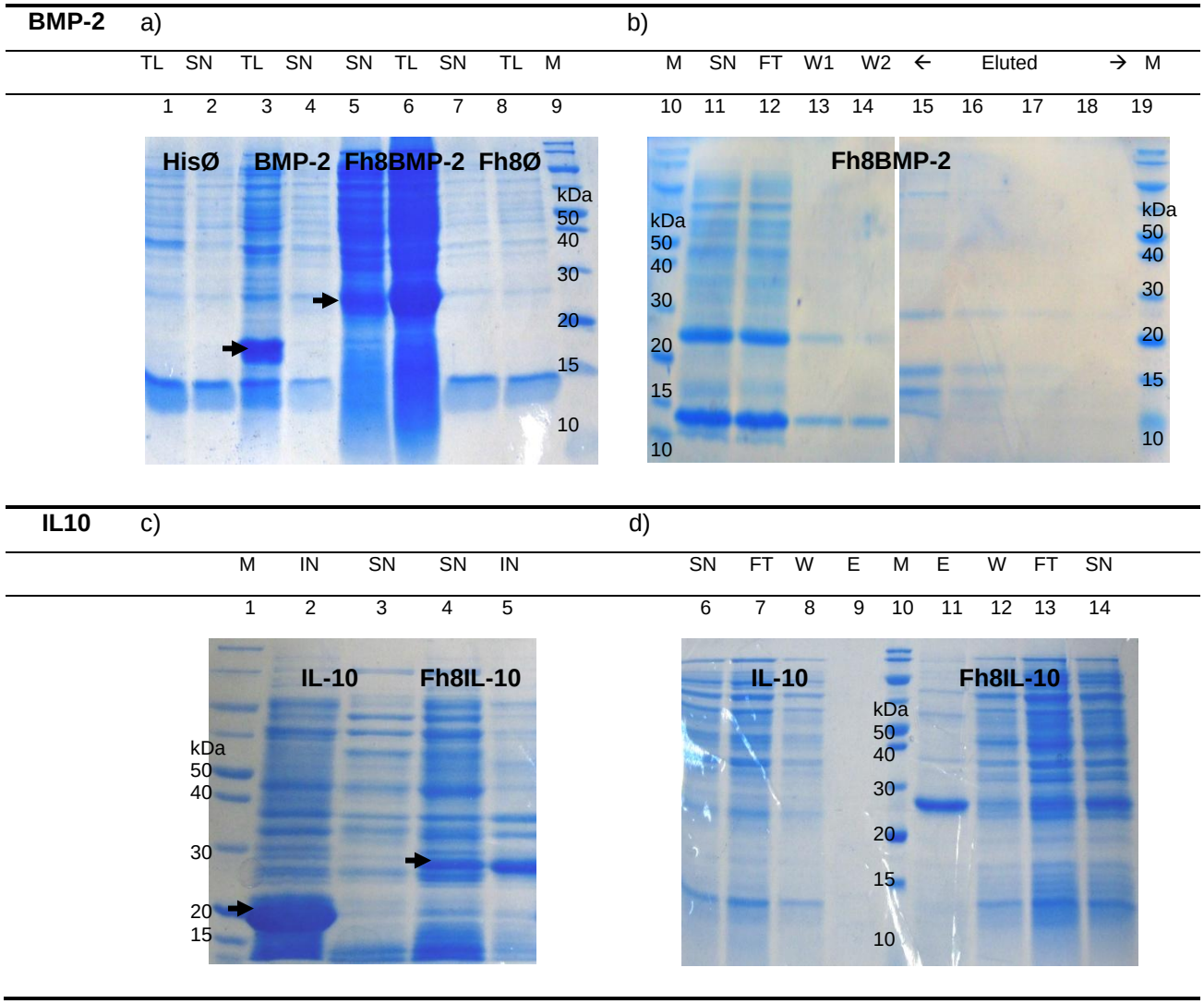


Figure 5.1. Evaluation of BMP-2 and IL-10 proteins solubility in small-scale cultures. (**a**) Comparison of total cell extract and soluble fraction of the *E. coli* harboring pETM11BMP-2, pETMFh8BMP-2 and corresponding pETM11 and pETMFh8 empty plasmids. (**b**) Ni-NTA pool downs of Fh8BMP-2. (**c**) Comparison of insoluble and soluble fractions of the *E. coli* harboring pETM11IL-10 and pETMFh8IL-10. (**d**) Ni-NTA pool downs of IL-10 and Fh8IL-10. TL – total lysate fraction (cell extract); SN – supernatant (soluble) sample; IN – insoluble sample; FT – flow-through sample; W, W1 and W2 – washing samples; E and Eluted – Eluted samples from Ni-NTA pool downs. M – PageRuler broad unstained protein marker (Thermo Scientific).

5.3.2. Fh8BMP-2 and Fh8IL-10 proteins purification

E. coli 500 mL cultures producing soluble Fh8BMP-2 and Fh8IL-10 fusion proteins were further purified by nickel affinity chromatography (IMAC, Figure 5.2), and by hydrophobic interaction through the Fh8 tag (Fh8-HIC, Figure 5.3).

As observed in Figure 5.2, identical results were obtained for the two recombinant proteins: Fh8BMP-2 or Fh8IL-10 supernatant fractions loaded onto the nickel column (lane 1 – Figure 5.2.a, and lane 2 – Figure 5.2.b) were mainly collected at the flow-through (lane 2 – Figure 5.2.a, and lane 3 – Figure 5.2.b) and washing (lanes 3 and 4 – Figure 5.2.a, and lanes 4 and 5 – Figure 5.2.b) samples, presenting low affinity for the nickel resin. This weak interaction revealed a low protein amount in the eluted samples, as observed in lanes 6-11 and 7-9 of Fh8BMP-2 and Fh8IL-10, respectively. Several optimizations of the IMAC protocol were conducted to improve the Fh8BMP-2 and Fh8IL-10 purification by testing different imidazole concentrations and buffer compositions, but no positive results were obtained (data not shown).

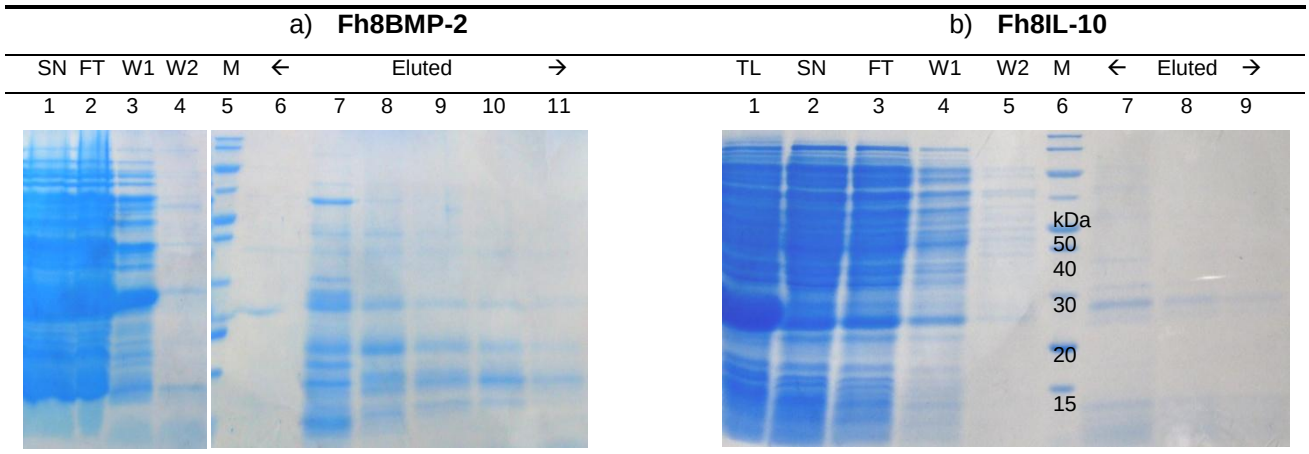


Figure 5.2. Fusion protein purification from the *E. coli* extract of a 500 mL culture by IMAC: (**a**) Fh8BMP-2 purification. (**b**) Fh8IL-10 purification. In both SDS-PAGE: SN – supernatant samples, FT – flow-through samples, W1 and W2 – washing samples, Eluted – eluted samples, M – PageRuler broad unstained protein marker (Thermo Scientific).

Taking into account the low purification efficiency achieved with nickel affinity, a novel strategy for Fh8BMP-2 and Fh8IL-10 proteins purification was performed, using the Fh8 tag as purification handle via HIC. When purified by this methodology, both fusion proteins interacted with the hydrophobic resin, presenting higher protein amounts in the eluted samples than by IMAC (Figure 5.3).

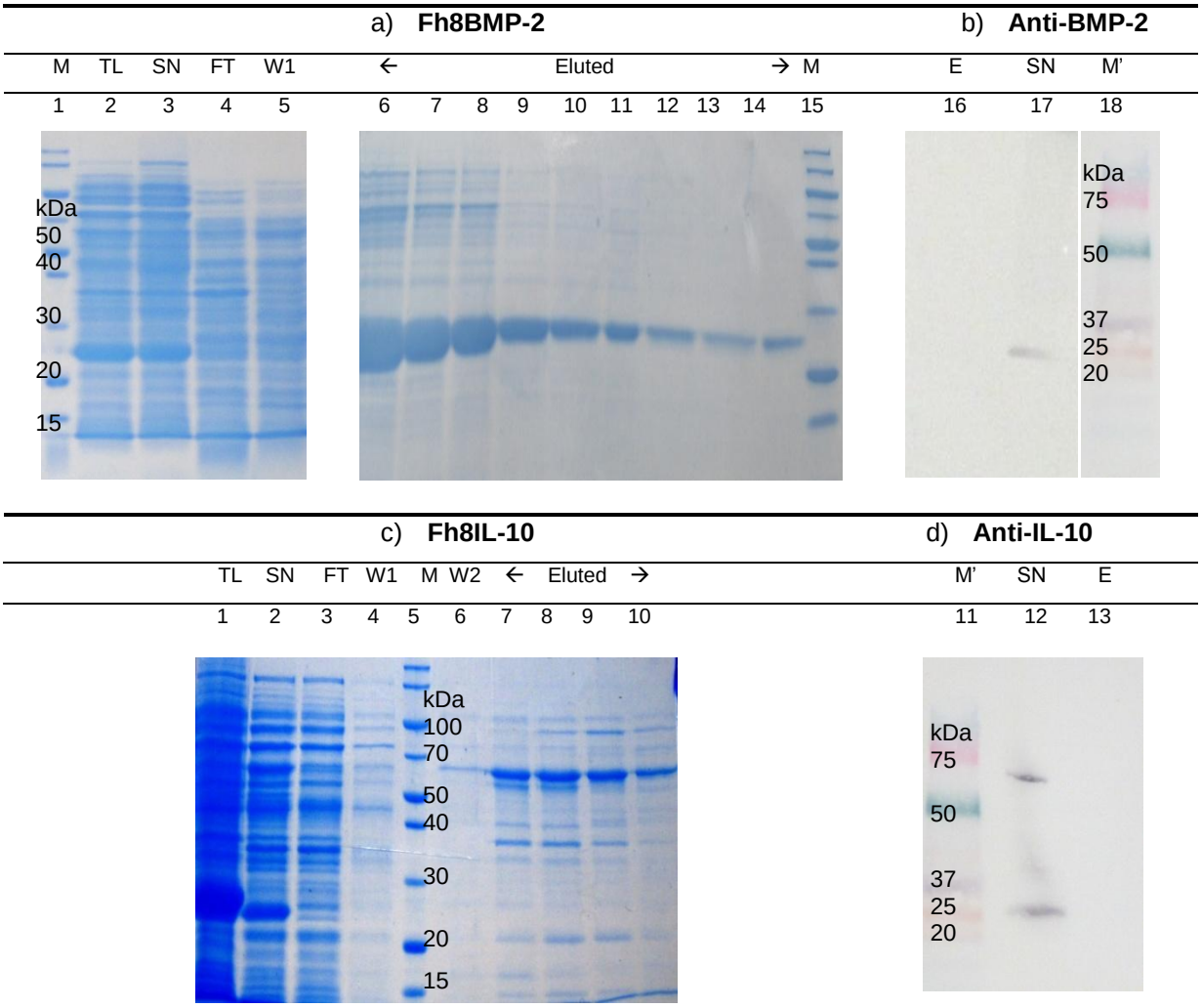


Figure 5.3. Fusion protein purification from the *E. coli* extract of a 500 mL culture by Fh8-HIC: (**a**) Fh8BMP-2 purification. (**b**) Western blot detection of supernatant sample of Fh8BMP-2. (**c**) Fh8IL-10 purification. (**d**) Western blot detection of supernatant sample of Fh8BMP-2. In all SDS-PAGE: TL – total lysate samples diluted 1:2, SN – supernatant samples diluted 1:2, FT – flow-through samples diluted 1:2, W1 and W2 – washing samples diluted 1:2, Eluted – eluted samples. M – PageRuler broad unstained protein marker (Thermo Scientific). M' – Precision Plus Protein Kaleidoscope standards (Bio Rad).

In the Fh8-HIC purification of Fh8BMP-2 protein (Figure 5.3.a), two gel bands of approximately 25 and 50 kDa were observed in the supernatant (lane 3), flow-through (lane 4) and washing (lane 5) samples. The Fh8BMP-2 eluted samples presented only the gel band of 25 kDa together with some protein contaminants from the *E. coli* extract (lanes 6-14, Figure 5.3.a).

The soluble Fh8BMP-2 loaded onto the HIC column was identified by Western blot but the purified fusion protein was, however, not detected (Figure 5.3.b). This soluble fusion protein was recovered with a final production yield of 31 mg per litre of *E. coli* culture. Eluted samples from lanes 6, 7 and 8 were purified further by a second Fh8-HIC methodology, improving the purity level (ratio between Fh8BMP-2 and other contaminants) of the obtained Fh8BMP-2 (data not shown).

The supernatant fraction of Fh8IL-10 protein for Fh8-HIC purification (lane 2, Figure 5.3.c) presented two gel bands of approximately 29 and 60 kDa that were detected by Western blot (Figure 5.3.d). The gel band of 60 kDa was mainly observed in the eluted Fh8IL-10 together with some protein contaminants from the *E. coli* extract (lanes 7-10, Figure 5.3.c). The eluted Fh8IL-10 was not identified by Western blot (Figure 5.3.d)

The soluble Fh8IL-10 was recovered from the HIC resin with a final production yield of, approximately, 6 mg per litre of *E. coli* culture.

5.3.3. Biophysicial characterization

The hydrodynamic radius of Fh8BMP-2 and Fh8IL-10 fusion proteins purified by Fh8/HIC and subsequent estimation of the molecular weight showed an oligomerization tendency of both fusion proteins. The Fh8BMP-2 resulted in a hydrodynamic radius of 3.43±0.0924 nm, corresponding to a globular protein of 60 kDa. As Fh8BMP-2 presents a molecular weight of 25 kDa in SDS-PAGE gel, the estimated molecular weight obtained in DLS analysis could correspond to a dimer of the Fh8BMP-2 molecule. The hydrodynamic radius measured in the Fh8IL-10 was 7.93±1.09 nm, estimating a globular protein with a molecular weight of 423 kDa. This molecular weight shows a high oligomerization state of Fh8IL-10.

Figure 5.4 presents the secondary structure analysis of both Fh8BMP-2 and Fh8IL-10 fusion proteins. The spectrum of Fh8BMP-2 exhibited one minimum at 213 nm, resembling the shape of a β-sheet predominant structure. The spectrum of Fh8IL-10 shows two minima at 209 and 219 nm, revealing a mainly alpha helical structure.

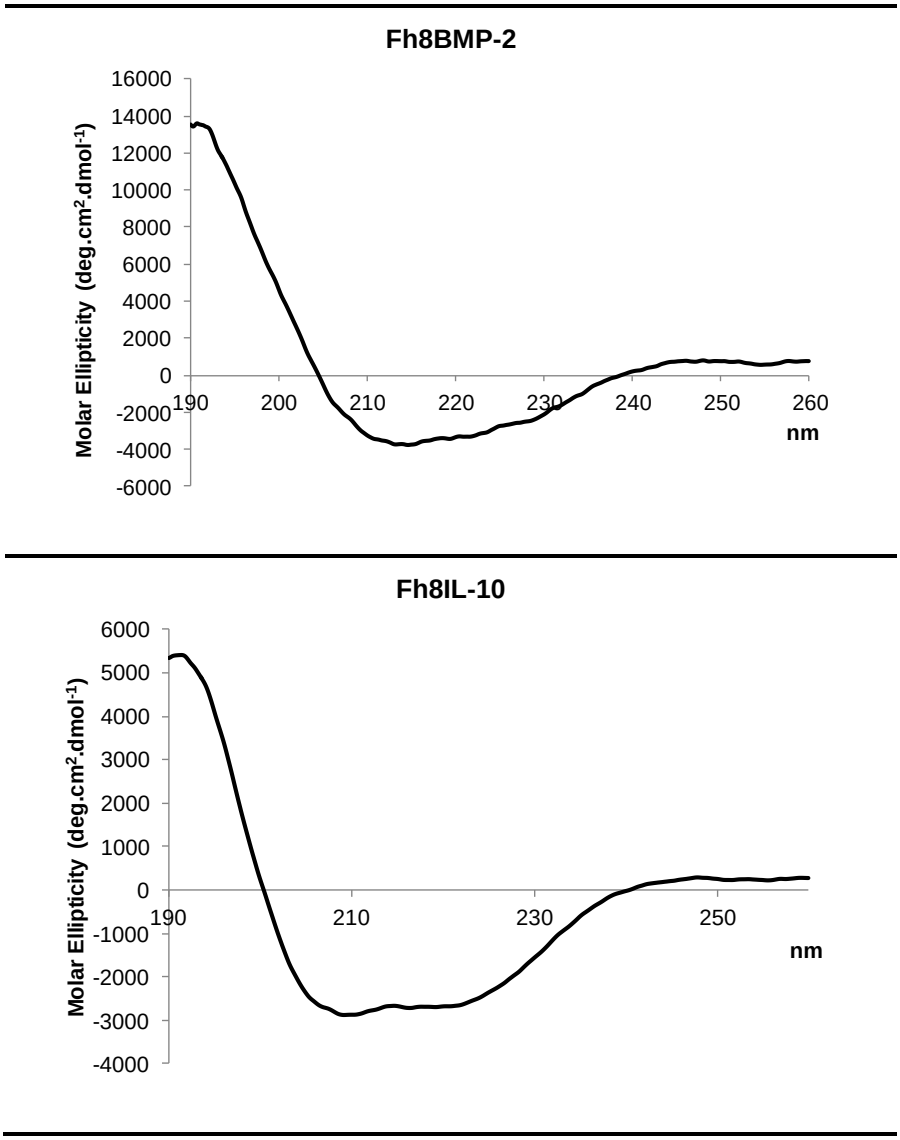


Figure 5.4. Secondary structure analysis by circular dichroism: spectra were measured between 190 and 260 nm, at 20 ºC, in a J-815 circular dichroism spectropolarimeter (Jasco), as described in the Material and Methods section.

5.3.4. Fh8BMP-2 bioactivity: evaluation of C2C12 myoblasts cells differentiation into osteoblasts

The alkaline phosphatase protein is a marker enzyme frequently used to measure the differentiation of diverse cells into osteoblasts (Zhang et al., 2010), and the C2C12 myoblast cells have been used in several differentiation studies with recombinant BMP-2 produced in *E. coli* (Bessa et al., 2008c; Ihm et al., 2008; Sharapova et al., 2010; von Einem et al., 2010).

Several *in vitro* experiments were performed to conclude about Fh8BMP-2 activity, testing different culture volumes, well-plates and protein amounts, but the same result was obtained for all the trials: the Fh8BMP-2 recombinant protein did not promote C2C12 cells for differentiation into osteoblasts, as observed by the low levels of alkaline phosphatase (ALP) per µg of Fh8BMP-2 protein (Figure 5.5.a, RFU – relative fluorescence units), and identical morphology of C2C12 cells cultured with or without Fh8BMP-2 fusion protein (Figure 5.5.b).

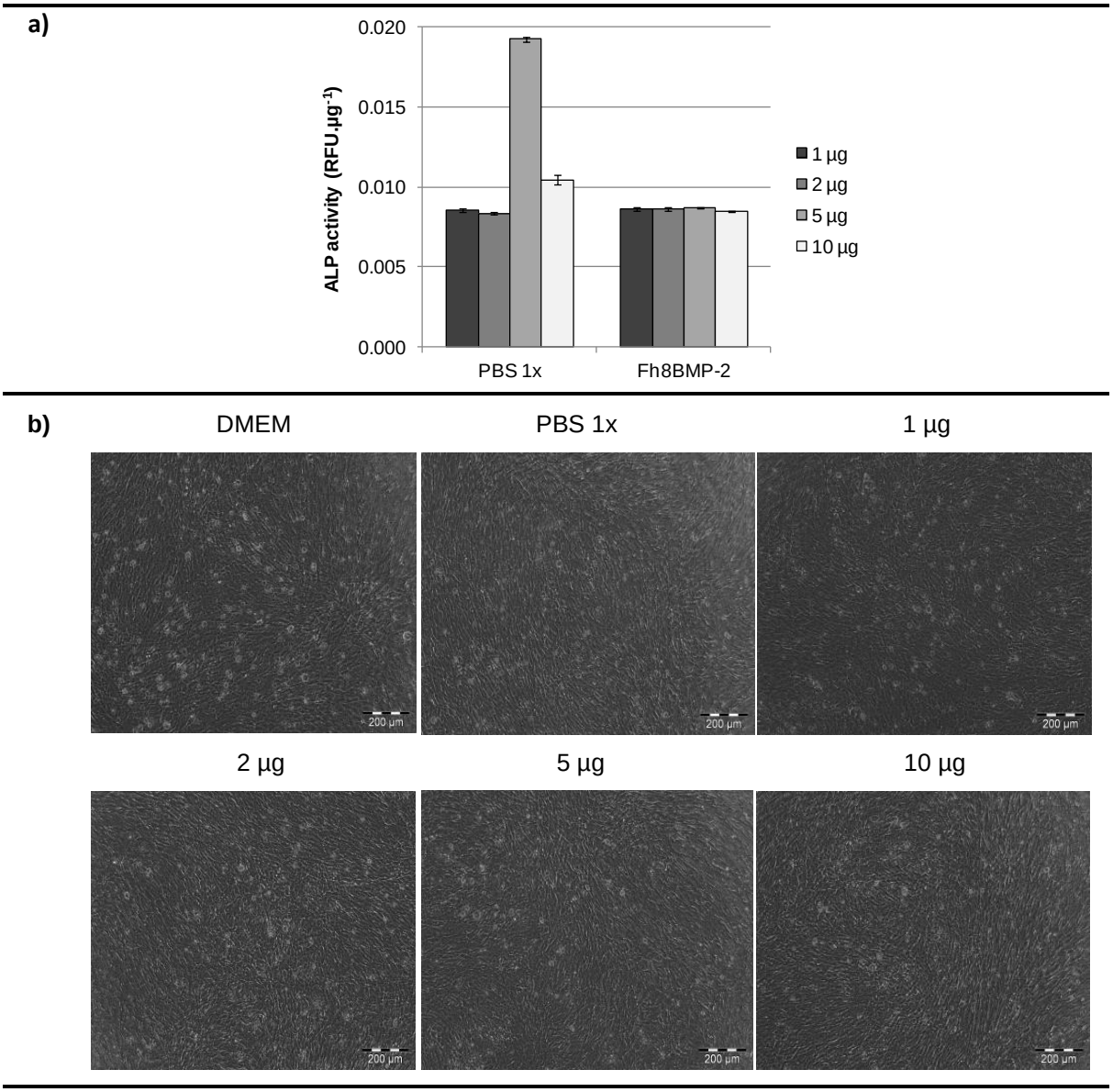


Figure 5.5. Biological activity of Fh8BMP-2. (**a**) Alkaline phosphatase (ALP) activity in C2C12 cells. (**b**) Comparison of C2C12 cell culture morphology between control wells (DMEM and PBS 1x), and Fh8BMP-2 wells, in which 1, 2, 5 and 10 µg of the purified fusion protein were added. Cell cultures were observed with 10x amplification in an Olympus BX51 microscope.

5.3.5. Fh8 tag removal and purification of cleaved BMP-2

Taking into account bioactivity results, the removal of Fh8 tag from the fusion protein was attempted using the TEV protease, which was placed between the fusion tag and the target proteins (Chapter 2) (Costa et al., 2012). An initial optimization of the cleavage buffer and protease concentration was performed (Figure 5.6.a), resulting in the selection of a TEV dilution of 1:20, and 50 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA buffer supplemented with 0.1% Tween 20 (buffer C). Even using these optimized conditions, the BMP-2 protein was removed from the fusion protein with low cleavage efficiency, as observed in Figure 5.6.b and 5.6.c.

In the IMAC purification (Figure 5.6.b), the BMP-2 cleaved protein was expected to be collected in the flow-through and washing samples, and the Fh8 tag and TEV protease were expected to be collected in the eluted samples. However, as observed in this figure, the flow-through contained the Fh8 tag, and washing samples contained both cleaved BMP-2 and Fh8BMP-2 proteins, resulting in a recovery of the cleaved protein of 18% and 20% in the washing samples (W1 and W2, estimated by densitometry in the Image Lab 2.0 software (Bio Rad), using the Molecular Imager Chemidoc XRS+ system (Bio Rad)).

The IEX purification was attempted to further purify the cleaved BMP-2 (Figure 5.6.c). As observed in this figure, the cleaved BMP-2 was not collected pure in the eluted sample 2 (E2), representing 42% of the total eluted content (estimated by densitometry). Taking into account that no pure BMP-2 was obtained after Fh8 tag removal, the bioactivity assay of cleaved BMP-2 was no longer conducted.

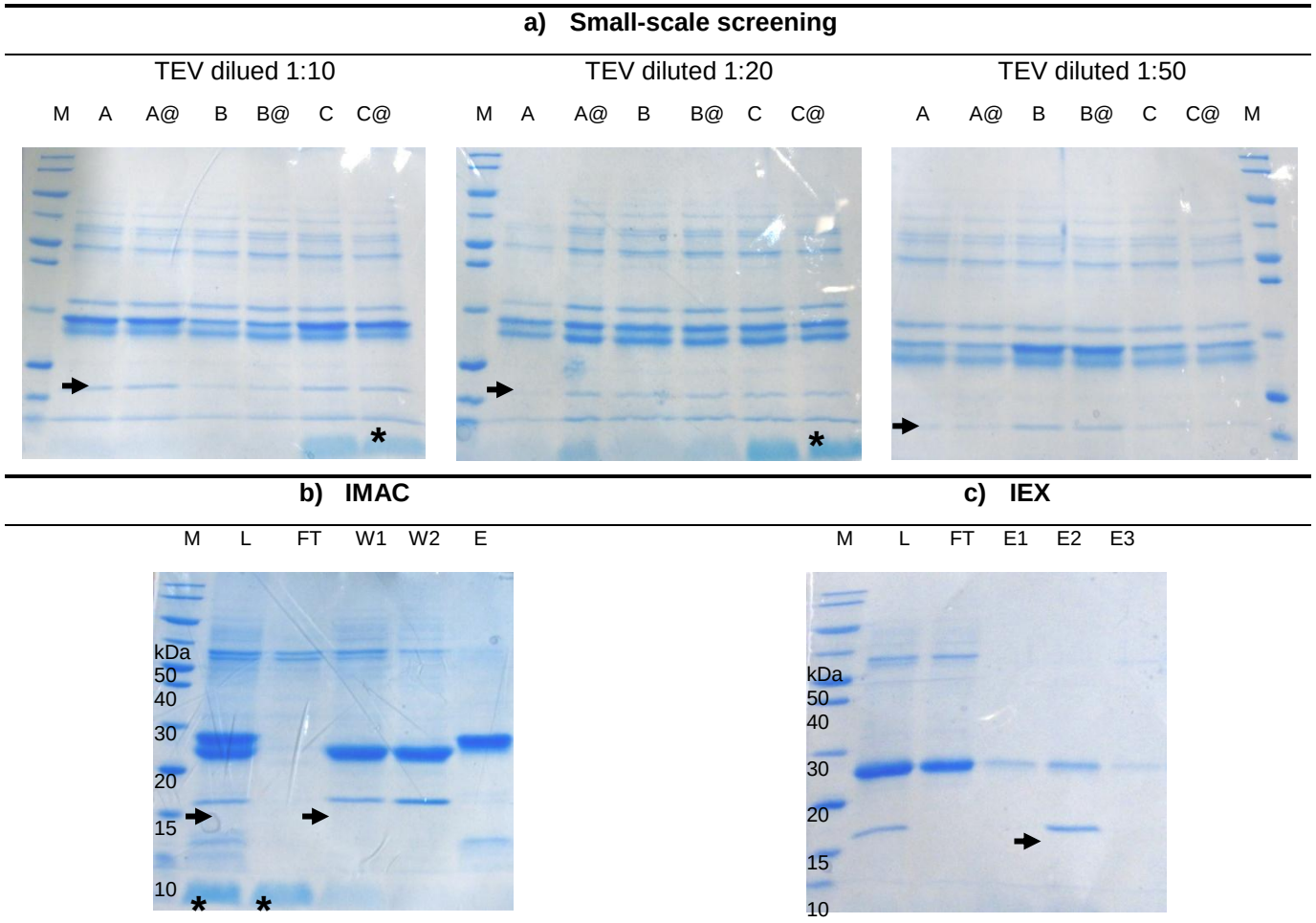


Figure 5.6. Tag removal from the Fh8BMP-2 fusion protein. (**a**) Small-scale screening of Fh8 tag removal using different TEV concentrations and buffers (see Material and Methods section). (**b**) IMAC purification of the cleaved BMP-2 protein. (**c**) IEX purification of the washing samples from IMAC purification. In all SDS-PAGE: M - PageRuler broad unstained protein marker (Thermo Scientific). A, B and C – TEV digested Fh8BMP-2 using different buffers, as referred in Material and Methods section. A@, B@ and C@ – Digested Fh8BMP-2 after centrifugation. L – loaded sample onto the IMAC or IEX column. FT – flow-through sample. W1 and W2 – washing samples 1 and 2. E, E1, E2, E3 – Eluted samples.

5.4. Discussion

A novel strategy for the *E. coli* soluble production of BMP-2 and IL-10 recombinant proteins is presented in this work, using the Fh8 tag as solubility enhancer partner (Chapter 2) (Costa et al., 2012) and purification handle (Chapter 4). The difficulty in expressing BMP-2 and IL-10 as soluble proteins in *E. coli* is easily recognized by the several reports describing the optimization of inclusion bodies solubilisation and renaturation procedures (Bessa et al., 2008c; Klompus et al., 2008; Carvalho et al., 2010; von Einem et al., 2010; Zhang et al., 2010; Zhang et al., 2011), and by the lack in literature of reports describing BMP-2 and IL-10 soluble production using native conditions (Ihm et al., 2008).

In this work, the solubility evaluation of both Fh8BMP-2 and Fh8IL-10 fusion proteins was conducted in two steps: a first small-scale screening, in which the Fh8-tagged BMP-2 and IL-10 proteins were compared to the corresponding His-tagged proteins using Ni-NTA pool downs, and a scale-up production, in which two purification methodologies – the IMAC and Fh8-HIC – were tested, using the Fh8BMP-2 and Fh8IL-10 fusion proteins.

The direct solubility comparison of BMP-2 and IL-10 proteins fused to the Fh8 tag and fused to the His_6 tag revealed that these proteins were only soluble expressed when the Fh8 tag was present. The soluble expression of Fh8BMP-2 and Fh8IL-10 was also confirmed by immunodetection, revealing a monomer molecular weight for Fh8BMP-2 and a monomer and dimer molecular weights for Fh8IL-10.

Fh8BMP-2 and Fh8IL-10 fusion proteins were soluble expressed in the *E. coli* BL21 (DE3) Codon Plus-RIL after an o/n induction at 18 ºC. This strain and this induction condition were selected among other strains and temperatures, in an initial comparison screening. Apart from the major contribution of Fh8 tag as solubility enhancer, the initial optimization of host cell and induction condition was also important for the soluble expression of BMP-2 and IL-10 fusion proteins in *E. coli*. In fact, as reported in other works (Dyson et al., 2004; Makino et al., 2011; Vernet et al., 2011; Pacheco et al., 2012), the selection of *E. coli* strains engineered with extra copies of rare codons, like the BL21 Codon Plus-RIL, together with a lower induction temperature may improve

the translation efficiency and the correct folding of target proteins that are known to be difficult to express in the bacterial host, like BMP-2 and IL-10.

The lower total expression levels observed in total lysate and insoluble fractions of Fh8BMP-2 and Fh8IL-10 fusion proteins, respectively, indicate a reduced expression rate of Fh8-fused BMP-2 and IL-10 proteins compared to His_6-tagged BMP-2 and IL-10 proteins. This lower expression was also observed in previous work (Chapter 2) (Costa et al., 2012) and it may contribute to the solubility shift promoted by Fh8 tag, opposing the insolubility of His_6-tagged BMP-2 and IL-10 proteins.

The inability to purify Fh8BMP-2 and Fh8IL-10 fusion proteins by nickel affinity via the His_6 tag at scale-up led to the hypothesis that some steric hindrance exists in the structure of both fusion proteins, occluding the exposure of the His_6 tail. Here, the Fh8 partner presented the additional advantage of acting as a purification handle as well. Both Fh8BMP-2 and Fh8IL-10 fusion proteins were successfully purified by Fh8-HIC methodology, presenting an organized secondary structure similar to the human BMP-2 and IL-10 previously reported structures (Zdanov et al., 1995; Scheufler et al., 1999; Hillger et al., 2005; Carvalho et al., 2010; Gilde et al., 2012). These results supported the Fh8 utility as an efficient affinity tag described in Chapter 4. The estimation of Fh8BMP-2 and Fh8IL-10 molecular weights, by measuring their hydrodynamic radius, corroborated the existence of dimer and oligomer forms for both fusion proteins.

The lack of immunodetection in the eluted Fh8BMP-2 and Fh8IL-10 samples was, however, unexpected, and it may indeed be a consequence of the higher oligomerization state of fusion proteins after passing through the HIC column. Previous studies (Fraga et al., 2010) demonstrated that the Fh8 molecule has a large solvent-exposed hydrophobic region and, it also undergoes conformational changes in the presence of calcium, promoting Fh8 interaction with other targets. So, the soluble Fh8BMP-2 and Fh8IL-10 loaded onto the HIC resin may already present an oligomerized form, but the structure of the fusion proteins at this stage is still exposing the specific epitopes for anti-BMP-2 or anti-IL-10 detection.

As the binding and washing of the HIC column was conducted in the presence of calcium, similar to the Fh8 molecule alone (Chapter 4), Fh8BMP-2 and Fh8IL-10 molecules will probably hold the calcium ions trapped into their structure during all purification steps, including the elution step. This contact with calcium may increase the availability for interaction, leading to a higher oligomerization state of both Fh8BMP-2

and Fh8IL-10 proteins. Moreover, the alkaline pH of elution buffer (pH 10) may be favoring the formation of disulfide bridges, as previously described in other works (Vallejo and Rinas, 2004b; Bessa et al., 2008c). All of these together can direct the both fusion proteins to a tridimensional structure that may not favor the exposure of specific epitopes for anti-BMP-2 or anti-IL-10 detection.

Unfortunately, the Fh8BMP-2 fusion protein did not promote the C2C12 cell differentiation into osteoblasts. Native BMP-2 is only biologically active in its homodimeric form, which is stabilized by the cystine knot between the two monomers (Scheufler et al., 1999). Hence, the formation of disulfide bonds is essential to the native and recombinant BMP-2 activity, as demonstrated by other works (Vallejo et al., 2002; Long et al., 2006; Ihm et al., 2008; Sharapova et al., 2010). Even when the protein is soluble, correct disulfide bridges must occur to obtain a natively folded BMP-2 (von Einem et al., 2010).

In spite of being directly soluble produced in the *E. coli* cytoplasm when fused to the Fh8 tag, the folded Fh8BMP-2 fusion protein is probably presenting a different rearrangement from that of native BMP-2, since proper disulfide bridges may not occur in the cytoplasm environment. In addition, the estimated dimers of this fusion protein may mostly result from the Fh8 oligomerization rather than the natively folded BMP-2.

Difficulties found in Fh8 removal from the Fh8BMP-2 fusion corroborated the steric hindrance suggested by the unsuccessful IMAC purification of this fusion protein.

5.5. Conclusion

A novel strategy for the production of soluble BMP-2 and IL-10 recombinant proteins was successfully established in this work. By taking advantage of the Fh8 tag solubility enhancing and purification properties, both BMP-2 and IL-10 proteins were highly soluble produced in *E. coli* as fusion proteins, without using any solubilization and renaturation procedures.

Fh8BMP-2 and Fh8IL-10 fusion proteins presented ordered secondary structures, and they were produced in monomer, dimer and oligomer forms that are essential for their biological activity.

All together, this work demonstrated for the first time a simple, cost-effective and directly soluble production of BMP-2 and IL-10 as fusion proteins in *E. coli*. However, some issues with the immunodetection of purified proteins, the biological inactivity of Fh8BMP-2 and difficulties in the tag removal suggest a structural complexity of the soluble Fh8-fused proteins that must be circumvented.

Acknowledgements

We thank Dr. Hüseyin Besir, Protein Expression and Purification Facility Core, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany, for supplying the pETM11 and pETHisTEV expression vectors, and also Dr. Frederico Silva, head of the Protein Production and Purification Unit from the Institute for Molecular and Cell Biology (IBMC), University of Porto, for the circular dichroism experiments and corresponding analyses. We also thank Doctor José Lacomba, head of the Tissue Engineering Group from Complutense University of Madrid, Spain, for the help with *in vitro* cell culture trials.

5.6. References

Abarrategi, A., Moreno-Vicente, C., Martinez-Vazquez, F.J., Civantos, A., Ramos, V., Sanz-Casado, J.V., Martinez-Corria, R., Perera, F.H., Mulero, F., Miranda, P. and Lopez-Lacomba, J.L. (2012) Biological properties of solid free form designed ceramic scaffolds with BMP-2: *in vitro* and *in vivo* evaluation. Plos One 7.

Asadullah, K., Sterry, W. and Volk, H.D. (2003) Interleukin-10 therapy - Review of a new approach. Pharmacological Reviews 55, 241-269.

Azari, K., Doll, B.A., Sfeir, C., Mu, Y. and Hollinger, J.O. (2001) Therapeutic potential of bone morphogenetic proteins. Expert Opinion on Investigational Drugs 10, 1677-1686.

Berzofsky, J.A., Ahlers, J.D. and Belyakov, I.M. (2001) Strategies for designing and optimizing new generation vaccines. Nature Reviews Immunology 1, 209-219.

Bessa, P.C., Casal, M. and Reis, R.L. (2008a) Bone morphogenetic proteins in tissue engineering: the road from laboratory to clinic, part II (BMP delivery). Journal of Tissue Engineering and Regenerative Medicine 2, 81-96.

Bessa, P.C., Casal, M. and Reis, R.L. (2008b) Bone morphogenetic proteins in tissue engineering: the road from the laboratory to the clinic, part I (basic concepts). Journal of Tissue Engineering and Regenerative Medicine 2, 1-13.

Bessa, P.C., Pedro, A.J., Klosch, B., Nobre, A., van Griensven, M., Reis, R.L. and Casal, M. (2008c) Osteoinduction in human fat-derived stem cells by recombinant human bone morphogenetic protein-2 produced in *Escherichia coli*. Biotechnology Letters 30, 15-21.

Bessho, K., Kusumoto, K., Fujimura, K., Konishi, Y., Ogawa, Y., Tani, Y. and Iizuka, T. (1999) Comparison of recombinant and purified human bone morphogenetic protein. British Journal of Oral & Maxillofacial Surgery 37, 2-5.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72, 248-54.

Carvalho, V. (2010) Development of carrier systems for the controlled release of interleukin-10. Ph.D. Thesis. Universidade do Minho, Braga, Portugal.

Carvalho, V., Castanheira, P., Faria, T.Q., Goncalves, C., Madureira, P., Faro, C., Domingues, L., Brito, R.M.M., Vilanova, M. and Gama, M. (2010) Biological activity of heterologous murine interleukin-10 and preliminary studies on the use of a dextrin nanogel as a delivery system. International Journal of Pharmaceutics 400, 234-242.

Costa, S.J., Almeida, A., Castro, A., Domingues, L. and Besir, H. (2012) The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology. Applied Microbiology and Biotechnology.

Demain, A.L. and Vaishnav, P. (2009) Production of recombinant proteins by microbes and higher organisms. Biotechnology Advances 27, 297-306.

Dyson, M.R., Shadbolt, S.P., Vincent, K.J., Perera, R.L. and McCafferty, J. (2004) Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. Bmc Biotechnology 4.

Fraga, H., Faria, T.Q., Pinto, F., Almeida, A., Brito, R.M.M. and Damas, A.M. (2010) FH8-a small EF-hand protein from *Fasciola hepatica*. Febs Journal 277, 5072-5085.

Gilde, F., Maniti, O., Guillot, R., Mano, J.F., Logeart-Avramoglou, D., Sailhan, F. and Picart, C. (2012) Secondary structure of rhBMP-2 in a protective biopolymeric carrier material. Biomacromolecules 3620−3626.

Granjeiro, J.M., Oliveira, R.C., Bustos-Valenzuela, J.C., Sogayar, M.C. and Taga, R. (2005) Bone morphogenetic proteins: from structure to clinical use. Brazilian Journal of Medical and Biological Research 38, 1463-1473.

Hillger, F., Herr, G., Rudolph, R. and Schwarz, E. (2005) Biophysical comparison of BMP-2, ProBMP-2, and the free pro-peptide reveals stabilization of the pro-peptide by the mature growth factor. Journal of Biological Chemistry 280, 14974-14980.

Hubel, K., Dale, D.C. and Liles, W.C. (2002) Therapeutic use of cytokines to modulate phagocyte function for the treatment of infectious diseases: Current status of granulocyte colony-stimulating factor, granulocyte-macrophage colony-stimulating factor, macrophage colony-stimulating factor, and interferon-gamma. Journal of Infectious Diseases 185, 1490-1501.

Ihm, H.J., Yang, S.J., Huh, J.W., Choi, S.Y. and Cho, S.W. (2008) Soluble expression and purification of synthetic human bone morphogenetic protein-2 in *Escherichia coli*. Bmb Reports 41, 404-407.

Israel, D.I., Nove, J., Kerns, K.M., Kaufman, R.J., Rosen, V., Cox, K.A. and Wozney, J.M. (1996) Heterodimeric bone morphogenetic proteins show enhanced activity *in vitro* and *in vivo*. Growth Factors 13, 291-300.

Klompus, S., Solomon, G. and Gertler, A. (2008) A simple novel method for the preparation of noncovalent homodimeric, biologically active human interleukin 10 in *Escherichia coli*-Enhancing protein expression by degenerate PCR of 5 ' DNA in the open reading frame. Protein Expression and Purification 62, 199-205.

Kwan, S.Y. (2011) The Flip Side of Osteoimmunity: Crosstalk among stem cells, BMP-2 and innate immune cells, and the control of osteoblastogenesis. In: Department of Biological Sciences, Vol. Paper 46. Doctor of Philosophy. Carnegie Mellon University, Pennsylvania.

Laemmli, U.K. (1970) Cleavage of Structural Proteins during Assembly of Head of Bacteriophage-T4. Nature 227, 680-685.

Long, S.N., Truong, L., Bennett, K., Phillips, A., Wong-Staal, F. and Ma, H.W. (2006) Expression, purification, and renaturation of bone morphogenetic protein-2 from *Escherichia coli*. Protein Expression and Purification 46, 374-378.

Makino, T., Skretas, G. and Georgiou, G. (2011) Strain engineering for improved expression of recombinant proteins in bacteria. Microbial Cell Factories 10.

Makrides, S.C. (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. Microbiological Reviews 60, 512-538.

Moore, K.W., Malefyt, R.D., Coffman, R.L. and O'Garra, A. (2001) Interleukin-10 and the interleukin-10 receptor. Annual Review of Immunology 19, 683-765.

Oliveira, A.F., Gemming, S. and Seifert, G. (2011) Conformational analysis of aqueous BMP-2 using atomistic molecular dynamics simulations. Journal of Physical Chemistry B 115, 1122-1130.

Pacheco, B., Crombet, L., Loppnau, P. and Cossar, D. (2012) A screening strategy for heterologous protein expression in *Escherichia coli* with the highest return of investment. Protein Expression and Purification 81, 33-41.

Pestka, S., Krause, C.D., Sarkar, D., Walter, M.R., Shi, Y.F. and Fisher, P.B. (2004) Interleukin-10 and related cytokines and receptors. Annual Review of Immunology 22, 929-979.

Pullerits, T. (2002) Cytokine modulation for anti-allergic treatment. Current Pharmaceutical Design 8, 1845-1853.

Scheufler, C., Sebald, W. and Hulsmeyer, M. (1999) Crystal structure of human bone morphogenetic protein-2 at 2.7 angstrom resolution. Journal of Molecular Biology 287, 103-115.

Sharapova, N.E., Kotnova, A.P., Galushkina, Z.M., Lavrova, N.V., Poletaeva, N.N., Tukhvatulin, A.E., Semikhin, A.S., Gromov, A.V., Soboleva, L.A., Ershova, A.S., Zaitsev, V.V., Sergienko, O.V., Lunin, V.G. and Karyagina, A.S. (2010) Production of the recombinant human bone morphogenetic protein-2 in *Escherichia coli* and testing of its biological activity in vitro and in vivo. Molecular Biology 44, 923-930.

Shea, C.M., Edgar, C.M., Einhorn, T.A. and Gerstenfeld, L.C. (2003) BMP treatment of C3H10T1/2 mesenchymal stem cells induces both chondrogenesis and osteogenesis. Journal of Cellular Biochemistry 90, 1112-1127.

Szpalski, M. and Gunzburg, R. (2005) Recombinant human bone morphogenetic protein-2: a novel osteoinductive alternative to autogenous bone graft? Acta Orthopaedica Belgica 71, 133-48.

Terpe, K. (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Applied Microbiology and Biotechnology 72, 211-222.

Urist, M.R. (1965) Bone - Formation by Autoinduction. Science 150, 893-899.

Vallejo, L.F., Brokelmann, M., Marten, S., Trappe, S., Cabrera-Crespo, J., Hoffmann, A., Gross, G., Weich, H.A. and Rinas, U. (2002) Renaturation and purification of bone morphogenetic protein-2 produced as inclusion bodies in high-cell-density cultures of recombinant *Escherichia coli*. Journal of Biotechnology 94, 185-194.

Vallejo, L.F. and Rinas, U. (2004a) Optimized procedure for renaturation of recombinant human bone morphogenetic protein-2 at high protein concentration. Biotechnology and Bioengineering 85, 601-609.

Vallejo, L.F. and Rinas, U. (2004b) Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. Microbial Cell Factories 3.

Vernet, E., Kotzsch, A., Voldborg, B. and Sundstrom, M. (2011) Screening of genetic parameters for soluble protein expression in *Escherichia coli*. Protein Expression and Purification 77, 104-111.

von Einem, S., Schwarz, E. and Rudolph, R. (2010) A novel TWO-STEP renaturation procedure for efficient production of recombinant BMP-2. Protein Expression and Purification 73, 65-69.

Yu, Y.Y., Lieu, S., Lu, C.Y. and Colnot, C. (2010) Bone morphogenetic protein 2 stimulates endochondral ossification by regulating periosteal cell fate during bone repair. Bone 47, 65-73.

Zdanov, A., Schalkhihi, C., Gustchina, A., Tsang, M., Weatherbee, J. and Wlodawer, A. (1995) Crystal-structure of interleukin-10 reveals the functional dimer with an unexpected topological similarity to Interferon-gamma. Structure 3, 591-601.

Zhang, H.B., Wu, J., Zhang, Y., Fu, N., Wang, J. and Zhao, S.J. (2010) Optimized procedure for expression and renaturation of recombinant human bone morphogenetic protein-2 at high protein concentrations. Molecular Biology Reports 37, 3089-3095.

Zhang, W.C., Xiao, W.H., Wei, H.M., Zhang, J. and Tian, Z.G. (2006) mRNA secondary structure at start AUG codon is a key limiting factor for human protein expression in *Escherichia coli*. Biochemical and Biophysical Research Communications 349, 69-78.

Zhang, Y.H., Ma, Y.S., Yang, M.Y., Min, S.J., Yao, J.M. and Zhu, L.J. (2011) Expression, purification, and refolding of a recombinant human bone morphogenetic protein 2 in vitro. Protein Expression and Purification 75, 155-160.

Chapter 6

Characterization of two novel Fh8 variants and evaluation of the mutation effects on Fh8 tag oligomerization and solubility enhancer activity

Abstract

The solubility enhancer and purification handle Fh8 tag is a Ca^{2+}-sensor protein that undergoes conformational changes upon calcium binding, exposing a large hydrophobic region for target interaction. The only cysteine residue of Fh8 sequence is solvent-exposed, becoming available for oxidation by covalent binding to other molecules. Both these features were previously suggested to be involved in the Fh8 oligomerization activity. Hence, two novel mutant Fh8 tags, Cys36Ala and Cys36Tyr mutants, were developed in this work, and characterized biophysically to evaluate the role of the cysteine residue in Fh8 oligomerization and to study the effect of the mutations on the Fh8 tag properties as solubility enhancer. The novel mutant Fh8 tags were fused to frutalin and 12-kDa *Cryptosporidium parvum* oocyst wall proteins, and Fh8Cys, Fh8Ala and Fh8Tyr fusion proteins were analyzed and compared in the presence and absence of calcium, and upon tag removal.

Fh8Ala or Fh8Tyr fusion proteins presented a reduced calcium-dependent conformational change, and less oligomer forms than those proteins fused to the wild type Fh8 tag, thus, suggesting an important role of the cysteine residue in Fh8 oligomerization. Since Fh8 mutant-fused proteins and corresponding cleaved proteins presented identical solubility and secondary structure as the Fh8-fused and corresponding cleaved ones, they are reported in this work to be structurally advantageous tags over Fh8Cys.

6.1. Introduction

The Fh8 fusion partner is a novel solubility enhancer tag that stands among the well-described best fusion partners, MBP, NusA, and Trx for the improvement of protein solubility in *Escherichia coli* (Chapter 2) (Costa et al., 2012). This fusion tag is also a promising purification handle for target protein recovery, owning to its calcium-binding properties via hydrophobic interaction chromatography (Chapter 4).

Before being applied as a fusion tag, the Fh8 protein was primarily identified as an 8-kDa calcium-binding recombinant protein (GenBank ID AF213970) extracted from the parasite *Fasciola hepatica* with great potential for the diagnosis of parasite infections, and consequent vaccine and drug development (Silva et al., 2004).

Calcium-binding proteins (CaBPs) are usually grouped between Ca^{2+}-sensors, which transduce calcium signals and display calcium-dependent conformational changes, and Ca^{2+}-buffers that modulate the shape and duration of calcium signals, undergoing minimal structural changes upon calcium binding (Lewit-Bentley and Rety, 2000; Bhattacharya et al., 2004; Gifford et al., 2007; Chazin, 2011).

Biochemical and structural characterization of Fh8 from *F. hepatica* revealed structural similarities with calmodulin (CaM) and troponin C (TnC) that are representative proteins of the Ca^{2+}-sensor group of CaBPs, and thus led to Fh8 classification as a Ca^{2+}-sensor protein (Castro, 2001; Fraga et al., 2010). Fh8 is structurally organized into two helix-loop-helix EF-hand motifs, which are involved in calcium coordination. The two EF-hands are covalently bound by a linker between the exit helix of EF-1 and the entering helix of EF-2, and their stability is maintained by an antiparallel β-sheet formed by two stretches of the calcium binding loops (Fraga et al., 2010).

In the absence of calcium, the Fh8 molecule presents already a large hydrophobic region that acts as a target-binding surface. When calcium is present, Fh8 switches from a closed to an open conformation, reorienting the four helices, and exposing a larger hydrophobic region. The Fh8 protein forms dimers in this loaded state, being found in monomer form in the apo-state.

The only cysteine residue of Fh8 sequence was located on protein surface, and due to its location and consequent availability for oxidation by covalent binding to other molecules, it was suggested to be involved in Fh8 dimerization (Fraga et al., 2010).

Previous studies of the Fh8 protein by directed mutagenesis in the cysteine residue (conducted by the research team at the National Health Institute Doutor Ricardo Jorge) highlighted an additional contribution of calcium ions in the Fh8 dimerization, and revealed a key role of the cysteine residue in the Fh8 stability.

Consistent with these observations/hypotheses, and taking into account the utility of Fh8 molecule as fusion tag, we propose here two novel variant Fh8 tags, the Fh8Ala and Fh8Tyr mutants, to decrease the oligomerization state of the Fh8Cys wild type tag, and to understand further the contribution of the cysteine residue in calcium-binding activity and solubility of the Fh8 fusion proteins.

The wild type and mutant Fh8 molecules are studied as fusion tags in *Escherichia coli* and biophysically characterized in the presence and absence of calcium, using Frutalin (FTL), an α-D-galactose binding lectin from the plant seed *Artocarpus incisa* (Oliveira et al., 2009), and 12-kDa *Cryptosporidium parvum* oocyst wall protein (CP12) (Yao et al., 2007) as target fusion proteins.

6.2. Material and Methods

6.2.1. General

Cloning PCRs used the Phusion High-Fidelity DNA Polymerase (New England Biolabs) with an annealing temperature of 55 ºC, according to the manufacturer's instructions. Colony PCRs were conducted using the NZYTaq DNA polymerase (Nzytech) with an annealing temperature of 55 ºC and with the T7 forward and reverse universal primers. Plasmid DNA extractions were performed using the PlasmidPure™ Miniprep Kit (Sigma) and the QIAquick DNA gel extraction kit or QIAquick PCR purification kit (Qiagen) were used for DNA purification. *Xba*I, *Nco*I and *Xho*I restriction enzymes were from New England Biolabs. All the DNA ligations were carried out with the Rapid DNA Ligation kit (Roche).

For plasmid maintenance and protein expression, different antibiotics (diluted 1000x) were used depending on the strain and plasmid requirements. Antibiotic stock solutions were prepared, filtered through 0.2 µm and stored at -20 ºC in the following concentrations: kanamycin 50 $mg.mL^{-1}$, and chloramphenicol 10 $mg.mL^{-1}$.

6.2.2. Construction of pETMFh8 Ala and Tyr mutant vectors

Fh8Ala and Fh8Tyr mutants were obtained from the modification of the Cys36 residue from Fh8 tag sequence into Ala36 and Tyr36, respectively (see Appendix 6.7.1, Figure A6.1).

Specific primers were design to insert the Ala or the Tyr point mutation into the previously constructed pETMFh8 tag vector (Chapter 2) (Costa et al., 2012), conducting three PCR as follows: PCR-I used the T7 forward universal primer and the Fh8 mutant (Fh8Ala or Fh8Tyr) reverse primer (Table 6.1) to modify and amplify the initial part of the Fh8 tag sequence containing the Cys36 residue; in PCR-II the Fh8 mutant (Fh8Ala or Fh8Tyr) forward primer (Table 6.1) and the T7 reverse universal primer were used to modify and amplify the final part of the Fh8 tag sequence containing the Cys36 residue. The Fh8 mutant forward and reverse primers were designed to have 13 nucleotides in common that will be essential in PCR-III for the amplification of the complete sequence of Fh8Ala/Tyr mutants using the T7 forward and reverse universal primers (Table 6.1). Final PCR products contained the His_6 tag/Fh8 mutant tag/TEV site sequence/Multiple cloning site to be inserted into the pETMFh8 tag vector. The pETMFh8Ala and pETMFh8Tyr plasmids were obtained by the ligation between the *Xba*I-*Xho*I digested purified PCR-III products and pETMFh8 plasmid.

E. coli DH5α competent cells were transformed with the constructed pETMFh8Ala/Tyr mutant plasmids and the resulting clones were analyzed by colony PCR. The novel pETMFh8Ala and pETMFh8Tyr fusion vectors were confirmed by sequencing with both T7 forward and reverse universal primers.

6.2.3. Cloning of *frutalin* and *cp12* target genes into new mutant vectors

The target genes used in this work were previously cloned into the pETMFh8 vector (Chapter 2) (Costa et al., 2012). The *frutalin* gene (*ftl*) was cloned into the novel mutant vectors by DNA ligation of *Nco*I-*Xho*I digested *ftl* from pETMFh8FTL minipreps and

*Nco*I-*Xho*I digested pETMFh8Ala/Tyr mutant plasmids. The *cp12* gene was cloned into the novel mutant vectors using an identical strategy as shown in Table 6.1. By using the pETMFh8CP12 plasmid as template, the Ala or Tyr point mutations were obtained by three PCR using the same primers designed for the construction of the mutant vectors. The final pETMFh8AlaCP12 and pETMFh8TyrCP12 plasmids were obtained by the ligation between the *Xba*I-*Xho*I digested purified PCR-III products and pETMFh8 plasmid.

E. coli DH5α competent cells were transformed with the constructed plasmids and the resulting clones were analyzed and confirmed as previously mentioned.

Table 6.1. List of primers used for the cloning of Fh8 mutants and target proteins:

PCR	Primer	Sequence	Comments
I	T7_FWD	5'-TAATACGACTCACTATAGGG-3'	-
	Fh8Ala_RV	5´-TTGGAGTCC*AGAGG**A<u>GC</u>**TTTTG*– 3’	The modified codon [TGT(Cys)→GCT(Ala)] is in bold and mutated nucleotides are underlined. The forward and reverse primer matching nucleotides are in italic.
	Fh8Tyr_RV	5´-TTGGAGTCC*AGAGG**<u>GT</u>A**TTTTG*– 3’	The modified codon [TGT(Cys)→TAC(Tyr)] is in bold and mutated nucleotides are underlined. The forward and reverse primer matching nucleotides are in italic.
II	Fh8Ala_FWD	5´- GCTGATGATT*CAAAA**<u>GC</u>T**CCTCT*-3´	See the above comment for Fh8Ala-RV
	Fh8Tyr_FWD	5´- GCTGATGATT*CAAAA**T<u>AC</u>**CCTCT*-3´	See the above comment for Fh8Tyr-RV
	T7_RV	5'-GCTAGTTATTGCTCAGCGG-3'	-
III	T7_FWD	as above	
	T7_RV	as above	

6.2.4. Protein expression and purification

Fh8Cys/Ala/Tyr fusion proteins were expressed in 3 L cultures (6x500 mL in 2 L flasks), using *E. coli* BL21 (DE3) Codon Plus-RIL strain for CP12s expression and *E. coli* Roseta (DE3) for FTLs expression. Cell growth, induction, and harvesting were conducted using an identical protocol as previously described for the Fh8CysFTL or CP12 fusion proteins (Chapter 2) (Costa et al., 2012) with the following additional step: after induction, three 5 mL samples were taken from each culture for dry weight estimation and the remaining cells were harvested for 25 minutes, at 4 ºC and 4000 rpm.

For cell lyses, bacterial pellets were resuspended in 60 mL of lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 20 mM imidazole, supplemented with 1x complete free EDTA protease inhibitor (Roche), 5 mM MgCl (Sigma), 5 $\mu g.mL^{-1}$ DNAse (Sigma) and 1 $mg.mL^{-1}$ lysozyme (Sigma)), and incubated at room temperature for 10 minutes with agitation. The cell suspension was lysed by sonication (Branson 450 Sonifier) and the supernatant fraction was collected at 10000 rpm, 4 ºC for 25 minutes. Total lysate and supernatant aliquots were taken and stored at 4 ºC.

A 5 mL prepacked Histrap HP column (GE Healthcare) was used for protein purification. Supernatant fractions were filtered through 0.45 µm and loaded onto the Histrap column. The purification protocol followed the manufacturer's instructions, using the following buffers: binding and washing buffers (50 mM Tris pH 8.0, 150 mM NaCl, 20 mM Imidazole), and elution buffer (50 mM Tris pH 8.0, 150 mM NaCl, 300 mM Imidazole). Columns regeneration and storage was also performed according to the manufacturer's instructions.

The flow-through, washing and eluted samples were stored at 4 ºC and analysed by SDS-PAGE. The eluted samples were pooled together and divided either for dialysis with storage buffer (50 mM Tris pH 8.0, 150 mM NaCl) at 4 ºC, o/n or for dialysis with binding buffer combined with HisTEV (EMBL) digestion. The cleaved CP12 and Frutalin fusion proteins were purified from the corresponding Fh8 tags and the HisTEV protease by nickel affinity chromatography using the same protocol as above. The purified cleaved proteins were further dialyzed with storage buffer at 4 ºC, o/n.

All the fusion proteins and cleaved proteins were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and quantified by Bradford assay and

by absorbance at 280 nm. Protein stocks of 0.5-1 mg.mL^{-1} were filtered through 0.22 µm and stored at -20 ºC for subsequent functional and structural analyses.

6.2.5. Western blot

Purified samples of FTL and CP12 fusion proteins were transferred to nitrocellulose membranes using a sandwich system. The Western blotting analysis was conducted with the polyclonal mouse antibodies anti-FTL and anti-CP12 diluted 1:1000 as primary antibodies, and anti-mouse IgG HRP (Biorad) diluted 1:1000 together with protein G HRP (Biorad) diluted 1:2000 as conjugates. Polyclonal antibodies anti-FTL and anti-CP12 were produced using the H partner methodology described in Chapter 3. The substrate 4-chloro-naftol (Sigma) was used for membrane revelation.

6.2.6. Analytical size exclusion and DLS

Analytical size exclusion of Fh8Cys/Ala/Tyr-tagged FTL and CP12 proteins and corresponding cleaved proteins was carried out in a Superose 12, 10/300 GL column (GE Healthcare) using the running buffer: 50 mM Tris pH 8.0, 150 mM NaCl, and 100 µg in 200 µL of each protein. Ribonuclease A (1.64 nm and 13.7 kDa), chymosin (2.09 nm and 25 kDa), ovalbumin (3.05 nm and 43 kDa), bovine serum albumin (3.55 nm and 67 kDa), aldolase (4.81 nm and 158 kDa), and catalase (5.22 nm and 232 kDa) were also analyzed in the Superose 12 column to be further used as standard proteins for calibration of size exclusion analyses.

The size calibration curve (see Appendix 6.7.2, Figure A6.2.a) was obtained by plotting each standard protein value of $\sqrt{-\log(Kav)}$ versus the corresponding Stoke radius (in nm). The K_{av} indicates the ratio between the elution volume of a given molecule and the total available volume of the column, and can be determined by the following equation: $Kav = \frac{Ve-Vo}{Vt-Vo}$, where *Ve* is the elution volume, *Vo* is the void volume and *Vt* is the total column volume.

The molecular weight calibration curve (see Appendix 6.7.2, Figure A6.2.b) was obtained by plotting the *Kav* versus *Log MW*.

The elution volume of each fusion and cleaved proteins was used for the estimation of size and molecular weight, using the above-mentioned calibration curves.

DLS measurements were conducted in Malvern Zetasizer Nano ZS using a protein concentration of 0.5 mg.mL^{-1}, previously filtered through 0.2 µm.

In both size exclusion and DLS experiments, Fh8Cys/Ala/Tyr-tagged FTL and CP12 samples were incubated in running buffer supplemented with 1 mM EDTA or 5 mM $CaCl_2$, at 4 ºC, overnight. Cleaved proteins were incubated in running buffer supplemented with 1 mM EDTA.

6.2.7. Frutalin functional assay

The hemagglutination assay of Fh8Cys/Ala/Tyr-tagged FTL proteins and corresponding cleaved FTL proteins was conducted as described elsewhere (Oliveira et al., 2009), using a initial protein concentration of 0.1 mg.mL^{-1} in storage buffer. The hemagglutination activity of fusion proteins and corresponding cleaved proteins was also evaluated in storage buffer supplemented with 5 mM $CaCl_2$ or 1 mM EDTA (proteins were incubated at 4 ºC, overnight, in this buffer prior to the analysis). Negative and positive controls were also performed with the Fh8 tag alone and with native Frutalin, respectively.

6.2.8. CD spectroscopy

The Fh8Cys/Ala/Tyr-tagged FTL and CP12 proteins as well as their corresponding cleaved proteins were further analyzed concerning its secondary structure. The circular dichroism (CD) spectra of purified recombinant proteins (20 µg per protein in 10 mM sodium phosphate buffer, pH 8.0) were collected with a 1 mm path length cuvette, between 190 and 260 nm, set up to 1 nm band width, continuous scan mode at 200 nm/min. The presented spectra are the average of 3 scans with the average buffer control spectrum subtracted. Spectra were acquired in a J-815 circular dichroism spectropolarimeter (Jasco), at 20 ºC.

The results are expressed in terms of molar ellipticity (mean residue ellipticity, θ) in deg.cm^2.dmol^{-1}, according to the equation:$(\theta) = \frac{\theta obs}{10 \times Cr \times l}$, where θ_{obs} is the observed ellipticity in mdeg, C_r is the mean residue molar concentration (mol/L), and l is the cuvette path length in cm. The mean residue molar concentration is obtained as follows:

$Cr = \frac{1000 \times n \times C'}{MW}$, where n is the number of residues of the protein, C' is the protein concentration (g/mL), and *MW* is the protein molecular mass (g/mol).

The spectra were deconvoluted in the Dichroweb server using the CONTIN analysis program and the SP175 reference dataset of proteins.

6.2.9. Fluorescence spectroscopy

The differential scanning fluorimetry (DSF) experiment was performed in quadruplicate, using the CFX96^{TM} Real-time PCR detection system (Biorad) and the Sypro Orange (Sigma) as reporter dye. Briefly, a 2.5x solution of Sypro Orange was added to the wells of a 96-well thin-wall PCR white plate (Biorad) together with 20 µM of fusion and cleaved FTL proteins, previously incubated in storage buffer in the presence of 1 mM EDTA or 5 mM $CaCl_2$. Control experiments were also conducted using water alone, dye alone, and proteins alone. The plates were sealed with Optical-Quality Sealing Tape (Biorad) and heated from 20 to 95 ºC with 30 seconds holding time every 0.5 ºC. The changes in fluorescence were collected using 545/585 nm wavelengths for excitation/emission, respectively.

The fluorescence of 8-anilino-1-naphthalenesulfonic acid ammonium salt (ANS) was measured in a Horiba spectramax fluorimeter using 20 µM of Fh8Cys/Ala/Tyr-fused proteins and 300 µM of ANS, in the presence of 1 mM EDTA or 20 mM $CaCl_2$. Fluorescence emission spectra were measured in the range of 400-700 nm, with an excitation wavelength for ANS of 385 nm, 3 nm excitation and 5 nm emission slits, and an integration time of 2 seconds.

6.3. Results

6.3.1. Novel Fh8 mutants cloning and fusion protein expression and purification

The sequencing results of the novel Fh8 mutant plasmids showed a successful cloning strategy: the Fh8Ala plasmid contains a GCT (Ala) instead of TGT (Cys) at the amino acid position 36 and the Fh8Tyr plasmid has a TAC codon (Tyr) instead of TGT (Cys) at the expected position. These two novel Fh8 fusion plasmids presented 99% identity to the Fh8Cys sequence, as observed by the nucleotide BLAST analysis (see Appendix 6.7.1, Figure A6.1).

The FTL and CP12 codifying genes were also successfully cloned into the novel Fh8Ala/Tyr mutant vectors, presenting 95% identity with the *Artocarpus integrifolia* jacalin isolectin (GenBank: L03797.1), and 100% identity with the *Cryptosporidium parvum* Iowa II hypothetical protein (GenBank: XM625821.1), respectively. *ftl* and *cp12* genes cloned into the novel pETMFh8Ala/Tyr mutant plasmids presented 100% identity with the same genes previously cloned into the pETMFh8 vector (Chapter 2) (Costa et al., 2012).

All the Fh8 fusion proteins were soluble expressed when using either the Fh8Cys tag (wild type tag) or the Fh8Ala/Tyr mutant tags, but FTL recombinant proteins were obtained with higher purity than the CP12 recombinant proteins, as shown by the SDS-PAGE analysis in Figure 6.1. All the fusion proteins in study were also detectable by anti-FTL or anti-CP12 polyclonal antibodies, as observed in the Western blot analysis presented in Figure 6.1.

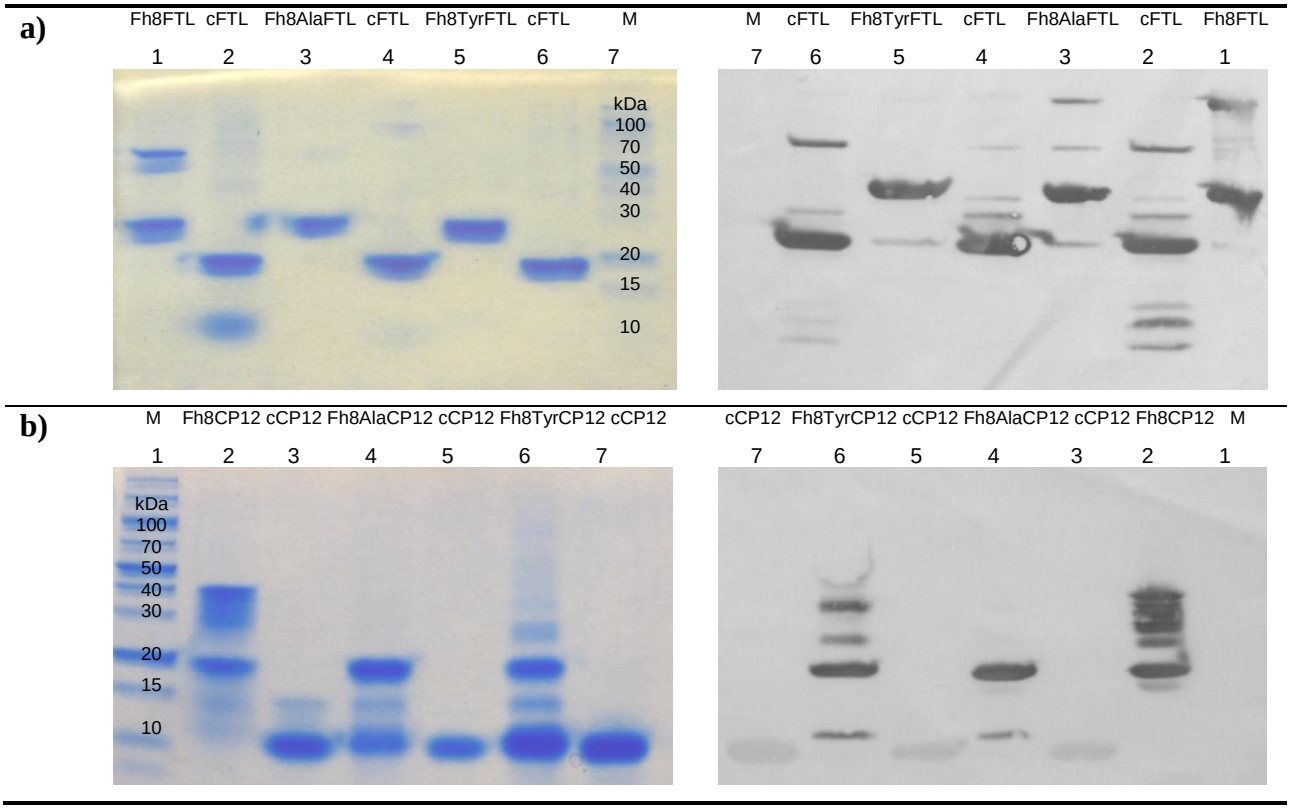


Figure 6.1. SDS-PAGE (left panel) and Western blot (right panel) analyses of (**a**) Fh8FTL fusion proteins, and (**b**) Fh8CP12 fusion proteins, and corresponding cleaved proteins after tag removal using the Tobacco Etch Virus (TEV) protease. In all images, each Fh8-fused protein is loaded aside of the matching cleaved protein, which is represented by a *c* preceding the protein's name. The same numbered lanes in the SDS-PAGE and Western blot images were used for the same loaded samples. M – PageRuler broad unstained protein marker (Thermo Scientific).

Frutalin recombinant proteins fused to the three Fh8 tags presented similar protein amounts after IMAC purification (mg of purified proteins per litre of *E. coli*), but some differences were found in the dry weight estimation of these cultures (Figure 6.2). The Fh8Ala/Tyr mutant FTL (Figure 6.2.a) or CP12 (Figure 6.2.b) cultures presented higher dry weight (g/L) than the original Fh8FTL or Fh8CP12 cultures. As the volumetric yields (mg of protein per litre of *E. coli* culture) were similar among each group of three fusion proteins, both Fh8FTL and Fh8CP12 presented higher final production yields (mg per g of dry weight of *E. coli*) than the corresponding mutant fusion proteins.

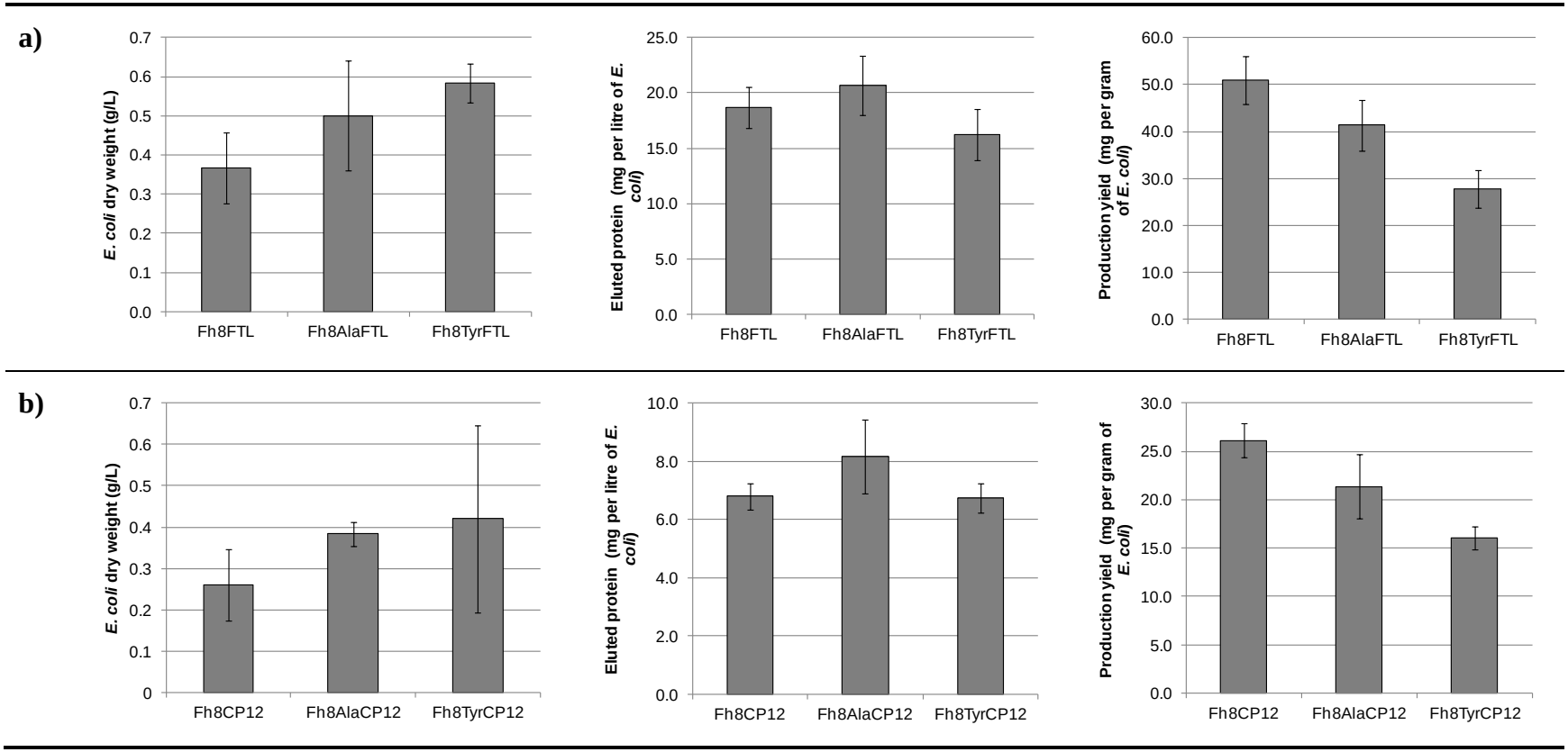


Figure 6.2. Comparison of the production results of (**a**) Fh8FTL fusion proteins, and (**b**) Fh8CP12 fusion proteins. Production results are represented by the *E. coli* dry weight estimation (g of dry cell per litre of *E. coli* culture), the Eluted protein (volumetric yield: protein mg per litre of *E. coli* culture), and the Production yield (protein mg per gram of *E. coli* culture).

6.3.2. Size estimation of Fh8-fused and cleaved proteins: characterization of protein oligomerization in apo- and calcium-loaded states

Table 6.2 summarizes the analytical size exclusion results for Fh8CP12 fusion proteins and corresponding cleaved proteins.

Table 6.2. Stokes radius (R, in nm) and corresponding protein molecular weight (MW, in kDa) of Fh8CP12 and cleaved CP12 proteins, estimated by analytical size exclusion:

Sample	**1 mM EDTA**		**5 mM $CaCl_2$**	
	R (nm)	MW (kDa)	R (nm)	MW (kDa)
Fh8CP12	3.5	68	3.2	55
	2.8	38	2.6	32
	2.1	23	1.9	18
Fh8AlaCP12	2.6	34	2.4	29
	1.6	15	1.7	16
Fh8TyrCP12	2.4	28	2.4	28
	1.5	13	1.8	17
Cleaved CP12 (Fh8)	1.9	18	-	-
Cleaved CP12 (Fh8Ala)	1.8	17	-	-
Cleaved CP12 (Fh8Tyr)	1.8	17	-	-

Comparing the Stokes radius (in nm) between proteins dissolved in EDTA buffer and proteins dissolved in $CaCl_2$ buffer, no considerable differences were observed among the three Fh8CP12 fusion proteins.

Fh8CP12 presented three protein populations close to the expected molecular weight for monomer (18.5 kDa), dimer (37 kDa), and possible tetramer (74 kDa) forms. This tetramer form was not observed for Fh8Ala/TyrCP12 fusion proteins, which presented only two mainly populations close to the expected molecular weight for monomer and dimer forms.

In all three fusion proteins, the dimer conformation was predominant, as observed by the highest peak intensity at the corresponding elution volume in size exclusion chromatograms (see Appendix 6.7.3, Figure A6.3).

Cleaved CP12 proteins have an expected molecular weight of about 9 kDa and were collected with an elution volume matching a predominant dimer form.

The analytical size exclusion was also performed for Fh8FTL fusion proteins and corresponding cleaved proteins, but a delay in elution was observed possibly due to a protein interaction with the size exclusion matrix. Thus, the obtained elution volumes could not be used for size and molecular weight determination, and a different approach was further conducted.

Figure 6.3 presents the hydrodynamic radius (comparable to Stokes radius in the previous analysis) of Fh8FTL fusion proteins, obtained in the DLS analysis.

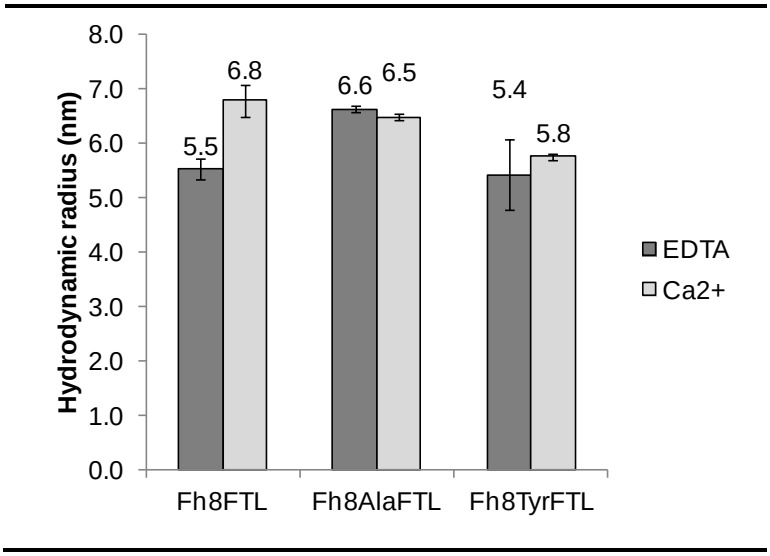


Figure 6.3. Hydrodynamic radius of the three Fh8FTL fusion proteins, measured by DLS. *EDTA* refers to proteins in the presence of 1 mM of this chelating agent. *Ca2+* refers to proteins in the presence of 5 mM $CaCl_2$.

Here, no differences were observed for the EDTA or $CaCl_2$ dissolved Fh8Ala/TyrFTL proteins, but a larger radius was obtained for Fh8FTL in the presence of calcium, compared to the same molecule in the presence of EDTA.

The hydrodynamic radius of all three Fh8FTL fusion proteins (wild type and mutants) revealed a high oligomerization state (between 100-200 kDa), estimated using the Zetasizer software, and assuming the three Fh8FTL as globular proteins.

Cleaved FTL proteins presented a hydrodynamic radius of 3.5±0.16 nm, resulting in a globular protein with a molecular weight of, approximately, 65 kDa. Taking into account the expected molecular weight (17.5 kDa), cleaved FTL proteins were possible obtained in the tetramer form.

6.3.3. Protein surface hydrophobicity in apo- and calcium-loaded states

The hydrophobic probe ANS was used to evaluate the hydrophobic surface exposure of Fh8 fusion proteins, thus, verifying if the Fh8 used in the fusion context exhibited the Ca^{2+}-sensor property previously reported for Fh8 protein alone (Fraga et al., 2010). Figure 6.4 presents the fluorescence measurements of the three Fh8CP12/FTL fusion proteins in the presence of 1 mM EDTA or 20 mM $CaCl_2$.

As observed in Figure 6.4.a, ANS fluorescence dramatically increased for Fh8FTL in the presence of calcium, when compared to the same protein in the presence of EDTA. The other Fh8Ala/TyrFTL fusion proteins showed little to no shift in ANS fluorescence between the EDTA and calcium state.

A similar result was obtained for Fh8CP12 fusion proteins (Figure 6.4.b). Among the three Fh8CP12 fusion proteins, the Fh8CP12 (wild type tag) promoted the highest shift of ANS fluorescence in the presence of calcium, comparing to that fluorescence obtained in the presence of EDTA. Fh8TyrCP12 promoted a slight shift in ANS fluorescence when incubated with calcium, and no considerable differences in ANS fluorescence were observed for the Fh8AlaCP12 in the presence of calcium or EDTA.

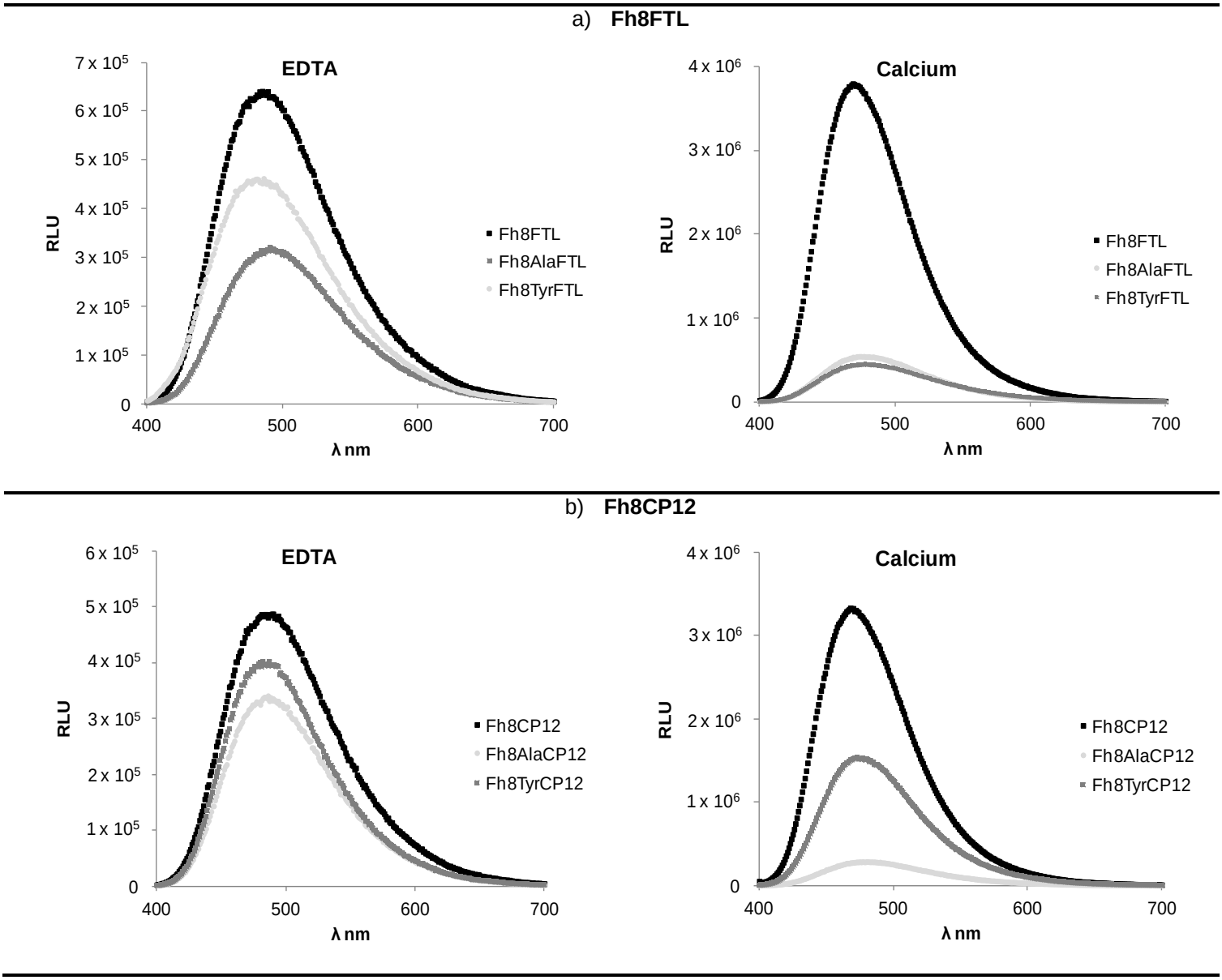


Figure 6.4. Comparison of the hydrophobic surface exposure between Fh8Cys, Fh8Ala and Fh8Tyr-fused proteins, in the presence of EDTA (1 mM) or Calcium (20 mM $CaCl_2$), by ANS fluorescence spectroscopy (see Material and Methods section). (**a**) Fh8FTL fusion proteins. (**b**) Fh8CP12 fusion proteins. In all images, *Fh8* is used for Fh8Cys-fused proteins, *λ nm* refers to the wavelength in nm, and *RLU* refers to relative luminescence units, which in this case represents the counts per second (CPS) .

6.3.4. Biological activity and stability of FTL recombinant proteins

Table 6.3 resumes the hemagglutination activity of the three purified Fh8FTL fusion proteins and corresponding cleaved proteins, presented into "hemagglutination units" (HU) that are equivalent to the reciprocal of the last dilution presenting clot formation, and into "specific activity" ($\mu g.mL^{-1}$), which represents the minimal protein concentration required to promote visible agglutination.

Table 6.3. Hemagglutination activity results of the different fusion FTL and cleaved FTL proteins:

Buffer	Sample	HU	Specific activity ($ug.mL^{-1}$)
50 mM Tris 150 mM NaCl pH 8.0	NC	-	-
	Fh8Cys, Fh8Ala, Fh8Tyr	-	-
	Fh8FTL	64	1.56
	Fh8AlaFTL	4	25.0
	Fh8TyrFTL	4	25.0
	cleavedFh8FTL	64	1.56
	cleavedFh8AlaFTL	4	25.0
	cleavedFh8TyrFTL	16	6.25
50 mM Tris 150 mM NaCl pH 8.0 **+ 5 mM $CaCl_2$**	Fh8FTL	128	0.781
	Fh8AlaFTL	128	0.781
	Fh8TyrFTL	128	0.781
	cleavedFh8FTL	64	1.56
	cleavedFh8AlaFTL	64	1.56
	cleavedFh8TyrFTL	64	1.56

As observed in Table 6.3, with no calcium or no EDTA addition, both Fh8FTL and cleaved FTL presented identical HUs and specific activity. The Fh8Ala/Tyr FTLs and corresponding cleaved proteins presented lower HUs than the original Fh8FTL and cleaved FTL proteins. Curiously, after calcium addition, all the three Fh8-fused FTL proteins showed identical HU and specific activity, and higher HUs than the same proteins without calcium. An identical result was also obtained for cleaved FTL proteins in the presence of calcium.

Fusion proteins lost their agglutination activity when incubated with 1 mM EDTA.

In general, all three Fh8FTL fusion proteins as well as corresponding cleaved proteins were very stable, as determined by the DSF assay (Figure 6.5). Here, similar melting temperatures were obtained for proteins in the presence of EDTA or $CaCl_2$. The three cleaved FTL proteins presented also identical melting temperatures (of about 59 ºC). Interestingly, this analysis revealed that Fh8TyrFTL was the most stable fusion protein, shifting to an unfolded structure only after 61 ºC.

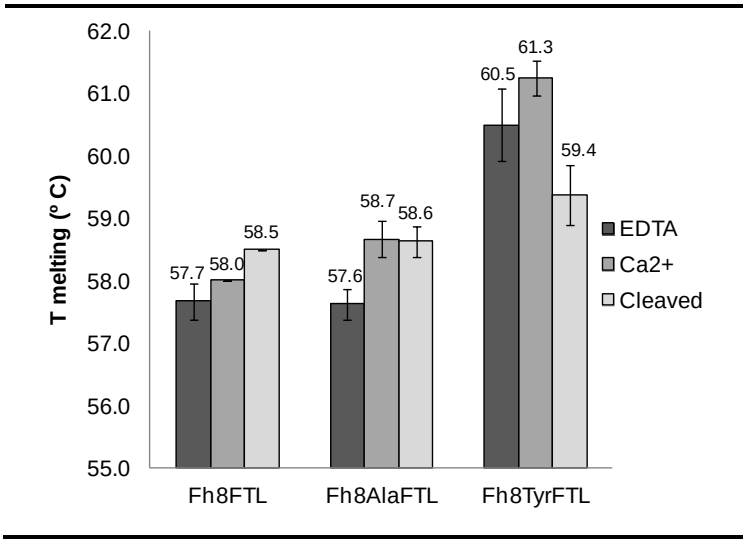


Figure 6.5. Thermal stability results of Fh8Cys, Fh8Ala and Fh8Tyr-fused FTL proteins, obtained by DSF (see Material and Methods section). *EDTA* refers to proteins in the presence of 1 mM of this chelating agent, *Ca2+* refers to proteins in the presence of 5 mM $CaCl_2$, and *Cleaved* refers to FTL proteins after tag removal using TEV protease. *Fh8* is used for Fh8Cys-fused proteins.

6.3.5. Secondary structure of Fh8-fused and cleaved proteins

The analysis of secondary structure from all Fh8FTL fusion proteins (Figure 6.6.a) resembled a predominant β-sheet conformation with a minimum molar ellipticity at about 218 nm. The obtained spectra were not, however, pure β-strands, revealing a possible mixed helical structure for the three fusion proteins (22-35% helical, 40-48% β-strand). The three cleaved FTL proteins presented identical structure, mainly β-sheet, (8-27% helical, 42-75% β-strand) with a minimum molar ellipticity at around 219 nm, as observed in Figure 6.6.b.

All three Fh8CP12 fusion proteins (Figure 6.6.c) presented an alpha-helical structure (63-80% helical, 10-12% β-turn), with two minima at about 205 and 220 nm. The three cleaved CP12 presented the same spectra, resembling an undefined structure (random coil), with a minimum at around 200 nm, as observed in Figure 6.6.d.

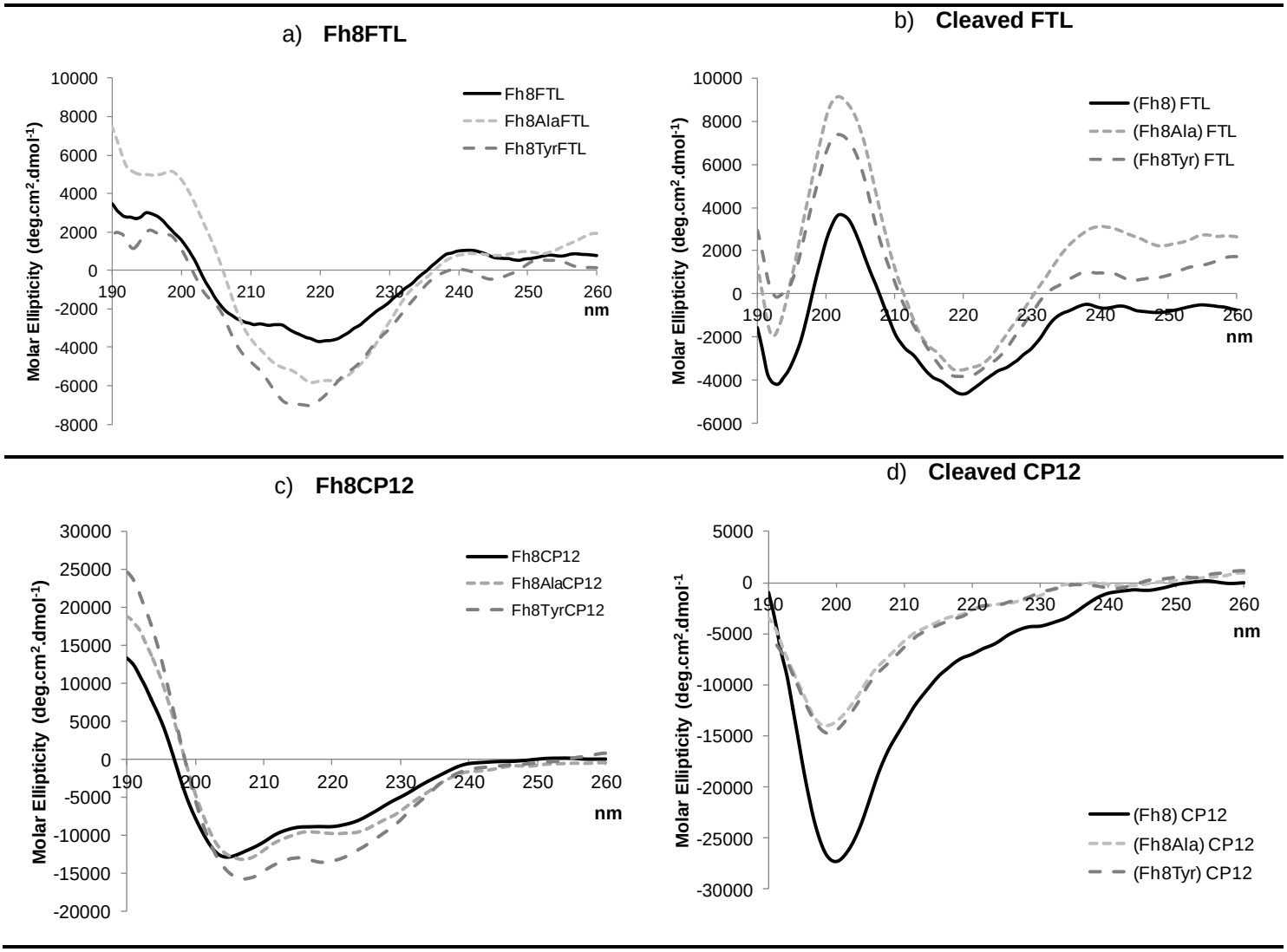


Figure 6.6. Secondary structures analyzed by CD of Fh8Cys, Fh8Ala and Fh8Tyr-fused proteins and corresponding cleaved proteins. (**a**) Fh8FTL fusion proteins, (**b**) cleaved FTL proteins, (**c**) Fh8CP12 fusion proteins, and (**d**) cleaved CP12 proteins. Results are presented in Molar Ellipticity (deg.cm^2.dmol^{-1}), calculated as mentioned at the Material and Methods section. In the four images, *Fh8* refers to Fh8Cys-fused proteins; *(Fh8)*, *(Fh8Ala)* and *(Fh8Tyr)* refer to cleaved proteins from the equivalent fusion proteins, and *nm* refers to the wavelength in nm.

6.4. Discussion

Two novel variants of the solubility enhancer Fh8 tag are characterized in this work by performing single point mutations at the only cysteine residue of Fh8, located at position 36. Cys36Ala and Cys36Tyr mutants, studied here for the first time as fusion tags, were designed to reduce the Fh8 tag oligomerization state, and also to understand the cysteine role in calcium-binding activity and solubility of the Fh8 fusion proteins. Interesting findings of this work suggested that the Cys36 residue is, indeed, involved in the calcium-dependent exposure of Fh8 hydrophobic patch, as discussed below.

Fh8Cys, Fh8Ala and Fh8Tyr mutant tags were fused to FTL and CP12 recombinant proteins and their structural characteristics were compared by size exclusion, DLS, ANS fluorescence and CD, to assess the calcium effect on the three types of fusion tags.

In both FTL and CP12 target proteins, Fh8 mutant tags presented the solubility enhancer effect of the Fh8Cys wild type tag, though, yielding less protein amounts per gram of *E. coli* culture than the latter. The size characterization results obtained from analytical size exclusion and DLS demonstrated that Fh8 mutant-fused proteins exhibit less calcium-dependent shifts in molecule radius and protein aggregation than the Fh8Cys-fused proteins.

Upon calcium binding, Fh8 mutant-fused proteins promoted little to no change in fluorescence of the hydrophobic probe ANS, differing from the Fh8Cys-fused proteins that presented a considerable calcium-dependent increase in ANS fluorescence, therefore, showing the Fh8-sensor protein behavior. The ability to bind ANS was previously used in other works (Ababou and Desjarlais, 2001; Ababou et al., 2001; Bunick et al., 2004; Fraga et al., 2010) as a qualitative indication of the presence of an open-state conformation. Hence, the similar ANS fluorescence between apo-state (proteins in the presence of 1 mM EDTA) and loaded-state (proteins in the presence of 20 mM $CaCl_2$) of Fh8 mutant-fused proteins, together with their lower ANS fluorescence compared to that of Fh8Cys-fused proteins, suggest that Fh8Ala and Fh8Tyr fusion proteins may no longer be adopting the full open-state conformation. However, as ANS fluorescence in the presence of EDTA is similar among Fh8Cys, Fh8Ala and Fh8Tyr fusion proteins, these ones may already be exposing a large

hydrophobic surface in the apo-state, like it was reported for the Fh8 molecule (Fraga et al., 2010).

Fh8FTL fusion proteins were characterized further by their ability to promote hemagglutination of rabbit erythrocytes, by their thermal stability and secondary structure. Contrary to what it was reported for native FTL (Moreira et al., 1998), the divalent calcium showed to have an important role in recombinant FTL biological activity, as the agglutination promoted by the three Fh8 fusion proteins and corresponding cleaved FTL proteins was completely abolished by 1 mM EDTA, a calcium chelating agent. Without calcium addition, the higher agglutination activity observed for Fh8CysFTL and for the corresponding cleaved FTL over Fh8 mutant-fused FTL and matching cleaved FTL proteins, can be correlated with an easier calcium accommodation already in the *E. coli* cytosol by Fh8CysFTL molecule, opposite to Fh8 mutant-fused proteins.

Curiously, when 5 mM $CaCl_2$ was added, all three fusion proteins and corresponding cleaved FTL proteins presented identical agglutination units and specific activities, possible due to a saturation of the calcium-binding sites of the three Fh8 fusion proteins, together with a possible identical calcium-effect on cleaved FTL proteins. Actually, other galactose-binding proteins were already reported to require this divalent ion for its activity (Suzuki et al., 1990; Sampaio et al., 1998; Dutta et al., 2005). In the specific case of recombinant FTL, little is known about calcium influence in its activity, but the different final structure arrangement between native FTL, which undergoes post-translation modifications, and recombinant FTL produced in *E. coli* (non-glycosylated protein) (Oliveira et al., 2009) may contribute for the observed calcium-effect in the latter protein. The addition of calcium did not considerable improve the thermal stability of the three Fh8FTL fusion proteins, which were already very stable without calcium.

An important finding came out from the characterization conducted with FTL: the removal of Fh8 tags did not hamper FTL biological activity and stability, and the obtained secondary structure of cleaved proteins was similar to that reported for native FTL (Moreira et al., 1998; Campana et al., 2002).

In spite of presenting little to no calcium-dependent conformational changes in contrast to the Fh8 fusion proteins, Cys36Ala and Cys36Tyr mutations did not modify Fh8FTL and Fh8CP12 secondary structures.

Fh8 mutant-fused proteins have also showed the presence of dimer forms, independently of the presence of calcium. These dimers were not unexpected, as FTL and CP12 target proteins exhibited oligomer forms in the unfused state, and the Fh8 molecule was previously reported to present a large solvent-exposed hydrophobic area in the Ca^{2+}-free state that is available for interaction with other targets (Fraga et al., 2010). Moreover, previous studies of the Fh8 protein using identical mutations showed also dimer forms for Fh8Ala and Fh8Tyr proteins.

Altogether, our results revealed that the novel Fh8 mutant tags are functional as solubility enhancer tags and promote less calcium-dependent protein oligomerization than the Fh8Cys wild type tag.

Mutations conducted in this study corroborated the previous observations by the research team with Fh8 mutant proteins, and demonstrated that the Fh8 tag dimerization is not exclusively due to the solvent-exposed cysteine residue, but this residue is probably playing an important role in Fh8 calcium-dependent conformational changes due to its location. Cys36 is in the linker between the two EF-hands, a region that presents different arrangements in apo- or calcium-loaded state, becoming close to the hydrophobic patch in the latter state (Fraga et al., 2010).

Besides Cys36, other key residues may be involved in Fh8 tag oligomerization, as for instance, residues responsible for calcium coordination in the two existent loops between each EF-hand helix, and polar residues located at the loop side chains. In other calmodulin-like proteins, these residues were determinant for calcium-induced conformational changes (Permyakov et al., 2000; Ababou and Desjarlais, 2001; Ababou et al., 2001; Bunick et al., 2004; Xiong et al., 2010).

Results from this work suggest that the Fh8 oligomerization can be important for its solubility enhancing mechanism, because Fh8Ala and Fh8Tyr-fused proteins are still soluble produced in oligomer forms, and are probably exposing a large hydrophobic surface in the apo-state identical to the Fh8Cys-fused proteins. Thus, similar to what it was proposed for other fusion tags (Fox et al., 2001; Nallamsetty and Waugh, 2006; Nallamsetty and Waugh, 2007), the Fh8 tag may present an intrinsic chaperone-like mechanism, by which its exposed hydrophobic patches (in both apo- and calcium-loaded states) interact with partially folded target proteins. This Fh8 hydrophobic interaction with target proteins can prevent their self-aggregation, and change their folding pathway to a soluble structure.

To better clarify if the oligomerization of Fh8 is essential for the solubility enhancer activity, and to further investigate the mechanisms behind this activity, new mutants at the loop residues of EF-hands and corresponding side chains should be designed and evaluated.

6.5. Conclusions

Two novel variants of the Fh8 solubility enhancer tag were successfully obtained in this work by mutation of the only cysteine residue to alanine and tyrosine.

The new Fh8Ala and Fh8Tyr fusion tags did not completely abolished the Fh8 tag oligomerization, but they highlighted the importance of the cysteine residue in the calcium-dependent conformational change of the Fh8 tag. The two Fh8 mutant tags preserved the solubility enhancer activity of the wild type tag, offering less oligomerization and calcium-dependent conformational changes to the fusion proteins. Hence, both Fh8Ala and Fh8Tyr fusion tags provide less complexity to the overall structure of the fusion protein, and are more attractive to be used in the fusion context.

In order to investigate further the mechanism by which the Fh8 tag promotes fusion protein solubility, additional mutants shall be designed to completely block the Fh8 oligomerization and to study its consequent effect in the solubility enhancer activity.

Aknowledgements

We thank Dr. Hüseyin Besir, Protein Expression and Purification Facility Core, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany, for kindly providing the pET24TEV vector for HisTEV expression, and also, Dr. Paula Magalhães, head of the Cell Culture and Genotyping Service (CCGen) from the Institute for Molecular and Cell Biology of the University of Porto, for the experimental help with the DSF analysis.

6.6. References

Ababou, A. and Desjarlais, J.R. (2001) Solvation energetics and conformational change in EF-hand proteins. Protein Science 10, 301-312.

Ababou, A., Shenvi, R.A. and Desjarlais, J.R. (2001) Long-range effects on calcium binding and conformational change in the N-domain of calmodulin. Biochemistry 40, 12719-12726.

Bhattacharya, S., Bunick, C.G. and Chazin, W.J. (2004) Target selectivity in EF-hand calcium binding proteins. Biochimica Et Biophysica Acta-Molecular Cell Research 1742, 69-79.

Bunick, C.G., Nelson, M.R., Mangahas, S., Hunter, M.J., Sheehan, J.H., Mizoue, L.S., Bunick, G.J. and Chazin, W.J. (2004) Designing sequence to control protein function in an EF-hand protein. Journal of the American Chemical Society 126, 5990-5998.

Campana, P.T., Moraes, D.I., Monteiro-Moreira, A.C.O. and Beltramini, L.M. (2002) Unfolding and refolding studies of frutalin, a tetrameric D-galactose binding lectin. European Journal of Biochemistry 269, 753-758.

Castro, A. (2001) Obtenção e caracterização de proteínas recombinantes homólogas de antigénios excretados/secretados pelo verme adulto de *Fasciola hepatica*. Ph.D. Thesis. Universidade do Porto, Oporto, Portugal.

Chazin, W.J. (2011) Relating form and function of EF-Hand calcium binding proteins. Accounts of Chemical Research 44, 171-179.

Costa, S.J., Almeida, A., Castro, A., Domingues, L. and Besir, H. (2012) The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology. Applied Microbiology and Biotechnology.

Dutta, S., Sinha, B., Bhattacharya, B., Chatterjee, B. and Mazumder, S. (2005) Characterization of a galactose binding serum lectin from the Indian catfish, *Clarias batrachus*: Possible involvement of fish lectins in differential recognition of pathogens. Comparative Biochemistry and Physiology C-Toxicology & Pharmacology 141, 76-84.

Fox, J.D., Kapust, R.B. and Waugh, D.S. (2001) Single amino acid substitutions on the surface of *Escherichia coli* maltose-binding protein can have a profound impact on the solubility of fusion proteins. Protein Science 10, 622-630.

Fraga, H., Faria, T.Q., Pinto, F., Almeida, A., Brito, R.M.M. and Damas, A.M. (2010) FH8-a small EF-hand protein from *Fasciola hepatica*. Febs Journal 277, 5072-5085.

Gifford, J.L., Walsh, M.P. and Vogel, H.J. (2007) Structures and metal-ion-binding properties of the Ca^{2+}-binding helix-loop-helix EF-hand motifs. Biochemical Journal 405, 199-221.

Lewit-Bentley, A. and Rety, S. (2000) EF-hand calcium-binding proteins. Current Opinion in Structural Biology 10, 637-643.

Moreira, R.A., Castelo-Branco, C.C., Monteiro, A.C.O., Tavares, R.O. and Beltramini, L.M. (1998) Isolation and partial characterization of a lectin from *Artocarpus incisa* L. seeds. Phytochemistry 47, 1183-1188.

Nallamsetty, S. and Waugh, D.S. (2006) Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. Protein Expression and Purification 45, 175-182.

Nallamsetty, S. and Waugh, D.S. (2007) Mutations that alter the equilibrium between open and closed conformations of *Escherichia coli* maltose-binding protein impede its ability to enhance the solubility of passenger proteins. Biochemical and Biophysical Research Communications 364, 639-644.

Oliveira, C., Costa, S., Teixeira, J.A. and Domingues, L. (2009) cDNA Cloning and functional expression of the alpha-D-galactose-binding lectin Frutalin in *Escherichia coli*. Molecular Biotechnology 43, 212-220.

Permyakov, S.E., Cherskaya, A., Senin, I.I., Zargarov, A.A., Shulga-Morskoy, S.V., Alekseev, A.M., Zinchenko, D.V., Lipkin, V.M., Philippov, P.P., Uversky, V.N. and Permyakov, E.A. (2000) Effects of mutations in the calcium-binding sites of recoverin on its calcium affinity: evidence for successive filling of the calcium binding sites. Protein Engineering 13, 783-790.

Sampaio, A.H., Rogers, D.J. and Barwell, C.J. (1998) A galactose-specific lectin from the red marine alga *Ptilota filicina*. Phytochemistry 48, 765-769.

Silva, E., Castro, A., Lopes, A., Rodrigues, A., Dias, C., Conceicao, A., Alonso, J., da Costa, J.M.C., Bastos, M., Parra, F., Moradas-Ferreira, P. and Silva, M. (2004) A recombinant antigen recognized by *Fasciola hepatica*-infected hosts. Journal of Parasitology 90, 746-751.

Suzuki, T., Takagi, T., Furukohri, T., Kawamura, K. and Nakauchi, M. (1990) A Calcium-Dependent Galactose-Binding Lectin from the *Tunicate Polyandrocarpa-Misakiensis* - Isolation, Characterization, and Amino-Acid-Sequence. Journal of Biological Chemistry 265, 1274-1281.

Xiong, L.W., Kleerekoper, Q.K., Wang, X. and Putkey, J.A. (2010) Intra- and interdomain effects due to mutation of calcium-binding sites in calmodulin. Journal of Biological Chemistry 285, 8094-8103.

Yao, L., Yin, J., Zhang, X., Liu, Q., Li, J., Chen, L., Zhao, Y., Gong, P. and Liu, C. (2007) *Cryptosporidium parvum*: identification of a new surface adhesion protein on sporozoite and oocyst by screening of a phage-display cDNA library. Exp Parasitol 115, 333-8.

6.7. Appendices

6.7.1. Sequence alignment of Fh8 mutant and Fh8 wild type tags

Fh8Ala mutant

```
>gb|AF213970.1|AF213970  Fasciola hepatica putative calcium-binding protein mRNA, complete
cds
Length=210

 Score =  360 bits (398),  Expect = 2e-96
 Identities = 202/204 (99%), Gaps = 0/204 (0%)
 Strand=Plus/Plus

Query  1    CCTAGTGTTCAAGAGGTTGAAAAACTCCTTCATGTTCTCGATCGCAACGGTGACGGGAAG  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4    CCTAGTGTTCAAGAGGTTGAAAAACTCCTTCATGTTCTCGATCGCAACGGTGACGGGAAG  63

Query  61   GTTTCTGCCGAGGAGTTGAAAGCCTTCGCTGATGATTCAAAAGCTCCTCTGGACTCCAAT  120
            ||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Sbjct  64   GTTTCTGCCGAGGAGTTGAAAGCCTTCGCTGATGATTCAAAATGTCCTCTGGACTCCAAT  123

Query  121  AAGATCAAGGCTTTCATTAAGGAACACGATAAAAACAAGGATGGCAAGCTTGATTTGAAA  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  124  AAGATCAAGGCTTTCATTAAGGAACACGATAAAAACAAGGATGGCAAGCTTGATTTGAAA  183

Query  181  GAACTCGTTTCGATTTTGTCATCA  204
            |||||||||||||||||||||||||
Sbjct  184  GAACTCGTTTCGATTTTGTCATCA  207
```

Fh8Tyr mutant

```
>gb|AF213970.1|AF213970  Fasciola hepatica putative calcium-binding protein mRNA, complete
cds
Length=210

 Score =  360 bits (398),  Expect = 2e-96
 Identities = 202/204 (99%), Gaps = 0/204 (0%)
 Strand=Plus/Plus

Query  1    CCTAGTGTTCAAGAGGTTGAAAAACTCCTTCATGTTCTCGATCGCAACGGTGACGGGAAG  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4    CCTAGTGTTCAAGAGGTTGAAAAACTCCTTCATGTTCTCGATCGCAACGGTGACGGGAAG  63

Query  61   GTTTCTGCCGAGGAGTTGAAAGCCTTCGCTGATGATTCAAAATACCCTCTGGACTCCAAT  120
            ||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Sbjct  64   GTTTCTGCCGAGGAGTTGAAAGCCTTCGCTGATGATTCAAAATGTCCTCTGGACTCCAAT  123

Query  121  AAGATCAAGGCTTTCATTAAGGAACACGATAAAAACAAGGATGGCAAGCTTGATTTGAAA  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  124  AAGATCAAGGCTTTCATTAAGGAACACGATAAAAACAAGGATGGCAAGCTTGATTTGAAA  183

Query  181  GAACTCGTTTCGATTTTGTCATCA  204
            |||||||||||||||||||||||||
Sbjct  184  GAACTCGTTTCGATTTTGTCATCA  207
```

Figure A6.1. Nucleotide BLAST (http://blast.ncbi.nlm.nih.gov/) of Fh8 tag sequence and sequenced Fh8Ala and Fh8Tyr mutants. The "subject" sequence refers to the Fh8Cys tag and the "query" sequence refers to the Fh8 mutant tag. In the Fh8Ala mutant tag sequence, the codon TGT (cys) was changed to GCT (ala). In the Fh8Tyr mutant tag sequence, the codon TGT (cys) was changed to TAV (tyr). Both point mutantions are highlighted in a red box.

6.7.2. Calibration curves of analytical size exclusion

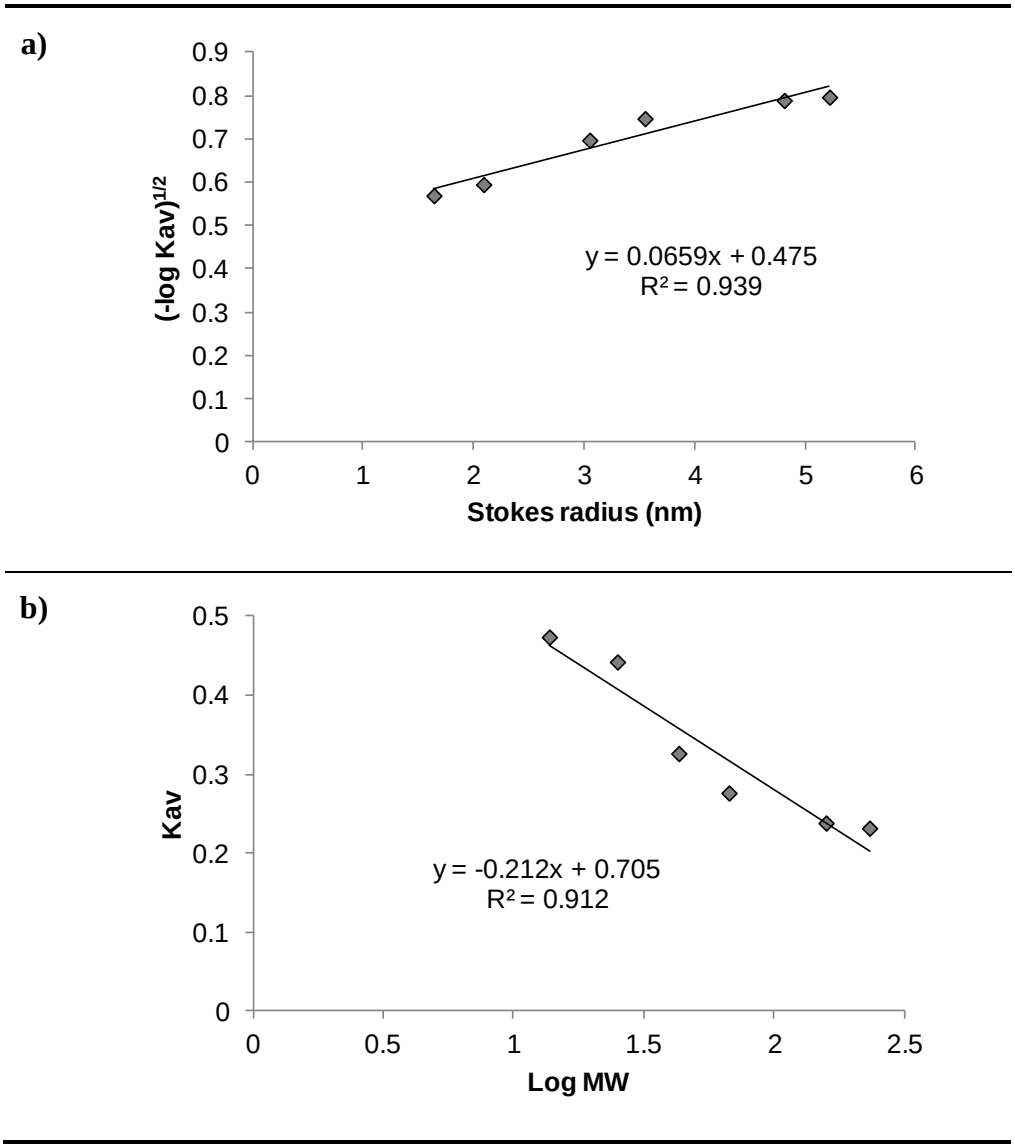


Figure A6.2. Calibration curves for size estimation using analystical size exclusion. (**a**) Calibration curve of Stokes radius. (**b**) Calibration curve of molecular weight. Both curves were obtained using the following standards: Ribonuclease A (1.64 nm and 13.7 kDa), chymosin (2.09 nm and 25 kDa), ovalbumin (3.05 nm and 43 kDa), bovine serum albumin (3.55 nm and 67 kDa), aldolase (4.81 nm and 158 kDa), and catalase (5.22 nm and 232 kDa).

6.7.3. Size exclusion chromatograms

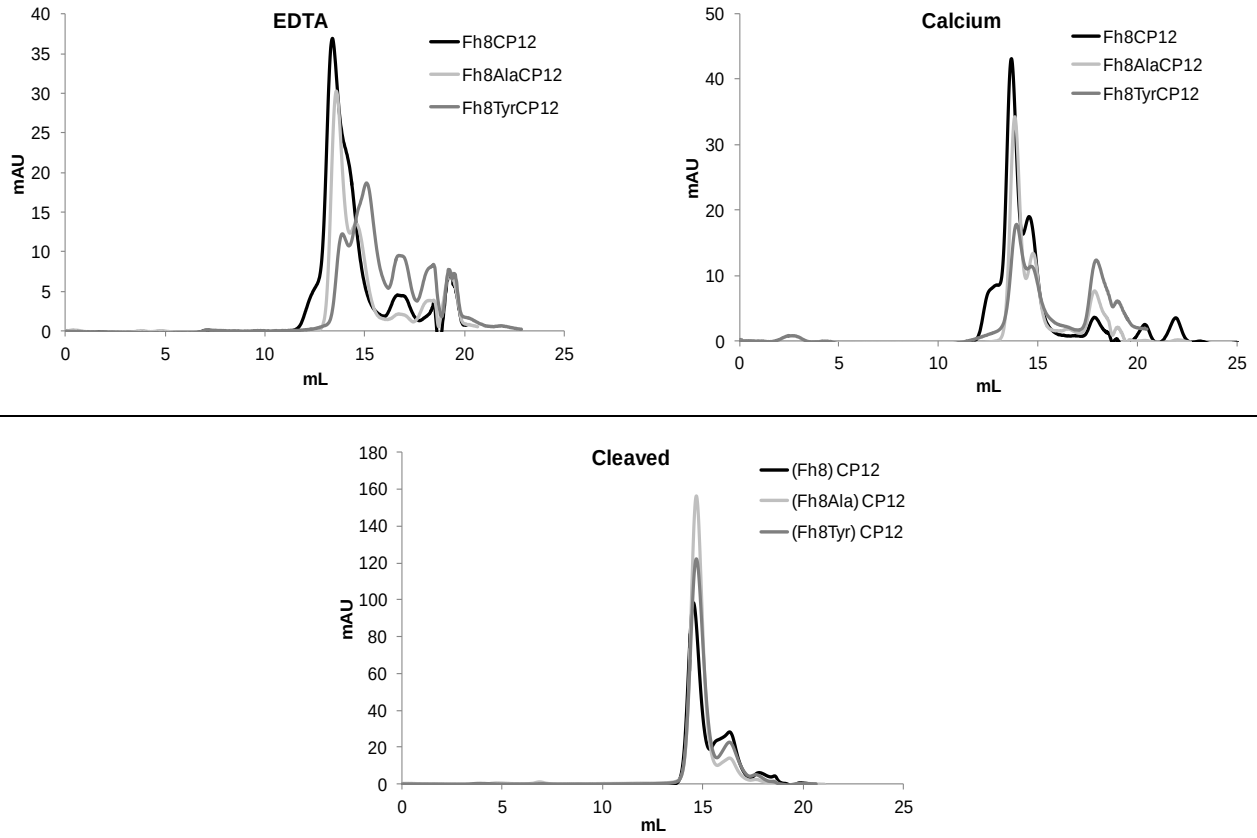


Figure A6.3. Size exclusion chromatograms of Fh8CP12 fusion proteins in the presence of 1 mM EDTA or 5 mM $CaCl_2$, and CP12 cleaved proteins in the presence of 1 mM EDTA.

Chapter 7

Conclusions and future perspectives

A novel and unique fusion system for simple and inexpensive soluble protein overexpression and purification in *E. coli* was developed in this work. Results obtained here led to the following conclusions:

- The H tag improves protein expression in *E. coli* but it does not function as solubility enhancer tag (Chapter 2). Nevertheless, this novel fusion partner presents attractive features for the production of immunogens and corresponding polyclonal antibodies (Chapter 3).
- The H tag (1 kDa) is the lowest fusion partner used so far for protein and antibody production, making it potentially advantageous over other fusion tags. The H tag facilitates a simple, rapid, and adjuvant-free production "from gene to antibody". In addition, polyclonal antibodies produced using the H tag can be directly applied for the immunnodiagnosis of several parasite infections. Although the H tag is here mostly applied for the production of recombinant parasite immunogens, it is also foreseen to be a promising tool for the immunoproduction of other proteins of interest as well.
- The Fh8 is an effective solubility enhancer partner and a robust purification handle (Chapters 2 and 4). This dual functionality turns this tag into a valuable tool for efficient recombinant protein production in *E. coli*.
- The Fh8 is ranked among the best solubility enhancer tags as Trx, MBP or NusA, being easily removed from the target protein without compromising the solubility of the latter (Chapter 2).
- Besides being one of the few existent fusion tags to offer the combined feature of enhancing protein solubility and purification, the Fh8 performs both actions in a unique inexpensive manner. Fh8 is potentially advantageous over MBP or GST affinity tags because it uses its natural calcium-binding properties and mild

conditions for hydrophobic interaction chromatography (Chapter 4) instead of expensive resins, harsh buffers and additional compounds for protein purification.

- The Fh8-HIC technology offers the potential feature of eliminating most of *E. coli* endotoxins in a single-step purification of biologically active recombinant proteins. The dual His_6-Fh8 tagging can also be explored when a more stringent and efficient removal of contaminating proteins is required.
- The low molecular weight of Fh8 (8 kDa) is, indeed, a great advantage over other large fusion partners for recombinant protein production in *E. coli*. The Fh8 is thus an excellent candidate for testing expression, solubility and purification in high throughput screenings next to the other well-known fusion tags.
- When applied to the production of proteins of interest, namely FTL, which has a potential biomedical interest in cancer diagnostic, the Fh8 tag presented two major advantages for its recombinant production in *E. coli*: it increased FTL expression and solubility, and the Fh8 removal from the fusion protein did not affect the solubility, native secondary structure and bioactivity of FTL (Chapters 2 and 6).
- The Fh8 fusion partner is also a promising tool for the economical and rapid *E. coli* soluble expression and purification of two difficult-to-express proteins: the bone morphogenetic protein-2 (BMP-2) and interleukin-10 (IL-10) proteins (Chapter 5). Both these proteins have relevant biomedical applications that emphasize the need for their efficient recombinant production. The Fh8 fusion system allows for the first time the soluble production of BMP-2 and IL-10 proteins in *E. coli* without any solubilization and renaturation process. For both proteins, the Fh8-HIC purification system is superior to the His_6-IMAC, overcoming limitations observed with the latter. The Fh8-HIC purification strategy allows both BMP-2 and IL-10 proteins to be purified in a single-step and low cost methodology. At the end of this purification, Fh8BMP-2 and Fh8IL-10 proteins are potentially ready for *in vitro* trials as they present ordered secondary structures and protein oligomerecy, and may contain low levels of endotoxins.
- The Fh8 tag may, however, address some obstacles to the biological activity of BMP-2 protein. This protein is only effective when presenting a specific dimeric conformation. The Fh8 can thus maintain the BMP-2 in solution long enough to undergo into a soluble folding pathway but, in spite of presenting an ordered

structure and dimeric form, this fusion protein may be in a stable conformation different from that of native BMP-2, which is crucial for its activity. In addition, the Fh8 is not easily removed from the fusion protein probably due to steric hindrance found in the complex structure. Results from Chapter 5 led to the conclusion that the structure of Fh8 fusion partner requires optimization for the production of oligomeric target proteins, in order to interfere as less as possible with the structure of the latter.

- Two novel variants of Fh8 fusion partner – Fh8Ala and Fh8Tyr – are structurally advantageous proteins over Fh8Cys to be used in the fusion context (Chapter 6) because they showed less calcium-dependent conformational changes and less oligomeric forms while performing as efficient solubility enhancer tags. Fh8 Ala and Fh8Tyr mutations led to the conclusion that the Cys residue is indeed involved in the Fh8 oligomerization, but other residues are surely contributing to this state.

The novel Fh8 fusion partner overcomes several issues related to the recombinant protein production in *E. coli*: by using a straightforward methodology, this novel system increases protein expression levels, promotes protein solubility and low cost purification, and helps for protein immunogenicity (the H tag). This novel fusion system offers the great advantage of combining these four abilities into the two lowest molecular weight fusion partners described so far. The aim of this work was thus achieved, highlighting the Fh8 fusion partner potential to move forward the recombinant protein production field.

In order to develop further the knowledge and utility of the Fh8 fusion system, the following strategies are proposed as future work:

- Other variants of Fh8 fusion partner shall be investigated to completely abolish the Fh8 oligomerization state, and to better clarify if its solubility enhancer activity is dependent of this state. Directed mutagenesis can be conducted in residues responsible for calcium-coordination between the EF-hands, and also in hydrophobic residues that are already solvent-exposed without the presence of calcium (apo-state). In other calcium-binding proteins, these residues showed to be involved in the Ca^{2+}-sensor activity, therefore being also here of most interest for the discovery of novel Fh8 variant fusion partners.

- The Fh8 fusion partner shall be considered for co-expression with the target protein of interest in order to evaluate if it still improves protein solubility. This co-expression strategy may be useful for certain applications that prohibit the use of fusion partners but require high amounts and/or soluble protein. Moreover, if the Fh8 improves target protein solubility by co-expression, it may help to clarify its mechanism of action (for instance, it may function as a chaperone or as a chaperone-magnet).
- A wide systems biology analysis can be conducted to track and evaluate the potential mechanisms underlying the protein soluble overexpression promoted by the N-terminal Fh8 fusion partner. The differential screening of transcriptional and proteomic profiles in *E. coli* expressing soluble Fh8 fusion proteins, expressing only soluble non-fused proteins, and expressing insoluble target proteins or fusion proteins, may highlight the pathways by which the Fh8 fusion partner acts as an effective solubility enhancer tag.
- An in depth understanding of the H tag mechanism of action shall be elucidated by the evaluation of the cellular components involved in the overall immunological response promoted by this fusion partner. *In vitro* trials shall be conducted to estimate the contribution of different cells from the immune system in the enhanced humoral response developed when using the H tag. Additional insights can be taken by the comparison between the H tag and other immunopotentiating agents.
- The proteolytic activity of the Fh8 fusion partner may be explored to create a novel self-cleaving tag. The Fh8 protein has demonstrated in previous studies to have an intrinsic and calcium-dependent autocatalytic activity. This feature shall be further investigated in the fusion context in order to advance the use of Fh8 for single-step protein solubility, purification and tag removal.

While this work in this thesis applies to the use of Fh8 and H for recombinant protein production in bacterial host systems, it is hoped that the novel fusion system presented here will apply to other hosts, as for instance, eukaryotes and mammalian cells.