Gonçalo Duarte Gomes Teixeira

Licenciado em Bioquímica

Chemical and biological synthesis of an engineered affinity protein

Dissertação para obtenção do Grau de Mestre em Biotecnologia

Orientador: Ana Cecília Afonso Roque, PhD,
FCT-UNL, Portugal
Co-orientador: Olga Iranzo Casanova, PhD, iSm2,
Aix-Marseille Université, France

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro 2017

LOMBADA

Chemical and biological synthesis of an engineered affinity protein
Gonçalo Teixeira

2017

Gonçalo Duarte Gomes Teixeira

Licenciado em Bioquímica

Chemical and biological synthesis of an engineered affinity protein

Dissertação para obtenção do Grau de Mestre em Biotecnologia

Orientador: Ana Cecília Afonso Roque, PhD,
FCT-UNL, Portugal
Co-orientador: Olga Iranzo Casanova, PhD, iSm2,
Aix-Marseille Université, France

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro 2017

Chemical and biological synthesis of an engineered affinity protein

“Copyright”

Gonçalo Duarte Gomes Teixeira

Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

“There are no secrets to success.

It is the result of preparation, hard work, and learning from failure.”

Collin Luther Powell

Agradecimentos

Em primeiro lugar, gostaria de dedicar um enorme agradecimento à minha orientadora, Professora Cecília Roque, por ter-me recebido neste projeto e dentro do seu fantástico grupo de trabalho. Queria ainda agradecer por todo o conhecimento que me transmitiu ao longo deste ano e, sobretudo, por ter sempre depositado uma enorme confiança nos seus alunos e ensiná-los a nunca desistir perante os constantes desafios que a ciência nos oferece.

À minha co-orientadora, Dra. Olga Iranzo, por toda a atenção e todo o suporte que me permitiu uma fácil adaptação num período de dois meses na Universidade de Marselha. Agradeço tudo o que aprendi sobre a síntese de péptidos, e ainda todos os conselhos que me permitiram melhorar, no laboratório e no dia-a-dia.

Um especial agradecimento à Dra. Ana Pina, por toda a ajuda durante o ano, mas sobretudo por uma reta final a um ritmo alucinante. Mesmo quando tudo corria mal, conseguimos fazer que corresse bem. Tudo acabou por compensar no final. Muito obrigado por tudo o que aprendi, pelos conselhos, constante atenção e acompanhamento, mesmo quando exigia um esforço adicional, principalmente fora de horas.

Gostaria ainda de agradecer a todas as pessoas que fazem parte do grupo de Engenharia Biomolecular. A todos os que me ajudaram direta ou indiretamente no meu trabalho, pelos vossos conselhos, valores, bom-ambiente e espírito de entreajuda. Não poderei deixar de agradecer a todas as pessoas que me apoiaram durante os dois meses que passei em Marselha, pela constante preocupação e ajuda. Queria também deixar um agradecimento à Leonor, por disponibilizar a autoclave e outros materiais necessários durante a expressão de proteínas.

Aos meus amigos de Coimbra, por fazerem-me olhar para as origens com um orgulho característico e tornarem as minhas visitas sempre especiais. Aos amigos da FCT, obrigado pela amizade e por terem tornado a minha adaptação a uma nova faculdade muito mais fácil.

Agradeço aos meus pais, ao meu irmão e a toda a família, por acreditarem sempre em mim, no meu trabalho e por darem sempre os conselhos mais acertados. Por toda a ajuda, constante apoio e compreensão quando poucas vezes podia ir visitá-los. Tudo compensou e valeu a pena o esforço.

Por fim, queria deixar um agradecimento especial à Sofia, por toda a ajuda, carinho, paciência e companhia. Por tornar tudo muito mais fácil e por conseguir encontrar sempre solução para os problemas que pareciam ser mais complicados.

Abstract

Protein engineering is an area of major interest for novel biotechnological and biopharmaceutical applications. As alternative to antibodies and other affinity ligands, several protein scaffolds have been explored. In this context, WW domains are proteins with high potential to become affinity reagents. They have approximately 40 amino acids, including two tryptophan residues spaced by 20 to 22 amino acids. Their structure present three antiparallel β-sheets that promote protein-protein interactions.

This work focuses on the detailed study of a previously selected 42 residue WW domain (CW3S) for binding Human Serum Albumin (HSA). The CW3S domain was produced chemically and biologically. The successfully produced protein by chemical synthesis was characterized through Mass Spectrometry and Circular Dichroism. To show the binding affinity for the target, several tests were performed, including binding tests after immobilization in a solid-support, MicroScale Thermophoresis (MST) and Multiparametric Surface Plasmon Resonance (MP-SPR). Overall, CW3S was successfully synthesized showing an affinity constant of 5.0×10^{6} M^{-1} for the interaction with HSA.

The biological production of the CW3S peptide was performed by bacterial expression (in *E. coli*) and then purified. The expression of CW3S was attempted alone or as a fusion protein with GFP. In the latter, BL21 Rosetta cells strain were used as expression host and the protein was purified by mixed-mode chromatography with A4C7 synthetic ligand chromatography, by gel filtration. The conditions for protein expression and purification need further optimizations.

Keywords: WW domains, CW3S, synthesis, expression, affinity, fusion protein.

Resumo

A engenharia de proteínas é uma área de grande interesse para o desenvolvimento de produtos biotecnológicos e biofarmacêuticos. Diversos *scaffolds* proteicos têm sido estudados como alternativa a anticorpos e outros ligandos de afinidade. Neste contexto, os domínios WW são proteínas com potencialidade para serem desenvolvidos como reagentes de afinidade. Apresentam aproximadamente 40 aminoácidos, contendo dois resíduos de triptofano separados por 20 a 22 aminoácidos. A sua estrutura possui três folhas-β antiparalelas que promovem interações com proteínas.

Este trabalho trata o estudo detalhado de um domínio WW de 42 resíduos (CW3S), previamente selecionado para ligar à Albumina de Soro Humano (HSA). O domínio CW3S foi produzido química e biologicamente. Através de síntese química, o péptido foi produzido com sucesso e caracterizado por Espectrometria de Massa e Dicroísmo Circular. Para demonstrar a afinidade e ligação à HSA foram realizados testes de afinidade, como um teste de ligação após imobilização do péptido num suporte, *MicroScale Thermophoresis* (MST) e *Multiparametric Surface Plasmon Resonance* (MP-SPR). Deste modo, o péptido CW3S foi sintetizado com sucesso, obtendo uma constante de afinidade de 5.0×10^{6} M^{-1}, no que diz respeito à interação com a HSA.

A produção biológica do CW3S foi realizada por expressão em *E. coli*, tendo o péptido sido posteriormente purificado. A expressão do CW3S foi realizada com o péptido sozinho e incorporado numa proteína de fusão, com a *Proteína Verde Fluorescente* (GFP). As células hospedeiras utilizadas foram *Rosetta* da estirpe BL21. Finalmente, a proteína de fusão foi purificada através de vários processos cromatográficos, incluindo o ligando sintético A4C7 e também filtração-gel. As condições para a expressão de proteína e purificação necessitam de posterior otimização.

Palavras-chave: Domínios WW, CW3S, síntese, expressão, afinidade, proteína de fusão.

Index of contents

Index of figures

Index of tables

Abbreviations

ABD - Albumin-binding domain

AI - After induction

Amp - Ampicilin

APS - Ammonium persulfate

Arg - Arginine

BCA - Bicinchoninic acid assay

BI - Before induction

bp - Base pairs

BSA - Bovine serum albumin

CBM - Carbohydrate binding molecule

CD - Circular Dichroism

CL - Cloranphenicol

CV - Column volume

Cys - Cysteine

DDT - Dichlorodiphenyltrichloroethane

DIEA - *N*,*N*-Diisopropylethylamine

DMF - *N*,*N*-Dimethylformamide

DNA - Deoxyribonucleic acid

DTNB - 5,5'-dithio-bis-[2-nitrobenzoic acid]

E. coli - *Escherichia coli*

EDTA - Ethylenediaminetetraacetic acid

EK - Enterokinase

ELISA - enzyme-linked immunosorbent assay

Fast-AP - Thermosensitive Alkaline Phosphatase

FDM - Fluorescein Diacetate 5-Maleimide

GdnHCl - Guanidinium chloride

GF - Gel-filtration

GFP - Green Fluorescent Protein

GST - Glutathione-S-transferase

HSA - Human serum albumin

HBTU - 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate

HEPES - 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid, N-(2-Hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid)

His - Histidine

HPLC - High-performance liquid chromatography

IgG - Immunoglobulin G

IPTG - Isopropyl β-D-1 thiogalactopyranoside

LB - Luria-Bertani medium

Lys - Lysine

mAb - Monoclonal antibodes

MP-SPR - Multi-Parametric Surface Plasmon Resonance

MRE - Molar residue ellipticity

MS - Mass Spectrometry

MST - MicroScale Thermophoresis

NGF - Nerve growth factor

NHS - N-hydroxysuccinimide

PC - Pellet Low-speed centrifugation

PCR - Polymerase chain reaction

PEG - Polyethylene glycol

Pro - Proline

PU - Pellet Ultracentrifugation

SC - Supernatant Low speed centrifugation

SDS - Sodium dodecyl sulfate

SpA - *Staphylococcus aureus*

SPPS - Solid-phase peptide synthesis

SPR - Surface Plasmon Resonance

SU - Supernatant Ultracentrifugation

Taq - *Thermus aquaticus*

TCEP - tris(2-carboxyethyl)phosphin

TEMED - Tetramethylethylenediamine

TFA - Tetrafluoroacetic acid

TNF-α - Tumor necrosis factor alpha

tRNA - Transfer ribonucleic acid

Trp - Tryptophan

Tyr - Tyrosine

UV - Ultraviolet

Vis - Visible

Chapter 1

Literature Review

1.1 – Background

The engineering of proteins is a revolutionizing field in several biopharmaceutical and biotechnological applications [1]. Antibodies can still be considered as gold standard affinity reagents. Their large surfaces can establish various types of interactions (e.g. hydrogen bonds, hydrophobic interactions in the same region), being naturally evolved affinity proteins with high specificity to their targets [2]. Conventionally, antibodies were used as diagnostic, therapeutic, detection, sensing and purification agents. In affinity purification, protein engineering has been crucial in the production of new affinity reagents, such as antibodies, antibody-based architectures [3] and also new natural biological ligands [4]. Artificial proteins, such as protein domains, started being plausible alternatives to antibodies. However, engineered proteins have been reaching great interest, due to their less complexity compared to antibodies, the capability to recognize ligands with high specificity and affinity, and lower production cost [5]. Therefore, protein engineering may develop a relevant role in biotechnological applications, for instance, in affinity chromatography. This has been a major interest for the biopharmaceutical industry, in which engineered versions of protein A (derived from *Staphylococcus aureus* – SpA), for instance, are nowadays the model-type monoclonal antibodies (mAb) purification agents. However, the currently available affinity resins for laboratory or industrial applications are still expensive for a meaningful proportion of all production costs (60–80%) [6]. Moreover, there has been a crescent number and variety of new molecules and particles entering the market, adding more pressure in the development of purification steps [2].

Over the years, efforts have been made in order to reduce costs of purification and downstream processing. However, the development of an affinity reagent for purification must take into account diverse conditions, such as: the introduction of proper chemistry towards immobilization in a solid-support, stability to adverse chemical conditions (e.g. organic solvents and extreme pH), and also stability and maintenance of binding capacity with resin using conditions ("cleaning-in-place", sterilization in place, resin reutilization). For instance, protein-based complexes have a tendency to be less stable under harsh conditions, in resin washes and elution cycles. Thus, this may lead to ligands leakage and a decreased resin life-time [2,7].

A significant number of affinity ligands with different structures have been explored, from natural protein binders (e.g. antibodies) to small synthetic and pseudobiospecific ligands (Fig. 1.1) [2,8]. They can be characterized by their size, complexity, specificity, and cost of production, for example. Small synthetic and pseudobiospecific ligands, for instance, are very simple and affordable-cost reagents to synthesize. They are less complex are more stable under harsh conditions [8–10]. On the contrary, natural protein binders are more complex molecules and consequently more expensive to produce [6,11]. They are also more unstable in chemical

immobilization strategies, in cleaning and elution conditions, and tend to be subjected to protease attack [12–14]. For example, antibodies production relies on the expression of eukaryotic cells systems. Thus, their purification process most likely depends on protein A resins, being a crucial factor to increase their cost of production [11]. Antibodies denaturation can be caused by adverse pH oscillations, organic solvents and high concentrations of salt, which may lead to product contamination problems in purification processing, where these molecules are immobilized in chromatographic supports [7,12].

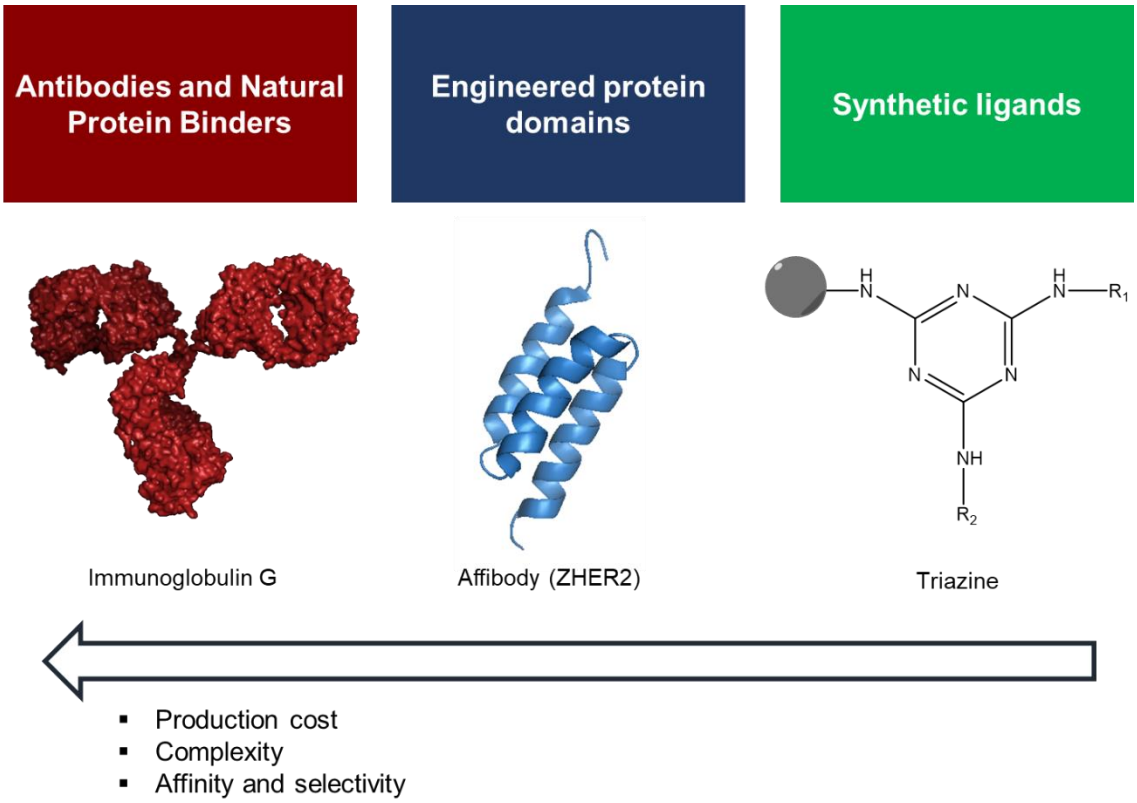

Fig. 1.1 - Different affinity reagents used in bioseparation and their properties variation (PDB code of Immunoglobulin G and Affibody: 1IGT and 2KZJ).

1.2 – Protein scaffolds as affinity reagents

Protein domains can be derived from Immunoglobulin (Ig) fragments and non-Immunoglobulins (non-Ig) structures, for the development of protein affinity reagents [2]. Recently, protein scaffolds have been used to generate combinatorial libraries to further develop affinity reagents to variable targets. This is possible through yeast, phage, ribosome, or cell free display systems [15]. Thus, this may allow the selection of scaffolds in order to adapt to a specific recognition region in a target protein [5]. These libraries were performed at first in Ig fragments by randomizing non-conserved regions. These considered mutations could be based on structural

and functional information of the target, or simply by amino acid combination [15]. Therefore, the objective of combinatorial libraries relies on improving a specific function, while maintaining the protein structure [16]. Thus, this started to be a challenging factor in antibodies engineering, due to their lack of stability and high production cost. Consequently, libraries based on non-Ig structures started to have the interest to produce, due to protein domains advantages referred in 1.1. Their affordable cost of production, for instance, relies on simple synthesis processes, via chemical synthesis or biological production (protein expression in *E. coli*) [5,13].

However, protein scaffolds to be used as affinity reagents for purification purposes must have ideal characteristics. For instance, these proteins should be robust and have small size (less than 20 kDa) [17]; include well-defined regions, for both diversity and stability, i.e. conserved regions to maintain the folding, and variable regions, which can be partially or totally randomized, not affecting the overall structure of the protein; contain a recognition area to promote interaction; not include cysteines and disulphide bridges; be easy to synthesize (chemically or biologically); be robust to an unnatural component introduction; permit an easy, oriented and stable conjugation in solid supports; and keep their stability in the final application conditions. Moreover, modifications can not be limited only to the loops (like in antibodies) or other flat surfaces. Consequently, these scaffolds can be selected to adapt to the target contact surface, promoting crucial interactions between them [5,18].

Protein scaffolds are categorized by various authors in two different families: the globular proteins and the repeat proteins. Domains that are constituted by one peptide chain are classified in globular proteins, thus, repeat proteins chains include two or more repeats in one domain [2]. There is a great variety of protein scaffolds produced by research groups and biotechnological industries, however, just a few have been studied as affinity reagents for purification purpose, next reviewed on this chapter.

Thus, non-Ig based protein scaffolds put together a set of important characteristics that justify their crescent use in biotechnological and pharmaceutical industries. High stability under different thermal and pH conditions, simple production systems (expression or chemical synthesis) and efficient interactions with the targets are the most searched and desired qualities of these proteins. Therefore, they can lead to favourable types of immobilization and also to better affinity purification yields. Currently, affibodies are the most explored and commercialized scaffold for affinity purification purposes. However, different protein scaffolds can have the same potential use.

Globular proteins

Affibody

Affibodies (Fig. 1.2A) are originated from a modified structure of the B domain of SpA, the Z domain, a three-helix structure with 58 residues (6 kDa), without disulphide bonds and independent folding [19]. This domain is easily synthesized through different approaches (biologically expressed and chemically produced) [20]. This structure has evolved by phage and ribosome display against variable targets. It includes a α-helix region, which typically can be randomized [21]. This region is the interaction surface with the targets in the natural protein A. Affibodies have already been studied for therapeutic solutions for some diseases, for example, by Affibody AB, a biotechnological company. Thus, affibody scaffolds have been engineered to target relevant receptors involved in human diseases, for example: cancer, breast cancer marker ErbB2 (Her2) with a 22 pM affinity [22]; and rheumatoid arthritis, TNF-α [23]. Affibodies were also explored in biotechnological applications, such as in sensors, diagnostic, immobilized in microarrays to capture proteins, and in other proteins fusion partner [21,24]. Affibodies also caused the generation of HOT Start DNA polymerases, by binding to Taq DNA polymerase. It was also demonstrated the binding and elution of Human Apolipoprotein A-1 [25]. Moreover, it was possible to capture SpA fusions and dimers in a picomolar range affinity by an anti-protein A affibody with a nanomolar range [26]. Furthermore, affibody-based resins have been used to purify several other targets involved in various diseases. These examples include Alzheimer amyloid β-peptides [27], transferrin [28], human serum albumin (HSA) [8,27,29] or transthyretin [30]. Consequently, this was explored with the purpose of depleting high abundant proteins and allowed the characterization of samples for the presence of low abundant proteins through mass spectrometry [2,30].

Affibody scaffolds were easily immobilized on solid supports, for instance, with cysteine residue groups or His-tags [27]. In affibodies with a free sulfhydryl group from a cysteine residue, they can be immobilized in a resin with an iodoacetyl group (Sulfo-Link coupling, Pierce) [31]. They have also been explored as tags for recombinant fusion proteins purification, through positive or negative ion exchange column [32,33].

There are commercially available affibodies already immobilized in a resin, or including a cysteine residue present for immobilization, such as: Anti-EGFR Affibody, Anti-IgE Affibody, Anti-TNF-a Affibody, or Anti-Interleukin (IL-8), can be purchased through Affibody AB or Abcam, UK. These affinity scaffolds have been used in liquid chromatography columns for depletion of highly abundant proteins and are commercialized in pre-packed column format MARS-7 and MARS-14 columns from Agilent Technologies [2].

Affilin

Affilin (Fig. 1.2B) is derived from human protein g-B-crystallin, present in vertebrates eye lens. This is a protein with 176 amino acids which fold into several β-sheets and presents high stability, highly important for therapeutic applications, but specifically for affinity purification purposes. Human g-B-crystallin structure was randomized and phage display was performed against the nerve growth factor (proNGF) recombinant pro-form. Thus, it was demonstrated the characterization of this affilin for purification of this target. Hence, the antiproNGF affillin was immobilized in an epoxy-activated polymethacrylate resin. This support is extremely robust to regeneration conditions (pH range 1.6–12.5) and the binding to the target is maintained after the referred treatments [34]. Moreover, affilins were already selected to bind the Fc fragment of IgG (10^{-8} M, 10 nM), estradiol, testosterone (100–200 nM) [35], and extradomain B of fibronectin. The latter is a target which expression is practically exclusive in tumour tissues (30 and 200 pM) [36]. This protein scaffold has been explored for different applications as therapeutics and diagnostic. The capture of biomarkers for diagnostic using different solid-supports, spacer arms, and ligand formats (mono or multivalent scaffolds) are important examples of these applications.

Affimer

The Affimer scaffold (Fig. 1.2C) is derived from a protease inhibitor Stefin A (cystatin A). It is a small size (10 kDa) protein, with high thermal (80ºC) and chemical (pH 2–12) stability, and a conformation with two loops. Besides this structure, the loops can be engineered in 12–36 positions [37,38]. These characteristics lead to an increasing interest in affimer studying as affinity reagents. Affimer-based binders can be applied in therapeutic applications, sensing (ELISA or Western-blot), diagnostic and also for affinity purification [39]. These applications are being explored by Avacta Life Sciences company. Regarding affinity purification, the enrichment of Complement C3b (CC3B3) from human serum was achieved. In order to simplify the immobilization in solid-supports procedure, affimers have been selected against the target and modified to have a cysteine residue or a His-tag.

Albumin-binding domain

The albumin-binding domain (ABD) (Fig. 1.2D) is a small domain constituted by 46 amino acids with a three-helix folding. ABD is derived from Streptococcal protein G and was obtained by phage display to increase affinity towards Human Serum Albumin (HSA), from a nanomolar to femtomolar range. This domain was engineeringly randomized in 15 amino acids, belonging to the ABD-HSA interaction region. Resulting from phage display, one of the designs developed

was applied in HSA purification [31], and in the extension of therapeutic proteins half-life, by the production of bivalent domains that can bind to HSA and Tumor Necrosis Factor-a (TNF-α) [40]. The HSA-binder was produced with a cysteine, allowing its immobilization on cross-linked agarose in order to promote HSA capture from human serum.

Carbohydrate binding module

The carbohydrate binding module (CBM) (Fig. 1.2E) is a small domain (18 kDa) derived from a xylanase from *Rhodothermus marinus*. It is a small domain (18 kDa) with a thermally stable structure that binds to soluble poly or oligosaccharides of xylose and glucose. CBM was randomized in 12 residues through a phage display library of 106 clones, against different targets. These targets included xylan cellulose, mannan, and a human glycoprotein (a monoclonal IgG4 antibody). CBM was immobilized in silica to use in the chromatographic separation of saccharides, employing an isocratic gradient of temperatures between 25 and 65ºC. The support capacity could be maintained with treatments at 65ºC for 160 h [41]. Therefore, this work might indicate a conceivable application for purification of various targets using immobilization of CBM in solid-supports.

Nanofitins

Nanofitins (also named as Affitins) (Fig. 1.2F) are derived from protein Sac7d, a double strand DNA-binding protein from *Sulfolobus acidocaldarius*. Nanofitin is a 66 residues protein (7 KDa), without cysteines, and very thermally and chemically stable, since *Sulfolobus acidocaldarius* is a thermophile organism [42]. Therefore, nanofitins are well expressed in *E. coli*, i.e. a simpler production and also affordable [43]. Ribosome display was performed in this protein, being randomized at the surface in two short loops and used against different targets: Lysozyme (for glycosidase activity inhibition) [44], secretin PulD protein [45], and also human IgG. In IgG, it was verified the binding of C3 selected clone to IgG was inhibited by SpA, suggesting the recognized epitope is on the Fc domain of IgG. Another clone (D1) was not inhibited by SpA to bind IgG, thus, demonstrating another binding site on Fc [43]. Due to their high stability, nanofitins have been explored by the company Affilogic, leading to biomedical applications: inflammatory bowel disease treatment by anti-TNF-α therapeutics, via oral administration, is one example. Nanofitins can be immobilized in a variety of surfaces (e.g., glass slides, resins, or magnetic supports) with different approaches, which is important to extend their applications potential. Their immobilization can be performed in solid-supports using His-tags, phosphorylated serine residues, and cysteine residues. Nanofitins affinity purpose can be applied

in detection in microarrays [46,47] and purification in solid-support chromatographic resins or membranes [48].

Recently, anti-lysozyme and antiIgG affitins were produced and immobilized using a cysteine residue, and Sulfo-SMCC chemistry in a magnetic support. This strategy successfully allowed the capture of lysozyme and human IgG from *E. coli* supernatant and human plasma, respectively. This process also was executed with high purity (>95%). Hence, the bioseparation potential can be explored in the future, in the field of vaccines purification. Glycoconjugates, protein antigens, and viruses, now supported by the European project DiViNe.

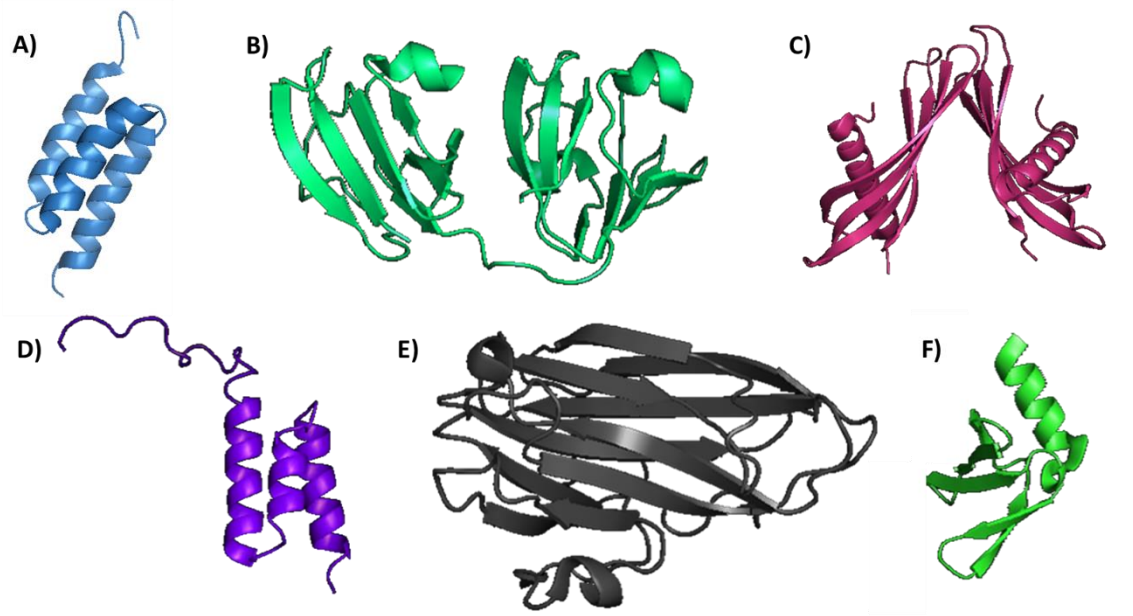

Fig. 1.2 - Globular proteins used in bioseparation. (A) Affibody (PDB code 5VZX). (B) Affilin (PDB code 2JDG). (C) Affimer (PDB code 5A0O). (D) Albumin Binding Domain (ABD) (PDB code 1GJT). (E) Carbohydrate Binding Module (CBM) (PDB code 1K45). (F) Nanofitin (PDB code 4CJ2, chain C).

Repeat Proteins

Armadillo repeat proteins

The armadillo repeat proteins (Fig. 1.3A) are a protein family which is encoded by the Armadillo *locus*, a DNA region in *Drosophila embryogenesis*. The stability of the structures were computationally optimized, applying N and C-terminal modifications with repeats, so the hydrophobic core could be protected from solvent exposure [49]. From this, the obtained structures had high thermal stability, and generally could be used at neutral to high pH [50].

The native structures have been used to capture different families of proteins. For instance, armadillo repeats were expressed with FLAG tag and immobilized in anti-FLAG antibody resins.

These resins showed the armadillo proteins could capture Striatin, SG2NA, and Zinedin (neuronal proteins) [51]. In a different example, it was randomized a designed armadillo repeat protein in 6 positions in each 42 amino acid repeat, producing a variety per repeat. This protein was selected against peptide neurotensin through ribosome display. Neurotensin was immobilized in streptavidin beads and incubated with an *E. coli* extract that included the armadillo repeat. Thus, a very specific binding could be verified, besides the low affinity observed (range of 7 mM) [52].

Repebody

Repebody (Fig. 1.3B) derives from variable lymphocyte receptors (VLRs), a three to six repeating Leucine-rich repeat (LRR) modules assembly. These modules hold 20–29 residues protein with a β-strand-turn-α-helix structure. These scaffolds are characterized by a high soluble expression in *E. coli*, and strong thermal (70–85ºC) and pH (3–12) stability. Repebodies have been currently engineered in order to obtain binders with different affinity against a target [53] basing on their modular structure. Therefore, phage display has been performed against interleukin-6 (IL-6), and lysozyme [54] and lately was used for the purification of IgG. The structure was immobilized in an N-hydroxysuccinimide (NHS)-activated resin and the purification of IgG performed at low pH. This strategy allowed the recovery of more than 96% of the protein of interest with high purity (95%). This support was also very stable to regeneration and allowed reutilization [55].

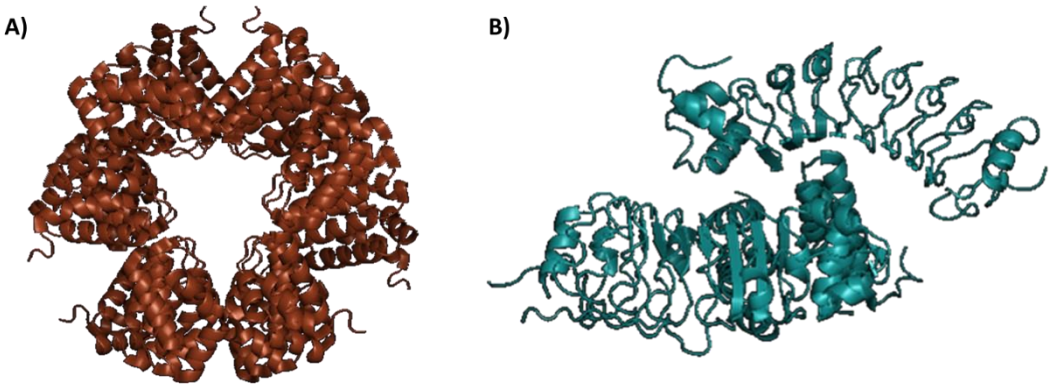

Fig. 1.3 - Repeat proteins used in bioseparation. (A) Armadillo Repeat Protein (PDB code 4DB9). (B) Repebody bound to human C5a complex (PDB code 5B4P).

1.3 – WW domains

WW domains are proteins with 38 to 40 residues in length, containing two tryptophan residues spaced by approximately 20 amino acids, which can develop an important role in protein function and stability [56,57]. These domains are assembled in a compact three-stranded and antiparallel

β-sheet structure [58]. This structure can promote protein-protein interactions, especially in proline-rich sequences, being frequently found in recognition regions. These interactions commonly occur in signalling pathways, although they are known as weak [59]. They are one of the smallest proteins which can fold into a monomer in solution, without disulphide bridges and cofactors [60]. This absence of cysteine residues was a good purpose for primary studies of WW domains as affinity reagents for purification purposes.

WW domains can be classified in two major groups and also in two minor groups [56]. The first major group (Group I), can bind polypeptides with Pro-Pro-X-Tyr residues; Group II, binds Pro-Pro-Leu-Pro motifs; minor groups can bind polyproline motifs flanked by Arg or Lys (Group III); and ligands that include phospho-Ser-Pro or phospho-Thr-Pro groups (Group IV) [59,61,62]. Despite not requiring cysteine residues for their folding, WW domains strongly necessitate hydrophobic interactions between conserved residues (for instance, proline and tryptophan residues in the both N and C terminals) and hydrogen bond interactions between in β-sheets amino acids [60].

The regions with higher variability in WW domains are the loops. *Macias et al* developed a WW prototype sequence as a protein scaffold, that allowed further use of this domain as a base for increasing, engineering and randomizing the loop sequences [63]. This would turn WW domains promising and easy engineered proteins for further binders selection and exploiting several biotechnological applications. Moreover, these domains respect most of characteristics that a potential affinity ligand should have. For instance, they do not include cysteines and disulphide bridges, they contain a recognition region for protein interaction, allow unnatural component incorporation and include well-defined regions, for diversity and stability, including variable regions that can be randomized. Therefore, WW domains can be highly explored as affinity reagents for purification purposes.

1.4 – Objectives

The design of a WW prototype sequence allowed the development of a novel WW domain with tailored affinity and selectivity towards relevant targets. WW domains are scaffolds in which the loop regions can be easily engineered and randomized towards a possible binder molecule, through an *in vitro* selection process. The target chosen to this process was human serum albumin (HSA), due to its functional properties as a carrier protein. HSA is an abundant protein in human plasma and possesses interesting biomedical and biotechnological properties [64]. Its carrier function allows the transportation of relevant compounds, such as small drugs, fatty acids and hormones [65]. HSA also accumulates in inflammation regions and can be used as an important diagnostic tool [66]. In addition, protein domains that bind to HSA had been developing interest. These domains can be natural albumin binders or derived from Ig or non-Ig fragments further engineered to improve affinity towards HSA, by *in vitro* or *in vivo* evolution techniques [67].

In a previous work [68], a library of WW randomized clones was built and screened against HSA. From this process, a 42 residue sequence was selected, named CW3 (WW domain clone 3) (Fig. 1.4). This sequence was engineered in order to overcome problems in peptide production, such as the presence of methionine and cysteine residues, that could promote disulphide bridges and the peptide oxidation. In addition, a cysteine residue was added to the N-terminal with the aim of promoting immobilization in a matrix. The sulphide group in cysteine could be used to facilitate immobilization for further chromatography and affinity purposes. The chemical synthesis of CW3S has been attempted in a previous work [68]. However, the expected molecular weight was never achieved, indicating that a Serine residue was not efficiently coupled during synthesis process. Nevertheless, characterization, immobilization and further affinity studies were performed, demonstrating a promising domain with positive affinity ligand properties (binding rate against HSA of 88%, in chromatography binding tests, capturing 0.291 μg protein/mg support.; k_a=8.37x10^6 M^{-1}) [68].

The objective of this work was to show that the chemical and biological production of CW3S was possible and that the produced protein showed binding for HSA. As such, this work is focused on the production of the CW3S peptide through chemical and biological routes, its characterization and study of affinity towards HSA. In this matter, several experiments were optimized and new affinity strategies were tried.

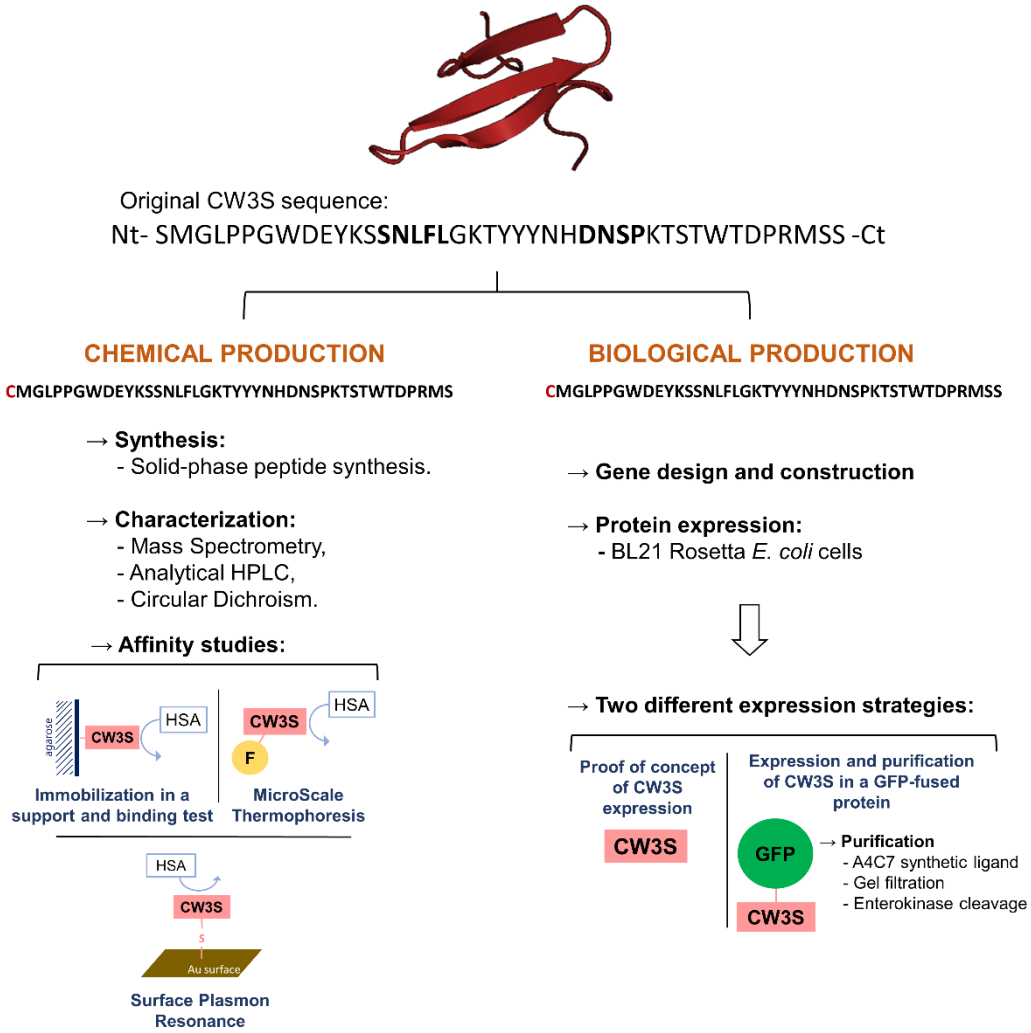

Fig. 1.4 - Representative scheme of the work planning. Optimization of the previous CW3S sequence obtained from phage display. Peptide chemical production: synthesis, characterization and affinity studies performed. Peptide biological production: techniques and strategies used in cloning, expression and purification (PBD code of WW prototype: 1E0M).

Chapter 2

Chemical production, characterization and affinity studies of CW3S, as a promising affinity reagent

Chapter 2

This chapter approaches the chemical synthesis, purification and characterization of the CW3S domain and the study of its potential as an affinity reagent for HSA.

The CW3S was chemically produced using Fmoc-based solid-phase peptide synthesis (SPPS) (Fig. 2.1). Fmoc-based SPPS strategy is based in the use of Fmoc-protected N-terminal amino acids and a solid support (resin). These amino acids contain orthogonal side chain protecting groups when needed. Namely, protecting groups that are resistant to the basic conditions used to remove the Fmoc in each coupling step. These protections are used to avoid side chain reactions that will reduce the overall yield [69]. The amino acids are assembled on the resin from the C-terminal to the N-terminal. After the peptide is fully assembled, final deprotection and cleavage from the resin is carried out under acidic conditions. The crude peptide obtained is then purified by reversed-phase HPLC and characterized by Mass Spectrometry (MS) methods. Its purity is usually determined by analytical HPLC. The overall yield and the quality of the crude peptide obtained by SPPS can be improved by using microwave technology, special resins with better swelling properties (i.e. PEG-based resins) and the use of special protected amino acids that can improve the yield of coupling steps in long or difficult peptide sequences.

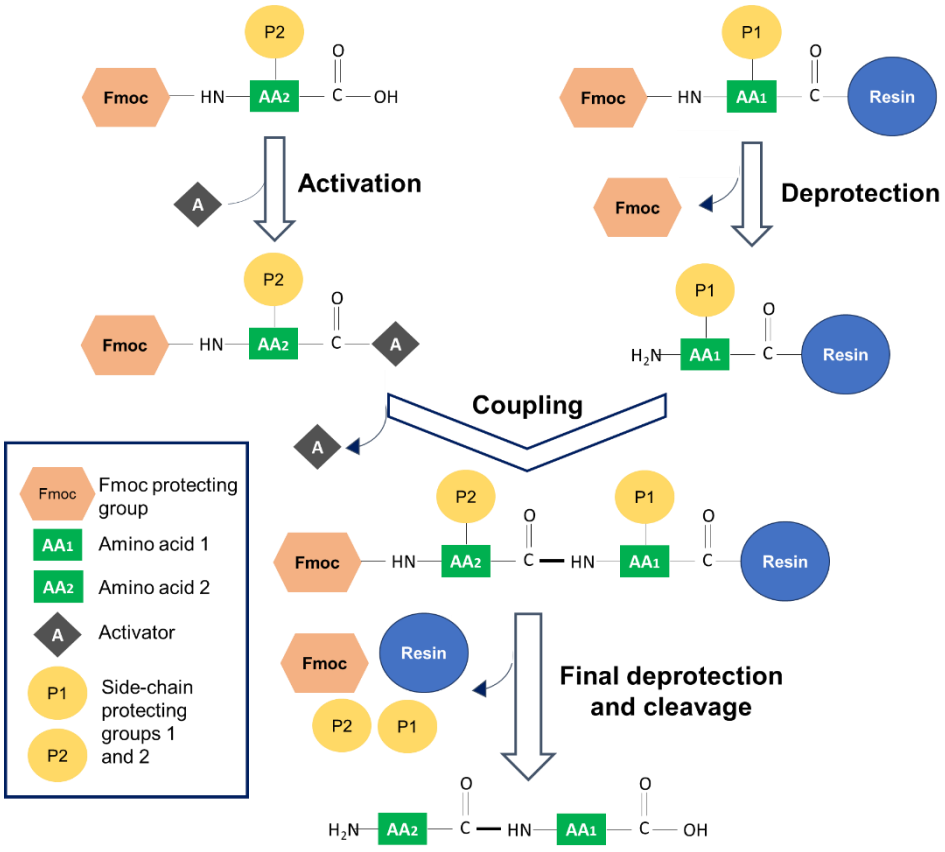

Fig. 2.1 – Scheme of solid-phase peptide synthesis (SPPS), adapted from Sigma-Aldrich.

CW3S has 41 residues and the sequence is shown in Fig. 1.4. The C-terminal Serine was depleted from the previous produced CW3S. This decision was taken considering that the mass spectrometry results of the previous synthesis indicated that one Serine was missing and the fact that this Serine is at the N-terminal and therefore the chances to affect the final structure of the peptide are minimal. Thus, the synthesized sequence is:

Acetyl-CMGLPPGWDEYKSSNLFLGKTYYYNHDNSPKTSTWTDPRMS-Amide

The chemically produced CW3S scaffold was purified by reversed-phase preparative HPLC, characterized by MS and Circular Dichroism (CD) spectroscopy and its purity determined by reversed-phase analytical HPLC. Afterwards, the peptide was immobilized in a solid support and finally, affinity studies were carried out using different techniques including chromatographic tests, MicroScale Thermophoresis and Multiparametric Surface Plasmon Resonance.

2.1 – Materials and Equipment

All Fmoc protected amino acids, 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), NOVA-PEG Rink amide resin and pseudoproline units (Fmoc-Lys(Boc)-Thr(ΨMe,Mepro)-OH and Fmoc-Ser(tBu)-Ser(ΨMe,Mepro)-OH), were purchased from EMD Biosciences/Merck Biosciences (Darmstadt, Germany). Solid-phase synthesis was performed in Biotage® Initiator+ Alstra™ Automated Microwave Peptide Synthesizer. Mass Spectrometry (ESI-MS) was performed in positive mode using a Waters Synapt G2 HDMS (Manchester, UK), and Circular Dichroism in a Chirascan Circular Dichroism Spectrometer. For peptide immobilization, agarose Sepharose CL-6B (GE Healthcare Life Sciences), Sulfo-SMCC (Pierce, ThermoScientific) and Bond-Breaker tris(2-carboxyethyl)phosphine (TCEP) (Thermo Fisher Scientific) were used. Protein quantification in binding studies was performed with QuantiPro BCA Assay Kit from Sigma-Aldrich. For MST and peptide labelling, Fluorescein Diacetate 5-Maleimide (Sigma-Aldrich) and PD10 desalting column (GE Healthcare Life Sciences) were used. MST assays were performed with Monolith NT.115 Standard Treated Capillary K002 in NT.115 Nanotemper, with filter BLUE/RED (for fluorescein, blue filter was used). In MP-SPR, Au-surface slides are provided by Bionavis and the assay is performed in MP-SPR Navi 200 (Bionavis).

2.2 – Methods

2.2.1 – Solid-phase peptide synthesis (SPPS)

The peptide chemical synthesis (0.25 mmol scale) was performed in a Biotage® Initiator+ Alstra™ Automated Microwave Peptide Synthesizer using the Rink Amide NOVA PEG resin (0.53 mmol/g substitution). The resin was added in a 10 mL reaction vessel where the synthesis took place. Standard Fmoc chemistry protocols were employed. Namely, the amino acids (4 equivalents) were coupled (75° C, 5 minutes) using HBTU (3.9 equivalents) as coupling agent, DIEA as base (8 equivalents) and DMF as solvent. For aspartate, histidine and cysteine the coupling reaction was carried out at 50ºC for 13 minutes. Double coupling was employed on arginine, proline, threonine, tryptophan and pseudoproline amino acids. The removal of the Fmoc protecting groups was always done by treatments with 20% piperidine in DMF solution (room temperature, 13 minutes). After the 23rd amino acid, the peptide was acetylated in the N-terminal using 10% acetic anhydride in DMF (room temperature, 10 minutes). After full assembly, the peptide was manually deprotected and cleaved from the resin by treatment with the mixture TFA/thioanisole/1,2 ethanedithiol/anisole (%v/v=90:5:3:2) for 3 h at room temperature and under nitrogen. The resin was filtered out and rinsed with TFA. The filtrate and rinses were combined and reduced under a nitrogen stream. Cold diethyl ether was added to precipitate the crude peptide which was recovered by centrifugation (8000 xg for 4 min). The pellet collected was dissolved in the minimum amount of water and lyophilized.

The crude CW3S was purified by reversed-phase preparative HPLC in a Jupiter C18 15µm 300Å column (250 x 21.2 mm) using solvent A (water/TFA, 99.9:0.1 v/v), solvent B (acetonitrile/water/TFA, 90:9.9:0.1 v/v) and a linear gradient of 20-60% solvent B for 30 minutes with a flow rate of 10 mL/min. The peptide was eluted from the column with a retention time of 20.65 minutes. The purity of the CW3S was determined by reversed-phase analytical HPLC using a Jupiter C18 15µm 300Å column (250 x 4.6 mm) and the same solvents and linear gradient of 0-100% solvent B, with a flow rate of 1 mL/min. The HPLC chromatograms were monitored at 220 nm and 280 nm. The pure peptide was characterized by Mass Spectrometry (ESI-MS) in positive mode using a Waters Synapt G2 HDMS (Manchester, UK) equipped with an ESI source employing the following parameters: ESI capillary voltage: +2.8 kV; extraction cone voltage: +20 V; desolvation gas (N_2) flow: 100 L.h^{-1}; source temperature: 35°C. The sample was dissolved in 300 µL of water and diluted 1:10 in 0.5% formic acid in methanol. This final solution was introduced into the ionization source at a flow rate of 10 µL/min using a syringe pump. These experiments were performed at Spectropole, the Analytical Facility of Aix-Marseille University. Calculated ESI Data (m/z): 4831.4 ± 0.5 $[M +1H]^{1+}$, 2416.2 ± 0.5 $[M + 2H]^{2+}$, 1611.1 ± 0.5 [M

$+3H]^{3+}$, 1208.6 ± 0.5 $[M +4H]^{4+}$ and 967.1 ± 0.5 $[M +5H]^{5+}$ and molecular weight of 4830.4 Da (using Peptide Molecular Weight Calculator https://pepetide2.com).

2.2.2 – Circular Dichroism (CD)

The CD spectra were collected under a constant flow of nitrogen on a qCD Chirascan spectrometer equipped with a thermostated cell holder, using a 1 mm path length quartz cell. The stock peptide solution was prepared with Milli-Q water previously purged with nitrogen and its concentration was determined by UV-Vis spectroscopy at 280 nm with 6 M GdnHCl, using the extinction coefficients of tryptophan (5690 $M^{-1}.cm^{-1}$) and tyrosine (1280 $M^{-1}.cm^{-1}$) [70]. Peptide concentration was also determined by the Ellman's test [71] (quantification of free cysteines by their reaction with the reagent DTNB, using the extinction coefficient of TNB, the product formed with DTNB and a free sulfhydryl group reaction: 13650 $M^{-1}.cm^{-1}$) using UV-Vis spectroscopy and monitoring the absorbance at 410 nm.

All the assays for the CD spectra of CW3S were carried out using 50 µM peptide in 10 mM sodium-phosphate, 100 mM NaCl pH 7 buffer. The spectra for the stability experiments were recorded in the wavelength range of 200-300 nm at 4ºC, 23ºC and 88ºC, with 2 to 5 minutes of equilibration in each temperature. The temperature denaturation studies were performed between 4ºC and 88ºC (unfolding curve from 4 to 88ºC and refolding from 88 to 4ºC) with a step of 1 nm, tolerance of 2ºC and 1.5 min equilibration time between temperatures. Ellipticity was monitored at 230 nm. The following parameters were used during data acquisition: 0.2 nm scale, continuous scanning, 3 repetitions of each spectrum with 3 sec/nm. The molar residue ellipticities (MRE, deg cm^2 $dmol^{-1}$) were calculated by Equation 2.1, where Θ_{abs} is the ellipticity in millidegrees, C the molar concentration, l the cell path length in cm, *residue* the total number of residues in the peptide sequence and *10* the conversion factor to dmol. The fitting of the unfolding results in a two-state model was performed using *Greenfield et al, 2007* [72].

$$MRE = \frac{\Theta abs}{C \times l \times residue \times 10} \quad \text{(Equation 2.1)}$$

2.2.3 – Immobilization in solid support

The CW3S peptide was immobilized in a chromatographic support (cross-linked agarose), using the peptide N-terminal cysteine with the Sulfo-SMCC chemistry (Pierce, ThermoScientific). Sulfo-SMCC is a cross-linking agent which can link an aminated support with a molecule containing a thiol group, in this case a cysteine.

2.2.3.1 – Solid support epoxy-activation and amination

The solid support used to immobilize the peptide was Sepharose CL-6B (GE Healthcare Life Sciences). 10 g of Sepharose CL-6B were epoxy-activated. For this, it was followed the method described in the literature [73,74] with few adaptations. Firstly, the resin was washed with distilled water in a glass filter funnel and drained. Next, the agarose was resuspended in water (1 mL/g) and in 10 M NaOH (0.040 mL/g), and incubated 30 minutes at 30ºC with orbital shaking (230 rpm). Then, epichlorohydrin (99%) was added in the proportion of 0.072 mL/g moist agarose. The slurry was incubated for 3h at 36ºC (230 rpm) and at the end washed with water at vacuum. The amount of epoxy groups in agarose was determined by mixing 3 mL of 1.3 M of sodium thiosulfate with 1 g of agarose, and incubated 20 minutes at room temperature. Then, the solution was titrated with 0.1 M HCl until the pH reached approximately 7. The amount of epoxy groups was determined from the amount of HCl needed in order to maintain neutrality: Volume (mL)/10 [73]. The total amount of epoxy-groups was about 24 µmol/ g agarose.

The amination of the epoxy-activated agarose was performed with 5 molar excess conditions of 1,4-diaminobutane (Sigma-Aldrich). The slurry was incubated overnight at 40ºC with 200 rpm of orbital shaking. The resin was washed about 10x the total volume with distilled water. The content of amines was determined using Kaiser test (24 µmol/ g agarose) [75].

2.2.3.2 – Immobilization of CW3S in the solid aminated support

To immobilize the peptide in the aminated support, the following solutions were used: 35% of 50 mM HEPES, 1 mM EDTA pH 7.5 in DMF, as the immobilization buffer; and 50 mM HEPES pH 7.5 to dissolve Sulfo-SMCC. All buffers were de-aerated with a nitrogen stream before use. The resin was previously washed with immobilization buffer. Sulfo-SMCC was used according to the manufacturer instructions (Thermo Scientific). It was used a 0.5 µmol peptide: 1 µmol aminated support ratio and a 5 molar excess of 10 mg/mL of Sulfo-SMCC. The cross-linking agent and the aminated support were firstly incubated 30 minutes at room temperature with agitation (35 rpm). It was prepared a 5 mg/mL peptide solution with Bond-Breaker TCEP reducing agent in a 1:100 ratio. The CW3S with reducing agent solution was incubated 30 minutes at room temperature in dark. Before adding the peptide solution to the activated support, the latter was washed with immobilization buffer, in order to remove the excess of Sulfo-SMCC that did not react. Then, the immobilization reaction of the reduced peptide solution with the Sulfo-activated support was incubated for 1h at room temperature with agitation (35 rpm). The resulting solution was washed with immobilization buffer to remove the peptide excess that did not react, and it was added 10 mg/mL of L-cysteine, as a blocking agent to the unreacted sites in the support (30 minutes incubation, at room temperature and rotary agitation of 35 rpm). Finally, the flow-

through and washes were quantified by fluorescence (280-340 nm) to determine the amount of peptide that was immobilized in the solid support. The immobilized peptide was stored overnight in solution at 4ºC.

2.2.4 – Affinity studies

The affinity studies consisted in the assessment of binding between the immobilized peptide (CW3S-Ag) and two different targets: HSA and IgG. Two buffer solutions were employed, namely PBS 1x (10 mM sodium-phosphate, 150 mM NaCl, pH 7.4) for both binding and elution, and 0.1 M NaOH in 30% isopropanol as regeneration buffer.

Firstly, CW3S-Ag was washed with water and regenerated 3 times with regeneration buffer. These washes were alternated with 3 distilled water washes. Then, the resin was equilibrated with binding buffer by several washes until the Abs_{280} was inferior to 0.005. The targets were added to 50 mg of CW3S-Ag in duplicates, in a concentration of 66 µg/mL (20 µg target in a total volume of 0.3 mL). Negative controls were performed in the same method, with only aminated agarose instead. The immobilized peptide was incubated with the different targets at 4ºC, for 2h (manually mixed every 30 minutes until 1h incubation). Afterwards, the flow-through was collected and the resin was washed 8 times with binding buffer at 4ºC. The elution was performed with the same buffer, but 6 times at 23ºC. The loading, flow-through, washes and elutions were quantified by QuantiPro BCA Assay Kit from Sigma-Aldrich. This assay was performed according manufacturer indications. Micro BCA solvents Q_A, Q_B and Q_C were mixed in this order with the proportion of 25:25:1. 150 µL of sample with 150 µL of micro BCA reagent were mixed in a Starsted 96-well transparent microplate, that was incubated for 1h at 60ºC. The absorbance at 560 nm was then measured. Binding (%) and elution (%) were determined by Equations 2.2 and 2.3.

$$Binding\ (\%) = \frac{mg\ bound}{mg\ loaded} \times 100 \text{ (Equation 2.2)}$$

$$Elution\ (\%) = \frac{mg\ eluted}{mg\ bound} \times 100 \text{ (Equation 2.3)}$$

2.2.5 – MicroScale Thermophoresis (MST)

MST is a technique to quantify biomolecular interactions, based in the concept of thermophoresis, which is the directed movement of molecules in a temperature gradient, that depends on a variety of molecular properties (e.g. size, charge, hydration shell or conformation). In order to detect and quantify the directed movement of molecules, CW3S peptide labelled with

a fluorescent agent (Fluorescein Diacetate 5-Maleimide) was tested for binding HSA, the ligand molecule.

2.2.5.1 – CW3S labelling with Fluorescein Diacetate 5-Maleimide (FDM)

CW3S labelling was performed using the Fluorescein 5-Maleimide protocol, provided by Thermo Fisher Scientific. A 15 fold molar excess of FDM was used into 1 mg/mL of CW3S, giving a final concentration of 3.105 mM of the fluorophore. Since FDM is not soluble in aqueous media, for concentrations higher than 1 mM, the fluorophore could be dissolved in DMF to a higher concentration of 6.21 mM (stock solution) and then diluted in 1:2 in the required labelling buffer (20 mM sodium phosphate buffer, 150 mM NaCl, 5 mM EDTA, pH 7.5) to reach the desired solution 3.105 mM in 50% DMF, 20 mM sodium phosphate buffer, 150 mM NaCl, 5 mM EDTA, pH 7.5.

To separate the labelled protein from not-labelled FDM and CW3S, a PD10 desalting column was used, with a 5 KDa cut-off. The equilibration buffer used was the binding buffer to perform MST experiments, PBS (10 mM sodium phosphate, 150 mM NaCl) pH 7.4. The obtained fractions were quantified by A_{280}, A_{495}, fluorescence ($\lambda_{excitation}$ = 345 nm and $\lambda_{emission}$ = 485 nm) and BCA assay. The determination of the protein concentration and labelling degree was achieved using the following equations, provided by FDM's supplier:

$$[Protein](M) = \frac{A_{280} - (A_{495} \times 0.3)}{\varepsilon} \quad \text{(Equation 2.4)}$$

$$mol\ (FDM)\ /\ mol\ (peptide) = \frac{A_{max}\ labelled\ peptide}{\varepsilon' \times peptide\ concentration\ (M)} \times dilution\ factor \quad \text{(Equation 2.5)}$$

BCA Reagents A and B were mixed with the proportion of 50:1. 25 µL of sample with 200 µL of BCA solution were mixed in a Starsted 96-well transparent microplate, that was incubated for 30 minutes at 37ºC. The plate was read with absorbance at 560 nm.

2.2.5.2 – MST assays

The MST experiment was performed using 0.1 µM of labelled CW3S, per assay. The maximum concentration of HSA used was 1 µM. The assay buffer used was PBS pH 7.4, and as system settings, the experiment proceeded with auto-detect excitation power of 80% and medium MST-power. 16 samples were prepared, all of them with the same concentration of CW3S (0.1 µM). The ligand was applied in serial dilutions of 1:2 from the first to the last sample, being the

maximum concentration 1 µM. These samples were transferred to 16 standard capillaries (Monolith NT.115 Standard Treated Capillary K002) (V ≈ 4 µL) and the signal read in Monolith NT.115 (Nanotemper).

2.2.6 – Multi-Parametric Surface Plasmon Resonance (MP-SPR)

The MP-SPR assay was performed using a gold (Au) surface chip modified in-house. The chip was washed for 10 minutes in a previously boiled solution of 30% ammonia (NH_4OH), 30% hydrogen peroxide (H_2O_2) in distilled water, in 1:1:5 proportion. Then, the Au sensor was rinsed in milliQ water and dried with a nitrogen air steam. For peptide immobilization, it was prepared: 10 µM of CW3S with TCEP reducing agent (1:100) and 1 M ethanethiol. For the target, seven concentrations up to 50 µM of HSA (0, 0.1, 0.5, 5, 10, 20 and 50 µM) were also prepared. All samples were dissolved in PBS pH 7.4. In the following assay, the liquid chamber was used. A solution of CW3S (250 µL of 10 µM) was injected in the chamber and incubated for 15 minutes. This step was repeated to confirm that the Au-surface was modified and saturated. Then, 250 µL of 1 M ethanethiol blocking agent was added and incubated for 15 minutes. After each incubation, the Au-surface was washed with PBS pH 7.4 for 10 minutes, before the following injection. Then, the target samples were injected. In each 250 µL injection, HSA was incubated for 10 minutes, followed by 10 minutes in regeneration conditions (addition of 0.1 M NaOH in 30% isopropanol) and a 10 minutes wash with PBS pH 7.4. The assay was performed with a flow-rate of 20 µL/min at 20ºC.

2.3 – Results and discussion

2.3.1 – CW3S synthesis and characterization

Chemical synthesis is a strategy already used for other protein domains production. Affibodies (through native chemical ligation) [20] and also hPin1 [76,77] and hYAP65 [78] WW domains are examples of protein domains in which this approach was positively used. Additionally, solid-phase peptide synthesis (SPPS) and related chemical strategies are commonly used in proteins to introduce markers, labels and residues that facilitate, for instance, immobilization or detection [69,77]. The CW3S peptide is a WW domain with 41 amino acids and its chemical synthesis had to be well-designed to obtain good yields. This research group has already explored different SPPS strategies to improve the yield of the chemical synthesis of other WW domains [77,78]. The use of microwave technology promotes more efficient amino acid coupling reactions [79]. In addition, PEG resins (hydrophobic resins) can facilitate difficult chemical synthesis processes [80]. In large peptides or peptides with tendency to aggregate, such as WW domains, the introduction of pseudo-proline units improve the yields of the production [77,81]. Pseudo-prolines are oxazolidine dipeptides described as secondary structure disruptors, which can overcome errors and problems in the synthesis of difficult peptides. In this case, the introduction of these units was possible since the sequences of CW3S includes Serine and Threonine residues, for which pseudo-proline units are available. Double-coupling reactions were included in several amino acids for more efficient coupling reactions. Moreover, to reduce the formation of truncated sequences and obtain crude peptides of better quality, capping reactions were also performed after the 23rd amino acid residue. The peptide was fully deprotected and cleaved from the resin using standard protocols and purified by reversed-phase preparative HPLC.

Fig. 2.2A shows the preparative HPLC chromatogram of the crude peptide. The peak at 20.86 minutes was collected and lyophilized. The purity of the collected peak was determined by analytical HPLC (Fig. 2.2B) and it was greater than 95%. ESI-MS analysis confirmed the identity of the CW3S peptide (Fig. 2.2C).

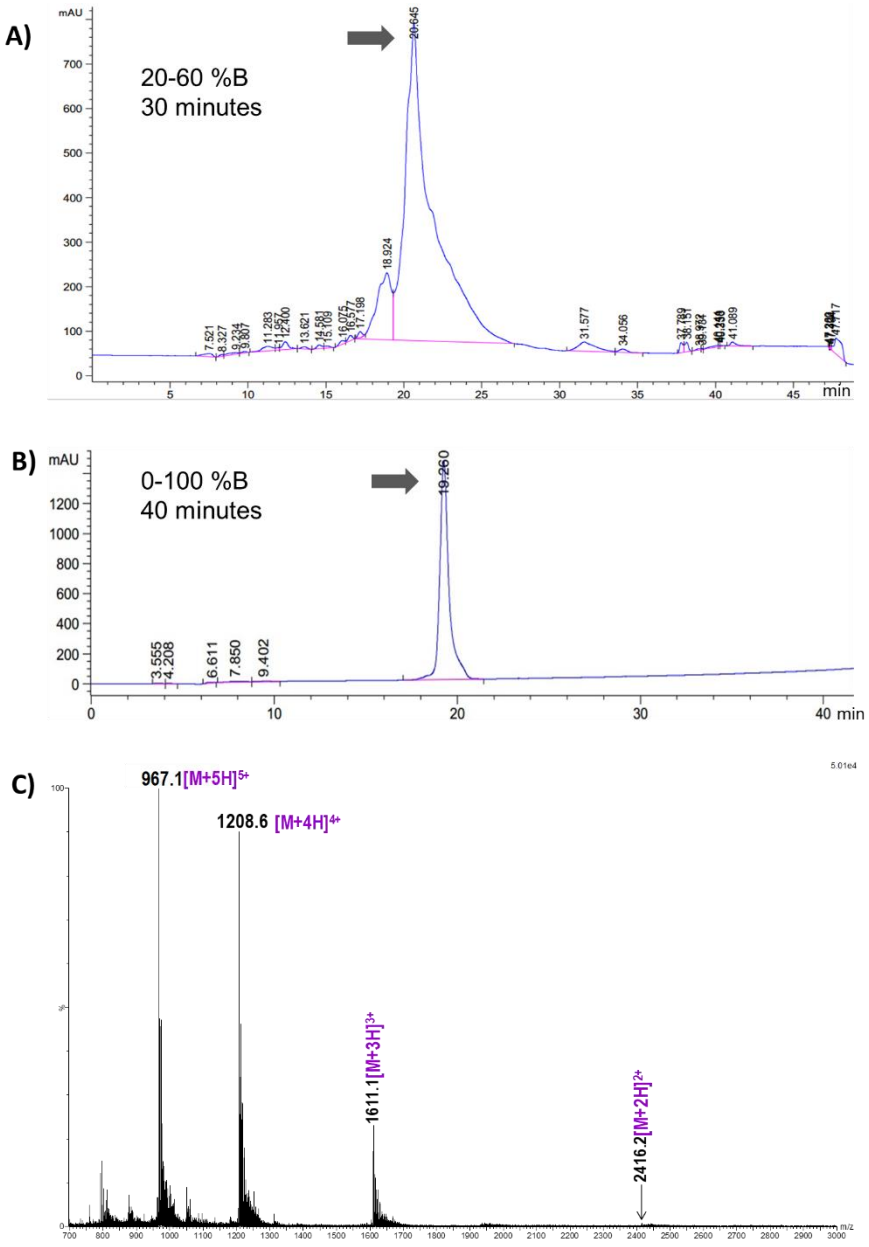

Fig. 2.2 – HPLC purification and MS identification of the crude peptide. (A) Preparative HPLC chromatogram of the crude peptide. Collected peak is identified by an arrow (retention time: 20.65) and the gradient used represented in the figure. (B) Corresponding analytical HPLC chromatogram of the collected and purified peak. (C) Mass spectrum of the collected peak obtained by ESI-MS. [M+2H]2+ 2416.2 (calculated)/2416.2 (determined), [M+3H]3+ 1611.1 (calculated)/1611.1 (determined), [M+4H]4+ 1208.6 (calculated)/1208.6 (determined), [M+5H]5+ 967.1 (calculated)/967.1 (determined).

The results shown in Fig. 2.2 corroborate the success of the chemical synthesis of the CW3S peptide. The use of PEG-based resin (Rink Amide Nova-PEG), the incorporation of two pseudo-prolines and the fully optimized SPPS protocol (coupling and capping reactions) contributed to this achievement [77,80,81].

To determine the peptide folding and characterize its thermal stability, CD spectroscopy was performed. The conditions used in the CD assays were the same as those previously applied in the study of other WW domains (hPin1 [77] and hYAP65 [78]). WW domains have a characteristic CD signal showing a maximum positive ellipticity at 230 nm and a maximum negative ellipticity at 206 nm [57]. The CD spectrum of the CW3S peptide was recorded in Far-UV region at different temperatures, namely 4°C, 23ºC and 88°C (Fig. 2.3A). At 4 °C, the CD spectrum shows only the positive ellipticity at 230 nm. At 88°C this ellipticity is not observed indicating that the peptide is unfolded. At 23ºC, the 230 nm signal is observed but with lower intensity, which is indicative of an intermediate state between 4ºC and 88ºC. After the peptide was denatured at 88ºC, the solution was cooled down to 4°C and the CD spectrum was recorded again (Fig. 2.3A). The data suggests that the peptide is able to refold back after denaturation. However, the ellipticity at 230 nm is lower than that observed before denaturation.

The temperature denaturation studies were performed between 4 °C and 88 °C (melting curve) by monitoring the change in signal at 230 nm (Fig. 2.3B). For the melting temperature (Tm) determination, the experimental data were fitted to a two-state model as described in methods. The Tm obtained was 28.95ºC. This value is lower in comparison with the Tm values of previously studied WW domain (54.2 ± 6.3 °C; R^2=0.97) [68] and to the WW prototype sequence (44.2 ± 0.2 °C; R^2=0.99) [63]. Thus, the determined Tm indicates a decrease in the thermal stability of CW3S. Specially, there is a significant difference between the 42 residue CW3S produced in previous work, and this one (41 residues). The reason of this difference is currently not fully understood since the simple depletion of the N-terminal amino acid is not most likely the origin. However, this Tm is comparable to other explored WW domains, in this case the 38 residue version of hYAP65 which has Tm values of 31.4 ± 2.9 °C (R^2=0.99) and 28.5 ± 2.6 °C (R^2=0.99) (produced through two different chemical synthesis strategies) [78].

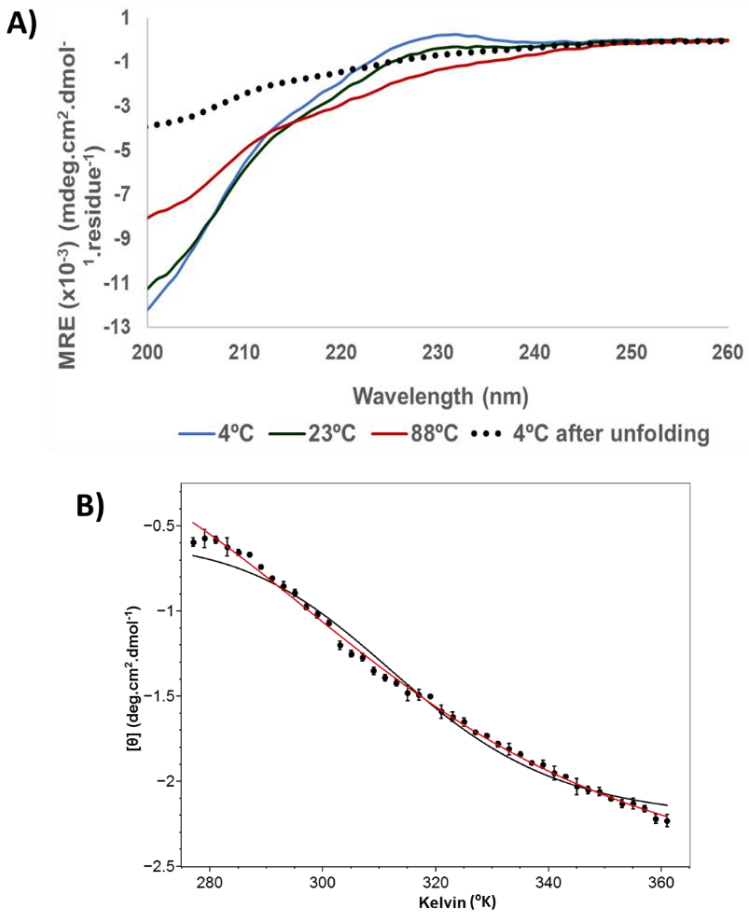

Fig. 2.3 – CW3S folding and thermal stability analysed through Circular Dichroism. (A) Far-UV CD spectra at 4ºC, 23ºC, 88°C and refolding at 4ºC. (B) Fitting to a two state-model curve for determination of melting temperature (Tm) between 4ºC and 88°C.

2.3.2 – Immobilization of CW3S in a chromatographic support and binding assessment

The purification of HSA is possible through depletion, capture and separation processes. Different commercial kits are available, that include synthetic ligands, Ig/non-Ig fragments, which typically operate, at high salt concentrations in binding and elution conditions. The latter can be especially harsh for proteins. Moreover, available synthetic ligands present lower specificity towards HSA, and antibody-based resins are too expensive. Therefore, the possibility to use CW3S as an affinity ligand for HSA selective recovery is a promising strategy, presenting less expensive materials and possibly higher specificity. In addition, binding to IgG was also tested in this assay to explore the selectivity of the CW3S ligand.

The solid support chosen to CW3S immobilization was cross-linked agarose (Sepharose CL-6B), using the N-terminal cysteine residue of the peptide. Sulfo-SMCC was already successfully used for immobilization in cross-linked agarose, providing a chemistry for an oriented immobilization in the chromatographic support [77]. In the present study, the immobilization

yield of CW3S was 85% with a final density of $4.96x10^{-3}$ µmol peptide/mg support. This value was within the range of previous immobilization procedures in CW3S peptide [68].

According to data previously obtained, the binding experiment was performed at 4ºC and the elution at 23ºC, in PBS pH 7.4 [68,77]. Fig. 2.4 and Table 2.1A show that the immobilized CW3S (CW3S-Ag) bound about 29.4 % of the IgG loaded (0.12 µg peptide/mg support) and 17.2% of the HSA loaded (0.07 µg peptide/mg support). In the latter, these values are similar to the binding of unfunctionalized agarose to HSA (negative control).

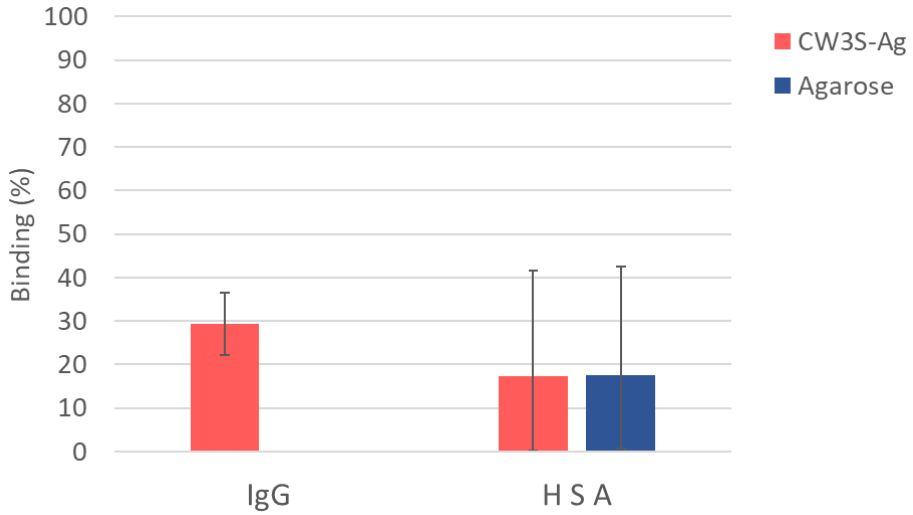

Fig. 2.4 – Binding studies of the CW3S peptide immobilized in agarose solid support (CW3S-Ag) with Immunoglobulin G (IgG) and Human Serum Albumin (HSA), in pink. Blue columns correspond to negative control assay, i.e., binding between unfunctionalized agarose and both same target proteins.

Table 2.1 – A) CW3S-Ag and Agarose binding to IgG and HSA (%), determined by Equation 2.2, and amount of target that bound per amount of support (µg target/mg support). B) IgG and HSA elution (%), determined by Equation 2.3, and amount of eluted target per protein bound (µg target/µg bound).

	A) Binding (CW3S to target)		B) Elution of IgG and HSA	
Target	**CW3S-Ag**	**Agarose**	**CW3S-Ag**	**Agarose**
IgG	29.4% ± 7.2 0.12 ug/mg support	0.0% ± 0.0	5.0% ± 0.3 0.05 ug/ug bound	0.0% ± 0.0
HSA	17.2% ± 24.4 0.07 ug/mg support	17.6% ± 24.9 0.06 ug/mg support	1.7% ± 2.4 0.03 ug/ug bound	38.1% ± 53.8 1.81 ug/ug bound

In previous tests, HSA binding to CW3S-Ag was 88% [68]. Despite the binding conditions (binding and elution temperatures, binding buffer and regeneration buffer) were the same between the preliminary data [68] and this work, the immobilization procedure was quite different. The CW3S peptide from this work, which differs from the previous work in one serine residue, showed serious difficulties to solubilize during the immobilization protocol, due to its highly hydrophobic

character. It was impossible to solubilize the peptide in the aqueous immobilization buffer, even combined with complementary strategies as vortex and sonication. The CW3S was only dissolved in 65% DMF, 35% 50 mM HEPES, 1 mM EDTA pH 7.5. The presence of this organic solvent could change completely the peptide properties and consequently affect its binding capacity towards HSA. It is therefore possible to assume that the peptide was not properly folded under those conditions, despite it was possible to immobilize it in agarose. As such, CW3S hydrophobicity was a major challenge and limitation in our protocols. Precipitation frequently occurred in the presence of aqueous media, even in relatively low concentrations (superior to 1 mM). Therefore, it was not possible to assess CW3S-Ag binding to HSA in the standard and required conditions (peptide concentration, solvents, pH). The adopted solution was the dissolution in organic solvent, however additional affinity tests need to be tried, namely techniques that do not necessitate high concentrations of peptide.

2.3.3 – MST: interaction between fluorescently labelled CW3S and HSA

MicroScale Thermophoresis (MST) is a technique to quantify biomolecular interactions, based in the concept of thermophoresis. Thermophoresis is the directed movement of molecules in a temperature gradient, that depends on a variety of molecular properties (e.g. size, charge, hydration shell or conformation). Therefore, MST shows high sensitivity to any molecular properties changes, allowing a precise quantification of molecular events. In this experiment, an infrared laser induces a temperature gradient in a glass capillary, a high-power LED (light-emitting diode) as a source of light excitation, a fluorescence detector, and a sample tray of 16 capillaries [82,83]. The dissociation constant (K_d) is deduced from the binding curve, derived by plotting the normalized fluorescence (F_{Norm}) at a given time against each ligand concentration [84–86]. The fluorescence source could be intrinsic (e.g., indole side chain of tryptophan), from attached dyes (e.g., Fluorescein, Alexa546) or from fluorescent protein (e.g., GFP, YFP). The fluorescence excitation/detection spectrum is also important in order to choose the appropriate MST apparatus. Different labelling approaches can be used in attaching dyes to a target protein, for example, the specific fluorophore molecule to the amine group of lysine, thiol group of cysteine, fusion protein (such as GFP), or even unnatural amino acids [87,88].

In this case, CW3S was labelled with Fluorescein Diacetate 5-Maleimide (FDM) through the maleimide chemistry, similar to Sulfo-SMCC referred in 2.3.2. Maleimide groups have the capacity to bind to the sulfhydryl group of the peptide cysteine, making the peptide detectable by fluorescence in the MST assay. FDM is a molecule effective for labelling sulfhydryl-containing compounds, in which maleimide groups react predominantly with sulfhydryls at pH 6.5-7.5, producing a stable thioether bond. FDM was the most adequate dye to use in this case, and already

available in the laboratory. FDM has an excitation wavelength of 492 nm and an emission of 530 nm according to the manufacturer indications.

After the labelling reaction, the mixture was separated by gel filtration and the fractions obtained from PD10 elution were analysed by A_{280}, A_{495}, fluorescence ($\lambda_{excitation}$ = 345 nm and $\lambda_{emission}$ = 485 nm) and BCA assay (A_{560}) (Fig 2.5). A_{280} shows the presence of peptide, while A_{495} and fluorescence detect the fluorophore FDM. BCA gives information about the existence of peptide ligations, that can trace the linkage between the peptide and FDM. PD10 desalting column has a 5 kDa cut-off, meaning that molecules with molecular weights superior to 5 kDa are eluted first. With this, it was expected that CW3S-FDM would be eluted first, with a molecular weight of 5.34 kDa, followed by non-labelled peptide and lastly FDM non-conjugated. In Fig 2.5., through the BCA and A_{280} profiles, it was possible to observe that CW3S-FDM was eluted first. The elution of non-labelled peptide was also possible to obverse in A_{280} profile, corresponding to the second peak. However, FDM was not detected after CW3S elution in A_{495}. It is possible that FDM was almost all conjugated with the peptide, thus there was no trace of it detectable in the chromatogram. In addition, FDM might present residual A_{280}, which can mean that non-conjugated FDM was eluted in this absorbance peak, with CW3S. For a more correct analysis, the BCA assay was performed and the ligation of the peptide and the marker was corroborated. Afterwards, the fraction corresponding to the BCA A_{560} peak (7 mL) was collected. Through Equation 2.4, a concentration of 0.28 mM of CW3S was determined and using Equation 2.5 it was possible to determine the degree of labelling of 0.134 mmol of FDM per mol of peptide.

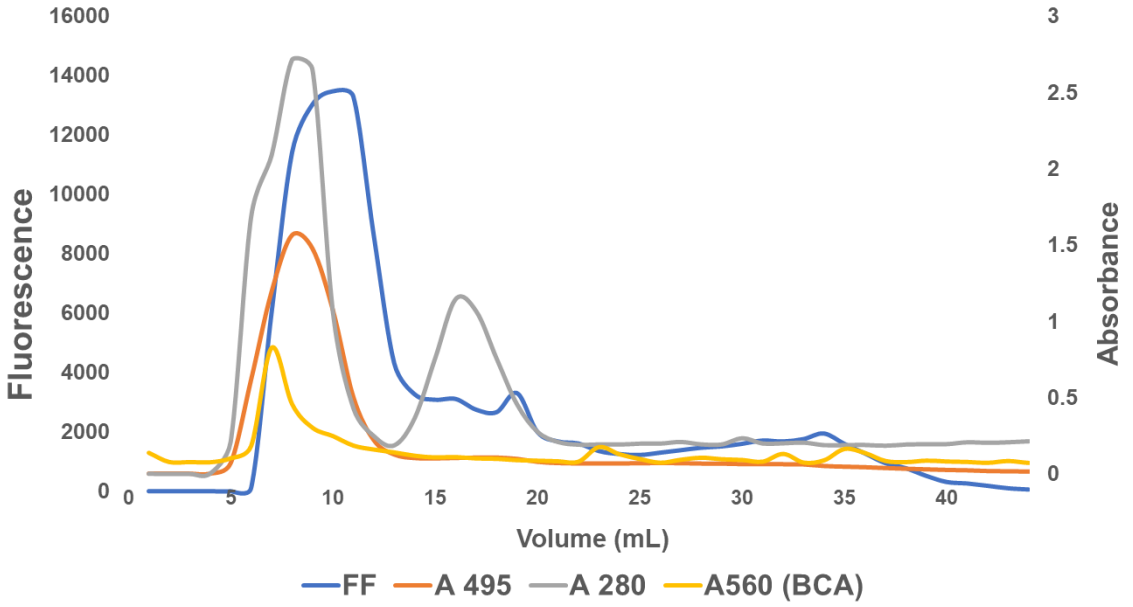

Fig. 2.5 – Chromatogram obtained in PD10 desalting column, regarding CW3S labelling with Fluorescein Diacetate 5-Maleimide. FF: fluorescence signal (λexcitation = 345 nm and λemission = 485 nm); A495: Absorbance 495 nm signal; A280: Absorbance 280 nm signal; A560: Absorbance 560 nm signal for BCA assay.

MST presents several advantages, namely that the affinity studies are performed in solution, with no need to immobilization. Also, the time required for optimization and K_d determination, as well as sample preparation and data processing, is low in comparison to other affinity interaction evaluation techniques. Furthermore, MST is described as an economical technique, efficient and convenient for a rapid screening of a series of molecules [89].

Three MST assays were performed with a fixed concentration of CW3S-FDM of 0.1 µM and an increasing HSA concentration, per assay. MST software analysed different parameters in order to obtain a final dose-response curve and respective dissociation constant (K_d). In those parameters, are included the fluorescence intensity and homogeneity of all capillaries, in order to confirm that every sample presents the labelled protein in a proper concentration and fluorescence ratio. Other parameter is the MST trace, which is an overview that provides information about aggregation of the labelled molecule and photobleaching properties of the samples. It shows the fluorescence intensity of all capillaries over time, reflecting the MST signal of the fluorescent target molecule in presence of increasing ligand concentrations. For the three assays performed, three dose-response curves were obtained (Fig. 2.6), relating HSA concentration (M) and an optimized fluorescence norm (‰), an indicator of thermophoretic property of binding partners. MST software allowed the fitting of these data in order to obtain a K_d (M) of the experiment interaction. K_a (M^{-1}) was also determined by the inverse of K_d. The values obtained for K_d and K_a were 2.0×10^{-7} M and 5.0×10^{6} M^{-1}, respectively. These values are within the range of previous by

assessed affinity constant (Table 2.2) [68] and also within the range of a favourable affinity interactions ($1x10^3$ to $1x10^{11}$ M^{-1}). The previous dissociation and affinity constants were obtained with GFP-fused CW3S (biologically produced), through ELISA. In our work, MST results for the chemically synthesized CW3S demonstrated K_a in the same order of magnitude (Table 2.2) [68]. In addition, a negative control was performed, to confirm that the labelling agent was not binding to HSA. The assay indicated no binding between the two compounds, and consequently a fitted dose-response curve was not possible to obtain. It is important to refer that this was the first time that a WW domain was studied by MST, and there was no previous work in this technique described in the laboratory. Therefore, these conditions were achieved through several optimizations. Nevertheless, different approaches can be tried. For example, using different labelling molecules, or labelling HSA instead and vary the CW3S concentration, or even trying different buffers in solution. Another complementary test would be performing MST in the same conditions, targeting IgG. Finally, these experiments were also marked by the absence of problems in the peptide solubility, since the concentrations of peptide and ligand used were very low. Therefore, MST provides a major advantage by not requiring substantial amounts of sample such as other affinity techniques, giving efficient and rapid results. However, protein labelling could be a challenge in certain cases.

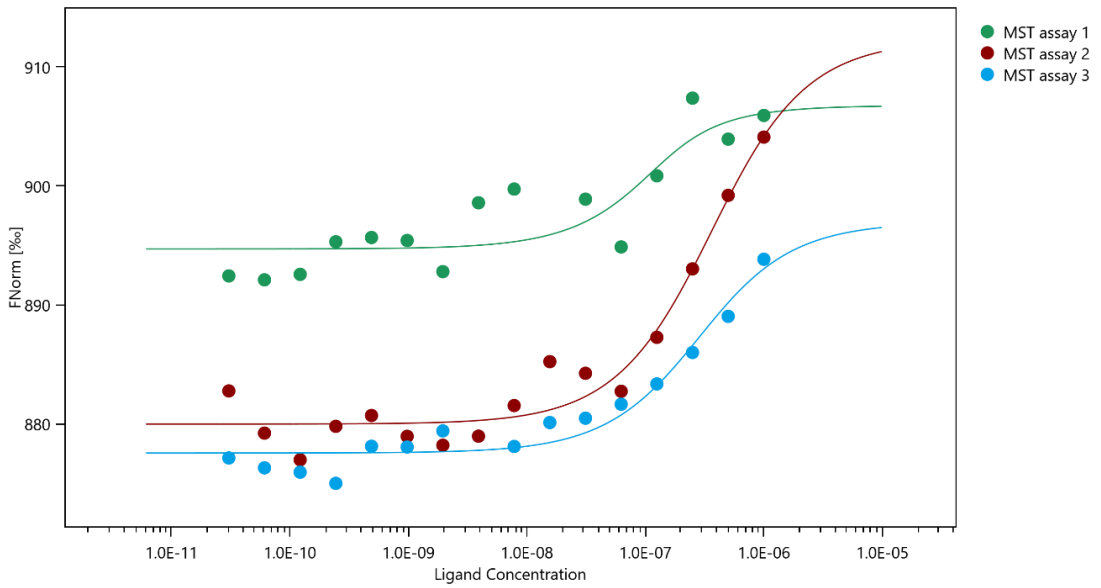

Fig. 2.6 – Dose-response curve fitted from MST assay, including the three assays performed. 0.1 µM of labelled CW3S was used in each assay, with a variation of HSA concentration (maximum fixed at 1 µM). Auto-detect excitation power of 80% and medium-MST power were used as equipment settings. Dissociation constant (K_d) deduced was $2.0x10^{-7}$ M and the affinity constant (K_a) $5.0x10^6$ M^{-1}.

Table 2.2 – Affinity and dissociation constant comparison between the previously biologically produced CW3S fused to GFP (through ELISA) and the chemically synthesized CW3S marked with FDM (through MST).

	Method	Affinity constant (K_a)	Dissociation constant (K_d)
GFP-CW3S with HSA [68]	ELISA	$8.37x10^{6}$ M^{-1}	$1.19x10^{-7}$ M
CW3S-FDM with HSA	MST	$5.0x10^{6}$ M^{-1}	$2.0x10^{-7}$ M

2.3.4 – MP-SPR: CW3S immobilization in Au-gold surface and binding against HSA

Surface plasmon resonance (SPR) is an optical technique for measuring the refractive index of thin layers adsorbed on a metal. A fraction of the incident light energy at a defined angle can interact with the delocalized electrons in the metal film (plasmon), promoting the reduction of the intensity of the reflected light. The SPR experiment is based that, at certain conditions, surface plasmons on a metallic film can be excited by photons, converting a photon into a surface plasmon, depending on the refractive index of the adsorbate. SPR is as an economical and convenient technique to characterize interactions between a protein and a small molecule, since the protein is not consumed during analyses and can be reused for various cycles [90,91].

In this strategy, a non-modified Au-chip was used as the immobilization surface for CW3S. Since this chip presents surface thiol groups, it was possible to use CW3S cysteine residue to bind the surface through the sulfhydryl linkage. The blocking agent used was ethanethiol, another thiol reagent, used to prevent non-specific binding, i.e. to avoid the target molecule to bind directly to the Au surface. A crescent concentration of ligand was used: 0, 0.1, 0.5, 5, 10, 20, 50 µM and finally 1 mM. Fig. 2.7 represents the SPR signal variation with time, including the periods of peptide, blocking agent and target addition in both affinity and control chambers. The signal was measured in two wavelengths, 670 nm and 785 nm. In Fig. 2.7, only 670 nm results are presented, since 785 nm demonstrate the same signal variations. In this assay, the affinity chamber includes the addition of peptide towards immobilization in the Au-surface, and further ethanethiol and ligand additions. The control chamber includes the same compounds, however in a surface not modified with peptide. CW3S was added to the Au-surface twice, in the same 10 µM concentration (at 15.15 min and 37.46 min) to ensure the surface saturation. In both additions, the peptide adsorption is well observed through the signal increase. Provider information demonstrates that an increment of 0.05 deg is sufficient to conclude a molecule was absorbed. Therefore, after the second peptide addition, and before the blocking agent, it is possible to conclude the peptide was successfully immobilized. It was impossible to observe adsorption of the peptide on the surface, and significant desorption of the peptide, i.e., after the peptide additions, the signal did not return to the baseline, which would indicate that the peptide had been

desorbed and not linked to the surface. Then, the blocking agent was added and practically a complete desorption was verified. Possibly, all Au-surface was already filled with immobilized peptide, thus not presenting free thiol groups to bind ethanethiol. Moreover, CW3S is a high size molecule and its immobilization could cause obstruction to ethanethiol binding. Afterwards, different HSA concentrations were tested against immobilized CW3S. In concentrations up to 50 µM, HSA did not bind to the immobilized peptide. It is well observed that after the injection of the target, it desorbs completely, i.e. the signal returns to the baseline with no significant variation between the initial and the final signals. However, in the control chamber, there is indication of a possible binding of HSA to the unmodified Au-gold surface. Since in this range of target concentrations, HSA did not bind to CW3S-Au, a higher concentration of 1 mM HSA was tried. Despite a small increase in the baseline (Δ=0.051), most of HSA had desorbed. According to supplier (BioNavis) instructions, affinity interactions start in variation (Δ) values superior to 0.05. This is an indicator that in future work, higher target concentrations should be tried. It is important to refer that there was not reported work regarding this technique towards this domain. All conditions were optimized for the first time for CW3S immobilization and further binding studies. For instance, it is unknown if the regeneration conditions used (0.1 M NaOH in 30% isopropanol) can harm the Au-surface binding. Complementary work must be performed in SPR for this purpose. Firstly, in the same experimental conditions, higher concentrations of target should be performed, in order to confirm HSA affinity towards CW3S. In addition, different buffer and pH conditions can also be experimented in HSA binding, as well as a different flow rate. Experiments without regeneration of the Au-surface can also be performed. Moreover, CW3S immobilization improvement could be explored with the presence of linkers that enable the binding between the peptide and the surface. As in chromatography affinity tests (2.3.2), CW3S interaction with IgG could also be studied.

SPR can obtain rapid results and is a suitable and cost-effective technique, requiring small amounts of sample volumes. However, each assay needs rigorous standardization, which on the other hand ends up in using considerable amounts of precious ligand and peptide. In addition, the technique does not reach temperatures under 20ºC, an important factor that should be explored, and which could influence CW3S folding and consequently its binding capacity. Moreover, SPR requires the use of proper washing steps and regeneration conditions, a correct flow rate to avoid mass transfer limited binding, the use of correct reference for data subtraction and zeroing, and finally requires multiples or replicates for each assay [91,92].

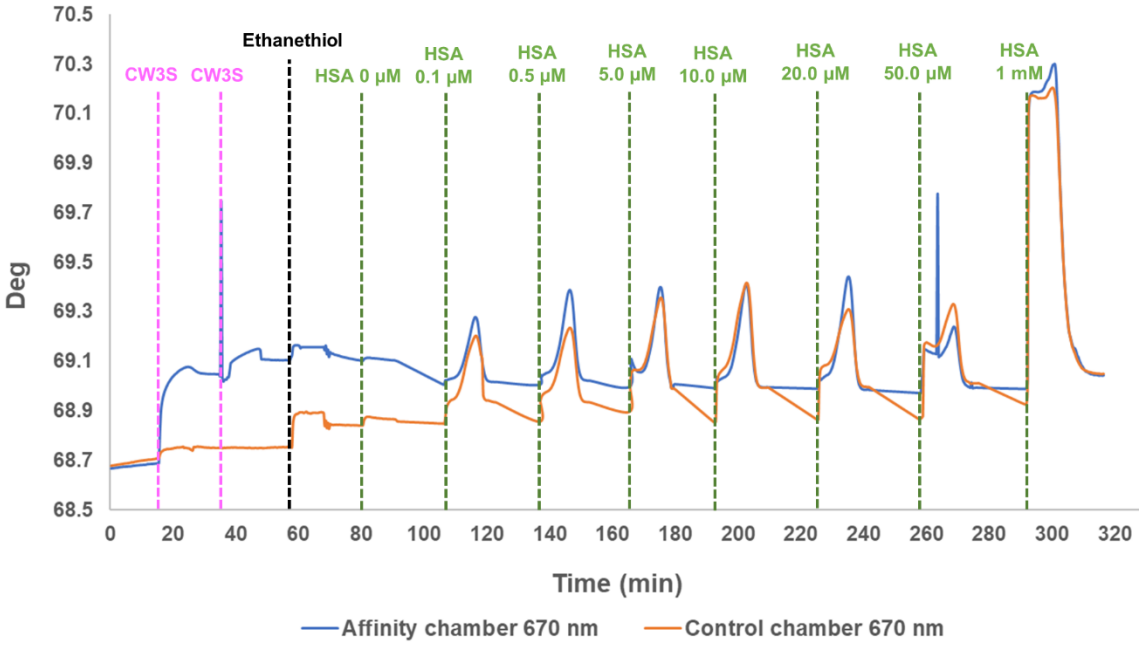

Fig. 2.7 – MP-SPR signal (deg) obtained with time (min), in affinity and control chambers, at a wavelength of 670 nm. Additions of CW3S, ethanethiol and all concentrations of HSA are represented in time (min): CW3S 15.15 min and 37.46 min; Ethanethiol 57.57 min; HSA: 0 µM 80.12 min, 0.1 µM 106.13 min, 0.5 µM 136.34 min, 5.0 µM 165.04 min, 10 µM 192.55 min, 20 µM 225.20 min, 50 µM 257.61 min, 1 mM 291.61 min.

2.4 – Concluding remarks

This chapter aimed at (i) the chemical production, purification and characterization of CW3S peptide; and (ii) the evaluation of the binding between CW3S and the target protein HSA. Positive results were accomplished regarding peptide synthesis, purification and characterization. The optimizations applied in the SPPS protocol were successful and the CW3S peptide was obtained with a yield of 3.3 %.

Three different affinity techniques were performed to assess the binding between CW3S and the target HSA, namely affinity chromatography, MST and MP-SPR. MST was the only technique which allowed the determination of a K_a comparable with previous data [68] (K_a=5.0x10^6 M^{-1}). Despite necessitating a labelling dye conjugated to CW3S, MST provides better assay conditions to a highly hydrophobic peptide. Although it uses the same binding buffers (PBS pH 7.4), the range of low concentrations used in MST avoids peptide precipitation and solubility problems. Moreover, MST does not require immobilization in a solid support and the assay is performed in solution. Still, labelling CW3S could be a challenge, since its hydrophobic character can affect solubility and promote precipitation during the removal of the labelling agent and the peptide not conjugated excess through PD10 desalting column elution. These issues could also be overcome by using low concentration of peptide in conjugation.

The immobilization of CW3S in a chromatographic resin required high concentrations of peptide in aqueous solution for immobilization (≈1.5 mM). The need to include organic solvents in the immobilization buffer constitution (about 65%) to improve solubility, compromised previously optimized conditions for CW3S binding against HSA. CW3S folding and consequently its binding regions could have been affected due to harsh conditions. Despite these results indicate that CW3S does not present affinity towards HSA, a conclusion can not be taken, since the peptide was dissolved in a completely different solution. On the contrary, MP-SPR uses low concentrations (10 µM), although 10 times superior to MST. Nonetheless, MP-SPR parameters, binding conditions and concentrations must be optimized. As referred in 2.3.4, there are no record of experiments with CW3S in SPR, thus these parameters were tried for the first time, based on studies with other protein domains.

Additional studies and different techniques need to be performed with appropriate conditions (namely low concentrations in aqueous buffers) for CW3S solubility. However, MST results indicate CW3S is still a promising scaffold for affinity and biotechnological applications.

Chapter 3

Biological production of CW3S peptide in bacterial cells

Chapter 3

This chapter focuses on the biological production of a WW domain (CW3S) in bacterial cells (Rosetta (DE3)) and further purification. The sequence of the CW3S peptide produced biologically includes the C-terminal Serine residue [68], since it was depleted during solid-phase peptide synthesis only due to coupling limitations. Therefore, CW3S presents a total of 42 residues, a molecular weight of 5 kDa and the corresponding encoding gene presents a total of 172 bp, presenting the sequence: (N-terminal) NH_2- CMGLPPGWDEYKSSNLFLGKTYYYNH-DNSPKTSTWTDPRMS-COOH (C-terminal).

Two different strategies are approached in this chapter. The first one consists in the expression of CW3S domain alone in *E. coli*, using the expression vector pAP006 (6245 bp). The second one uses Green Fluorescent Protein (GFP) as a reporter tag fused at the N-terminal of CW3S, and the expression vector pAP004 (6199 bp) to facilitate the monitorization of the protein expression during production. GFP is a protein obtained from *Aequorea victoria* jellyfish, that has fluorescent properties, when a chromophore is formed due to molecular oxygen contact [93].

3.1 – Materials and Equipment

The plasmids pAP004 and pAP006 and the competent cells *Escherichia coli* Rosetta BL21(DE3) were available in the Group (UCIBIO, Portugal). pEX-A128 was ordered from Eurofins (Genomics). Nzy5α, NzyStar and MiniPrep kit were purchased from NZYTech. Restriction enzymes (NheI, KpnI and HindIII) were purchased from ThermoScientific. DNase I was obtained from Roche (Lisbon, Portugal). For DNA purification, E-gel from Invitrogen-Thermo Scientific was used. The chromatographic materials cross-linked agarose (SepharoseTM CL-6B) and HiLoad 16/600 Superdex 75 column were obtained from GE Healthcare (Uppsala, Sweden). A4C7 resin was available from previous group work [94]. Brand Black immunograde 96-well microplates and 96-well UV half area (Greiner) were supplied by VWR International (Lisbon, Portugal). The 96-well transparent microplates were acquired from Sarstedt (Lisbon, Portugal).

3.2 – Methods

3.2.1 – Transformation of pEX-A128, pAP004 and pAP006 plasmids into *E. coli* Nzy5α competent cells

pEX-A128 was firstly designed with the sequence of interest and then ordered and purchased to Eurofins Genomics. pAP004 and pAP006 are pET21c plasmids used for molecular cloning.

Regarding the transformation, 1 µL of plasmid was added to 50 µL of Nzy5α competent cells. For positive and negative controls, it was used 20 µL of competent cells. For the positive control, it was added the plasmid control provided by NzyTech (pCCC). These preparations were incubated in ice for 30 minutes, followed by 40 seconds in a water bath at 42ºC (heat-shock treatment), and finally 2 minutes in ice. It was added room-temperature LB media to each preparation, up to 1 mL of total volume, followed by 1h of incubation at 37ºC and 225 rpm orbital agitation. Afterwards, it was spread 50 µL and 150 µL from each sample to previously prepared LB/Agar/Ampicillin (Amp) plates. The remaining volume of each sample was centrifuged (2 minutes/5000 rpm), most of the supernatant discarded and the pellet resuspended in a small volume. This was also spread in LB/Agar/Amp plates. The cells were grown on the plates overnight, at 37ºC and then stored at 4ºC.

3.2.2 – Isolation, quantification and integrity of plasmid DNA from bacterial cells

The pre-inoculums of the plasmids were prepared in duplicates (2 for each plasmid) by adding one isolated colony to LB media with 100 µg/mL Amp. In high-copy vectors (pEX-A128) it was used 6 mL of LB media, however in low-copy vectors (pAP004 and pAP006) it was used 12 mL.

The isolation of the transformed plasmids was performed using NzyMiniPrep kit from NzyTech, according to the manufacturer indications. The following volumes are referred according to high copy vectors. For low-copy, the volumes of A1, A2 and A3 buffers were the double of high-copy. The pre-inoculums were centrifuged for 2 minutes at 11000 rpm, the supernatant discarded and the pellets resuspended in 250 µL of buffer A1. Then it was added the same volume of A2 and incubated for 4 minutes at room-temperature. After the addition of 300 µL of A3 buffer, the mixture was centrifuged for 10 minutes at 11000 rpm. The supernatant was collected to silica-gel based columns which were centrifuged for 2 minutes at 11000 rpm. The flow-through was discarded. Next, it was added to the columns 500 µL of AY buffer (previously heated) and 600 µL of A4 buffer with ethanol. In both buffer additions, the solutions were centrifuged for 2 minutes at 11000 rpm and the flow-through discarded. Next, to dry the columns, they were centrifuged for 3 minutes at 11000 rpm. The columns were placed in clean 1.5 mL eppendorfs to proceed to the elution. The first elution was performed by adding 30 µL of pre-warmed milliQ water, incubation for 1 minute in a water bath at 42ºC and then a 2 minutes centrifugation at 11000 rpm. The flow-through was collected to each eppendorf. The second elution was performed by the same method, however with 50 µL of milliQ water instead. The obtained DNA was stored at -20ºC.

The plasmid DNA was quantified in Nanodrop 1000 (Thermo Scientific).

In order to evaluate the purity and integrity of each plasmid, a 0.8% agarose gel was run for pEX-A128, pAP004 and pAP006 plasmids. The gel was performed by adding the corresponding amount of ultra-pure grade agarose (NzyTech) to TAE buffer 1x (40 mM tris(hydroxymethyl)aminomethane, 20 mM glacial acetic acid and 1 mM EDTA). The gel dye used was Green Safe Premium, about 3 µL/100 mL agarose solution. The samples used in the gel were NzyDNA Ladder III DNA Marker, the first and second elutions of the extracted pEX-A128, pAP004 and pAP006 plasmids. For the plasmids, it was used 2 µL of plasmid DNA diluted in 2 µL of milliQ water and 1 µL of Loading Dye from NzyTech (proportion of 1 volume of dye to 5 volumes of sample), performing a total volume of 5 µL loaded in the gel. The electrophoresis was run for 60 minutes at 100 V. For gel revelation, it was used a UV Transiluminator (Bio-Rad Gel Doc XR+).

3.2.3 – Digestions regarding the expression of CW3S peptide: pEX-A128 and pAP006 double-digestions with NheI and HindIII

The purified plasmids pEX-A128 and pAP006 were digested with NheI and HindIII restriction enzymes, according to the manufacturer protocols, from Thermo Scientific. The digestion of pEX-A128 was prepared with 4 µg of plasmid DNA, in a total volume of 40 µL. It was also added 4

μL of fast digestion buffer 10x, 2 μL of each restriction enzyme and milliQ water up to 40 μL. The digestion of pAP006 was prepared with 2 μg of plasmid DNA, with 2 μL of fast digestion buffer 10x, 1 μL of each restriction enzyme and milliQ water up to 20 μL. The digestion reactions occurred at 37ºC for 3h, plus the thermal inactivation of NheI and HindIII (65ºC/5 minutes and 80ºC/10 minutes, respectively). The reaction was carried out in a T-100 Thermal Cycler from Bio-Rad.

Double digested pAP006 plasmid (dd_pAP006) was also dephosphorylated with Fast-AP (Thermosensitive Alkaline Phosphatase), from Thermo Scientific, to prevent recircularization and self-ligation of the transformation plasmid. The method was carried out as referred by the provider protocol. Fast-AP was added to dd_pAP006 in a proportion of 1 μL/10 μL double digestion volume. The mixture was incubated for 30 minutes at 37ºC, plus 20 minutes at 80ºC (Fast-AP enzyme inactivation) in the T-100 Thermal Cycler from Bio-Rad.

Finally, it was performed an electrophoresis in agarose gel to confirm the release of the insert that contain the gene of interest in pEX-A128, and also to evaluate the digestion of the transformation plasmid, pAP006. For pEX-A128 it was used a 2% agarose gel, and for pAP006 a 0.8% gel. The gels were accomplished as referred in 3.2.2. Regarding the 2% gel, both NzyDNA Ladder III DNA Marker and NzyDNA Ladder V DNA Marker were used. Undigested pEX-A128 and double digested pEX-A128 (dd_pEX-A128) were loaded in the gel. For the first one, 1 μL of purified plasmid was diluted in 3 μL of milliQ water and 1 μL of loading dye. As for dd_pEX-A128, 2 μL of plasmid were diluted in 2 μL of milliQ water and 1 μL of loading dye. For the 0.8% gel, only NzyDNA Ladder III DNA Marker was used. For both undigested pAP006 and dd_pAP006 samples, it was added a volume proportion of 1 μL of plasmid, 3 μL of milliQ water and 1 μL of loading dye. In both gels, the final volume of every sample was 5 μL. The Loading Dye and gel revelation were used as referred in 3.2.2.

3.2.4 – Digestions regarding the expression of CW3S peptide coupled to GFP: pEX-A128 and pAP004 double-digestions with KpnI and HindIII

The purified plasmids pEX-A128 and pAP004 were digested with KpnI and HindIII restriction enzymes, according to the manufacturer protocols, from Thermo Scientific. The digestion of pEX-A128 was prepared with 2 μg of plasmid DNA, in a total volume of 20 μL. It was also added 2 μL of fast digestion buffer 10x, 1 μL of each restriction enzyme and milliQ water up to 20 μL. The digestion of pAP004 was prepared with 3 μg of plasmid DNA, with 3 μL of fast digestion buffer 10x, 1.5 μL of each restriction enzyme and milliQ water up to 30 μL. The digestion reactions occurred at 37ºC for 3h, plus the thermal inactivation of KpnI and HindIII (80ºC/5

minutes and 80ºC/10 minutes, respectively). The reaction was carried out in a T-100 Thermal Cycler from Bio-Rad.

Double digested pAP004 (dd_pAP004) was dephosphorylated as referred in 3.2.3 for dd_pAP006.

Finally, it was performed an electrophoresis in agarose gel to confirm the release of the insert that contain the gene of interest in pEX-A128, and also to evaluate the digestion of the transformation plasmid, pAP004. The gels were accomplished as referred in 3.2.2. In the gel, it was loaded NzyDNA Ladder III DNA Marker, double-digested pEX-A128 (dd_pEX-A128) and dd_pAP004. For both plasmid samples, 1 µL of purified plasmid was diluted in 3 µL of milliQ water and 1 µL of loading dye. The Loading Dye and gel revelation were used as referred in 3.2.2.

3.2.5 – Purification of the inserts and double-digested plasmids by e-gel

The inserts released in both digestions of pEX-A128 and the double-digested transformation plasmids (dd_pAP004 and dd_pAP006) were purified before the ligation reaction. The equipment used was E-gel from Invitrogen-Thermo Scientific and the purification methodology was performed according to the supplier instructions. The purification method consisted in an electronic gel in which 700 ng of digested DNA were loaded in each upper-well, in milliQ water, up to 20 µL. For each e-gel electrophoresis, 5 µL of a DNA marker was loaded. The method chosen was "0.8% gel run". The different bands of each DNA sample were visible in the e-gel as in a usual electrophoresis. The bands of interest of each sample were collected in the collecting wells and quantified by Nanodrop before the ligation reactions.

3.2.6 – Strategies for cloning CW3S sequence with an expression vector

For cloning the gene encoding CW3S peptide in an expression vector, two strategies were considered. In the first approach, the gene that codifies for CW3S peptide (obtained by digestion of pEX-A128 with NheI and HindIII) was cloned with the dd_pAP006 expression vector, which is referred in 3.2.6.1. The second approach considered the cloning of the insert with the genes that codified CW3S peptide sequence, Enterokinase (EK) cleavage site and Green-Fluorescent Protein (GFP) – GFP-EK-CW3S. This insert was obtained by digestion of pEX-A128 with KpnI and HindIII. The expression vector considered was dd_pAP004. The cloning method is referred in 3.2.6.2.

3.2.6.1 – Cloning of CW3S sequence insert in dd_pAP006 expression vector

The cloning of the insert of interest in dd_pAP006 expression vector was performed with T4 DNA Ligase (Thermo Scientific), by using the supplier protocol. The amount of DNA of insert was determined by the following equation:

$$\frac{ng\ (vector) \times kb\ (insert\ size)}{kb\ (vector\ size)} \times \frac{insert}{vector}\ (molar\ ratio) = ng\ (insert) \quad \text{(Equation 3.1)}$$

To perform the ligation reaction, the amount of insert previously determined was added to other components to a final reaction volume of 20 µL. Therefore, they were considered two different vector-insert molar ratios (1:3 and 1:10) and a negative control reaction. In both cases, the following components were added into a PCR reaction tube, by this order: milliQ water up to 20 µL, 25 ng of vector, the determined amount of insert (depending on the ratio considered), 2 µL of 10x Reaction Buffer and 0.4 µL of T4 DNA Ligase (5U/µL). Regarding the negative control reaction, no the insert was added to the reaction mixture. The reactions mixtures were then incubated for 2h at 22ºC for, followed by 5 minutes at 70ºC for T4 DNA Ligase inactivation, in a T-100 Thermal Cycler from Bio-Rad. The cloning product was stored at -20ºC.

3.2.6.2 – Cloning of GFP-EK-CW3S sequence insert in dd_pAP004 expression vector

The methodology used for the cloning of GFP-EK-CW3S in the dd_pAP004 expression vector was the same as referred in 3.2.6.1.

3.2.7 – Transformation of the cloning products (ligation reaction) from 3.2.6.1 and 3.2.6.2 in NzyStar competent cells

10 µL of each cloning product was added to 100 µL of NzyStar competent cells. For positive and negative controls, it was used 50 µL of competent cells. For the positive control, it was added the plasmid control provided by NzyTech (pCCC). These preparations were incubated in ice for 30 minutes, followed by 40 seconds in a water bath at 42ºC (heat-shock treatment), and finally 2 minutes in ice. It was added room-temperature LB media to each preparation, up to 1 mL of total volume, followed by 1h of incubation at 37ºC and 225 rpm orbital agitation. Afterwards, it was spread 50 µL and 150 µL from each sample to previously prepared LB/Agar/Amp plates. The remaining volume of each sample was centrifuged (2 minutes/5000 rpm), most of the supernatant discarded and the pellet resuspended in a small volume. This was also spread in LB/Agar/Amp plates. The cells were grown on the plates overnight, at 37ºC and then stored at 4ºC.

3.2.8 – Restriction analysis – Insert release confirmation

After transformation, 10 colonies were selected: 5 for each insert-vector molar ratio. Afterwards, restriction analysis of the respective colonies was performed.

3.2.8.1 – Restriction analysis – CW3S cloning product

10 pre-inoculums, and the respective extraction and purification of DNA were performed, as in 3.2.2. 2 μg of each sample of plasmid DNA, in a final volume of 20 μL, were digested with NheI and HindIII restriction enzymes, as referred in 3.2.3.

In order to confirm the release of the insert, a 2% gel electrophoresis was performed. Both NzyDNA Ladder III DNA Marker and NzyDNA Ladder V DNA Marker were loaded in the gel, plus an undigested sample and all 10 products from ligation and digestion. The gel production, revelation and run, and all the samples dilutions were performed as in 3.2.3.

3.2.8.2 – Restriction analysis – GFP-EK-CW3S cloning product

10 pre-inoculums and the respective extraction and purification of DNA were performed, as in 3.2.2 and 3.2.8.1. The purified products of GFP-EK-CW3S cloning were digested with KpnI and HindIII restriction enzymes and prepared with 2 μg of DNA in a final volume of 20 μL. The digestion reactions components, methodology and equipment were performed as referred in 3.2.4 and for 2 μg DNA digestions.

In order to confirm the release of the insert, a 0.8% gel electrophoresis was performed. NzyDNA Ladder III DNA Marker was used, plus an undigested sample and all 10 products from ligation and digestion of GFP-EK-CW3S. The gel production, revelation and run, and all the samples dilutions were performed as in 3.2.4.

3.2.9 – Transformation of positive clones in *E. coli* Rosetta (DE3) competent cells

For transformation in Rosetta cells, one positive clone for each cloning strategy was selected. The clones were transformed, following the methodology referred in 3.2.7 for NzyStar competent cells, although for this case LB/Agar/Amp/Chloramphenicol (CL) plates were used.

3.2.10 – Expression of CW3S and GFP-EK-CW3S sequences in *E. coli*

For each expression strategy, a pre-inoculum with one isolated clone with 5 mL of LB media, 100 μg/mL Amp and 100 μg/mL CL was incubated for ≈9h at 37ºC with 225 rpm of orbital

agitation. 1 mL of the previous pre-inoculums was added to 50 mL of LB media containing 100 µg/mL Amp and 100 µg/mL CL, and incubated overnight at 37ºC with 225 rpm of orbital agitation. 10 mL of the respective inoculums were added to 1 L of LB media pH 7.5, 100 µg/mL Amp and 100 µg/mL CL, in 2 L Erlenmeyers, incubated at 37ºC with 225 rpm of orbital agitation. Once the culture reached an optimal density at 600 nm (OD_{600}) of approximately 0.6, the protein expression was induced with 1 mM of isopropyl β-D-1 thiogalactopyranoside (IPTG). Every 2h after induction, for 6h, aliquots were collected. The culture was kept overnight at 30ºC with 225 rpm agitation. After overnight time, two more aliquots were collected.

3.2.11 – Cellular fractionation

After the expression, the cell cultures were taken from the 2 L Erlenmeyers and were centrifuged at 6500 rpm for 15 minutes at 4ºC. The pellets were then resuspended in 15 mL PBS and incubated at -80ºC. In order to promote the fragility of the cellular membrane, three cycles of freeze/thaw cycles were performed. Afterwards, the cells were subject to four cycles of disruption under 4000 psi, in a French Press (ThermoScientific). The obtained disrupted cells were treated with 10% DNAaseI and incubated on ice for 15 minutes, followed by 30 minutes centrifugation at 10000 xg at 4ºC. The pellet was resuspended in 20 mL and the supernatant collected to ultracentrifuge at 204710 xg for 1h30min at 4ºC. The supernatant of the ultracentrifugation was collected and stored at -20ºC and the pellet resuspended in 20 mL PBS. Cellular fractionation for GFP-EK-CW3S expression was evaluated by GFP fluorescence, BCA assay and SDS-PAGE analysis. For CW3S expression, the analysis by fluorescence were not performed, as in this case, the GFP was not used as reporter tag. The fractions evaluated were the pellet and the supernatant of the centrifugation (PC and SC, respectively), and the pellet and the supernatant of the ultracentrifugation (PU and SU, respectively).

3.2.12 – Quantification of GFP-EK-CW3S by fluorescence intensity

The samples regarding protein expression and cellular fractionation of GFP-EK-CW3S were quantified by fluorescence intensity of GFP fusion protein. The fluorescence was read in a 96-well black microplate (Brand, VWR), using an excitation and emission wavelength of 485 nm and 535 nm, respectively. The gain considered for the readings was 41 and the volume use in the plate was 200 µL. The fluorescence was read in an Infinite Tecan Microplate Reader. GFP concentration was determined by the following equation:

$$Protein\ Fluorescence = 487016 \times [GFP] + 279.24 \quad \text{(Equation 3.2)}$$

3.2.13 – Quantification of total protein for both CW3S and GFP-EK-CW3S expression by BCA assay

BCA assay was performed according to supplier instructions Sigma-Aldrich. For BCA quantification, it was used 25 µL of each sample in a 96 well transparent flat-bottom microplate (Stardsted). A calibration curve of BSA was also prepared, in the range of 0.2 to 1 mg/mL. The BCA working solution was prepared by mixing reagent A with reagent B in a 50:1 ratio. 200 µL of BCA solution was added in each well containing the respective sample, followed by a 30 minutes incubation at 37ºC, with aluminium foil covering the microplate. Afterwards, the absorbance was immediately read at 560 nm in an Infinite Tecan Microplate Reader.

3.2.14 – SDS-PAGE Electrophoresis Analysis

The samples from the protein expression during time course and cellular fractionation were analysed by SDS-PAGE electrophoresis gel. For CW3S expression, the electrophoresis was carried out in a 12.5% polyacrylamide gel and also in a Tris-Tricine gel. For GFP-EK-CW3S, it was used only a 12.5% polyacrylamide gel.

3.2.14.1 – SDS-PAGE Electrophoresis in a 12.5% polyacrylamide gel

To produce the 12.5% acrylamide running gel, it was added 1.5 mL of 3M Tris Base pH 8.8-9.0; 4.16 mL of 30% acrylamide/bis solution 37:5:1; 0.1 mL of 10% SDS; 4.2 mL of distilled water; 0.076 mL of 10% APS; 0.005 mL of TEMED. For 5% acrylamide stacking gel, it was added 0.9 mL 0.5 M Tris Base, pH 6.6-6.8; 30% acrylamide/bis solution 37:5:1; 0.036 mL of 10% SDS; 1.88 mL of distilled water; 0.027 mL of 10% APS; 0.004 mL of TEMED. As protein molecular weight marker, it was used Precision Plus Protein Dual Xtra Standard (from Bio-Rad). The sample buffer used was 5 mL of 0.5 M Tris Base, pH 6.6-6.8; 2 mL of 100% glycerol, 4 mg of Blue Bromophenol; 8 mL of 10% SDS, 1 mL β-mercaptoethanol; distilled water up to 20 mL. 50 µL of sample buffer was added to time-course expression samples and 5 µL to cellular fractionation samples. The samples were boiled for 10 minutes, before they were loaded in the gel. The volume added of time-course expression protein samples was determined by the following equation:

$$V\ (mL) = \frac{1.2}{Sample\ OD_{600\ nm}} \quad \text{(Equation 3.3)}$$

The gel was run for approximately 90 minutes at 100 V. Then, the gel was stained with Coomassie blue (1 g of Coomassie blue R-250, 15 mL Acetic Acid Glacial, 90 mL Methanol, distilled water up to 200 mL) for 30 minutes with small agitation; and distained using a distaining

solution (75 mL Acetic Acid Glacial, 450 mL Methanol, distilled water up to 1000 mL). The revelation of the gel was performed in a UV Transiluminator.

3.2.14.2 – SDS-PAGE Electrophoresis in a Tris-tricine gel

To produce the Tris-tricine separating gel (amounts optimized for two gels), it was added 3.5 mL of distilled water; 5 mL of 3M Tris-HCl/SDS Base pH 8.45; 5 mL of 30% acrylamide/bis solution 37:5:1; 1.5 mL of 100% glycerol; 0.014 mL of 30% APS; 0.014 mL of TEMED. For the stacking gel, it was added 3.9 mL of distilled water; 1.55 mL of 3M Tris-HCl/SDS pH 8.45; 0.8 mL 30% acrylamide/bis solution 37:5:1; 0.014 mL of 10% APS; 0.014 mL of TEMED. As protein molecular weight marker, it was used Precision Plus Protein Dual Xtra Standard (from Bio-Rad). The sample buffer used was 5 mL of 1 M Tris-HCl, pH 6.8; 12 mL of 100% glycerol; 4 g of SDS; 1.5 g of DDT (Dichlorodiphenyltrichloroethane); 10 mg Coomassie Blue R350; distilled water up to 50 mL. 50 µL of sample buffer was added to time-course expression samples and 5 µL to cellular fractionation samples. The anode buffer was 2 M Tris pH 8.8 (2x) and the running buffer Tris/Tricine/SDS Running Buffer (1x). The samples were boiled for 10 minutes, before they were loaded in the gel. The volume added of time-course expression protein samples was determined by the Equation 3.3.

3.2.15 – Purification of GFP-EK-CW3S with A4C7 resin in AKTA system

The soluble fraction of the crude extract of GFP-EK-CW3S (supernatant of the ultracentrifugation – SU) was purified in the AKTA Pure System. A Tricorn 10/50 empty column was packed using agarose resin modified with Ugi Ligand A4C7 [94]. It was used 2.3 mL of resin of a 50% slurry. The resin was packed overnight under gravitational flow rate at 4°C. The column was connected to the AKTA Pure System using a flow rate of 0.3 mL/min. The following steps of column regeneration, equilibration, binding and elution ware carried by using a flow rate of 1 mL/min. Initially the resin was regenerated with 0.1M NaOH in 30% Isopropanol alternated with water, 10 CV each, with a total of 30 CV. To equilibrate the column, PBS buffer pH 7.4 was used in a total of 15 CV. The crude extract sample, previously filtered in a 0.2 µm filter, was then loaded into the column by using a 5 mL loop. The column was then washed with 15 CV of buffer pH 7.4 to remove unbound proteins. Protein elution was then performed by using 20 CV 0.1 mM glycine–NaOH pH 9.0. The flow-through, washes and elutions were collected in a 96-well deep well block in 2 mL independent fractions. All fractions were quantified by fluorescence and BCA assay (as referred in 3.2.13), giving a chromatogram of this purification step. From this

chromatogram, the fractions with higher fluorescence values and coincident higher absorbance were pooled into one fraction. This pool was stored at -20ºC for further SDS-PAGE analysis.

3.2.16 – Purification by Gel-Filtration in AKTA system

The pool fraction obtained from A4C7 resin chromatography was purified in the AKTA Pure System. A HiLoad 16/600 Superdex 75 (GE Healthcare) column was equilibrated with 2 CV of water and 1.5 CV of PBS pH 7.4, using a flow rate of 0.5 mL/min. The column was connected to the AKTA Pure System using a flow rate of 0.3 mL/min. Then, 2.5 mL of A4C7 purified sample was injected into the column through 5 mL injection loop. The extract sample, previously filtered in a 0.2 μm filter, was then loaded into the injection 5 mL loop. Purification was then performed at a flow rate of 1 mL/min. In the column regeneration step 1 CV of water was used with a flow rate of 1 mL/min, and lastly 1 CV of 20% Ethanol at a flow rate of 0.5 mL/min. All fractions were collected in a 96-well deep well block in 2 mL independent fractions. They were quantified by fluorescence and BCA assay (as referred in 3.2.13), giving a chromatogram of this purification step. From this chromatogram, the fractions with higher fluorescence values and coincident higher absorbance were put together in a pool, labelled as the product of Gel-Filtration purification. This pool was stored at -20ºC for future SDS-PAGE analysis.

3.2.17 – Cleavage of GFP-EK-CW3S with Enterokinase

The cleavage with Recombinant Enterokinase (Novagen) was performed according to manufacturer instructions, 1 unit of EK cleaves 50 μg of protein). The obtained pool fraction from 3.2.16 was concentrated in an Amicon (Merck Millipore) and buffer exchanged to EK-cleavage buffer (Tris-HCl pH 7.4, 50 mM NaCl, 2 mM $CaCl_2$). Protein concentration was then determined by Equation 3.2, The cleavage reaction was performed for 16h at 23ºC with gentle agitation.

3.2.18 – Purification by Gel-Filtration in AKTA system after cleavage with Enterokinase

Purification by gel-filtration was performed accordingly to 3.2.16.

3.2.19 – Quantification of the obtained CW3S expressed peptide by tryptophan (Trp) fluorescence

The samples obtained after the cleavage with Enterokinase, followed by gel-filtration purification, were quantified by fluorescence intensity of Trp amino acid. The fluorescence was

read in a 96-well black microplate from Brand, using an excitation and emission wavelength of 280 nm and 345 nm, respectively. The gain considered for the readings was 41 and the volume used in the plate was 200 µL. The fluorescence was read in an Infinite Tecan Microplate Reader.

3.3 – Results and discussion

3.3.1 – Design of the gene containing the CW3S sequence for cloning in expression vectors

For CW3S expression, the DNA fragment encoding for CW3S was designed. This fragment included CW3S encoding gene and the incorporation of restriction sites and other sequences (e.g. affinity tags and protease cleavage site) in order to facilitate molecular cloning and subsequent purification (Fig. 3.1).

Three polylinkers, were incorporated in the sequence (Fig. 3.1), each one compiling different restriction sites. The first polylinker (PI) introduced, was constituted by BamHI, KpnI, NdeI and SacI restriction sites, in this order (from 5' to 3'). Followed by PI, the sequences encoding for GFP reporter tag and an Enterokinase (EK) cleavage site (that allows the final recovery of the CW3S peptide in the purification process) were also incorporated. Then, the polylinker II (PII) was introduced, constituted by NheI and SalI restriction sites. After the PII, the gene encoding for CW3S and a STOP codon were incorporated. Next to the STOP codon, the last polylinker was introduced. Polylinker III (PIII) was constituted by XhoI, SmaI, BamHI, EcoRI and HindIII restriction sites, from 5' to 3'.

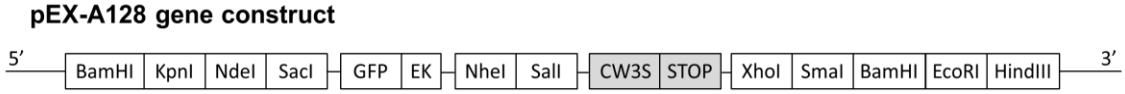

Fig. 3.1 – Design of the gene construct that includes the gene encoding for CW3S.

GFP reporter tag was incorporated in the CW3S peptide N-terminal in order to mask the peptide high hydrophobicity and also to facilitate and enable purification. Moreover, GFP inclusion can also promote obtaining CW3S without undesired extra amino acids in the fusion protein to purify, since they could influence and limit the peptide affinity towards possible targets in eventual future binding studies.

3.3.2 – Protein expression strategies

The three polylinkers allow performing two different protein expression strategies (Fig. 3.2). The first one, for the expression of the CW3S sequence itself (Fig. 3.2B). Through PII and PIII restriction sites, the cloning of the DNA fragment containing CW3S into pAP006 expression vector, was possible to conduct. The second approach consists in the expression of CW3S fused to a reporter tag that allow protein monitorization (GFP) and a protease cleavage site (EK) for

further recovery of CW3S, in purification (Fig. 3.2C). PI and PIII restriction sites allowed the cloning of the CW3S-fused DNA fragment into pAP006 expression vector.

Expression vectors pAP004 and pAP006 are pET21c plasmids. pET21c is an optimized system for cloning and expression of proteins in *E. coli*. Target genes are cloned in pET plasmids under control of bacteriophage T7 transcription and (optionally) translation signals. These plasmids can be transferred into expression hosts containing a chromosomal copy of the T7 RNA polymerase gene under lacUV5 control, thus the expression can be induced by the addition of IPTG (pET Manual System). pET21c is already presents a start codon (AUG). The suffix "c" refers to the recognition region of BamHI cloning site reading frame, GGATCC. Furthermore, accordingly to the supplier manual [95], this type of vector allows the expression levels in *E. coli* increase with fused protein sequences with efficient expressed amino terminal sequences. pET21c plasmids contain a marker for antibiotic resistance. In this case, it is the gene amp^r, that codifies for the plasmid resistance to Ampicillin (Amp). Thus, Amp is used in LB/Agar media plates (LB/Agar/Amp), in all transformations in competent cells, to allow the growth to only Amp-resistant DNA.

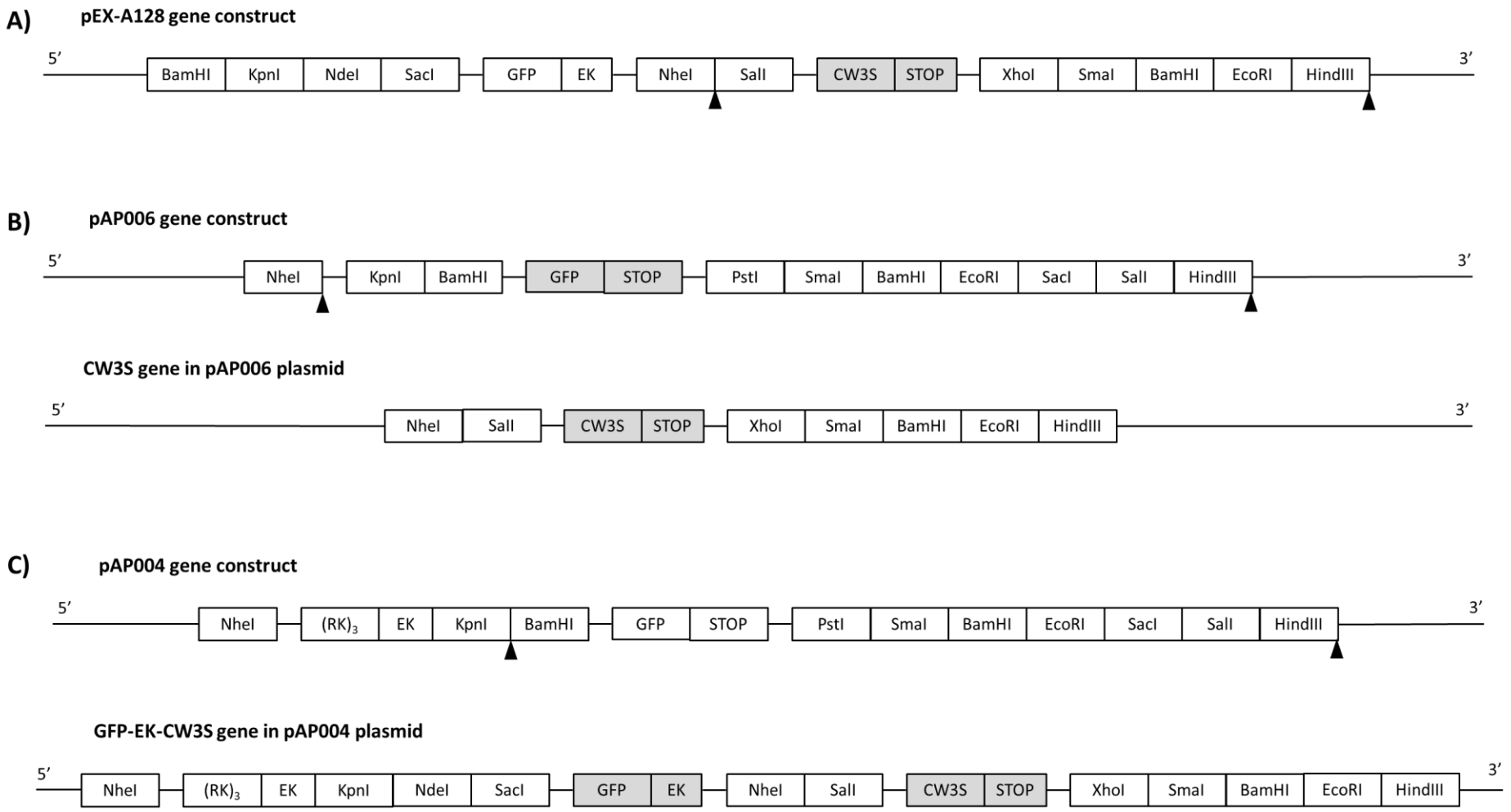

Fig. 3.2 – Representative design of the two strategies approached. (A) Plasmid that contains the encoding gene for CW3S, (B) gene construct of pAP006 expression vector and corresponding cloning strategy of the insert; (C) gene construct of pAP004 expression vector and corresponding cloning strategy of the insert.

3.3.3 – Molecular cloning and integrity evaluation of plasmid DNA

The obtained DNA fragments of the gene of interest (CW3S and GFP-EK-CW3S) were cloned in pAP006 and pAP004 expression vectors, respectively. At first, all considered plasmids – pEX-A128, pAP006, pAP004 – were successfully transformed in Nzy5α competent cells. The obtained plasmid DNA was isolated and purified and the final DNA concentration determined and presented in Table 3.1.

Table 3.1 – Quantification of the plasmid DNA purified by E-gel kit (Invitrogen).

Plasmid DNA	[DNA] (ng/µL)	Abs_{260}/Abs_{280}	Abs_{260}/Abs_{230}
pEX-A128	230.1	1.79	1.33
pAP006	265.5	1.92	2.28
pAP004	141.4	1.89	2.10

The quantification of the purified plasmid DNA presented acceptable concentrations and the values of Abs260/Abs280 and Abs260/Abs230 ratios revealed the good quality of the DNA. The Abs260/280 ratio shows the evaluation of nucleic acids (maximum absorbance at 260 nm) regarding the trace of proteins (maximum absorbance at 280 nm). Values in the order of 1.8 indicate a good quality and purified DNA. On the other hand, the Abs260/230 ratio may indicate other possible contaminations, for instance alcoholic contaminants (maximum absorbance at 230 nm). The expected values for Abs260/230 ratio are about 2.0 to 2.2 [96]. The observed results presented in Table 3.1 indicate normal values for the two absorbance ratios considered. Thus, it was possible to use the obtained concentrations for further cloning process.

The purified plasmid DNA integrity was evaluated and confirmed by a 0.8% agarose gel electrophoresis (Fig. 3.3). It is observed that the obtained DNA bands correspond to their correct molecular weight. The plasmid which includes the encoding gene for CW3S (pEX-A128) has 2450 bp and the expression vectors, pAP004 and pAP006 present 6199 bp and 6245 bp, respectively.

The two strategies approached will include the cloning of the gene that codifies for CW3S in pAP006 expression vector and the cloning of CW3S-fused sequence (GFP-EK-CW3S) in pAP004 expression vector. Two purified samples of pEX-A128 were chosen to use in both approaches, and further in each double-digested expression plasmid.

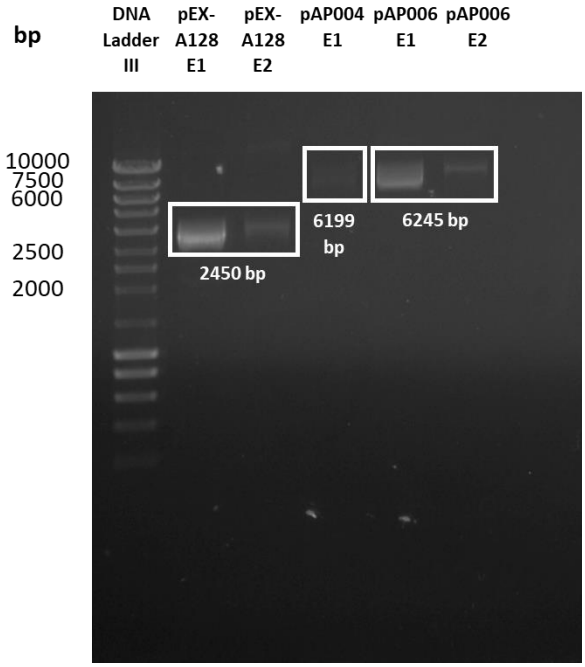

Fig. 3.3 – Integrity evaluation of the plasmid DNA extracted by Miniprep kit for the considered plasmids. 0.8% (w/v) agarose gel. Each sample was loaded with the total volume of 5 µL, including DNA Ladder III. E1 and E2 represent elution 1 and elution 2, respectively.

3.3.3.1 – CW3S cloning in pAP006 expression vector

At first, a pEX-A128 sample and pAP006 were digested with NheI and HindIII restriction enzymes. In Fig. 3.4A-B, the release of the respective fragments in each plasmid is shown. In Fig. 3.4.A, the obtained fragment from pEX-A128 digestion is the CW3S insert, presenting 172 bp (released from the 2450 bp pEX-A128). Due to the insert low size, it was performed an electrophoresis in a 2% agarose gel. The concentrated gel allowed a better visualization of the 172 bp DNA band. pAP006 double-digestion is shown in a 0.8% agarose gel electrophoresis (Fig. 3.4B), where the double-digested plasmid (dd_pAP006) is obtained, with 5443 bp. This weight corresponds to the non-digested pAP006 (6245 bp) without the released fragment (802 bp). Each band of interest was extracted and purified by e-gel. Nanodrop quantification of these samples presented the concentrations and purity ratios of the following Table 3.2.

Table 3.2 – Quantification of the DNA purified by E-gel kit (Invitrogen).

Plasmid DNA	[DNA] (ng/µL)	Abs_{260}/Abs_{280}	Abs_{260}/Abs_{230}
dd_pEX-A128	10.4	1.23	0.76
dd_pAP006	10.8	1.33	0.38

With the indicated concentrations (Table 3.2) and Equation 3.1, it was possible to proceed with the cloning reaction, using T4 DNA Ligase enzyme. Three vector:insert molar ratios were used: 1:3, 1:10 and a negative control of 0:1 (without insert). T4 DNA Ligase, beyond being the catalyst for the ligation reaction, can repair single-strand nicks of DNA to form a double helix structure [97].

The ligation products were successfully transformed in NzyStar competent cells. To ensure that the grown colonies actually had the insert incorporated in the digested plasmid, it was performed a colony screening by a restriction analysis assessment. This consisted in the selection, isolation and purification of 10 colonies (5 for 1:3 and other 5 to 1:10 molar ratios) and further digestion with the same restriction enzymes used before, NheI and HindIII. The DNA fragments release was assessed in a 2% agarose gel electrophoresis (Fig. 3.4C), where it was possible to confirm the release of the insert previously incorporated, and then collect one sample that was positively cloned. In Fig. 3.4C, only two clones did not confirm the insert ligation (clones 1 and 3 from 1:3 molar ratio). Clone 1 shows a pAP006 with different weight than the digested ones, and clone 3 did not show any DNA bands at all. All the other samples revealed the release the insert of interest: the insert CW3S with 172 bp. However, other small bands are observed in some clones, indicating that most likely the digestion was not specific in these DNA fragments. The chosen clone to proceed to the expression was clone 2 of 1:10 molar ratio. This one showed only the two DNA bands of interest. This clone was transformed in *E. coli* Rosetta (DE3) cells and then plated in LB/Agar/Amp/CL.

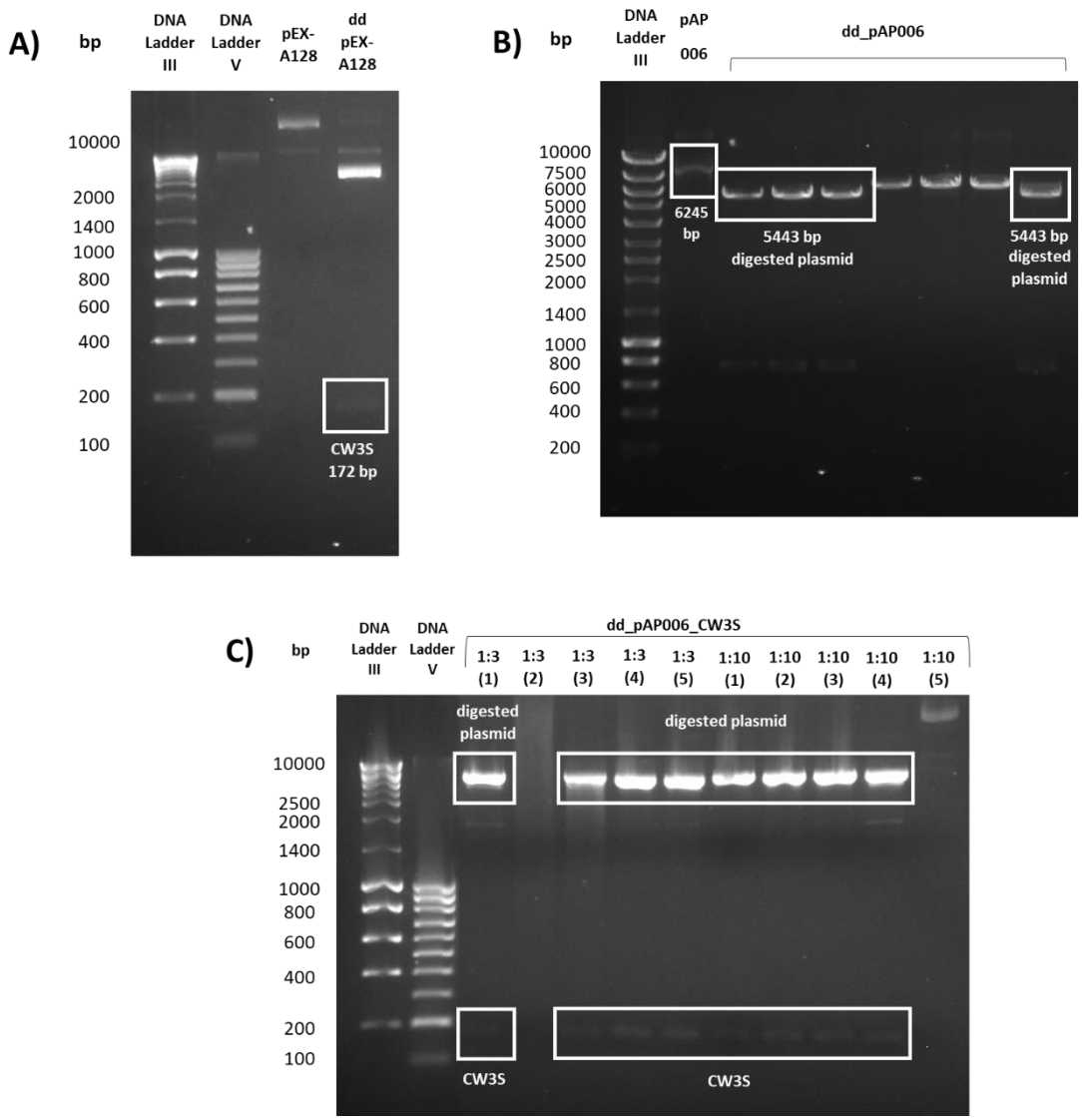

Fig. 3.4 – Hydrolysis with NheI and HindIII restriction enzymes, concerning the CW3S cloning into pAP006 expression vector. Each sample was loaded with the total volume of 5 µL, including DNA Ladders. (A) Hydrolysis of pEX-A128 plasmid in 2% (w/v) agarose gel. (B) Hydrolysis of pAP006 expression vector in 0.8% (w/v) agarose gel. (C) Insert (CW3S) release confirmation (ligation reaction confirmation) from pAP006, in a 2% (w/v) agarose gel.

3.3.3.2 – Cloning of the GFP-EK-CW3S insert in pAP004 expression vector

In the second cloning strategy, pEX-A128 and pAP004 were digested with KpnI and HindIII. In this case, the obtained insert pEX-A128 digestion is composed by GFP-EK-CW3S sequence. The band corresponding to the released insert can be observed in a 0.8% agarose gel electrophoresis (Fig. 3.5A) with 911 bp. In addition, the expression vector dd_pAP004 was completely digested, presenting in the gel the band corresponding to 5350 bp (Fig 3.5A).

Each band of interest was extracted and purified by e-gel. Nanodrop quantification of these samples presented the concentrations and purity ratios of the following Table 3.3.

Table 3.3 – Quantification of the DNA purified by E-gel kit (Invitrogen).

Plasmid DNA	[DNA] (ng/µL)	Abs_{260}/Abs_{280}	Abs_{260}/Abs_{230}
dd_pEX-A128	8.4	1.33	0.84
dd_pAP004	8.7	1.42	0.44

Using the previous concentrations (Table 3.3) and Equation 3.1, it was possible to proceed to the cloning reaction, using T4 DNA Ligase enzyme. Three insert:vector molar ratios were used: 1:3, 1:10 and a negative control of 0:1 (without insert). The ligation products were successfully transformed in NzyStar competent cells and the colony screening by restriction analysis was performed. Thus, the 10 products were double-digested with KpnI and HindIII and ran in a 0.8% agarose gel electrophoresis, in order to assess the DNA fragments release. In Fig. 3.5B, it is possible to conclude that the best clone was clone 5 of 1:10 molar ratio, indicating the correct molecular weights of the DNA bands corresponding obtained (insert GFP-EK-CW3S with 911 bp and dd_pAP004 with 5350 bp). However, most of the other clones screening showed incomplete digestions, in which is verified a division of dd_pAP004 band in two. In addition, clones 4 and 5 from 1:3 molar ratio and clone 4 from 1:10 did not even show the insert release, indicating that possibly the ligation in these samples was not successful. Still, clone 5 of 1:10 molar ratio was positive to proceed to further expression. Finally, it was transformed in *E. coli* Rosetta (DE3) cells and then plated in LB/Agar/Amp/CL.

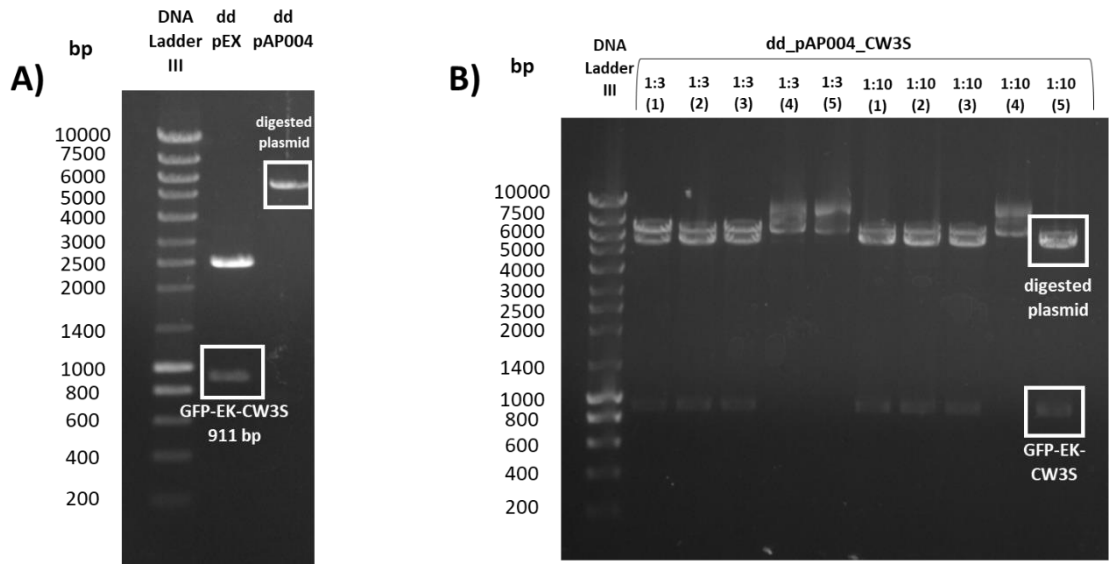

Fig. 3.5 – Hydrolysis with KpnI and HindIII restriction enzymes, concerning the GFP-EK-CW3S cloning into pAP004 expression vector. Each sample was loaded with the total volume of 5 µL, including DNA Ladders. (A) Hydrolysis of pEX-A128 and pAP004 plasmids in 0.8% (w/v) agarose gel. (B) Insert (GFP-EK-CW3S) release confirmation (ligation reaction confirmation) from pAP004, in a 0.8% (w/v) agarose gel.

3.3.4 – Expression of CW3S and GFP-EK-CW3S clone products in *E. coli* Rosetta (DE3) cells

Both expression strategies were performed in *E. coli* as cell host. The benefits of using *E. coli* as an expression host are currently well known. It has unparalleled fast growth kinetics [98], i.e. high cell density cultures are easily achieved using rich complex media, that can be made from easily available and not expensive components [99]. Moreover, transformation with exogenous DNA is fast and easy [100,101]. Concretely, both inserts obtained were cloned into the expression vectors, which in turn were expressed in *E. coli* Rosetta (DE3) cells. Rosetta host strains are BL21(DE3) derivatives designed to improve the expression of proteins by comprising codons rarely used in *E. coli* [102]. These strains include tRNAs for AGG, AGA, AUA, CUA, CCC, GGA codons on a compatible chloramphenicol-resistant plasmid. In Rosetta (DE3) pLysS, the rare tRNA genes are present on the same plasmids that carries the T7 lysozyme gene, for instance pET expression systems. DE3 shows that the host is a lysogen of λDE3, containing a chromosomal copy of the T7 RNA polymerase gene under control of the lacUV5 promoter. This *lac* promoter allows the induction of expression with a lactose analogue, IPTG [101].

The expression was performed using the clone products from 3.3.2, for CW3S and GFP-EK-CW3S. The latter, refers to the expression of the CW3S peptide using GFP as a reporter tag. This tag fluorescent properties made GFP a promising protein to perform fusion constructs and labelled-proteins [103], which led to an easier monitorization of protein expression [104] and other advantages, such as exploring protein stability [105], folding/unfolding mechanisms [106], and to monitor membrane protein topology and overexpression [93,107]. GFP is also very stable under harsh conditions and does not present significant toxicity to the host cell, which can be advantageous in protein expression [108]. Protein expression using GFP-fusion proteins is well documented in literature [106] and CW3S peptide is a novel protein domain that has been explored. Therefore, the two presented expression strategies consist in the proof of concept that CW3S can be expressed alone and, in a second approach, in the form of a to GFP fusion protein.

Since the peptide being expressed has an hydrophobic character, slower rates of protein expression can give newly transcribed recombinant proteins time to fold properly, and consequently an increase of cellular protein concentration with the proper folding. In this way, a few parameters were followed to overcome protein expression possible issues. Incubation temperature after induction was decreased from 37ºC to 30ºC, which would reduce aggregation, since it is favoured at higher temperatures due to hydrophobic interactions temperature dependence [101,109]. Moreover, the pH of the LB media was previously set to 7.5, thus the protein could be kept negatively charged (pI=7.08) and electrostatic repulsions prevent the protein from aggregation.

The cells were grown until OD600nm was about 0.6 to 0.8 and the expression induced with 1mM IPTG in each suspension. Aliquots were collected in different time-periods: the first one in the induction moment (t=0), representing the initial time of the expression profile; the following ones in every 2h until 6h; 18h; 20h; and at the end, 22h.

The following chapters carry the expression results regarding CW3S peptide (3.3.4.1) and GFP-EK-CW3S (3.3.4.2) fusion protein.

3.3.4.1 – CW3S expression in *E. coli* Rosetta (DE3) cells

The results concerning the 5 kDa CW3S peptide expression are represented in Fig. 3.6. According to this, CW3S peptide expression (with pAP006 as expression vector) was not conclusive.

At first, in Fig. 3.6A, the cellular optical density of the protein production increases with the induction time, which was expected. Regarding the expression procedure, as referred in 3.3.3, incubation temperature was reduced to 30ºC and the LB media pH adapted to 7.5, in order to prevent aggregation and favour the peptide correct folding [101,109].

The time-course of the expression and the cellular fractionation process were analysed by SDS-PAGE in a polyacrylamide 12.5% gel (Fig. 3.6B-C). Total protein quantification is presented in Table 3.4. In order to conclude there was protein expression, a difference between before induction (B.I.) and after induction (A.I.) samples has to be observed (Fig. 3.6B), as well as presence of insoluble and/or soluble protein (Fig. 3.6C). In both gels the 5 kDa band corresponding to CW3S expression was not shown. In Fig. 3.6B-C, bands with molecular weights lower than 10 kDa are very difficult to observe, meaning that SDS-PAGE protocol needs optimization. To ensure the visualization of bands of this weight range, a higher gel crosslinking density could be attempted. would lead to a tighter gel, promoting a slower migration of the bands and possibly a better separation of low molecular weight bands [110]. A similar approach for this issue was the production of Tris-tricine gels, to improve SDS-PAGE resolution (Fig. 3.6D-E) [111]. However, the gels did not run properly and this technique still needs optimization.

Table 3.4 – Total protein quantification (using BCA) of cellular fractionation samples.

Sample	Total Protein (mg/mL)
Supernatant Low speed centrifugation (SC)	1.24 ± 0.07
Supernatant Ultracentrifugation (SU)	1.15 ± 0.16

The absence of conclusion about CW3S expression clearly highlights the need of a reporter tag to its production. The peptide small weight would benefit if expressed in a fusion protein, for a better identification in the gels, monitorization over expression and easier purification (e.g. with affinity tags). Moreover, a hydrophobic peptide with tendency to form erroneous disulphide bonds (CW3S has cysteine and methionine residues) can lead to protein misfolding and aggregation into inclusion bodies [101]. Since cysteine oxidation takes place in the cell periplasm, there are engineered *E. coli* strains that allow expression in the cytoplasm. For this, documented Origami cells can be attempted [101]. Nevertheless, the second cloning strategy further discussed, consists in CW3S-fused to GFP protein expression.

The objective regarding this cloning strategy was to assess the proof of concept of a WW domain expression in *E. coli*. Unfortunately, due to the peptide small size and lack of optimization on peptide separation electrophoresis techniques, it is impossible to conclude about its expression in bacterial hosts.

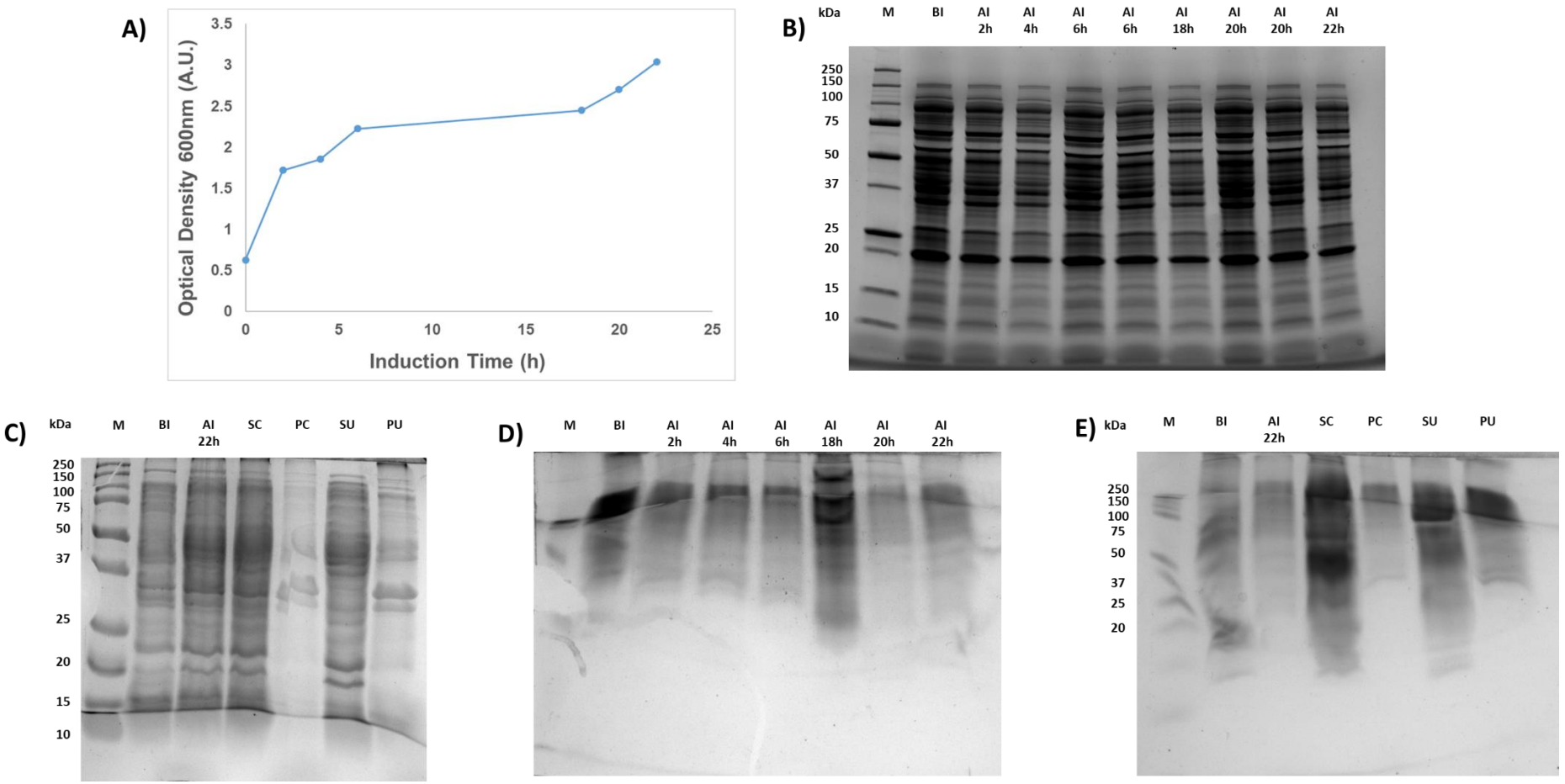

Fig. 3.6 – Analysis of the expression of CW3S. (A) Monitorization of the Optical Density at 600nm with induction time (induction with 1 mM IPTG). (B) Time-course expression evaluation by SDS-PAGE (12.5% polyacrylamide). (C) Cellular fractionation evaluation by SDS-PAGE (12.5% polyacrylamide). (D) Time-course expression evaluation by Tris-tricine SDS-PAGE. (E) Cellular fractionation evaluation by Tris-tricine SDS-PAGE; M: Precision Plus Protein Dual Xtra Prestained Protein Standard, BI: before induction, AI: after induction, SC: supernatant of low speed centrifugation, PC: pellet of low speed centrifugation, SU: supernatant of ultracentrifugation, PU: pellet of ultracentrifugation.

3.3.4.2 – CW3S expression in *E. coli* Rosetta (DE3) cells as a GFP-fusion protein

Fusion protein GFP-EK-CW3S (33 kDa) expression results are represented in Fig. 3.7. According to this, the fusion-protein (with pAP004 as expression vector) was expressed.

Firstly, in Fig. 3.7A, the cellular optical density of the protein production increases with the induction time, which was expected. Regarding the expression procedure, as referred in 3.3.3, incubation temperature was reduced to 30ºC and the LB media pH adapted to 7.5, in order to prevent aggregation and favour the peptide correct folding, as also performed for expression of CW3S itself [101,109]. However, GFP fluorescence intensity did not show any substantial increase as shown in OD_{600nm}. It was expected that fluorescence intensity would increase as the OD_{600nm} values, once the fusion protein was being expressed. Nonetheless, from t=0 to 2h induction, there is a fluorescence intensity significant increase, reaching a maximum at approximately 1362 F.U.. In general, from 2h to 22h A.I fluorescence intensity stays practically constant (1362 to 1326 F.U.).

The time-course of the protein expression and the following cellular fractionation were also analysed by SDS-PAGE in a polyacrylamide 12.5% gel (Fig. 3.7B-C).

According to Fig. 3.7B, the production of GFP-EK-CW3S over a total of 22h of induction was verified. By monitoring the band of GFP-EK-CW3S fusion protein of 33 kDa, it was possible to observe an increase of protein band intensity, between before induction (B.I.) and after induction (A.I.) aliquots. Also, according to SDS-PAGE gels of the Fig. 3.7C, it is possible to notice that the fusion protein was expressed in both soluble and insoluble forms. The soluble fractions correspond to the fractions obtained from the supernatants of the low speed centrifugation (SC) and the ultracentrifugation (SU). The SU fraction was the fraction with the highest fluorescence intensity (6100 F.U.) (Fig. 3.7D). The same sample presented a total protein concentration of 1.0 ± 0.29 mg/mL, where 1.2% ([GFP](mg/mL)/[Total Protein](mg/mL)*100) corresponds to GFP fusion protein (0.012 mg/mL). Total protein quantification regarding the cellular fractionation is represented in Table 3.5.

Table 3.5 – Total protein quantification (using BCA) of the cellular fractionation samples.

Sample	Total Protein (mg/mL)
Supernatant Low speed centrifugation (SC)	2.89 ± 0.15
Supernatant Ultracentrifugation (SU)	1.0 ± 0.29

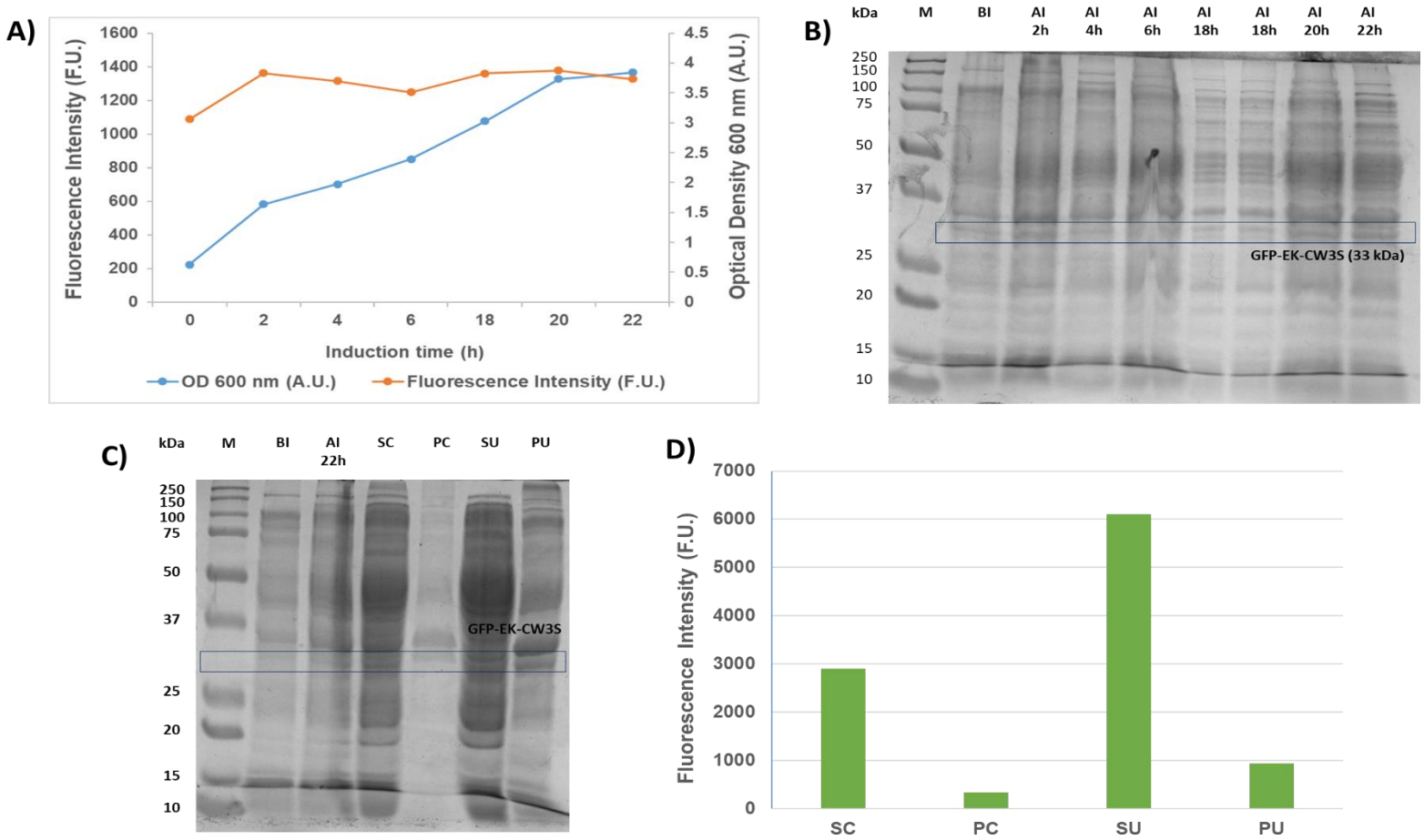

Fig. 3.7 – Analysis of the expression of GFP-EK-CW3S. (A) Monitorization of the Optical Density at 600nm and Fluorescence Intensity (λexcitation and λemission of 485 nm and 535 nm, respectively) with induction time (induction with 1 mM IPTG). (B) Time-course expression evaluation by SDS-PAGE (12.5% polyacrylamide). (C) Cellular fractionation evaluation by SDS-PAGE (12.5% polyacrylamide). (D) Fluorescence intensity of cellular fractionation samples. M: Precision Plus Protein Dual Xtra Prestained Protein Standard, BI: before induction, AI: after induction, SC: supernatant of low speed centrifugation, PC: pellet of low speed centrifugation, SU: supernatant of ultracentrifugation, PU: pellet of ultracentrifugation.

The SU fraction was firstly purified by affinity chromatography with Sepharose-CL6B functionalized with the affinity ligand A4C7 in a Tricorn 10/50 column. A4C7 is a synthetic ligand with high affinity towards GFP [94]. From this process, a chromatogram was obtained (Fig. 3.8A), overlaying fluorescence intensity and absorbance at 560 nm of BCA. It is verified that some contaminant proteins were firstly eluted in the washes with PBS 1x pH 7.4 (18 to 36 mL) and afterwards the GFP-EK-CW3S fusion protein in the elution with 0.1 mM glycine–NaOH pH 9.0 (74 to 82 mL). 76 to 80 mL elutions were collected to a pool aliquot, corresponding to the absorbance and GFP-fluorescence peaks. Purification through A4C7 ligand chromatography registered 100% binding to GFP-fusion protein, however only 15.8% was recovered. Basing on existent literature, it would be expected a higher recovery of GFP-fusion protein [94]. Thus, CW3S hydrophobicity character might be affecting A4C7 interaction to GFP. A4C7 only elution peak (Fig. 3.8A) was loaded in a 12.5% polyacrylamide SDS-PAGE gel (Fig. 3.8D). At first, it is seen that A4C7 chromatography did not purify in the best conditions. The lane was clearer than the loading sample, however it should have been more accurate purification process. It is observed that the elution purity is too low. A4C7 has a capacity of 1 mg GFP/g of resin [94]. The concentration loaded was 0.012 mg/mL of GFP, in a total amount of 0.18 mg of GFP introduced in the column, indicating that A4C7 was not saturated with loading sample.

Although the elution fraction from A4C7 affinity chromatography did not reveal high purity, the samples were still purified in in a gel-filtration in a HiLoad 16/600 Superdex 75 (GE Healthcare) column. Gel-filtration is a size-exclusion chromatography, in which molecules with higher molecular weight are eluted first. A chromatogram with fluorescence intensity and absorbance at 560 nm was also obtained (Fig. 3.8B). Both total protein (by BCA) and GFP-fluorescence (≈32 to 50 mL) were detectable. These volumes were collected to an aliquot, representing the maximum of GFP-fluorescence to use in the next purification. The fraction GFP concentration obtained after gel-filtration was 0.0014 mg/mL. Before the cleavage with EK, the pool sample from gel-filtration was exchanged to EK-cleavage buffer Tris-HCl pH 7.4, 50 mM NaCl, 2 mM $CaCl_2$. The concentrated sample used to perform EK-cleavage and to purify in the second gel-filtration presented now a GFP concentration of 0.0086 mg/mL. It was used 1 U of EK in excess to cleave GFP-EK-CW3S.

With EK-cleavage and following gel-filtration, it was expected to observe the separation of GFP, EK and CW3S, obtaining the final pure peptide. Fig. 3.8C represents the chromatogram obtained from the second gel filtration. The expected interpretation of this chromatogram is that GFP and EK are obtained firstly, and finally the CW3S peptide (5 kDa). Therefore, the observed peaks in this purification were analysed in a SDS-PAGE gel (12.5%), along with an aliquot of the product from A4C7 chromatography, the first gel filtration and a GFP-EK-CW3S loading sample (Fig. 3.8D). The referred peaks are represented by peak 1 (P1) and peak 2 (P2). The latter is a

pool of the last three small peaks obtained in the elution. Observing Fig. 3.8C profile, one possible interpretation was that GFP was eluted in the first peak (P1) and the CW3S peptide at the end of the elution (P2), respecting gel-filtration size principles.

From Fig. 3.8D, it is clear that the purification of CW3S peptide could not be concluded. The highlighted area corresponds to the 33 kDa band of GFP fusion protein. Regarding EK cleavage, it would be expected to observe a decrease in P1 band molecular weight, since in this step the GFP-fusion protein was cleaved. Therefore, the 33 kDa band would decrease to 28 kDa (GFP molecular weight). However, the band decrease was not verified, which can mean that EK cleavage was not efficient. Its protocol could require optimization for more effective fusion protein cleavage. Lastly, in P2, CW3S (5 kDa) is not observed. It could possibly have precipitated or is very diluted in the sample, since there is not a clear signal of CW3S presence in Fig. 3.8C. Moreover, if EK did not cleave effectively, most of CW3S is still present in GFP-fusion protein, which can be observed, however in a thinner band. As an attempt to observe CW3S 5 kDa band, a Tris-tricine gel was performed (Fig. 3.8E). This protocol still needs optimization and no possible conclusion could be taken.

CW3S is a highly hydrophobic peptide, which can be a challenge in purification purposes. Its tendency to precipitate requires the peptide detection at very low concentrations. Thus, the peptide could have precipitated during the process, thus compromising the purification efficiency. Therefore, different reporter tags could be experimented for CW3S expression and purification. Affinity tags can be detected along the expression and purification processes and can reach maximal solubility [101,112]. Glutathione-S-transferase (GST), for example, is a non-peptide fusion partner that has the extra advantage in performing as a solubility enhancer [113]. The reasons why it can act as solubility enhancer remain unclear [101]. From the most studied non-peptide fusion tags, GST has the lowest solubility enhancement capacity [114]. However, it can be used to purify the fused protein by affinity chromatography [115].

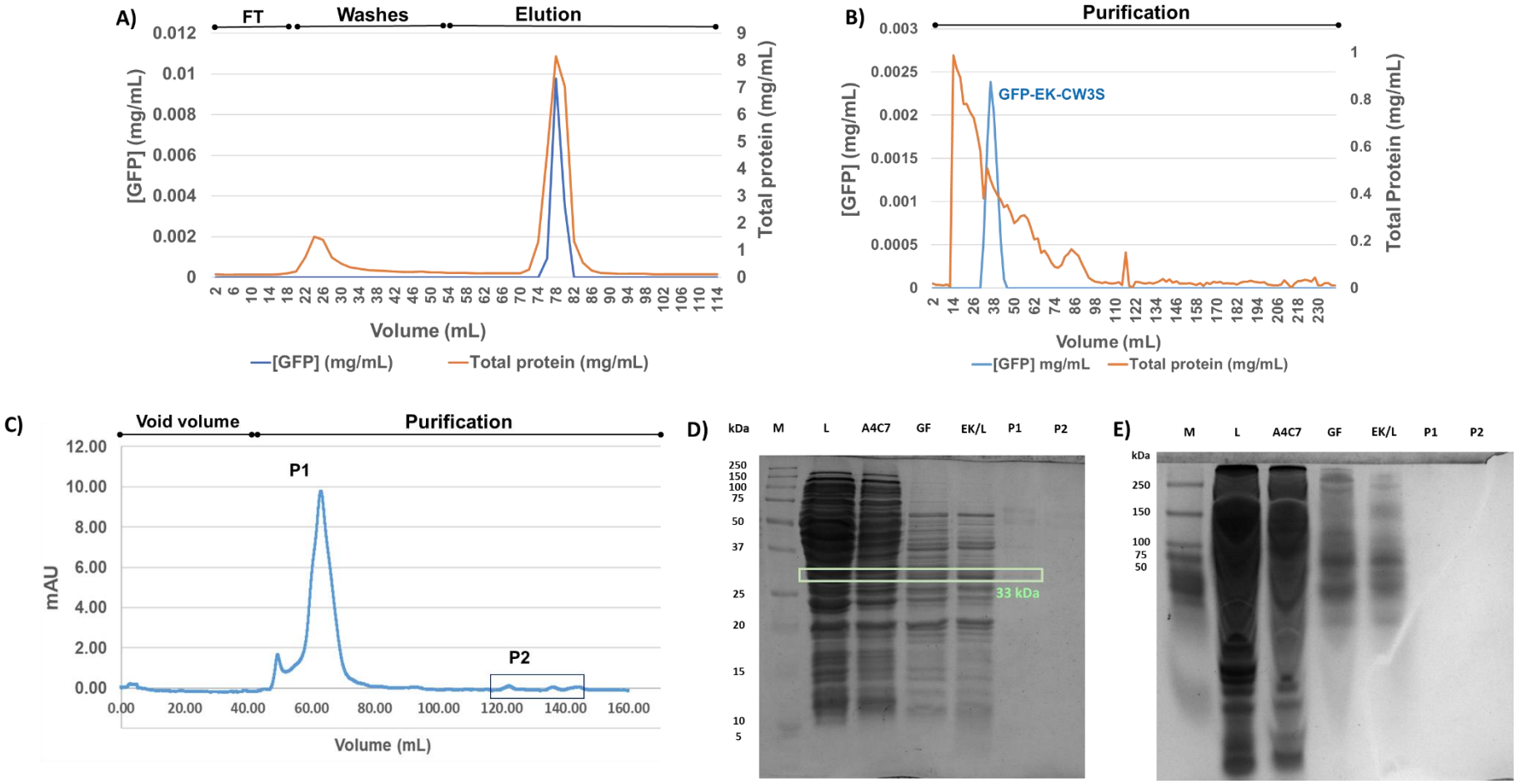

Fig. 3.8 – Purification of GFP-EK-CW3S fusion protein analysis. (A) Chromatogram of chromatography with A4C7 ligand. (B) Chromatogram of the first gel-filtration. (C) Chromatogram of the second gel-filtration (exported from AKTA Pure System). (D) Purification course evaluation through SDS-PAGE (12.5% polyacrylamide). (E) Purification course evaluation through Tris-tricine SDS-PAGE. M: Precision Plus Protein Dual Xtra Prestained Protein Standard, L: Loading, A4C7: Pool obtained from the elution of the chromatography with A4C7 ligand, GF: Pool obtained from the gel-filtration peak marked as GFP-EK-CW3S, EK/L: Loading of the second gel-filtration (after cleavage with EK), P1: First peak of the second gel filtration, P2: Pool of the final elution peaks of the second gel filtration.

3.4 – Concluding remarks

This chapter presented the differences between two expression strategies of the CW3S peptide in bacterial cells. In the first approach (CW3S expression), nothing could be concluded, since the 5 kDa peptide could not be identified in both time-course and fractionation gel results. However, from this chapter, the benefits in using fusion proteins in small peptides expression were clearly demonstrated and experimented. In the second strategy, GFP allowed a much easier monitorization and quantification of an intracellular fusion protein, without the need to lyse cells. Moreover, the final recovery of the peptide is much simpler. A4C7 ligand is a major advantage for GFP-fusion proteins, which then can be used in chromatography. Other techniques can be combined, such as enzymatic cleavage of a specific site (e.g. Enterokinase) and complementary chromatography. Despite the CW3S could not be identified at the end of the purification process, it is demonstrated that GFP-fusion protein expression and purification is an efficient and optimistic strategy for protein production.

Chapter 4

Final concluding remarks

Chapter 4

This work focused on the production of CW3S peptide and its assessment as a suitable affinity reagent to HSA, aiming to biotechnological or biopharmaceutical applications. CW3S production and respective studies involved a wide variety of techniques. The peptide was produced through two different strategies: solid-phase peptide synthesis (SPPS) and the peptide expression in bacterial cells.

Chemical production through SPPS brought several benefits to reduce errors in the production process, such as the possibility of optimizing a protocol, introducing double coupling and capping reactions. The incorporation of pseudo-proline units was also a determinant parameter for synthesis of a long peptide. Finally, insoluble resins (such as PEG-based resins) and the microwave technology promoted a more efficient production [77]. In general, SPPS presents various advantages: the use of filtration to separate the intermediate peptides from soluble reagents and solvents, excess reagents can be employed to drive reactions to completion, and finally, significant losses can be minimized as the peptide remains attached to the support (resin) during the synthesis. However, SPPS has some limitations. The formation of incomplete reactions, side reactions or impure reagents can accumulate throughout peptide chain assembly and increase the final product contaminations. Specially for long peptides (superior to 50 residues), the probability of occurring errors is quite significant, being SPPS not recommended for these products [69,81]. The 41 amino acid CW3S is far from being a short peptide, thus this production was a difficult synthesis. One error could compromise all peptide sequence and its final identity (molecular weight). Furthermore, long peptides tend to spend more volume of solvents, which increases their production cost. Despite all difficulties, CW3S chemically produced presented the correct molecular weight (through Mass Spectrometry), besides being thermally unstable (Tm=29ºC), as compared to previously prepared WW domains.

The biological production of CW3S involved several steps until obtaining the expressed peptide. This strategy allowed the design and manipulation of the gene that encodes for CW3S peptide. Therefore, this permitted the selection of the interest sequence and its introduction in an expression vector (cloning). The cloned vector was then transformed into the host of choice (*E. coli* bacterial cells), the expression was induced and, after concluding the peptide was expressed, it was ready for purification and characterization. On the contrary of SPPS, this process can be highly monitored and easily manipulated in each step. This could be positive for error analysis and troubleshooting in various stages, however, several errors can occur. The most common ones are cloning (gene insertion in the expression vector), poor growth of the cells, inclusion body formation, protein inactivity and, finally, not obtaining any protein [101]. In order to avoid these particular situations, this peptide was produced in Rosetta BL21 *E. coli* cells strain, that contain

rarely used codons in *E. coli*, including tRNAs for AGG, AGA, AUA, CUA, CCC, GGA codons on a compatible chloramphenicol-resistant plasmid. Moreover, this production was performed with a reporter tag (GFP), for having an easier monitorization of a small peptide (5 kDa is not easily detected in SDS-PAGE gels), through the fusion protein size (33 kDa) and also its fluorescent properties. Moreover, GFP facilitated the purification process, using a synthetic ligand with high affinity towards GFP. In addition, a protease cleavage site with EK allowed the separation of GFP from the interest peptide (CW3S). Despite the positive expression obtained of the fusion protein (0.012 mg/mL of GFP in the soluble fraction), the purification process was more complicated. CW3S highly hydrophobic characteristics promote its precipitation during the process, compromising its identification in the final elutions. In addition, from the purification profile (Fig. 3.8D), it was possible to conclude the unsuccessful purification process. In A4C7 ligand chromatography, the resin could have been highly saturated, requiring a lower concentration of the loading sample. Additionally, protease cleavage with EK also could have been ineffective. Therefore, biologically produced GFP-EK-CW3S purification require significant optimizations, such as the loading in A4C7 chromatography and EK cleavage protocol. Moreover, affinity tags such as GST, could be experimented for CW3S expression and purification to improve the peptide solubility.

Regarding CW3S potential affinity properties towards HSA, affinity techniques that require low peptide concentration need to be optimized in the future. From this analysis, it was demonstrated that CW3S hydrophobicity did not allow the use of techniques that require high peptide concentrations (e.g. 2 mM in aqueous buffer conditions), for instance the binding tests in agarose against HSA. However, both MST and MP-SPR must be further optimized in order to obtain significant results regarding the peptide interaction with HSA. Concretely, in MP-SPR the regeneration conditions must be completely changed, or even replaced with simple binding buffer (PBS 1x pH 7.4) washes, in order to attempt the saturation of CW3S-modified Au-surface and consequently, its interaction with HSA analysis. Finally, it would be interesting to try performing MST and MP-SPR in the same conditions to evaluate CW3S interaction with IgG, and the selectivity towards HSA.

Chapter 5

References

[1] A.T. Hanke, M. Ottens, Purifying biopharmaceuticals: Knowledge-based chromatographic process development, Trends Biotechnol. 32 (2014) 210–220. doi:10.1016/j.tibtech.2014.02.001.

[2] A.M.G.C. Dias, A.C.A. Roque, The future of protein scaffolds as affinity reagents for purification, Biotechnol. Bioeng. 114 (2016) 481–491. doi:10.1002/bit.26090.

[3] Á.M. Cuesta, N. Sainz-Pastor, J. Bonet, B. Oliva, L. Álvarez-Vallina, Multivalent antibodies: When design surpasses evolution, Trends Biotechnol. 28 (2010) 355–362. doi:10.1016/j.tibtech.2010.03.007.

[4] S. Gülich, M. Linhult, S. Ståhl, S. Hober, Engineering streptococcal protein G for increased alkaline stability, Protein Eng. 15 (2002) 835–842. doi:10.1093/protein/15.10.835.

[5] N. Sawyer, E.B. Speltz, L. Regan, NextGen protein design., Biochem. Soc. Trans. 41 (2013) 1131–6. doi:10.1042/BST20130112.

[6] P. Gronemeyer, R. Ditz, J. Strube, Trends in Upstream and Downstream Process Development for Antibody Manufacturing, Bioengineering. 1 (2014) 188–212. doi:10.3390/bioengineering1040188.

[7] M. Hedhammar, T. Gräslund, S. Hober, Protein engineering strategies for selective protein purification, Chem. Eng. Technol. 28 (2005) 1315–1325. doi:10.1002/ceat.200500144.

[8] X. Fang, W.W. Zhang, Affinity separation and enrichment methods in proteomic analysis, J. Proteomics. 71 (2008) 284–303. doi:10.1016/j.jprot.2008.06.011.

[9] C.R. Lowe, Combinatorial approaches to affinity chromatography, Curr. Opin. Chem. Biol. 5 (2001) 248–256. doi:10.1016/S1367-5931(00)00199-X.

[10] O. Boutureira, G.J.L. Bernardes, Advances in chemical protein modification, Chem. Rev. 115 (2015) 2174–2195. doi:10.1021/cr500399p.

[11] A. Skerra, Alternative non-antibody scaffolds for molecular recognition, Curr. Opin. Biotechnol. 18 (2007) 295–304. doi:10.1016/j.copbio.2007.04.010.

[12] A. Wörn, A. Plückthun, Stability engineering of antibody single-chain Fv fragments., J. Mol. Biol. 305 (2001) 989–1010. doi:10.1006/jmbi.2000.4265.

[13] H.K. Binz, P. Amstutz, A. Plückthun, Engineering novel binding proteins from nonimmunoglobulin domains., Nat. Biotechnol. 23 (2005) 1257–1268. doi:10.1038/nbt1127.

[14] M.J. Taussig, O. Stoevesandt, C.A.K. Borrebaeck, A.R. Bradbury, D. Cahill, C. Cambillau, A. de Daruvar, S. Dübel, J. Eichler, R. Frank, T.J. Gibson, D. Gloriam, L. Gold, F.W. Herberg, H. Hermjakob, J.D. Hoheisel, T.O. Joos, O. Kallioniemi, M. Koegl, Z. Konthur, B. Korn, E. Kremmer, S. Krobitsch, U. Landegren, S. van der Maarel, J. McCafferty, S. Muyldermans, P.-Å. Nygren, S. Palcy, A. Plückthun, B. Polic, M. Przybylski, P. Saviranta, A. Sawyer, D.J. Sherman, A. Skerra, M. Templin, M. Ueffing, M. Uhlén, ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome, Nat. Methods. 4 (2007) 13–17. doi:10.1038/nmeth0207-187a.

[15] C. Grönwall, S. Ståhl, Engineered affinity proteins-Generation and applications, J. Biotechnol. 140 (2009) 254–269. doi:10.1016/j.jbiotec.2009.01.014.

[16] Y.L. Boersma, A. Plückthun, DARPins and other repeat protein scaffolds: Advances in engineering and applications, Curr. Opin. Biotechnol. 22 (2011) 849–857. doi:10.1016/j.copbio.2011.06.004.

[17] M. Uhlén, S. Hober, Generation and validation of affinity reagents on a proteome-wide level, J. Mol. Recognit. 22 (2009) 57–64. doi:10.1002/jmr.891.

[18] D. Ponsel, J. Neugebauer, K. Ladetzki-Baehs, K. Tissot, High affinity, developability and functional size: The holy grail of combinatorial antibody library generation, Molecules. 16 (2011) 3675–3700. doi:10.3390/molecules16053675.

[19] A.C. Braisted, J.A. Wells, Minimizing a binding domain from protein A., Proc. Natl. Acad. Sci. 93 (1996) 5688–5692. doi:10.1073/pnas.93.12.5688.

[20] J. Lindgren, C. Ekblad, L. Abrahmsøn, A. Eriksson, L. Abrahmsén, A. Eriksson Karlström, A Native Chemical Ligation Approach for Combinatorial Assembly of Affibody Molecules, ChemBioChem. 13 (2012) 1024–1031. doi:10.1002/cbic.201200052.

[21] N. Per-a, P.Å. Nygren, Alternative binding proteins: Affibody binding proteins developed from a small three-helix bundle scaffold, FEBS J. 275 (2008) 2668–2676. doi:10.1111/j.1742-4658.2008.06438.x.

[22] M. Wikman, A.C. Steffen, E. Gunneriusson, V. Tolmachev, G.P. Adams, J. Carlsson, S. Ståhl, Selection and characterization of HER2/neu-binding affibody ligands, Protein Eng. Des. Sel. 17 (2004) 455–462. doi:10.1093/protein/gzh053.

[23] P.A. Löfdahl, O. Nord, L. Janzon, P.A. Nygren, Selection of TNF-α binding affibody molecules using a β-lactamase protein fragment complementation assay, N. Biotechnol. 26 (2009) 244–250. doi:10.1016/j.nbt.2009.06.980.

[24] C.I.L. Justino, A.C. Duarte, T.A.P. Rocha-Santos, Analytical applications of affibodies, TrAC - Trends Anal. Chem. 65 (2015) 73–82. doi:10.1016/j.trac.2014.10.014.

[25] K. Nord, E. Gunneriusson, M. Uhlén, P.A. Nygren, Ligands selected from combinatorial libraries of protein A for use in affinity capture of apolipoprotein A-1(M) and Taq DNA polymerase, J. Biotechnol. 80 (2000) 45–54. doi:10.1016/S0168-1656(00)00232-7.

[26] M. Lindborg, A. Dubnovitsky, K. Olesen, T. Björkman, L. Abrahmsén, J. Feldwisch, T. Härd, High-affinity binding to staphylococcal protein A by an engineered dimeric Affibody molecule, Protein Eng. Des. Sel. 26 (2013) 635–644. doi:10.1093/protein/gzt038.

[27] C. Gronwall, A. Jonsson, S. Lindstrom, E. Gunneriusson, S. Stahl, N. Herne, Selection and characterization of Affibody ligands binding to Alzheimer amyloid β peptides, J. Biotechnol. 128 (2007) 162–183. doi:10.1016/j.jbiotec.2006.09.013.

[28] C. Grönwall, A. Sjöberg, M. Ramström, I. Höidém-Guthenberg, S. Hober, P. Jonasson, S. Ståhl, Affibody-mediated transferrin depletion for proteomics applications, Biotechnol. J. 2 (2007) 1389–1398. doi:10.1002/biot.200700053.

[29] H. Orelma, L.O. Morales, L.-S. Johansson, I.C. Hoeger, I. Filpponen, C. Castro, O.J. Rojas, J. Laine, Affibody conjugation onto bacterial cellulose tubes and bioseparation of human serum albumin, RSC Adv. 4 (2014) 51440–51450. doi:10.1039/C4RA08882D.

[30] M. Ramstrom, Z. Aida, C. Grönwall, J. Hanrieder, J. Bergquist, S. Hober, Development of affinity columns for the removal of high-abundance proteins in cerebrospinal fluid, Biotechnol. Appl. Biochem. 52 (2009) 159–166. doi:10.1042/BA20080028.

[31] A. Jonsson, J. Dogan, N. Herne, L. Abrahmse, Engineering of a femtomolar affinity binding protein to human serum albumin, 21 (2008) 515–527. doi:10.1093/protein/gzn028.

[32] M. Hedhammar, T. Gräslund, M. Uhlén, S. Hober, Negatively charged purification tags

for selective anion-exchange recovery, Protein Eng. Des. Sel. 17 (2004) 779–786. doi:10.1093/protein/gzh092.

[33] M. Hedhammar, S. Hober, Zbasic-A novel purification tag for efficient protein recovery, J. Chromatogr. A. 1161 (2007) 22–28. doi:10.1016/j.chroma.2007.05.091.

[34] E. Fiedler, M. Fiedler, G. Proetzel, T. Scheuermann, U. Fiedler, R. Rudolph, Affilin™ Molecules Novel Ligands for Bioseparation, Food Bioprod. Process. 84 (2006) 3–8. doi:10.1205/fbp.05222.

[35] H. Ebersbach, E. Fiedler, T. Scheuermann, M. Fiedler, M.T. Stubbs, C. Reimann, G. Proetzel, R. Rudolph, U. Fiedler, Affilin-Novel Binding Molecules Based on Human γ-B-Crystallin, an All β-Sheet Protein, J. Mol. Biol. 372 (2007) 172–185. doi:10.1016/j.jmb.2007.06.045.

[36] S. Lorey, E. Fiedler, A. Kunert, J. Nerkamp, C. Lange, M. Fiedler, E. Bosse-Doenecke, M. Meysing, M. Gloser, C. Rundfeldt, U. Rauchhaus, I. Hanssgen, T. Gottler, A. Steuernagel, U. Fiedler, U. Haupts, Novel ubiquitin-derived high affinity binding proteins with tumor targeting properties, J. Biol. Chem. 289 (2014) 8493–8507. doi:10.1074/jbc.M113.519884.

[37] T. Hoffmann, L.K.J. Stadler, M. Busby, Q. Song, A.T. Buxton, S.D. Wagner, J.J. Davis, P. Ko Ferrigno, Structure-function studies of an engineered scaffold protein derived from stefin A. I: Development of the SQM variant, Protein Eng. Des. Sel. 23 (2010) 403–413. doi:10.1093/protein/gzq012.

[38] L.K.J. Stadler, T. Hoffmann, D.C. Tomlinson, Q. Song, T. Lee, M. Busby, Y. Nyathi, E. Gendra, C. Tiede, K. Flanagan, S.J. Cockell, A. Wipat, C. Harwood, S.D. Wagner, M.A. Knowles, J.J. Davis, N. Keegan, P. Ko Ferrigno, Structure-function studies of an engineered scaffold protein derived from Stefin A. II: Development and applications of the SQT variant, Protein Eng. Des. Sel. 24 (2011) 751–763. doi:10.1093/protein/gzr019.

[39] S. Straw, P.K. Ferrigno, Q. Song, D. Tomlinson, F. Del Galdo, Proof of concept study to identify candidate biomarkers of fibrosis using high throughput peptide aptamer microarray and validate by enzyme linked immunosorbant assay, J. Biomed. Sci. Eng. 6 (2013) 32–42. doi:10.4236/jbise.2013.68A2005.

[40] J. Nilvebrant, S. Hober, the Albumin-Binding Domain As a Scaffold for Protein Engineering, Comput. Struct. Biotechnol. J. 6 (2013) 1–8. doi:10.5936/csbj.201303009.

[41] R. Johansson, L.C. Gunnarsson, M. Ohlin, S. Ohlson, Thermostable carbohydrate-binding modules in affinity chromatography, in: J. Mol. Recognit., 2006: pp. 275–281. doi:10.1002/jmr.794.

[42] S.P. Edmondson, J.W. Shriver, DNA-binding proteins Sac7d and Sso7d from Sulfolobus, Methods Enzymol. 334 (2001) 129–145. doi:10.1016/S0076-6879(01)34463-4.

[43] G. Béhar, M. Bellinzoni, M. Maillasson, L. Paillard-Laurance, P.M. Alzari, X. He, B. Mouratou, F. Pecorari, Tolerance of the archaeal Sac7d scaffold protein to alternative library designs: Characterization of anti-immunoglobulin G Affitins, Protein Eng. Des. Sel. 26 (2013) 267–275. doi:10.1093/protein/gzs106.

[44] A. Correa, S. Pacheco, A.E. Mechaly, G. Obal, G. Béhar, B. Mouratou, P. Oppezzo, P.M. Alzari, F. Pecorari, Potent and specific inhibition of glycosidases by small artificial binding proteins (Affitins), PLoS One. 9 (2014). doi:10.1371/journal.pone.0097438.

[45] B. Mouratou, F. Schaeffer, I. Guilvout, D. Tello-Manigne, A.P. Pugsley, P.M. Alzari, F. Pecorari, Remodeling a DNA-binding protein as a specific in vivo inhibitor of bacterial secretin PulD., Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 17983–8. doi:10.1073/pnas.0702963104.

[46] M. Cinier, M. Petit, M.N. Williams, R.M. Fabre, F. Pecorari, D.R. Talham, B. Bujoli, C. Tellier, Bisphosphonate adaptors for specific protein binding on zirconium phosphonate-based microarrays, Bioconjug. Chem. 20 (2009) 2270–2277. doi:10.1021/bc9002597.

[47] F.F. Miranda, E. Brient-Litzler, N. Zidane, F. Pecorari, H. Bedouelle, Reagentless fluorescent biosensors from artificial families of antigen binding proteins, Biosens. Bioelectron. 26 (2011) 4184–4190. doi:10.1016/j.bios.2011.04.030.

[48] G. Béhar, A. Renodon-Cornière, B. Mouratou, F. Pecorari, Affitins as robust tailored reagents for affinity chromatography purification of antibodies and non-immunoglobulin proteins, J. Chromatogr. A. 1441 (2016) 44–51. doi:10.1016/j.chroma.2016.02.068.

[49] F. Parmeggiani, R. Pellarin, A.P. Larsen, G. Varadamsetty, M.T. Stumpp, O. Zerbe, A. Caflisch, A. Plückthun, Designed Armadillo Repeat Proteins as General Peptide-Binding Scaffolds: Consensus Design and Computational Optimization of the Hydrophobic Core, J. Mol. Biol. 376 (2008) 1282–1304. doi:10.1016/j.jmb.2007.12.014.

[50] P. Alfarano, G. Varadamsetty, C. Ewald, F. Parmeggiani, R. Pellarin, O. Zerbe, A. Plückthun, A. Caflisch, Optimization of designed armadillo repeat proteins by molecular dynamics simulations and NMR spectroscopy, Protein Sci. 21 (2012) 1298–1314. doi:10.1002/pro.2117.

[51] M. Breitman, A. Zilberberg, M. Caspi, R. Rosin-Arbesfeld, The armadillo repeat domain of the APC tumor suppressor protein interacts with Striatin family members, Biochim. Biophys. Acta - Mol. Cell Res. 1783 (2008) 1792–1802. doi:10.1016/j.bbamcr.2008.04.017.

[52] G. Varadamsetty, D. Tremmel, S. Hansen, F. Parmeggiani, A. Plückthun, Designed armadillo repeat proteins: Library generation, characterization and selection of peptide binders with high specificity, J. Mol. Biol. 424 (2012) 68–87. doi:10.1016/j.jmb.2012.08.029.

[53] Y. Lee, J.J. Lee, S. Kim, S.-C.C. Lee, J. Han, W. Heu, K. Park, H.J. Kim, H.-K.K. Cheong, D. Kim, H.-S.S. Kim, K.W. Lee, Dissecting the critical factors for thermodynamic stability of modular proteins using molecular modeling approach., PLoS One. 9 (2014) e98243. doi:10.1371/journal.pone.0098243.

[54] S.-C.S.-G. Lee, K. Park, J. Han, J. -j. Lee, H.-S.J. Kim, S. Hong, W. Heu, Y.J. Kim, J.-S. Ha, S.-C.S.-G. Lee, H.-K. Cheong, Y.H. Jeon, D. Kim, H.-S.J. Kim, Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering, Proc. Natl. Acad. Sci. 109 (2012) 3299–3304. doi:10.1073/pnas.1113193109.

[55] W. Heu, J. Choi, J. Lee, S. Jeong, H. Kim, Protein Binder for A ffi nity Puri fi cation of Human Immunoglobulin Antibodies, (2014).

[56] M. Sudol, T. Hunter, New wrinkles for an old domain, Cell. 103 (2000) 1001–1004. doi:10.1016/S0092-8674(00)00203-8.

[57] E.K. Koepf, H.M. Petrassi, M. Sudol, J.W. Kelly, WW: An isolated three-stranded antiparallel beta-sheet domain that unfolds and refolds reversibly; evidence for a structured hydrophobic cluster in urea and GdnHCl and a disordered thermal unfolded state., Protein Sci. 8 (1999) 841–853. doi:10.1110/ps.8.4.841.

[58] M.J. Macias, M. Hyvönen, E. Baraldi, J. Schultz, M. Sudol, M. Saraste, H. Oschkinat, Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide, Nature. 382 (1996) 646–649. doi:10.1038/382646a0.

[59] M.J. Macias, S. Wiesner, M. Sudol, WW and SH3 domains, two different scaffolds to recognize proline-rich ligands, FEBS Lett. 513 (2002) 30–37. doi:10.1016/S0014-5793(01)03290-2.

[60] M. Jäger, M. Dendle, A. a Fuller, J.W. Kelly, A cross-strand Trp-Trp pair stabilizes the hPin1 WW domain at the expense of function., Protein Sci. 16 (2007) 2306–2313. doi:10.1110/ps.072904107.

[61] X. Espanel, N. Navin, Y. Kato, M. Tanokura, M. Sudol, Probing WW domains to uncover and refine determinants of specificity in ligand recognition, Cytotechnology. 43 (2003) 105–111. doi:10.1023/B:CYTO.0000039913.56708.06.

[62] Y. Kato, K. Nagata, M. Takahashi, L. Lian, J.J. Herrero, M. Sudol, M. Tanokura, Common Mechanism of Ligand Recognition by Group II / III WW Domains, J. Biol. Chem. 279 (2004) 31833–31841. doi:10.1074/jbc.M404719200.

[63] M.J. Macias, V. Gervais, Structural analysis of WW domains and design of a WW prototype, (2000) 375–379.

[64] S. Sugio, A. Kashima, S. Mochizuki, M. Noda, K. Kobayashi, Crystal structure of human serum albumin at 2.5 A resolution., Protein Eng. 12 (1999) 439–46. doi:10.1093/PROTEIN/12.6.439.

[65] S. Lejon, I.-M. Frick, L. Bjorck, M. Wikstrom, S. Svensson, Crystal Structure and Biological Implications of a Bacterial Albumin Binding Module in Complex with Human Serum Albumin, J. Biol. Chem. 279 (2004) 42924–42928. doi:10.1074/jbc.M406957200.

[66] B. Elsadek, F. Kratz, Impact of albumin on drug delivery - New applications on the horizon, J. Control. Release. 157 (2012) 4–28. doi:10.1016/j.jconrel.2011.09.069.

[67] A. Skerra, Binding proteins from alternative scaffolds, 290 (2004) 3–28. doi:10.1016/j.jim.2004.04.006.

[68] A.M. Dias, Exploring New Protein-Based Scaffolds for Bioengineering Applications, Universidade Nova de Lisboa, 2015.

[69] W.C. Chan, P.D. White, Fmoc Solid Phase Peptide Synthesis, Oxford University Press, New York, 2000.

[70] T. Creighton E., Protein Structure: A Practical Approach, Second Edi, IRL Oxford Press.

[71] G.L. Ellman, Tissue Sulfhydryl Groups, (1959) 70–77.

[72] N. Greenfield, Using circular dichroism spectra to estimate protein secondary structure, Nat. Protoc. 1 (2007) 2876–2890. doi:doi: 10.1038/nprot.2006.202.

[73] S.F. Teng, K. Sproule, A. Husain, C.R. Lowe, Affinity chromatography on immobilized “biomimetic” ligands, J. Chromatogr. B Biomed. Sci. Appl. 740 (2000) 1–15. doi:10.1016/S0378-4347(99)00549-6.

[74] J.M. Haigh, A. Hussain, M.L. Mimmack, C.R. Lowe, Affinity ligands for immunoglobulins based on the multicomponent Ugi reaction, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 877 (2009) 1440–1452. doi:10.1016/j.jchromb.2009.03.010.

[75] E. Kaiser, R. Colescott, C. Bossinger, P. Cook, Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides., Anal. Biochem. 34 (1970) 595–598.

[76] A. Jonsson, J. Dogan, N. Herne, L. Abrahmsén, P.Å. Nygren, Engineering of a femtomolar affinity binding protein to human serum albumin, Protein Eng. Des. Sel. 21 (2008) 515–527. doi:10.1093/protein/gzn028.

[77] A.M.G.C. Dias, O. Iranzo, A.C.A. Roque, An in silico and chemical approach towards small protein production and application in phosphoproteomics, RSC Adv. (2015). doi:10.1039/C4RA16934D.

[78] A.M.G.C. Dias, O. Iranzo, A.C.A. Roque, Affinity adsorbents for proline-rich peptide sequences: a new role for WW domains, RSC Adv. 6 (2016) 68979–6. doi:10.1039/C6RA10900D.

[79] S.L. Pedersen, A.P. Tofteng, L. Malik, K.J. Jensen, Microwave heating in solid-phase peptide synthesis, Chem. Soc. Rev. 41 (2012) 1826–1844. doi:10.1039/C1CS15214A.

[80] B.G. De La Torre, A. Jakab, D. Andreu, Polyethyleneglycol-based resins as solid supports for the synthesis of difficult or long peptides, Int. J. Pept. Res. Ther. 13 (2007) 265–270. doi:10.1007/s10989-006-9077-5.

[81] I. Coin, M. Beyermann, M. Bienert, Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences, Nat. Protoc. 2 (2007) 3247–3256. doi:10.1038/nprot.2007.454.

[82] M. Jerabek-Willemsen, T. André, R. Wanner, H.M. Roth, S. Duhr, P. Baaske, D. Breitsprecher, MicroScale Thermophoresis: Interaction analysis and beyond, J. Mol. Struct. 1077 (2014) 101–113. doi:10.1016/j.molstruc.2014.03.009.

[83] M. Jerabek-Willemsen, C.J. Wienken, D. Braun, P. Baaske, S. Duhr, Molecular Interaction Studies Using Microscale Thermophoresis, Assay Drug Dev. Technol. 9 (2011) 342–353. doi:10.1089/adt.2011.0380.

[84] Y. Mao, L. Yu, R. Yang, L. Qu, P. de B. Harrington, A novel method for the study of molecular interaction by using microscale thermophoresis, Talanta. 132 (2015) 894–901. doi:10.1016/j.talanta.2014.09.038.

[85] S. Duhr, D. Braun, Why molecules move along a temperature gradient, Proc. Natl. Acad. Sci. 103 (2006) 19678–19682. doi:10.1073/pnas.0603873103.

[86] P. Baaske, C.J. Wienken, P. Reineck, S. Duhr, D. Braun, Optical thermophoresis for quantifying the buffer dependence of aptamer binding, Angew. Chemie - Int. Ed. 49 (2010) 2238–2241. doi:10.1002/anie.200903998.

[87] C.J. Wienken, P. Baaske, U. Rothbauer, D. Braun, S. Duhr, Protein-binding assays in biological liquids using microscale thermophoresis, Nat. Commun. 1 (2010) 100. doi:10.1038/ncomms1093.

[88] Y. Bourne, H. van Tilbeurgh, C. Cambillau, Protein-carbohydrate interactions, Curr. Opin. Struct. Biol. 3 (1993) 681–686. doi:10.1016/0959-440X(93)90050-U.

[89] T. Rogez-Florent, C. Foulon, A.S. Drucbert, N. Schifano, P. Six, S. Devassine, P. Depreux, P.M. Danzé, L. Goossens, C. Danel, J.F. Goossens, Chiral separation of new sulfonamide derivatives and evaluation of their enantioselective affinity for human carbonic anhydrase II by microscale thermophoresis and surface plasmon resonance, J. Pharm. Biomed. Anal. 137 (2017) 113–122. doi:10.1016/j.jpba.2017.01.023.

[90] A.J. Tudos, R.B.M. Schasfoort, Introduction to surface plasmon resonance, Handb. Surf. Plasmon Reson. (2008) 1–14. doi:10.1039/9781847558220.

[91] P. Pattnaik, Surface plasmon resonance: applications in understanding receptor-ligand interaction, Appl. Biochem. Biotechnol. 126 (2005) 79–92. doi:10.1385/ABAB:126:2:079.

[92] S.G. Patching, Surface plasmon resonance spectroscopy for characterisation of membrane protein–ligand interactions and its potential for drug discovery, Biochim. Biophys. Acta - Biomembr. 1838 (2014) 43–55. doi:10.1016/j.bbamem.2013.04.028.

[93] C.S.M. Fernandes, A.S. Pina, A.M.G.C. Dias, R.J.F. Branco, A.C.A. Roque, A theoretical and experimental approach toward the development of affinity adsorbents for GFP and

GFP-fusion proteins purification, J. Biotechnol. 186 (2014) 13–20. doi:10.1016/j.jbiotec.2014.06.016.

[94] A.S. Pina, A.M.G.C. Dias, F.I. Ustok, G. El Khoury, C.S.M. Fernandes, R.J.F. Branco, C.R. Lowe, A.C.A. Roque, Mild and cost-effective green fluorescent protein purification employing small synthetic ligands, J. Chromatogr. A. 1418 (2015) 83–93. doi:10.1016/j.chroma.2015.09.036.

[95] Novagen, pET System Manual, 1999. doi:10.1093/aje/kwq410.

[96] L. Cler, D. Bu, C. Lewis, D. Euhus, A comparison of five methods for extracting DNA from paucicellular clinical samples, Mol. Cell. Probes. 20 (2006) 191–196. doi:10.1016/j.mcp.2005.12.003.

[97] C. Aslanidis, P.J. de Jong, Ligation-independent cloning of PCR products (LIC-PCR)., Nucleic Acids Res. 18 (1990) 6069–74. doi:10.1093/nar/18.20.6069.

[98] G. Sezonov, D. Joseleau-Petit, R. D'Ari, *Escherichia coli* physiology in Luria-Bertani broth, J. Bacteriol. 189 (2007) 8746–8749. doi:10.1128/JB.01368-07.

[99] J. Shiloach, R. Fass, Growing *E. coli* to high cell density - A historical perspective on method development, Biotechnol. Adv. 23 (2005) 345–357. doi:10.1016/j.biotechadv.2005.04.004.

[100] B. Pope, H.M. Kent, High efficiency 5 min transformation of Escherichia coli, Nucleic Acids Res. 24 (1996) 536–537. doi:10.1093/nar/24.3.536.

[101] G.L. Rosano, E.A. Ceccarelli, Recombinant protein expression in Escherichia coli: Advances and challenges, Front. Microbiol. 5 (2014) 1–17. doi:10.3389/fmicb.2014.00172.

[102] G.J. Gopal, A. Kumar, Strategies for the production of recombinant protein in escherichia coli, Protein J. 32 (2013) 419–425. doi:10.1007/s10930-013-9502-5.

[103] P. Dedecker, F.C. De Schryver, J. Hofkens, Fluorescent proteins: Shine on, you crazy diamond, J. Am. Chem. Soc. 135 (2013) 2387–2402. doi:10.1021/ja309768d.

[104] M.A.J. Iafolla, M. Mazumder, V. Sardana, T. Velauthapillai, K. Pannu, D.R. McMillen, Dark proteins: Effect of inclusion body formation on quantification of protein expression, Proteins Struct. Funct. Genet. 72 (2008) 1233–1242. doi:10.1002/prot.22024.

[105] M.J.J. Moreau, I. Morin, P.M. Schaeffer, Quantitative determination of protein stability and ligand binding using a green fluorescent protein reporter system., Mol. Biosyst. 6 (2010) 1285–1292. doi:10.1039/c002001j.

[106] M.R. Abedi, G. Caponigro, A. Kamb, Green fluorescent protein as a scaffold for intracellular presentation of peptides, Nucleic Acids Res. 26 (1998) 623–630. doi:10.1093/nar/26.2.623.

[107] D. Drew, D.-J. Slotboom, G. Friso, T. Reda, P. Genevaux, M. Rapp, N.M. Meindl-Beinker, W. Lambert, M. Lerch, D.O. Daley, K.-J. Van Wijk, J. Hirst, E. Kunji, J.-W. De Gier, A scalable, GFP-based pipeline for membrane protein overexpression screening and purification., Protein Sci. 14 (2005) 2011–2017. doi:10.1110/ps.051466205.

[108] A. Explorer, A. Explorer, Green Fluorescent Protein: Properties, Applications, and Protocols, Second Edition, 2006.

[109] R.L. Baldwin, Temperature Dependence of the Hydrophobic Interaction in Protein Folding, Proc. Natl. Acad. Sci. 83 (1986) 8069–8072. doi:10.1073/pnas.83.21.8069.

[110] P. Barril, S. Nates, Introduction to agarose and polyacrylamide gel electrophoresis

matrices with respect to their detection sensitivities, Gel Electrophor. - Princ. Basics. (2012) 3–15. doi:10.5772/38573.

[111] H. Schägger, G. von Jagow, Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa., Anal. Biochem. 166 (1987) 368–379. doi:10.1016/0003-2697(87)90587-2.

[112] J. Nilsson, S. Stå, J. Lundeberg, M. Uhlén, P.-Å. Ke Nygren, REVIEW Affinity Fusion Strategies for Detection, Purification, and Immobilization of Recombinant Proteins The introduction of a recognition sequence for a, Protein Expr. Purif. 11 (1997) 1–16. doi:10.1006/prep.1997.0767.

[113] D.B. Smith, K.S. Johnson, Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase, Gene. 67 (1988) 31–40. doi:10.1016/0378-1119(88)90005-4.

[114] M. Hammarstrom, E.A. Woestenenk, N. Hellgren, T. Hard, H. Berglund, Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein, J. Struct. Funct. Genomics. 7 (2006) 1–14. doi:10.1007/s10969-005-9003-7.

[115] R.F. Ohana, L.P. Encell, K. Zhao, D. Simpson, M.R. Slater, M. Urh, K. V. Wood, HaloTag7: A genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification, Protein Expr. Purif. 68 (2009) 110–120. doi:10.1016/j.pep.2009.05.010.