New tools for biotechnology: Signal peptides for selected protein secretion

Steeve da Silva Lima

Mestrado em Biologia Celular e Molecular

Departamento de Biologia

2015

Orientador

Professora Doutora Paula Tamagnini, Professora Associada da Faculdade de Ciências da Universidade do Porto; Líder do grupo Bioengenharia e Microbiologia Sintética, IBMC, Universidade do Porto

Coorientador

Doutor Paulo Oliveira, Bolseiro de Pós-doutoramento, IBMC

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

“Si la vie te donne une centaine de raisons de pleurer, montre à la vie que tu as un millier de raisons de sourire.”

Mãe

“Whoever could make two ears of corn or two blades of grass upon a spot of ground where only one grew before, deserves better of Mankind, and do more essential service to his country than the whole race of politicians put together.”

The king of Broddingnag, Gullivers’s Travels, Jonathan Swift, 1727

"Etre homme, c'est précisément être responsable. C'est sentir, en posant sa pierre, que l'on contribue à bâtir le monde."

Saint-Exupéry

Aos meus Heróis

Agradecimentos

Ora aqui estou eu! Após uma translação completa em redor do Sol (mais uns pozinhos interestelares…), estou finalmente a escrever os agradecimentos deste longo, sinuoso e belo caminho que foi a dissertação de mestrado.

Queria começar por agradecer à Professora Paula Tamagnini, não só por me ter dado a possibilidade de realizar a dissertação no seio do seu laboratório, mas por me ter recebido de braços abertos no seu magnífico grupo de investigação "Bioengenharia e Microbiologia Sintética" em anos anteriores também, quando ainda me alimentava à base de papa. De salientar também os inúmeros ensinamentos sobre o fascinante mundo das Cianobactérias e da Biologia Sintética.

Ao Paulo, um obrigado do tamanho do mundo. Obrigado por me teres dado a possibilidade de, simultaneamente, errar e aprender. Garanto-te que tentei assimilar o máximo de ensinamentos que tinhas para oferecer. Obrigado por cada conselho, cada palmada nas costas, cada conversa (científica ou não), cada puxão de orelhas (os merecidos e imerecidos), pela preocupação e paciência. Considero-te um mágico do laboratório, um número 10 na gíria futebolística, com uma cartola repleta de ideias que distribuem jogo e fazem publicar.

Aos meus companheiros de bancada, Marina e Carlos, sempre pacientes e disponíveis para me auxiliar (e debater a última estreia cinematográfica), um muito obrigado. Ao (marreta do) José, meu parceiro de mestrado, dizer-lhe que tivemos acesos debates que fariam bater recordes de audiência. Um especial agradecimento à moderadora, Eunice, que também funcionou como parceira de mestrado e sempre me clareou as dúvidas sem pestanejar. À Catarina, à Rita, à Ângela, à Meri, ao Filipe, à Sara e à Gina, e restantes elementos do “grupo”, um sincero obrigado pelo fantástico ambiente laboratorial que criam, assim como pela constante disponibilidade demonstrada para me ajudar.

Às minhas amigas, Andreia e Sílvia, obrigado pelas inúmeras gargalhadas partilhadas na hora de almoço. À Gina, um obrigado pelas incontáveis e preciosas boleias.

À minha “Madrinha” (segunda mãe) e respetiva família, um obrigado por me apoiar, ajudar e integrar, especialmente neste complicado ano. À minha irmã, um obrigado por todo o apoio prestado, sem ti não teria sido possível. À Mariana, obrigado por continuares a ser o meu pilar (mesmo longe), a minha companheira nesta ininterrupta caminhada que é a vida. Tudo se torna mais fácil ao teu lado.

Por fim, queria agradecer aos meus verdadeiros heróis de todos os dias, ao meu Pai e à minha Mãe. Não há palavras para descrever o que significam para mim.

Abstract

Cyanobacteria are Gram-negative photosynthetic prokaryotes that use sunlight and CO_2 as energy and carbon sources, respectively. These features makes them attractive candidates to use in industrial/biotechnological applications, e.g. for biofuel production. Previously, the research group started to characterize the poorly known exoproteome of the cyanobacterium *Anabaena* sp. PCC 7120 and identified HesF (Heterocyst-specific attachment Factor), an exoprotein found in relatively high amounts in the medium of cells grown diazotrophically. Despite the abundance of this protein, no ordinary signal-peptide (SP) was found via bioinformatics analyses. Furthermore, HesF (Alr0267) was also found to be a type I secretion system (T1SS) substrate, and is generally recognized that T1SS-secreted proteins have a C-terminal signal peptide (SP) required for secretion, despite some studies described T1SS-secreted proteins with a N-terminal SP. In this work, we proposed to identity the amino acid sequence which determines HesF secretion, and to demonstrate its applicability for protein secretion. We focused on the N-terminus and C-terminus of the HesF sequence. Thus, the DNA sequences encoding the first HesF 80 amino acids (designated N80 *hesF*) and the last ones (designated C80 *hesF*) were used as a starting point. Subsequently, constructs containing N80 *hesF* portion or C80 *hesF* portion, or both simultaneously, were fused with *gfp* (green fluorescent protein reporter gene) and assembled with self-replicating plasmids pRL25C and pNir4. Hence, the cyanobacterium *Anabaena* sp. PCC 7120 was transformed with pRL25C or pNir4 based constructs using the triparental mating technique. Afterwards, a SDS-PAGE/Western blot approach with an antibody anti-GFP was performed to investigate the proteome and exoproteome of each transformant, in different appropriate growth conditions. Our results indicate that a putative signal peptide is probably present at the C-terminus of HesF protein, since a truncated chimeric protein comprising this portion was successfully secreted. Moreover, a specific post-translational cleavage process occur in the HesF analyzed portions. Overall, these results contributed to the promise of exploiting cyanobacteria as cell factories notably to produce and secrete selected proteins, as well as for the fundamental elucidation of the protein secretion process in cyanobacteria.

Keywords: *Anabaena*; HesF; Protein secretion; Exoproteome; Signal peptide.

Resumo

As cianobactérias são procariontes fotossintéticos Gram-negativos que usam respetivamente a luz solar e CO_2 como fontes de energia e carbono. Estas características fazem com que sejam atrativas para determinadas aplicações biotecnológicas/industriais, tais como a produção de biocombustíveis. Anteriormente, o grupo onde estou inserido iniciou a caracterização do exoproteoma de *Anabaena* sp. PCC 7120 e identificou HesF (Alr0267), uma exoproteína que aparece em quantidades elevadas no meio de cultura de células crescidas em condições fixadoras de azoto. Apesar da presença abundante no meio extracelular, HesF não possui qualquer péptido-sinal (SP) clássico, como foi demonstrado por uma análise bioinformática. Foi também demonstrado que HesF é uma exoproteína secretada pelo sistema de secreção tipo 1 (T1SS), no qual é geralmente reconhecida a presença de um SP localizado preferencialmente no C-terminal, apesar de alguns estudos descreverem também um SP N-terminal. Neste trabalho propõe-se avaliar que sequência aminoacídica determina a secreção de HesF e demonstrar a sua possível aplicabilidade para a secreção de proteínas específicas. Assim, este estudo incidiu nas porções N- e C-terminal da sequência aminoacídica de HesF. As sequências de DNA codificando os primeiros 80 aminoácidos de HesF (designado N80 *hesF)* e os últimos 80 aminoácidos (designado C80 *hesF*) foram usadas como ponto de partida. Constructos contendo N80 *hesF*, C80 *hesF* ou ambos, foram fundidos ao gene repórter *gfp* (proteína verde fluorescente), e inseridos nos plasmídeos replicativos pRL25C e pNir4. A cianobactéria *Anabaena* sp. PCC 7120 foi transformada com os constructos baseados em pRL25C e pNir4 utilizando a técnica de conjugação triparental. Subsequentemente foi feito um SDS-PAGE/Western blot usando um anticorpo anti-GFP, com o intuito de investigar os proteomas e exoproteomas dos mutantes obtidos e crescidos em condições apropriadas. Os resultados obtidos indicam que um péptido sinal parece estar presente na porção C-terminal de HesF, visto que foi secretada com sucesso uma proteína quimérica que possui esse fragmento. Adicionalmente, um processo de clivagem pós-tradução parece ocorrer nas porções de HesF analisadas. Os resultados obtidos contribuíram para a possibilidade de explorar as cianobactérias como fábricas celulares, nomeadamente para produzir e secretar proteínas de interesse, assim como para a elucidação do processo de secreção proteica em cianobactérias.

Palavras-chave: *Anabaena*; HesF; Secreção proteica; Exoproteoma; Péptido sinal.

Table of contents

List of tables and figures

List of abbreviations

ABC – ATP Binding cassette

Amp - Ampicillin

Amp^R - Resistance to ampicillin

bp - Base pairs

Cm - Chloramphenicol

Cm^R - Resistance to chloramphenicol

DNA - Deoxyribonucleic acid

dNTP - deoxyribonucleoside triphosphate

EDTA - Ethylenediaminetetraacetic acid

EPS - extracellular polymeric substances

GFP - Green fluorescent protein

HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid buffer

HesF - Heterocyst-specific attachment Factor

HgdD - heterocyst glycolipid deposition protein

HRP - Horseradishperoxidase

kbp - Kilobase pairs

Km - Kanamycin

Km^R - Resistance to Kanamycin

LB - Luria–Bertani medium

MFP - membrane fusion protein

NH_4Cl - Ammonium chloride

nirA - nitrite reductase A

Nm - Neomycin

Nm^R - Resistance to neomycin

OD - Optical density

OMF - outer membrane protein-channel or factor

ON - Overnight

PCR - Polymerase chain reaction

Pr – Promoter

r.p.m - Revolutions per min

RNA - Ribonucleic acid

SDS - sodium dodecyl sulphate

Sec - general secretory

SP - signal peptide

T1SS - type 1 secretion system

T3SS - type 3 secretion system

T4SS - type 4 secretion system

TAE - Buffer solution containing a mixture of Tris base, acetic acid and EDTA

Tat - twin-arginine translocation

Tris HCl - 2-Amino-2-(hydroxymethyl)propane-1,3-diol hydrochloride

VCBS - *Vibrio*, *Colwellia*, *Bradyrhizobium* and *Shewanella*

WT- Wild-type

Introduction

1. Cyanobacteria

Cyanobacteria are Gram-negative photosynthetic prokaryotes and believed to have been the pioneers of oxygenic photosynthesis, a distinctive process that has radically changed the atmosphere, providing the necessary conditions to expand Earth's colonization [1]. According to geological data and phylogenetic evidences, this group of bacteria, which is among the main primary producers of the aquatic environment, has its origins in the Archean geological period, about 3,000 million years ago [2-5]. Furthermore, it is generally accepted that present-day chloroplasts of algae and green plants are derived from cyanobacterial ancestors [6], evidencing its importance through the course of eukaryotic evolution [5]. Indeed, cyanobacteria are one of the largest and possibly most diverse group of bacteria. Through evolution, they have acquired a high complexity at the morphological level, existing in unicellular, colonial and filamentous forms, and at the physiological level, with several strains having the remarkable capacity to perform dinitrogen fixation [7-9]. Therefore, cyanobacteria have a cosmopolitan ecological distribution as a consequence of their minimal nutritional requirements - water, air, sunlight and a few minerals- and metabolic plasticity. As a result, they are distributed either through aquatic and terrestrial ecosystems, as well as extreme habitats (deserts, polar regions and hot springs), being some species even capable of establishing symbiotic associations with a variety of organisms (e.g. plants) [10-12]. To sum up, due to their cosmopolitan ecological distribution and their long evolutionary history, cyanobacteria represent one of the most successful group of microorganisms on Earth [13].

2. Cyanobacterial cell envelope

Cyanobacteria cell envelope, as any bacterial cell envelope, is the first protection barrier and contact with the surrounding environment [13]. Unlike cells of higher organisms, cyanobacteria are constantly exposed to a random and often aggressive environmental conditions [14]. To survive, cyanobacteria have evolved crucial biologically cell envelope-dependent strategies, such as cell–cell recognition, adhesion and communication [14, 15], exchange of biomolecules (e.g. DNA) [13], establishment of biofilms [16], symbiont infection and cell differentiation [13]. Regardless of being recognized as Gram-negatives, presenting an outer membrane and lipopolysaccharides, the cyanobacterial envelope is quite exclusive, possessing a mixture of Gram-negative and Gram-positive features [17]. Thus, distinguishing cyanobacteria cell envelope features are: the peptidoglycan layer thickness and cross-linking degree (similar to the Gram-positive bacteria) [18]; the lipopolysaccharide composition [19]; the outer membrane elements [20, 21]; the existence of extracellular matrices composed of specific glycolipids and polysaccharides coating the outer membrane [22, 23]. Additionally, cyanobacteria are able to synthesize and translocate extracellular polymeric substances (EPS) [23, 24] and proteins [12, 13, 25-34] from the cytoplasm to the extracellular space.

3. Protein secretion

3.1. Protein secretion as a vital process

Protein secretion is a housekeeping process for prokaryotic organisms to modify or otherwise influence their community and environment [26]. Indeed, by analysing the production of extracellular proteins and peptides, we can infer an organism‟s ecological strategy and function within its ecological niche [35, 36]. Hence, protein secretion is involved in crucial processes of bacterial life such as motility [37], nutrient acquisition [38-40], pathogenesis in the host organism and antibacterial activity [41-43].

It is worthwhile to notice that in Gram-positive monodermic bacteria (one membrane only), translocation across this unique membrane, in some cases, drives secretion [44]. Instead, in Gram-negative didermic bacteria (two membranes delimitating a periplasmic space), to reach the extracellular space, substrates have to be transported across the plasma membrane, the periplasm and the outer membrane. Thus, translocation across the inner membrane does not, in itself, lead to protein secretion [45]. In any case, once translocated across the outer membrane, a protein can remain anchored to the membrane, associate (non)-covalently with outer membrane components, assemble into macromolecular structures on the cell surface or be released into the surrounding environment [46] (see Figure 1).

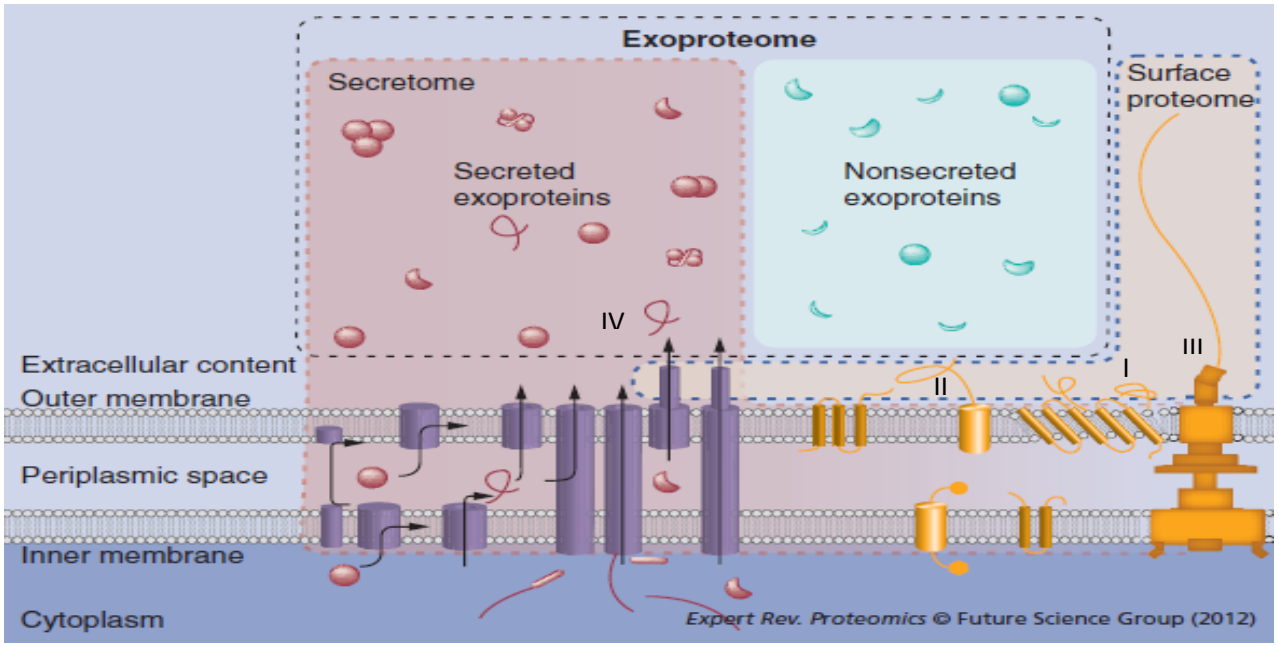

Figure 1. Exoproteome concept illustrated with a Gram-negative bacterium model. The exoproteome (black dotted line) includes the so-called 'secreted exoproteins' (secretome) and 'nonsecreted exoproteins' (among them, proteins from cellular lysis or fragments from the surface proteome due to abrasion). The secretome (red dotted line) includes the so-called 'secreted exoproteins', most membrane proteins here shown in orange, and the diverse secretion systems here drawn in purple. The surface proteome – surfaceome - (blue dotted line) includes the accessible fraction of the proteins inserted in the membrane in contact to the extracellular environment. Note that once translocated across the outer membrane, a protein can remain anchored to the membrane (I), associate (non)-covalently with outer membrane components (II), assemble into macromolecular structures on the cell surface (III) or be released into the surrounding environment (IV). From [47].

3.2. Protein secretion systems

The proper targeting of proteins to their ultimate compartments is an indispensable task of all cell types. As a result, the mechanisms of protein secretion have been extensively studied in bacteria (for reviews see [48, 49]). Thus, through the evolution, a variety of secretion systems were developed. In particular, Gram-negative bacteria possess several multicomponent secretion pathways to transport proteins to the extracellular medium across the inner membrane, the periplasm, and the outer membrane [50]. Protein translocation through the inner membrane is believed to be mainly driven by the Sec and Tat pathways (see e.g. [51] for a review of the cyanobacterial Tat system). However, to translocate proteins through the inner and outer membranes of Gram-negative bacteria, up to eight different translocation systems are described, three of which secrete proteins directly though the Gram-negative bilayer (T1SS [type 1 secretion system], T3SS and T4SS) [49, 52]. These systems are very different with respect to their functional mechanism and complexity, each one with advantages and limitations, concerning the number, size, folding state and fate of their substrates [53].

Type 1 secretion system is present in a wide range of Gram-negative bacteria, like cyanobacteria, and is remarkable for its apparent simplicity. T1SS are involved in the secretion of diverse types of biological substrates, directing for e.g. the secretion of a wide range of proteins of very different sizes

and activities [54]. T1SS consist of three proteins: an ABC transporter, a membrane fusion protein (MFP) and an outer membrane protein-channel or factor (OMF) of the TolC family (see Figure 2). Assembly of the tripartite complex is transient and induced upon binding of the substrate to the ABC protein (see Figure 2). It is generally recognized that T1SS-secreted proteins have a C-terminal signal peptide (SP) required for secretion, and that this signal interacts with the ABC protein [27, 55]. Although being absolutely required for secretion, no consensus signal peptide has been identified within the C-terminal signals [56, 57]. Moreover, some studies even described T1SS-secreted proteins with a N-terminal SP [58-60].

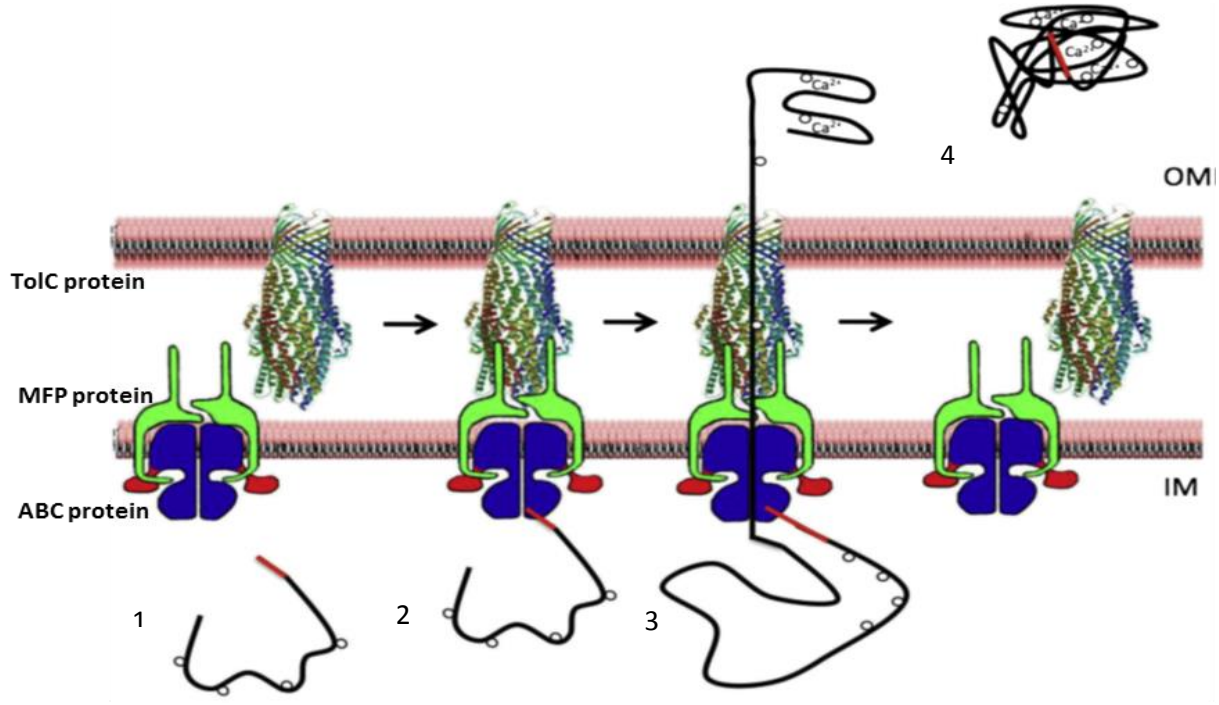

Figure 2. Current model of the coordination of a T1SS specific for proteins in time and space. The ABC transporter is shown in blue, the MFP protein in green and the outer membrane protein of the TolC family is shown attached to the OM (Outer Membrane). The substrate is shown in black in the unfolded state in the cytosol. The secretion sequence, here represented at the C-terminus, is highlighted in red. Step 1: The ABC transporter and the MFP form a static complex in the inner membrane in the absence of the substrate. Step 2: The secretion sequence of the substrate located in the extreme C-terminus interacts with the ABC transporter and/or MFP and triggers engagement of the OMF and formation of the channel-tunnel through the periplasm. Step 3: Stepwise translocation of the substrate in the unfolded state. Step 4: Refolding of the substrate in the extracellular space is helped by Ca^{2+} and resetting of the T1SS. OM – Outer membrane; IM – Inner Membrane; MFP – Membrane Fusion Protein; From [36].

3.3. Biotechnological applications of protein secretion

Prokaryotes are progressively used for the production of industrially relevant proteins, such as the therapeutic proteins insulin, interleukin-2 and human growth hormone. The facility with which bacteria accept and express new genes allows the use of these cells in a multitude of applications [61-63]. For instance, in the case of applications where purified recombinant proteins are used directly, secretion of these proteins could significantly reduce the complexity of a production process by eliminating the need for cell lysis and reducing the burden of removing host proteins. In addition,

secretion of highly expressed proteins minimizes formation of inclusion bodies, aids in folding, allows disulfide bond formation, reduces the effects of intracellular protein degradation and lessens the detrimental effects of cytotoxic proteins [64].

4. Cyanobacteria as cell factories

Cyanobacteria possess outstanding properties to be used in biotechnological applications. In particular, the ability to use sunlight and carbon dioxide as energy and carbon sources, respectively, together with faster growth rates (compared to plants), and the aptitude to be genetically engineered (compared to algae), make them attractive candidates that stand out from all other organisms so far used for these purposes. Therefore, cyanobacteria could be explored as cell factories to produce and secrete selected proteins [11, 12, 65-67]. Indeed, the use of cyanobacteria offers an alternative production approach, in which carbohydrate feedstock costs would be eliminated. Moreover, with the advent of the global warming, there is also an emergent demand to use processes that couple CO_2 capture and biotechnological applicability [66]. Ultimately, the expectation is that cyanobacteria will be developed to become the "green *E. coli*" [65].

4.1. Cyanobacterial protein secretion

Cyanobacterial protein secretion may be a valuable alternative for solving various exciting challenges, as it is exquisite explained by Oliveira et al., [12]. Thus, for instance, in bioremediation, they could secrete specific heavy metal chelators in order to detoxifying heavy-metal contaminated soils. Moreover, they could help in the cyanobacteria biomass recovery by promoting expression of specific exoproteins that will trigger cyanobacterial cells to adhere, aggregate and flocculate [12]. Furthermore, in biofuel production, the deconstruction of, e.g., polymeric lignocellulosic biomass to fermentable sugars is an important area with increasing interest. In that particular case, it was suggested that "polysaccharase" or "cellulase" secreting engineered cyanobacteria would contribute to degrading these complex polymeric substances to simpler sugars. As cyanobacteria are capable of fixing their own carbon, products of lignocellulosic deconstruction would not be used. Instead, more suitable fermenting bacteria could then convert the fully available recently released sugars into valuable biofuels [12].

As already mentioned, various studies corroborate that cyanobacteria have the capacity of secrete proteins to the extracellular space [12, 13, 26-34]. The exoproteome, i.e., the protein fraction found in the extracellular proximity of an organism (see again Figure 2), has been increasingly studied in cyanobacteria in the past few years. However, as pointed out by Oliveira et al. [12], not only is the number of cyanobacterial strains examined limited (*Nostoc*, *Anabaena, Synechocystis* and *Synechococcus*), as well as the growth conditions tested are restricted: N_2 fixing conditions, presence or absence of nitrate or ammonium chloride, and interaction with a heterotrophic bacteria [12, 13, 30-

32, 34]. It is worthwhile to notice that exoproteomes comprises actively exported proteins and those that are not actively secreted, but released by cell lysis or leakage [46] (see Figure 2). For instance, the exoproteomes of pure cultures are those proteins that can be detected after purification from the culture medium once cells have been removed [47].

5. *Anabaena* sp. PCC 7120, its exoproteome and HesF

Anabaena (also known as *Nostoc*) sp. PCC 7120 (see Figure 3) is a filamentous cyanobacterium which genome was fully sequenced and it is available (http://genome.microbedb.jp/cyanobase/Anabaena). It has become an important study model of bacterial cell differentiation, since it presents the ability, in diazotrophically conditions, to differentiate vegetative cells into heterocysts (metabolically highly active cells that perform dinitrogen fixation) [5, 68-70]. The fixation of atmospheric N_2 is accomplished through the nitrogenase-complex activity. In completely developed heterocysts, three occurrences ultimately render the cell microoxic and suitable for the function of the oxygen-sensitive nitrogenase: (i) the formation of extra envelope layers outside the Gram-negative cell wall [27], (ii) the absence of fully functional oxygen-evolving photosystem II [71] and (iii) an increase in O_2-consuming respiration [68]. The heterocysts provide the vegetative cells with their products (combined nitrogen) and obtain from the vegetative cells metabolites that serve as reductants for N_2 fixation and C skeletons for nitrogen assimilation [70].

Figure 3. Bright field optical micrographs of *Anabaena* sp. PCC 7120 wild-type (WT) grown under diazotrophic conditions. Presence of a heterocyst is highlighted by an arrow. Size bar, 50 µm.

The group were I am inserted, started recently to identify and to characterize *Anabaena* sp. PCC 7120 poorly studied exoproteome [12, 13]. From these analyses, 139 proteins belonging to 16 different functional categories have been identified. In particular, a protein designated HesF (Heterocyst-specific attachment Factor) was found accumulating in relatively high amounts in the medium of cells grown diazotrophically [13]. Functional characterization of HesF (Alr0267) led to determine that this exoprotein contributes for adherence and aggregation of this cyanobacterium filaments. Furthermore, HesF was found to be a type I secretion system substrate, since an HgdD (TolC-like) mutant failed to secrete HesF [13, 31]. Moreover, this result also indicate that even though a signal peptide (SP) could not be identified in the sequence of HesF, whatever addresses the protein

for secretion seems to be highly efficient. This observation suggests that HesF may possess a novel SP, which may be useful for biotechnological applications as well as for the fundamental elucidation of protein secretion in cyanobacteria.

6. Objectives

The goals of this work were:

i) To uncover the putative signal peptide(s) present in exoprotein HesF of *Anabaena* sp. PCC 7120 determining its secretion;

ii) To demonstrate the applicability of this possible new finding in biotechnology (secretion of proteins of interest);

Material and Methods

1. Organisms and standard growth conditions

The filamentous heterocyst-forming cyanobacterium *Anabaena* (also known as *Nostoc*) sp. PCC 7120 was grown in liquid BG11 medium [72], medium containing nitrate; or in $BG11_0$, medium devoid of combined nitrogen (in such conditions, atmospheric nitrogen is fixed); or in $BG11_0$ supplemented with 3 mM NH_4Cl and 10 mM HEPES pH 7.5 ($BG11_0$ + NH_4Cl). Cultures were grown either in Erlenmeyer flasks with orbital shaking or in glass gas washing bottles with aeration under a continuous light regime of 30–40 µmol photons m^{-2} s^{-1}, at 25 ºC.

Anabaena sp. PCC 7120 mutant strains were cultivated in the abovementioned media supplemented with 30 µg ml^{-1} neomycin. *Escherichia coli* strains (XL1-Blue, DH5α MCR, and HB101) were grown in either liquid or solid LB (Luria–Bertani) medium, supplemented with standard antibiotic concentrations, ampicillin (Amp, 100 µg mL^{-1}), chloramphenicol (Cm, 50 µg mL^{-1}) or kanamycin (Km, 50 µg mL^{-1}).

Anabaena sp. PCC 7120 as well as its derived mutant strains were also grown in solid BG11 (containing 1.5% Difco® Agar Noble), in the same conditions described above and supplemented with 30 µg mL^{-1} neomycin.

2. *Chlorophyll a* concentration determination

Anabaena sp. PCC 7120 culture growth was monitored by measuring the *Chlorophyll a* content. For that purpose, 1 mL of of culture was centrifuged at 14100 *g* for 1 minute (min). Nine hundred µL of the respective supernatant was discarded, being the pellet resuspended in an equal volume of 100% methanol. The suspension was then vigorously vortexed and left at 4 ºC overnight. The following day, samples were centrifuged at 14100 *g* for 1 minute (min), and the supernatant's absorbance measured at 663 nm in a Shimadzu® UVmini-1240 spectrophotometer. An extinction coefficient of 12,7 was used to calculate *Chlorophyll a* concentration, as deduced in [73].

3. *Anabaena* sp. PCC 7120 DNA extraction

To obtain *Anabaena* sp. PCC 7120 genomic DNA, it was followed a previously described method [74]. Thus, 50 mL of culture were harvested by centrifugation (4360 *g*) for 10 min. The resulted pellet was transferred to 1.5 mL Eppendorf tubes and an additional centrifugation step was performed to remove excess of growth medium. Then, the pellet was suspended in 300 µL of Resuspension Buffer (10 mM EDTA; 50 mM Tris HCl, pH 8.0) and mixed with 350 mg of 0.6 mm diameter glass beads (Sigma®), 25 µL of 10% SDS (sodium dodecyl sulphate), and 500 µL of 1:1 phenol-chloroform (v/v)

solution. The mixture was thoroughly vortexed for 30 seconds (sec) and incubated on ice for an equal period of time. These 1 min cycles were repeated in a total of 4 times. In the end the mixture was centrifuged at 13100 *g* for 10 min at 4 ºC. The resulting upper phase (aqueous phase) was then transferred to a new 1.5 mL Eppendorf tube and mixed with 500 µL chloroform. The sample was again centrifuged at 13100 *g* for 10 min at 4 ºC and the DNA present in the aqueous phase was precipitated with 2.5 volumes of ice cold 100% ethanol, with the help of one tenth of its volume of 3 M sodium acetate (pH 5.2). The mixture was incubated at -20 ºC for 1 hour (h), followed by a centrifugation step at 14000 *g* at 4 ºC for 30 min to pellet the genomic DNA. The supernatant was discarded and the pellet was washed with ice cold 70% ethanol. A centrifugation at 14000 *g* at 4 ºC for 10 minutes was performed, and the ethanol was again carefully removed, letting the pellet dry for 10 min. Finally, the dried pellet was dissolved in 50 µL of sterile deionized water. Nucleic acids concentration was determined using a Nanodrop ND-1000 (Thermo Scientific®), and standard electrophoresis (see "DNA electrophoresis, purification and quantification") was performed to evaluate DNA extraction quality.

4. DNA agarose gel electrophoresis, purification and quantification

Nucleic acids were separated by electrophoresis on 1% (w/v) agarose (Pronadisa) gels containing 1x TAE buffer [75] and stained with ethidium bromide. The GeneRuler™ DNA Ladder Mix (ThermoScientific®) was used as molecular weight marker.

DNA purification from gel or from enzymatic reactions was carried out using the NZYGelpure kit (Nzytech, Lda.), following the manufacturer's instructions. DNA concentration was determined using a Nanodrop ND-1000 (Nanodrop Technologies, Inc.).

5. PCR – Polymerase Chain Reaction

Two sets of PCR reactions were performed, using two different polymerases: a Thermo Scientific Phusion High-Fidelity DNA Polymerase, for preparative purposes (DNA constructions/Cloning steps); and a Promega GoTaq® DNA polymerase, for routinely laboratorial experiments. Both polymerases were used following the manufacturer's instructions. Thus, when using Phusion High-Fidelity DNA Polymerase, 20 µL PCR reactions contained: 1x HF (ThermoScientific) buffer, dNTP (deoxyribonucleoside triphosphate) mixture at a final concentration of 0.2 mM, 0.02 U μL^{-1} of Phusion DNA Polymerase, and 0.5 µM oligonucleotides (Table 1). PCR reactions were carried out in a thermocycler (Bio-Rad) using the following profile: 30 seconds denaturation step at 98 ºC; followed by 35 cycles of 10 seconds at 98 ºC, 30 sec of the appropriate annealing temperature and then 72 ºC for 30 sec for every kbp of the target DNA; at the end of the cycles another 5 min of extension step at 72 ºC. Alternatively, when using GoTaq® DNA polymerase, 20 µL PCR reactions contained: 1x GoTaq buffer (Promega), 2 mM magnesium chloride, 1 µM of oligonucleotides (see Table 1), 0.2 mM dNTP

mixture, and 1 U of GoTaq DNA polymerase (Promega). PCR reactions were performed following this profile: 2 min denaturation step at 95 ºC; followed by 35 cycles of 30 sec at 95 ºC, 30 sec of annealing temperature at 55 ºC and 72 ºC for extension (1 min for every kbp of the target DNA used); and ended by an extension step at 72 ºC for 7 min.

6. DNA Digestion and Ligation

All DNA digestions performed in this work were made accordingly to the manufacturer's instructions (ThermoScientific). In case of double restriction reactions, optimal reaction conditions were followed using the Double Digest tool (http://www.thermoscientificbio.com/webtools/doubledigest/).

All DNA ligation reactions (20 µL) were executed according to the instructions of the manufacturer (Thermo Scientific), i.e., using 1 U T4 DNA ligase (Thermo Scientific). Typically in a ligation reaction, 100 ng of a 3 kbp vector was used, while the amount of insert used was 3 times the molar concentration of the vector. Ligation reactions were performed in a thermocycler (Bio-Rad) over-night at 22 ºC, and inactivated at 65 ºC for 10 min.

7. Preparing *Escherichia coli* XL1-Blue/DH5α MCR competent cells

To prepare competent cells, a variation of the Hanahan protocol [76] was used. An *E. coli* XL1-Blue/ DH5α MCR previously prepared competent cells tube containing 110 µL was used as seed stock. That stock was grown over-night (ON) in 250 mL Erlenmeyer flasks containing 50 mL of LB medium at 25 ºC. Cells were then centrifuged at 1570 *g* at 4 ºC for 10 min. The supernatant was discarded and the pellet was suspended in 32 mL of ice cold CCMB80 buffer (10 mM KOAc, 80 mM $CaCl_2.2H_2O$, 20 mM $MnCl_2.4H_2O$, 10 mM $MgCl_2.6H_2O$, 10% glycerol; adjusted to pH 6.4) and incubated on ice for 20 min. Cells were centrifuged in the same conditions as described and again the supernatant was discarded. The pellet was suspended in 2 mL of ice cold CCMB80 buffer. The optical density measured at 600 nm (OD_{600}) of the cells' suspension was measured and adjusted to a final OD_{600} between 5 and 6. Competent cells were then transferred to 1,5 mL Eppendorf tubes as 110 µL aliquots and stored at -80 ºC.

8. Competent cells transformation and Plasmid Preparation

To transform *E. coli* XL1-Blue/ DH5α MCR competent cells, 7 µL of a given ligation/1 µL of a plasmid preparation was added to a 110 µL aliquot (previously thawed on ice for 15 min) and the mixture was incubated there for 30 min. Next, this mixture was heat shocked at 42 ºC for 45 sec in a water-bath, and quickly placed back on ice for 2 min. Previously warm LB medium was added to the mixture to a final volume of 1 mL and cells were left to recover in an orbital shaker at 37 ºC for 60 min. Finally, cells were plated onto LB-agar plates supplemented with the appropriate antibiotic(s) to select

for transformants, being the plates incubated ON at 37 ºC. To recover the plasmid DNA, single isolated colonies were inoculated in 5 mL of LB medium supplemented with the same antibiotic(s) and grown ON at 37 ºC with vigorous shaking (180 r.p.m.). Plasmid DNA was obtained with a GenElute™ plasmid miniprep kit (Sigma®) following the manufacturer's instructions. Lastly, a standard electrophoresis was performed to evaluate plasmid preparation quality.

9. DNA Sequencing

All DNA constructs (confirmed by plasmid restriction with appropriate enzymes) were sent to STAB VIDA Company for sequencing by the Sanger method.

10. Constructs (pRL25C and pNir4 based)

All plasmids used in this study are listed in Table 2. For a schematic representation of the generated constructions, refer to Figure 4. In order to keep expression of the generated chimeric proteins under the regulation of the *hesF* promoter, the respective DNA constructs were cloned in pRL25C [77], a self-replicating plasmid. With this purpose, the promoter region of *hesF* (Pr *hesF*) together with the designated N80 *hesF*'s portion (nucleotides corresponding to the first 80 amino acids of HesF) was fused with *gfp* (Fig. 4a). Alternatively, the promoter region of *hesF* only was fused with *gfp* and with *hesF*'s C80 portion (nucleotides corresponding to the last 80 amino acids of HesF) (Fig. 4b). In a third combination, the promoter region of *hesF* together with *hesF*'s N80 portion was fused with *gfp* and with *hesF*'s C 80 (Fig. 4c). Alternatively, in order to bypass *hesF*'s natural regulation, expression of the generated chimeric proteins was controlled by the *nirA* promoter, which is present in the self-replicating plasmid pNir4 [13, 78]. The respective DNA constructs were cloned downstream of the *nirA* promoter. With this purpose, and similarly to what was described for pRL25C based constructions, three constructions were prepared: *gfp* was fused with *hesF*'s N80 portion (Fig. 4d); *gfp* was cloned with *hesF*'s C80 portion (Fig. 4e); and *gfp* was fused with both *hesF*'s N80 and C80 portions (Fig. 4f). For both sets of DNA constructions, the following strategy was used. The coding sequence of GFP was amplified by PCR using appropriate oligonucleotides listed in Table 1 and an in-house generated pRL25C plasmid containing the fragment Pr *hesF*::*gfp* as template [13]. The Pr

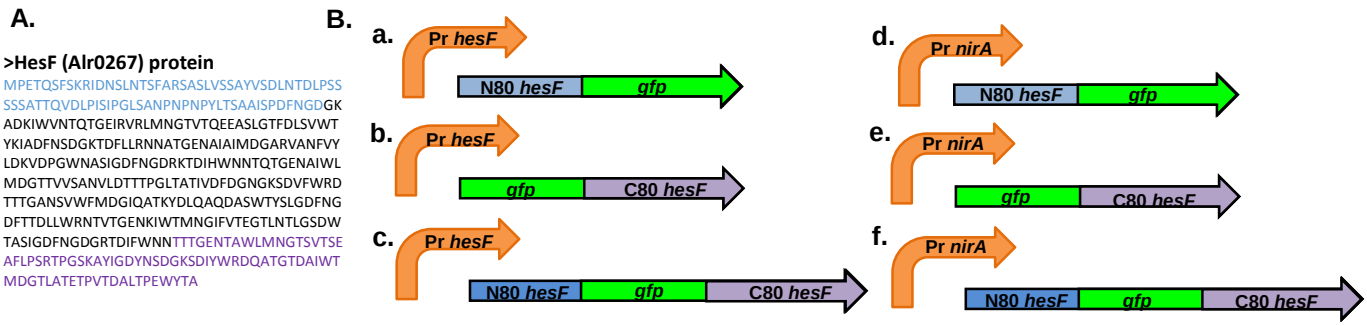

Figure 4. A. HesF (Alr0267) deduced amino acid sequence. B. Schematic representation of the constructs used to investigate a novel SP in the exoprotein HesF of *Anabaena* sp. PCC 7120, and where its location (N-terminal or C-terminal). Constructs **a.** to **c.** was assembled in pRL25C plasmid, and constructs **d.** to **f.** was built in pNir4 plasmid: **a.** Pr *hesF*::N80*hesF*::*gfp*, **b.** Pr *hesF*:: *gfp*::C80*hesF*, **c.** Pr *hesF*::N80*hesF*::*gfp*::C80*hesF*, **d.** Pr *nirA*::N80*hesF*:: *gfp*, **e.** Pr *nirA*:: *gfp*::C80he*sF* and **f.** Pr *nirA*::N80*hesF*:: *gfp*::C80*hesF*.

hesF, the Pr *hesF* with N80, N80 only, and C80 fragments, were obtained by PCR using the *Anabaena* sp. PCC 7120 genomic DNA as template and the appropriate oligonucleotides (Table 1). Subsequently, each fragment was digested with suitable enzymes (restriction sites are included in the oligonucleotides' sequence and listed in Table 1), followed by a purifying step. Assembly of each construct was carried out first in pBluescript SK+ plasmid (Agilent) and transformed in *Escherichia coli* XL1-Blue cells. Once the identities of the different constructs were determined by sequencing, constructs were excised from the pBluescript SK+ and cloned in the respective final plasmid (pRL25C or pNir4) and transformed into *E. coli* XL1-Blue. The pRL25C based constructs were ligated either in direct or reverse direction relative to the antibiotic resistance cassette. Afterward, the confirmed plasmid preparations were used to transform *E. coli* DH5α MCR cargo strain, already carrying the helper plasmid pRL623 [79]. Finally, the transformants containing pRL25C/pNir4 based constructs were ready to conjugate *Anabaena* sp. PCC 7120 using the triparental mating technique (see below).

Table 1. Oligonucleotides used to investigate HesF SP localization

Oligonucleotide Pair	Purpose of the Oligonucleotide Pair	Sequence 5' → 3'	Ta (ºC)	Target size (bp)	Restriction Site	Reference
alr0267_GFP_F	**N80 (including Pr *hesF*)**	aaaCTCGAGGGATCCccgatattctgcttgcataacc	67	1043	**XhoI/BamHI**	[13]
alr0267_N80_R		aaaGAATTCatcgccgttgaaatcaggtgaaa			**EcoRI**	This work
alr0267_GFP_F	**Pr *hesF***	aaaCTCGAGGGATCCccgatattctgcttgcataacc	64	833	**XhoI/BamHI**	[13]
alr0267_GFP_R		aaaGAATTCgctttgcgtttctggcat			**EcoRI**	
alr0267_C80_F	**C80**	aaaCTGCAGacaacaggcgagaacactgctt	64	353	**PstI**	This work
alr0267_Comp_R		aaaGGATCCtaaaacctcgcctcttgtgg			**BamHI**	[13]
alr0267_nirN_F	**N80nir**	aaaGAATTCatgccagaaacgcaaagct	65,2	258	**EcoRI**	[13]
alr0267_nirN80_R		aaaCTGCAGatcgccgttgaaatcaggtgaaa			**PstI**	This work
alr0267_nirC80_F	**C80nir**	aaaCCCGGGacaacaggcgagaacactgctt	64	347	**XmaI**	This work
alr0267_Comp_R		aaaGGATCCtaaaacctcgcctcttgtgg			**BamHI**	[13]
GFP_F	***gfp***	aaaGAATTCAtgagtaaaggagaaga	50	729	**EcoRI**	[13]
GFP_R		aaaGGATCCttatttgtatagttcatcc			**BamHI**	
GFP_F	***gfp*$_{noStop}$**	aaaGAATTCAtgagtaaaggagaaga	52	729	**EcoRI**	[13]
GFP_noStop_R		aaaCTGCAGtttgtatagttcatccatg			**PstI**	
GFP_nir_F	***gfp*nir**	aaaCTGCAGAtgagtaaaggagaaga	50	729	**PstI**	[13]
GFP_nir_R		aaaGGATCCttatttgtatagttcatcc			**BamHI**	
GFP_F	***gfp*nir$_{noStop1}$**	aaaGAATTCAtgagtaaaggagaaga	50	729	**EcoRI**	[13]
GFP_nir_noSTOP_R		aaaCCCGGGtttgtatagttcatccatg			**XmaI**	
GFP_nir_F	***gfp*nir$_{noStop2}$**	aaaCTGCAGAtgagtaaaggagaaga	50	729	**PstI**	[13]
GFP_nir_noSTOP_R		aaaCCCGGGtttgtatagttcatccatg			**XmaI**	
GFP_NR	**PCR confirmation and construct sequencing**	ctccactgacagaaaatttg	-	-	**-**	This work

Table 2. Plasmids used in this study.

Plasmid	Description	Reference	Resistance
pBluescript SK+	Regular cloning plasmid	Agilent	Amp^R
pRI25C	Mobilizable self-replicating plasmid in *Anabaena* sp. PCC 7120	[78]	Km^R and Nm^R
pNir4	Mobilizable and self-replicating plasmid in *Anabaena* sp. PCC 7120. Contains the *nirA* promoter.	[79]	Km^R and Nm^R
pRL443	Conjugal plasmid	[80]	Amp^R
pRL623	Helper plasmid carrying methylases that will protect DNA from all identified restriction		Cm^R
pRL25C::Pr*hesF*::*gfp*	Source of *gfp*	[13]	Km^R and Nm^R

11. *Anabaena* sp. PCC 7120 Transformation (Triparental mating)

Anabaena sp. PCC 7120 is not naturally transformable. Still, conjugation with *E. coli* can be used to transform this filamentous cyanobacterial strain. Briefly, the technique is based on the ability of an *E. coli* HB101 conjugal strain carrying a conjugal plasmid (pRL443), to transfer DNA via conjugation, with the aid of a second cargo DH5α MCR strain, which possesses a cargo plasmid (in this case, the generated pRL25C- and pNir4-based plasmids containing the constructions of interest) and a helper plasmid (pRL623) [11]. When *Anabaena* sp. PCC 7120 is mixed with such *E. coli* strains, the cyanobacteria will be able to receive a cargo plasmid. Thus, plasmids harbouring constructs of interest can be transferred into *Anabaena* sp. PCC 7120 by conjugation, using this well-established triparental mating technique [80]. To do so the *E. coli* strain harbouring both the plasmid of interest and the helper plasmid pRL623, and the *E. coli* HB101 harboring the conjugative plasmid pRL443, were grown overnight in LB medium supplemented with the appropriate antibiotics. The following day, cultures of both strains were inoculated in 25 mL of fresh LB medium supplemented with the appropriate antibiotics to an initial OD_{600} of 0.08. The cultures were then grown in an orbital shaker at 37 ºC for 5 h, at 120 r.p.m.. At the end of this period, cultures were centrifuged at 1570 *g* for 10 min and pellets washed twice with LB medium. After the last wash, *E. coli* cells were centrifuged and the pellet of the conjugal cells was suspended in 1 mL of LB medium. The conjugal cell suspension was then used to suspend the cargo cells, resulting in a cell suspension containing a mixture of *E. coli* strains. One additional washing step was carried out and the pellet containing both conjugal and cargo cells was finally suspended in 200 µL of LB medium. The cell suspension was left resting quietly in a tube rack for approximately 1 h, promoting appropriate conditions for the first conjugation event between the two *E. coli* strains, while the cyanobacterial strain was being prepared for the subsequent steps of the mating. A volume of 50 mL of the cyanobacterial strain were harvested in mid exponential phase by centrifugation for 10 min at 4360 *g*. Then the cyanobacterial pellet was washed and suspended in the

least volume possible of BG11 medium, and transferred to 2 mL tubes. A centrifugation step at 10000 g during 1 min was performed, and the resulting pellet suspended into 500 µL BG11 [11]. An additional step to disrupt the filaments was carried out, which consisted in passing the suspended pellet through a non-pyrogenic needle (0,5 x 16 mm diameter; Neolus®), in order to improve the chance of obtaining transformants. Then, cyanobacterial cell suspension was mixed with the *E. coli* mix, and mating was let to occur for 1.5 hour. After that the mixture, now containing all three strains, was plated on BG11 plates supplemented with 5% LB medium and plates were placed under standard cyanobacterial growth conditions. Approximately 30 h later plates were washed with 1.5 mL of BG11. The cells washed out from the plate were then transferred to a 1.5 mL Eppendorf and submitted to a centrifugation at 10000 g during 1 min, being the resulting pellet suspended into 500 µL BG11 and plated onto fresh BG11 plates supplemented with suitable antibiotic(s).

12. Microscopy

Cells of *Anabaena* sp. PCC 7120 wild-type and transformants were visualized under an Olympus CX31 bright field microscope directly from the culture medium, being the micrographs acquired with an Olympus DP25 camera and the Cell B software. In addition, using a Leica SP2 AOBS SE Laser Scanning Confocal microscope, transformants of *Anabaena* sp. PCC 7120 harboring pRL25C or pNir4 constructs, expressing a chimeric protein containing GFP, were studied. For that purpose, cells were loaded on a 1% low-melting-point agarose bed, respectively dissolved in BG11 or $BG11_0$, and covered with a clover slip. The GFP emission, collected between 500 and 540 nm, was observed when cells were exposed to an Ar Laser Beam of 488 nm, while cyanobacterial autofluorescence, acquired between 640 and 700 nm, was visualized after excitation at 633 nm, using a HeNe Laser. Wild-type cells were used to define the basal autofluorescence signal in the GFP channel, and the same acquisition settings were used throughout the various experiments.

13. Protein extraction and quantification

In order to obtain total protein cyanobacterial extracts, an aliquot of 2 mL of a pre-culture was diluted with fresh growth medium to a final culture volume of 200 mL (1:100 dilution) in 500 mL glass gas washing bottles. Cultures were grown with aeration with air (flux of approximately 1 L min^{-1}). Briefly, 50 mL of a respective culture was centrifuged for 10 min at 4360 *g*. Then, the resulted pellet was transferred to 1.5 mL Eppendorf tubes and submitted to an additional centrifugation step (10000 *g*, 90 sec). The final pellet was suspended with 750 µL of Protein Extraction Buffer [10 mM HEPES, 0.5% triton X-100, 10 mM EDTA, 2 mM DTT, pH 8.0, supplemented with protease inhibitor cocktail (Roche Diagnostics GmbH)]. After this step, cells were lysed by sonication (Branson Sonifier 250) with cycles of 20 sec sonication followed by incubation on ice for 20 sec (a total of 5 cycles were performed for each sample). Each cycle had a duty cycle of 50% and an output control of ± 4. Cell

debris and unbroken cells were separated from the extracts by a 10 min centrifugation at 10000 *g* at 4 ºC. Proteome extracts were stored at -20 ºC. Determination of the protein content was performed using the BCA™ Protein Assay (Pierce), with bovine serum albumin (BSA) as standard.

14. Exoproteome extraction and concentration

The isolation of the exoproteome of *Anabaena* sp. PCC 7120 was performed as described previously [13]. Briefly, 200 mL of cultures grown in BG11 or $BG11_0$ were collected at different time points (transformants with pRL25C based constructs: 48 h and 72 h after inducing nitrogen-fixing conditions; transformants with pNir4 based constructs: 6 days after grown the cells in BG11) by centrifugation (4360 *g*), to separate cells from the respective medium, at different time points of growth. Next, the resulted supernatant was filtered through 0.2 µm pore size filter to further remove any cell contaminant and concentrated by centrifugation at 4360 *g* with Amicon Ultra-15 centrifugal filter units (Merck Millipore) with a nominal molecular weight limit of 3 kDa. Concentrated exoproteome samples were then saved at −20 ºC until further analysis. Exoproteome samples were separated by electrophoresis on 12% (w/v) SDS-polyacrylamide gels (see below).

15. SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

Proteome and exoproteome samples were separated by electrophoresis on 12% (w/v) SDS-polyacrylamide gels. Electrophoresis were performed on a vertical mini Protean system (Bio-Rad®) following the method of Laemmli [81]. Samples were mixed with Laemmli buffer in a 4:1 proportion and heated at 95 ºC for 6 min. After being cooled to room temperature, samples were loaded on the gel, as well as the NZYColour Protein Marker II (Nzytech®). Protein separation was performed with a constant electric current of 20 mA per gel. After protein separation, the gel was incubated with fixing solution [40% (v/v) methanol and 7% (v/v) acetic acid] for at least one hour. Proteins were visualized using colloidal Coomassie Brilliant Blue G (Sigma). Staining was carried out in an orbital shaker (60 r.p.m.) overnight, after which the gel was destained with deionized water. After the gel was thoroughly destained, it was scanned in a Molecular Imager GS800 calibrated densitometer (Bio-Rad®).

16. Western Blot

After electrophoresis, gel separated proteome and exoproteome samples that did not proceed for Coomassie blue staining were transferred onto a nitrocellulose membrane using a semi-dry system. Protein transfer was performed using transfer solution [0,581% (w/v) of Tris, 0,293% (w/v) of Glicine, 0,375 % (v/v) SDS 100%, 20% (v/v) methanol 100%, pH 8.3] in a trans-blot® semi-dry electrophoretic

transfer cell. An electric current of 38 mA per gel to be transferred (0,8 mA/cm^2) was applied for 1 h and 20 min. Then, the nitrocellulose membrane containing the transferred proteins was blocked for 3 h in 35 mL of blocking solution [5% (w/v) powdered Nestlé milk, 0.05% (v/v) Tween, TBS 1X], with agitation. Later, the nitrocellulose membrane was incubated in 12 mL fresh blocking solution containing the monoclonal mouse anti-GFP antibody (Roche®) (1:4000), at 4 ºC, with agitation, ON. The following day, a sequence of three washing steps (one of 15 min and two of 10 min) were executed with TTBS washing solution [TBS 1X, 0.05% (v/v) Tween] with agitation (80 r.p.m.), next being the nitrocellulose membrane incubated, at room temperature with agitation, with 12 mL of a blocking solution containing the secondary antibody (goat Anti-mouse IgG; 1:4000) conjugated with Horseradishperoxidase (HRP) enzyme (Santa Cruz Biotechnology®). Then, an equal abovementioned procedure to wash the membrane was performed with TTBS Washing Solution, followed of two washing steps with TBS (1 X), 15 min each, with agitation. Finally, detection was performed on a ChemiDoc MP System (Life Science Research, Bio-Rad), using the Amersham ECL Western Blotting Detection Reagent (GE Healthcare Life Sciences) according to the manufacturer's instructions.

Results and discussion

This work was developed subsequently to the recent effort performed by the group were I am inserted to study *Anabaena* sp. PCC 7120 exoproteome [12, 13]. Indeed, due to the growing consciousness of the dynamic role of exoproteins on bacterial housekeeping process as homeostasis and development [16, 82], the curiosity on profiling the largely unknown exoproteome of cyanobacteria has been rising. Thus, our group started to address important aspects of the cyanobacterial exoproteome, namely by identifying and characterizing exoproteins of morphologically and metabolically different cyanobacterial strains. One of them, *Anabaena* sp. PCC 7120, had its poorly studied exoproteome characterized [12, 13]. From these analyses, 139 proteins belonging to 16 different functional categories have been identified. In particular, a protein designated HesF (Heterocyst-specific attachment Factor) was found accumulating in relatively high amounts in the medium of cells grown diazotrophically [13]. Moreover, even though a signal peptide (SP) could not be identified in the sequence of HesF, whatever addresses the protein for secretion seems to be highly efficient. This observation suggested that HesF may possess a unknown SP. Therefore, this study proposed to uncover the putative signal peptide present in exoprotein HesF of *Anabaena* sp. PCC 7120, as well as to demonstrate the applicability of this possible new finding in biotechnological applications (secretion of proteins of interest).

1. Investigation of HesF's Signal peptide localization

To determine the SP that promotes secretion of HesF to the extracellular space, both size and location of that determinant was studied. During the bioinformatics analysis of HesF's deduced amino acid sequence (408 amino acids long) [13], a protein domain (termed VCBS[1]) was found. This domain has been suggested to have a role in HesF's function promoting binding to extracellular carbohydrates [13]. The VCBS domain is present in multiple copies (6) in HesF, covering amino acids positions 75 to 376. Thus, we hypothesized that if the novel SP is present at the N-terminus, possibly it was located before the VCBS functional domain. Therefore, in this study, we focused on the first 80 amino acids (designated N80 portion) and the last ones (designated C80 portion) of the HesF sequence (available in CyanoBase). Thus, using DNA sequences encoding the first HesF 80 amino acids (N80 *hesF*) and the last ones (C80 *hesF*) as a starting point, constructs were prepared (see Figure 4), and assembled with the self-replicating plasmid pRL25C. The N80 *hesF* portion, C80 *hesF* portion, or both, were fused with reporter GFP (Green Fluorescent Protein) encoding gene. The

1. VCBS – the VCBS domain is composed of approximately 100 amino acid residues, and is found in multiple copies (up to 35) in long proteins from several species of *Vibrio*, *Colwellia*, *Bradyrhizobium* and *Shewanella* and hence the name VCBS (TIGRFAM database – TIGR01965, The J. Craig Venter Institute).

chimeric proteins were expressed under the regulation of hesF's native promoter (Pr *hesF*). Moreover, it is known that when *Anabaena* sp. PCC 7120 is cultivated in the presence of a source of combined nitrogen such as nitrate or ammonium, long filaments containing photosynthetic vegetative cells are formed. In the absence of combined nitrogen, the cyanobacterium differentiates heterocysts, which are nitrogen-fixing cells that provide fixed nitrogen to rest of the filament [70]. In Oliveira et al., [13], it was reported that HesF was highly upregulated upon transition from non-nitrogen-fixing to nitrogen-fixing conditions, and the highest transcript levels were detected towards the end of the heterocyst differentiation process. Thus, our study was performed following the expression of HesF in "native conditions", namely nitrogen-fixing conditions. However, it should be noted that in this condition only 5 to 10% of a culture would be heterocysts [70]. So, the expression of our constructs in "native conditions" could cause a weak chimeric protein expression, due to the presence of few heterocysts in the culture. In order to anticipate this fact and guarantee a high expression of our chimeric proteins, one additional promoter was used: *nirA*, as shown in Figure 4. The N80 *hesF* portion, C80 *hesF* portion, and both N80 *hesF* and C80 *hesF* fused with *gfp* gene were assembled under the control of the nitrate inducible promoter *nirA* present in self-replicating plasmid pNir4 [13, 78], enabling the bypass of the native expression mechanism and the expression of the chimeric proteins in non-nitrogen-fixing conditions [83].

1.1. Evaluation of the constructs' correct assembly

In order to verify the accurate assembly of constructs present in Figure 4, the digestion with appropriate enzymes which allow the excision of all the respective total insert was done and is shown in Figure 5. Regarding the pRL25C based constructs, a digestion with BamHI, enable the excision of these total inserts, as it can be seen in Figure 5. Therefore, the designated N80 construct, which possess the *hesF* promoter (821 bp) linked to N80 *hesF* (240 bp) and *gfp* (729 bp), has a expected size of 1790 bp. Moreover, the same evaluation was performed for the designated C80 construct, which possess a full expected size of 1888 bp (C80 *hesF* - 338 bp), and for the designated NC80 construct, with an excepted size of 2128 bp. Concerning the pNir4 based constructs, it is worthwhile to notice that the promoter *nirA* is contained in the backbone of this replicative plasmid, being the remain of the construction cloned dowstream. Consequently, the digestion of these constructs with EcoRI and BamHI only enable the separation of the total insert representing the N80 *hesF* or/and C80 *hesF* portions linked to *gfp*. Hence, the expected size is 969 bp for the N80nir construct, 1067 bp regarding the C80nir construct, and 1307 bp for the construct possessing both (Figure 5). All the constructs were also confirmed by DNA sequencing.

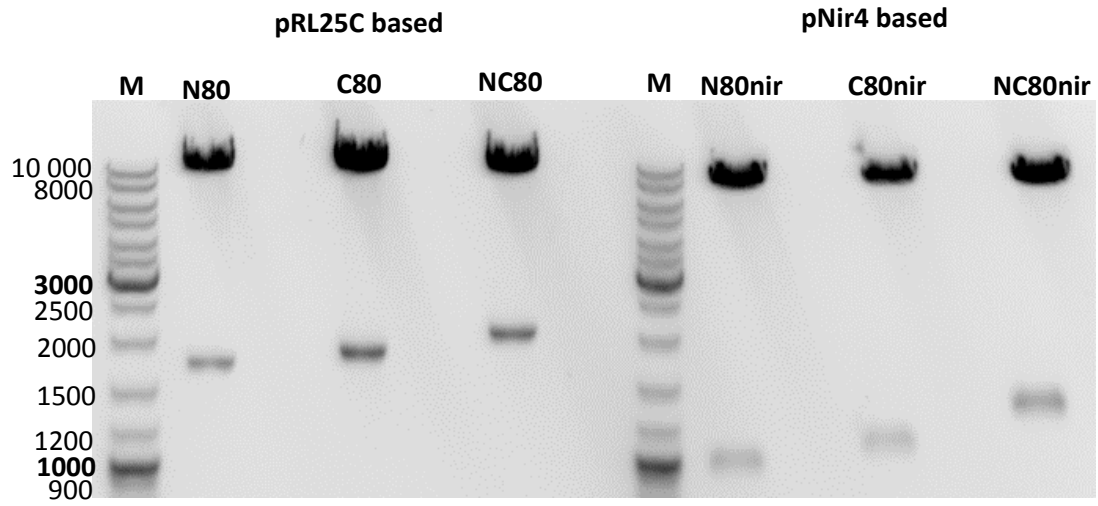

Figure 5. Ethidium bromide stained agarose gel depicting the confirmative digestion of the constructs used in this work. For the pRL25C based constructs, a digestion using BamHI was performed, while for the pNir4 based constructs, a double digestion with EcoRI and BamH was accomplished, in order to excise the respective total insert. Constructs assembled in pRL25C plasmid (10 kbp) are shown on the left of the panel, while pNir4 plasmid (8 kbp) based constructs are shown on the right side of the panel. The pRL5C total inserts are: **N80** – Pr *hesF*::N80*hesF*::*gfp* (1790 bp), **C80 –** Pr *hesF*:: *gfp*::C80*hesF* (1888 bp)**, NC80 –** Pr *hesF*::N80*hesF*::*gfp*::C80*hesF* (2128 bp). The pNir4 total inserts are: **N80nir -** N80*hesF*:: *gfp* (969 bp), **C80nir -** *gfp*::C80h*esF* (1067 bp) and **NC80** - N80*hesF*:: *gfp*::C80*hesF* (1307 bp). It is worthwhile to notice that the promoter *nirA* is contained in the backbone of pNir4 replicative plasmid. M, GeneRuler™ Ladder Mix (Thermo Scientific). The sizes in base-pairs (bp) of some of the ladders's DNA fragments are shown for reference.

1.2. Molecular analysis of the transformants

Afterwards, *Anabaena* sp. PCC 7120 was transformed with pRL25C or pNir4 based constructs using the triparental mating technique. In order to verify the success of this task, PCRs using as template the DNA previously extracted from each transformant with pRL25C or pNir4 based constructs were performed and are displayed in Figure 6 and 7, respectively.

Regarding the PCR analyses depicting the *Anabaena* sp. PCC 7120 transformation with pRL25C based constructs (Figure 6), two combinations of oligonucleotides were used to confirm the accomplishment of the respective conjugation and to distinguish each one, namely: 1 – alr0267_GFP_F + GFP_NR, and 2 – GFP_F + alr0267_CompR. Thus, using the first pair of oligonucleotides, a band with approximately 1161 bp concerning the N80 and NC80 transformants was reported as expected, while a band with 921 bp was obtained for the C80 transformant. Moreover, using the second pair of oligonucleotides, no band was visualized for the N80 transformant as anticipated, whereas a band with approximately 1067 bp was obtained regarding the C80 and NC80 transformants. Indeed, this experiment allowed the confirmation of *Anabaena* sp. PCC 7120 transformed with pRL25C based constructs.

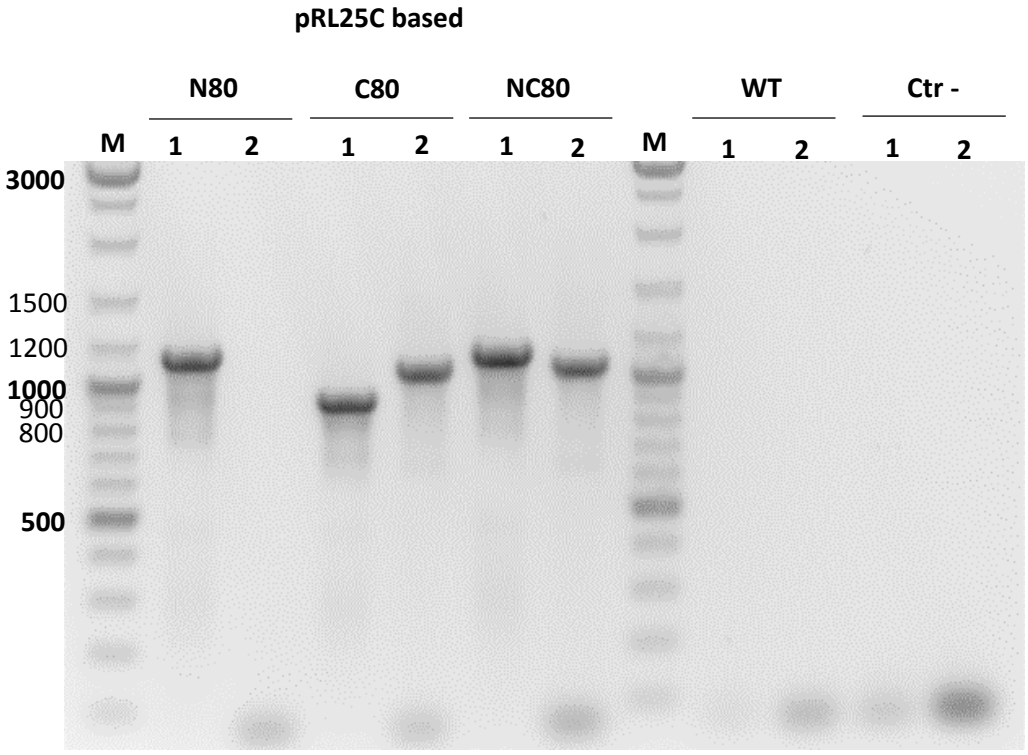

Figure 6. Ethidium bromide stained agarose gel depicting the PCR analyses to access the success of *Anabaena* sp. PCC 7120 transformation with pRL25C based constructs. PCRs were carried out by using DNA extracted from *Anabaena* sp. PCC 7120 wild-type (WT) or from each pRL25C transformant as the template, or in the case of negative controls (Ctr-) water was used instead. Two combinations of oligonucleotides were used, namely: **1** – alr0267_GFP_F + GFP_NR, and **2** – GFP_F + alr0267_CompR. Constructs assembled in pRL25C plasmid: **N80** – Pr *hesF*::N80*hesF*::*gfp*, **C80** – Pr *hesF*:: *gfp*::C80*hesF* and **NC80 –** Pr *hesF*::N80*hesF*::*gfp*::C80*hesF*. M, GeneRuler™ Ladder Mix (Thermo Scientific). The sizes in base-pairs (bp) of some of the ladders's DNA fragments are shown for reference.

Focusing now in the PCR analysis of the *Anabaena* sp. PCC 7120 transformants with pNir4 based constructs (Figure 7), a similar evaluation was performed using two combinations of oligonucleotides, namely: a – alr0267_nirN_F + GFP_NR, and b – GFPnir_F + alr0267_CompR. Using the first pair of oligonucleotides, a band with 340 bp was obtained for the designated N80nir and NC80nir transformants, while, as expected, no amplification was verified regarding the C80 transformant. Furthermore, using the second pair of oligonucleotides abovementioned, a 1067 bp band was observed for the C80nir and NC80nir transformants, whereas, as anticipated, no signal was reported for the N80nir transformant. These results sustained the success of *Anabaena* sp. PCC 7120 transformed with pNir4 based constructs.

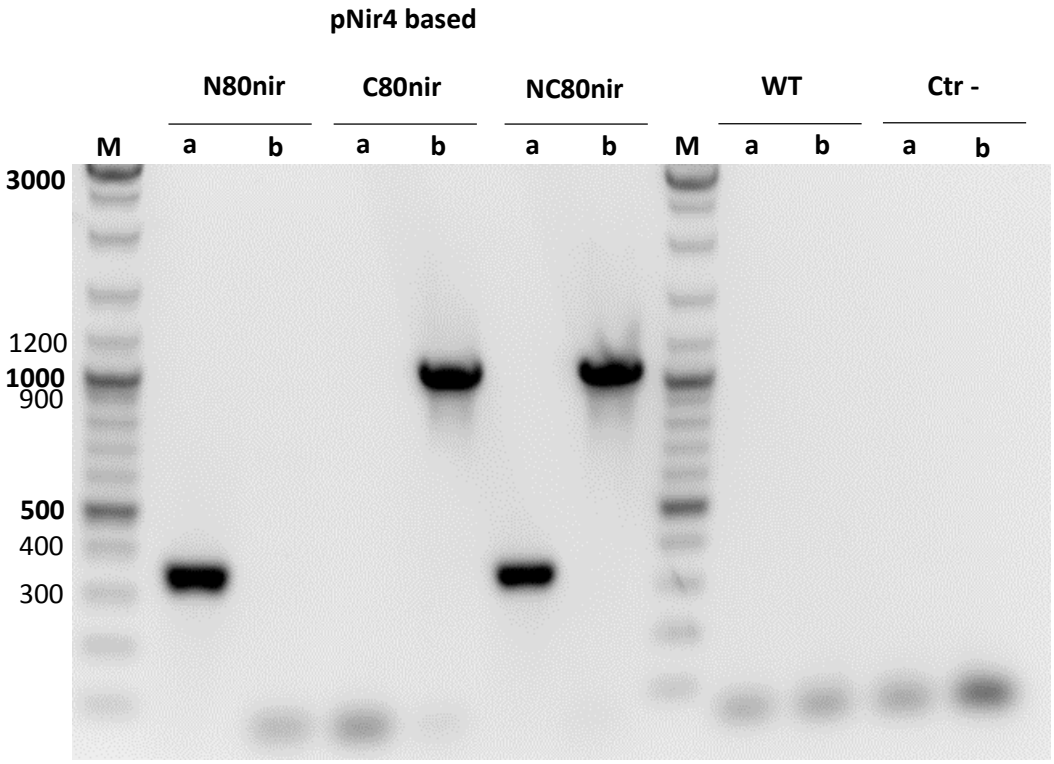

Figure 7. Ethidium bromide stained agarose gel depicting the PCR analyses to access the success of *Anabaena* sp. PCC 7120 transformation with pNir4 based constructs. PCRs were carried out by using DNA extracted from *Anabaena* sp. PCC 7120 wild-type (WT) or from each pNir4 transformant as the template, or in the case of negative controls (Ctr-) water was used instead. Two combinations of oligonucleotides were used, namely: a – alr0267_nirN_F + GFP_NR, and b – GFPnir_F + alr0267_CompR. Constructs assembled in pNir4 plasmid: , **N80nir -** Pr *nirA*::N80*hesF*:: *gfp*, **C80nir -** Pr *nirA*:: *gfp*::C80h*esF* and **NC80nir -** Pr *nirA*::N80*hesF*:: *gfp*::C80*hesF*. M, GeneRuler™ Ladder Mix (Thermo Scientific). The sizes in base-pairs (bp) of some of the ladders's DNA fragments are shown for reference.

1.3. Cellular localization of GFP-chimeric proteins

All constructs used in this study possess the *gfp* reporter gene. Consequently, a confocal microscopy analysis was performed (see Figure 8) in order to corroborate the transformation of *Anabaena* sp. PCC 7120 with the pRL25C or pNir4 based constructs, and to confirm the chimeric proteins expression. With this purpose, transformants *Anabaena* sp. PCC 7120 harboring pRL25C or pNir4 constructs, expressing a chimeric protein with the putative HesF N-terminal or C-terminal signal peptides, or with both fused with GFP and under the control of either the *hesF* or *nirA* promoters, were studied.

Focusing in pRL25C based transformants (Figure 8, panel A), they were previously grown in non-nitrogen fixing conditions (BG11), switched to nitrogen-fixing conditions ($BG11_0$) and analysed after 24 h, since in Oliveira et al., [13] it was reported that HesF was highly upregulated upon transition from non-nitrogen-fixing to nitrogen-fixing conditions, and the highest transcript levels were detected towards the end of the heterocyst differentiation process (24 h). Thus, the study was done following

the expression of HesF in “native conditions”, namely nitrogen-fixing conditions. Confocal microscopy was used to detect the respective chimeric fluorescent protein, and therefore to identify the cell type(s) possessing an active promoter. As expected and described in [13], in all transformants a strong GFP signal limited to the heterocysts was detected. In contrast, and also as anticipated, no GFP signal could be observed in vegetative cells.

Regarding pNir4 based tansformants (Figure 8, panel B), they were cultivated in non-nitrogen-fixing conditions (BG11 medium containing nitrate). With the control of the nitrate inducible promoter nirA [78] a bypass of the native expression mechanism is possible, enabling respective chimeric protein expression in vegetative cells [13, 83]. A strong GFP signal was visualized, as expected, in vegetative cells of all the transformants. Besides this fact, it is interesting to note that not all the filaments have a GFP signal, as well seen in the confocal micrograph of N80nir (Figure 8, panel B). Additionally, within some filaments of the transformants, the intensity of the GFP signal is variable, as depicted in micrograph of NC80nir (Figure 8, panel B). These evidences were already reported at confocal micrographs displayed in Desplancq et al., [83], where a pNir4::*gfp* vector was studied.

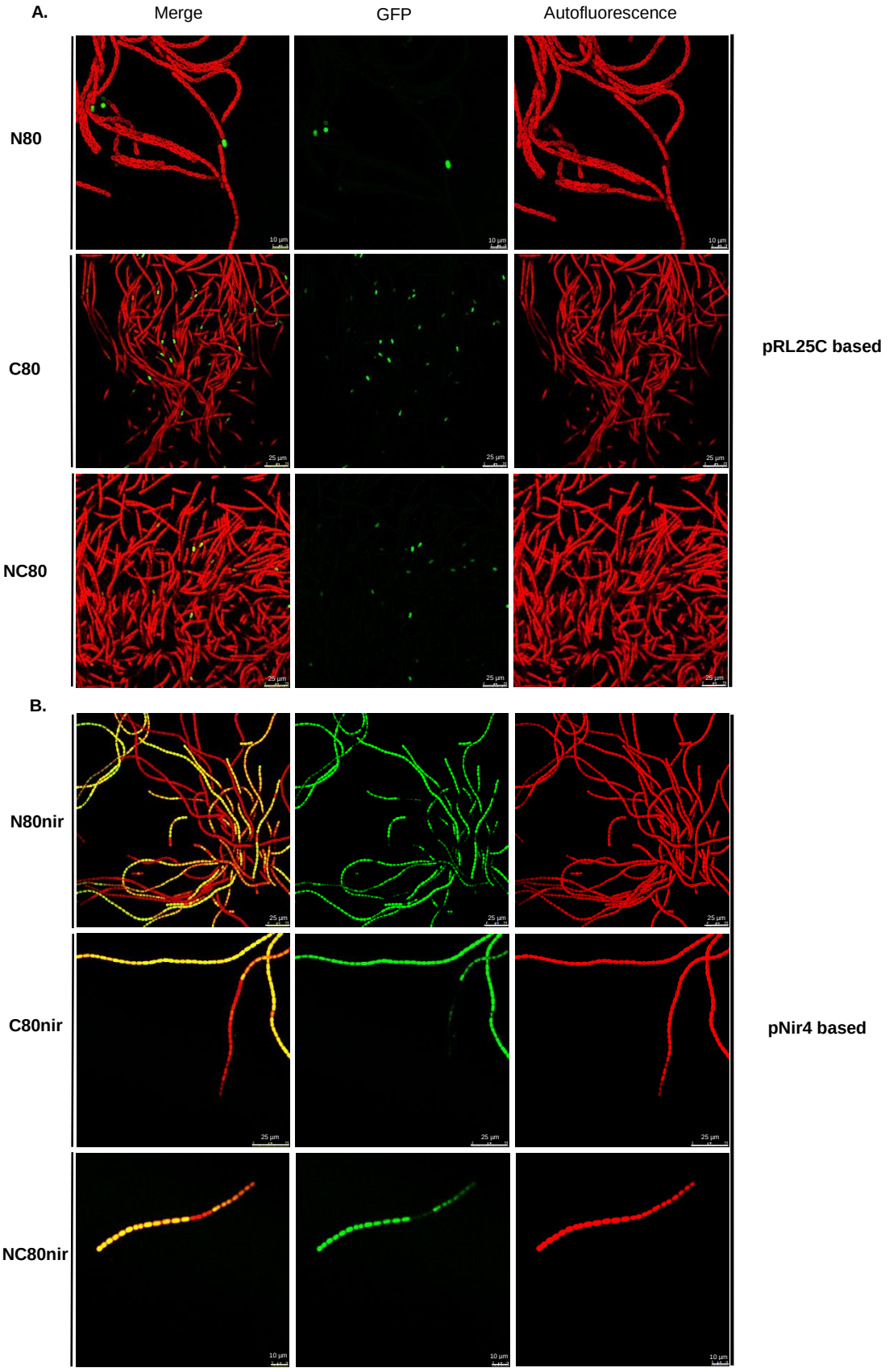

Figure 8. Localization of chimeric proteins expression. **A.** Confocal micrographs of N80, C80 and NC80 transformants filaments grown without combined nitrogen. The cyanobacterial autofluorescence is shown in the right column, while the GFP signal is depicted in the medium column. The result of merging the signals from both channels (cyanobacterial autofluorescence and GFP) is shown in the left column. **B.** Confocal micrographs of N80nir, C80nir and NC80nir transformants filaments grown with nitrate as source of combined nitrogen. The cyanobacterial autofluorescence is shown in the right column, while the GFP signal is depicted in the medium column. The result of merging the signals from both channels (cyanobacterial autofluorescence and GFP) is shown in the left column.

1.4. Proteome/Exoproteome analyses of *Anabaena* sp. PCC 7120 transformants

In order to uncover the putative signal peptide present in exoprotein HesF of *Anabaena* sp. PCC 7120 determining its secretion, as well as to demonstrate the applicability of this possible new finding in biotechnological applications (secretion of proteins of interest), a SDS-PAGE/Western blot approach with an antibody anti-GFP was performed. This allowed to investigate the proteome and exoproteome of *Anabaena* sp. PCC 7120 transformants grown in different growth conditions.

The conceivable contribution of cell lysis or leakage to the whole exoproteome is a concern that justifies cautious consideration when studying the exoproteome of a microorganism [12]. Moreover, as discussed by Oliveira et al., [12], some difficulties arise when studying the exoproteome, namely distinguish between exoproteins of interest transported to the medium as the result of active protein secretion via a specific secretion system (secretome) and those which end up accumulating in the extracellular space as the result of cell lysis, and, thus, contributing to the exoproteome [12]. Many markers have been used to show that extracellular proteins found in the growth medium were lysis-independent. For instance, studies in *E*.coli, refer that the absence of detectable nucleic acids in the culture supernatant implies that secretion of exoproteins did not result from cell lysis [84] , whereas in [85] the absence of activity of a cytoplasmic enzyme in the medium during recombinant protein excretion was used as an indicator that cell lysis did not occur. However, bacterial biofilm formation, which can possess DNA in its composition, was already reported [86-88]. Moreover, active secretion of DNA was described to aid bacterial biofilm formation [88]. Thus, it is clear that DNA can ending up accumulating in growth medium without cell lysis. Additionally, the evidence presented in [12] shows that *Anabaena* sp. PCC 7120 is capable of outer membrane vesicle (OMV) formation, which possibly can contain DNA [89] and that these vesicles are likely to contribute to the exoproteome profile. In view of all this information, it seems obvious the trouble to find a cell lysis marker which is valid for several organisms. Furthermore, in [12], to assess the contribution of cell lysis during the cultivation period, cells of *Anabaena* sp. PCC 7120 were cultivated in three different growth conditions with two types of systems - glass gas washing bottles with continuous aeration (like this study) and 1-L Erlenmeyer flasks in an orbital shaker with gentle shaking (100 r.p.m.) – being periodically sampled and the respective exoproteomes analysed. It was thought that culture aeration may result in filament shearing and ultimately lead to mechanical cell lysis, and hypothesized that as cultures get older, more cytoplasmic content ends up accumulating in the medium as a result of cell lysis. However, no significant differences could be observed in the overall exoproteome composition between these two conditions and along a cultivation period of two weeks [12]. Accordingly, in the study here presented, even being aware of this issue, exoproteome results obtained are truthfully believed to provide from protein secretion, and scarcely from cell lysis.

1.4.1. Analysis using a heterocyst specific system

Regarding transformants with pRL25C based constructs, namely designated N80, C80 and NC80, the following experience was performed. Cells were firstly grown during 5 days in non-nitrogen-fixing conditions (medium supplemented with ammonium chloride), and changed to nitrogen-fixing conditions and analysed after 48 h and 72 h. It is worthwhile to remember that, at nitrogen-fixing conditions, only 5 to 10% of a respective culture are heterocysts, as well as that a strong GFP signal was detected limited to the heterocysts's transformants (Figure 8, panel A). Being the chimeric proteins expression driven by the promoter *hesF*, which is upregulated upon transition from non-nitrogen-fixing to nitrogen-fixing conditions, only expression after 24 h (48 h and 72 h) was expected. Note that ExPASy Compute pI/Mw tool (http://web.expasy.org/compute_pi/) was used to determine approximately the expected molecular weight of each chimeric protein. Compute pI/Mw is a tool which allows the computation of the theoretical pI (isoelectric point) and Mw (molecular weight) for user entered sequences [90] Therefore, the designated N80 and C80 chimeric proteins has an expected size of 36/37 kDa, respectively, while a 44 kDa size was predicted for NC80 chimeric protein.

As anticipated, concerning proteome results (Figure 9, panel A and B), no differences were detected in the Coomassie brilliant blue stained SDS-polyacrylamide gel (Figure 9, panel A), since little amount of chimeric protein is expressed with this heterocyst-specific system in nitrogen-fixing conditions. Note that this gel (Figure 9, panel A) is shown essentially to confirm that an equal protein quantity was loaded in each lane. Furthermore, Western blot proteome analyses (Figure 9, panel B) confirmed that no chimeric protein expression occurred under non-nitrogen-fixing conditions. Interestingly and surprisingly, 72 h after shifting from non-nitrogen-fixing to nitrogen-fixing conditions, not only a band corresponding to the respective chimeric protein was observed, but a specific pattern of bands (Figure 9, panel B). It is relevant to highlight that all the lowers bands have an estimated size just above 27 kDa, the predicted size of the GFP alone. Moreover, the first bands of each transformant have the estimated size of the respective complete chimeric protein. These results indicate that a post-translational cleavage process may be occurring on N80 or/and C80 portions of the chimeric protein, whereas GFP apparently remains not affected. One more interesting observation was the high amount of signal present at the N80 transformant (Figure 9, panel B) compared to the others. This could be explained with the presence of a putative signal peptide located at the C-terminus (C80 HesF), enabling the chimeric protein secretion.

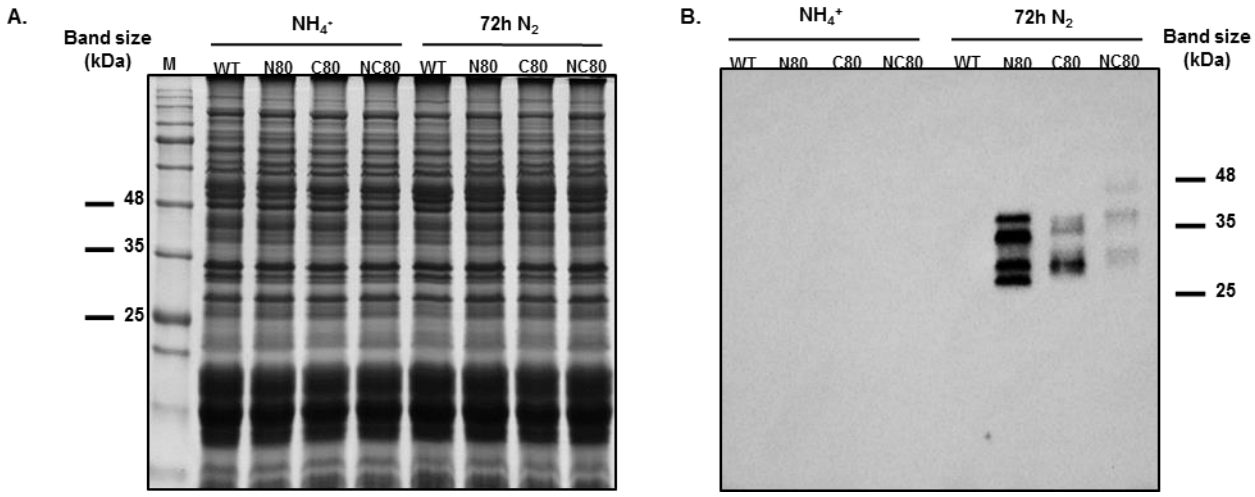

Figure 9. Proteome analysis concerning the transformants harboring pRL25C based constructs. **A**. Coomassie brilliant blue stained SDS-polyacrylamide gel depicting the proteomes of *Anabaena* sp. PCC 7120 wild-type (WT) and transformants with pRL25C based constructs (N80, C80 and NC80), collected in non-nitrogen-fixing conditions with ammonium chloride as nitrogen source (NH_4^+), and 72 h after combined nitrogen step down (72 h N_2). This gel is shown to confirm that an equal protein quantity was loaded in each lane. M, NZYColour Protein Marker II (Nzytech®). The sizes in kilodaltons (kDa) of some of the ladders's protein fragments are shown for reference to the left. **B.** Western blot result using an antibody anti-GFP depicting the proteomes of *Anabaena* sp. PCC 7120 WT and designated N80, C80 and NC80 transformants. The sizes in kilodaltons (kDa) of some of the ladders's protein fragments are shown for reference to the right.

The exoproteome of the transformants with pRL25C based constructs are depicted in Figure 10 (panel A and B) (data from after combined-nitrogen step down: 48 h and 72 h). In the Coomassie brilliant blue stained SDS-polyacrylamide gel (Figure 10, panel A), it was not possible to distinguish any extra band with the respective expected chimeric protein size (comparing to the lane regarding the wild-type) indicating chimeric protein secretion. Once again, it should be noted that only a few amount of chimeric protein is expressed in this heterocyst-specific system. Interestingly, analyzing the exoproteome Western blot results (Figure 10, panel B), it was possible to detect a signal in the C80 transformant corresponding lane (highlighted with an arrow), 48 h and 72 h after shifting from non-nitrogen to nitrogen-fixing conditions, whereas no band was observed for the others two transformants. These results indicate that a putative signal peptide may be present at the C-terminus of HesF protein, since a truncated chimeric protein comprising this portion was secreted. Indeed, HesF was found to be a type I secretion system substrate, since an HgdD (TolC-like) mutant failed to secrete HesF [13]. Moreover, as already mentioned, is generally recognized that T1SS-secreted proteins have a C-terminal signal peptide (SP) required for secretion [27, 55]. Note that the designated NC80 chimeric protein was not apparently secreted, despite possessing the C80 HesF fragment. This could be explained due to the possible loss of tridimensional stability by the chimeric protein in the extracellular medium. On the other hand, despite of three biological replicates achieved with equal results, it is worthwhile to notice that only one clone of each transformant was tested. Consequently,

it will be recommended to test others clones in order to confirm these results, and particularly to see if the NC80 transformant continues without chimeric protein secretion.

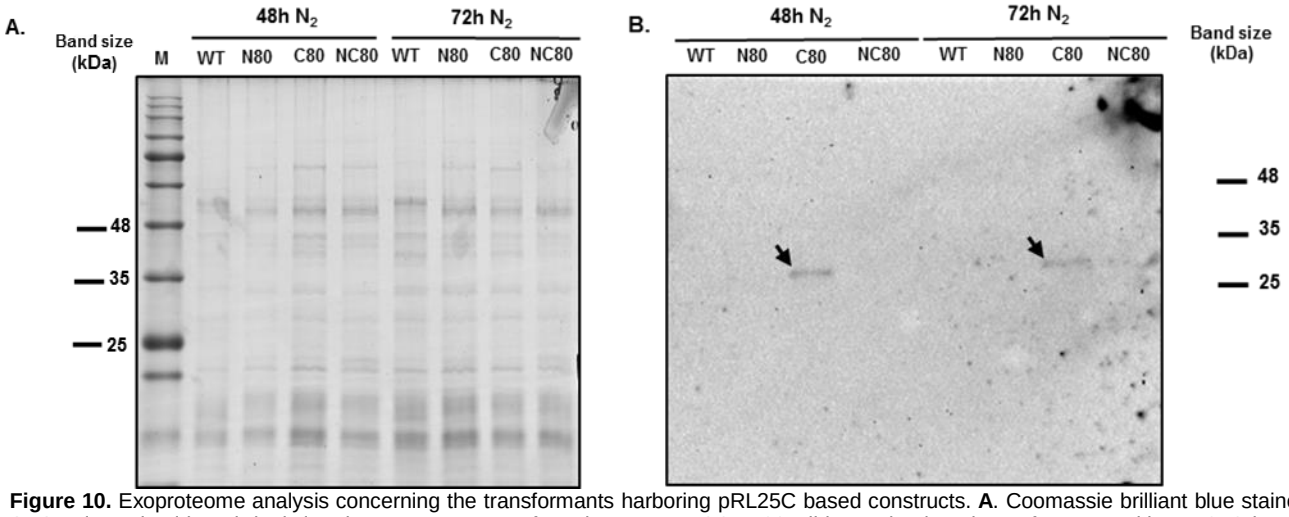

Figure 10. Exoproteome analysis concerning the transformants harboring pRL25C based constructs. **A**. Coomassie brilliant blue stained SDS-polyacrylamide gel depicting the exoproteomes of *Anabaena* sp. PCC 7120 wild-type (WT) and transformants with pRL25C based constructs (N80, C80 and NC80), collected in nitrogen-fixing conditions, 48 h and 72 h after combined nitrogen step down. This gel is shown to confirm that an equal protein quantity was loaded in each lane. M, NZYColour Protein Marker II (Nzytech®). The sizes in kilodaltons (kDa) of some of the ladders's protein fragments are shown for reference to the left. **B.** Western blot result using an antibody anti-GFP depicting the exoproteomes of *Anabaena* sp. PCC 7120 WT and designated N80, C80 and NC80 transformants, with abovementioned conditions. It is worthwhile to notice that a signal was detected regarding the C80 lanes (highlighted with an arrow). The sizes in kilodaltons (kDa) of some of the ladders's protein fragments are shown for reference to the right.

Others aspects that justify close consideration when scheming a protein secretion microorganism are the strength of the promoter used (Pr *hesF* in this study), the timing of expression (nitrogen-fixing conditions with Pr *hesF*) and the cell-type specificity (expressed in heterocysts of *Anabaena* sp. PCC 7120). Therefore, analyses of the *hesF* promoter by Oliveira et al., [13], concludes that both NtcA, a global transcription regulator belonging to the cyclic AMP receptor protein family [91], and HetR, a master regulator of heterocyst development [92], are essential for the gene's in vivo activation. Moreover, transcriptional studies showed that the *hesF* gene is transcribed at very low levels in a medium supplemented with NH_4Cl and strongly upregulated after combined nitrogen step-down in heterocysts only. These points are crucial when designing a protein secretion *Anabaena* [12].

1.4.2. Analysis using a vegetative cell specific system

Concerning transformants with pNir4 based constructs, i.e., designated N80nir, C80nir and NC80nir, the succeeding assay was accomplished. Cells were grown during 6 days in non-nitrogen-fixing conditions (medium BG11 containing nitrate), and analysed. It is worthwhile to recall that, with this condition and with the control of the nitrate inducible promoter *nirA* [78], a bypass of the native expression mechanism is possible, enabling respective chimeric protein expression [13, 83], which was further confirmed with the confocal micrographs above displayed (see Figure 8, panel B), where a strong GFP signal was visualized in vegetative cells of all the transformants, as expected.

Therefore, relating to proteome results (Figure 11, left side of panel A and B), no differences were perceived in the Coomassie brilliant blue stained SDS-polyacrylamide gel (Figure 11, panel A). Note that this gel is shown essentially to confirm that an equal protein quantity was loaded in each lane. Interestingly, observing the proteome Western blot results (Figure 11, left side of panel B), once again, not a single band corresponding to the respective chimeric protein was observed, but a specific pattern of bands. It is pertinent to highlight that, as proteome results of transformants with pRL25C based constructs, all the lowers bands have an estimated size just above 27 kDa, the predicted size of the GFP alone. Furthermore, the first bands of each transformant have the estimated size of the respective complete chimeric protein too. Additionally and interestingly, it was not possible to detect any signal in the proteome result regarding the NC80nir transformant, besides the previous detection of fluorescence. Thus, it appears that all the NC80nir chimeric protein produced was secreted. Moreover, these results indicate that a specific post-translational cleavage process may be occurring on N80 or/and C80 portions of the chimeric protein, whereas GFP apparently remains not affected. Another relevant observation was the high amount of signal present at the N80nir transformant (Figure 11, left side of panel B) compared to the others transformants. This could be explained with the absence of the putative signal peptide located at the C-terminus (C80 HesF) which putatively allow the chimeric protein secretion.

Exoproteome results of transformants with pNir4 based constructs are showed in Figure 11 (right side of panel A and B). Thus, in the Coomassie brilliant blue stained SDS-polyacrylamide gel (Figure 11, right side of panel A), it was not possible to distinguish any extra band with the respective chimeric protein expected size (comparing to the lane regarding the wild-type) indicating chimeric protein secretion. Surprisingly, analyzing the exoproteome Western blot results (Figure 11, right side of panel B), it was possible to detect strong signals in the NC80nir transformant corresponding lane, whereas no bands were reported for the N80nir and C80nir transformant lanes. Therefore, in this specific vegetative cell system, where exoprotein HesF is not naturally expressed, it seems that the portion C80 of HesF is also required to allow protein secretion. In addition, it appears that the presence of the N80 portion in the chimeric protein which possess both portions (N80 and C80), seems to aid in the protein secretion process in this specific vegetative cell system. However, it should be noted that

only one clone of each transformant was tested. Thus, despite of several biological replicates performed gave rise to similar results, it will be relevant to repeat these experiments with other clones to confirm the results obtained, and namely to see if the C80nir transformant continues without chimeric protein secretion.

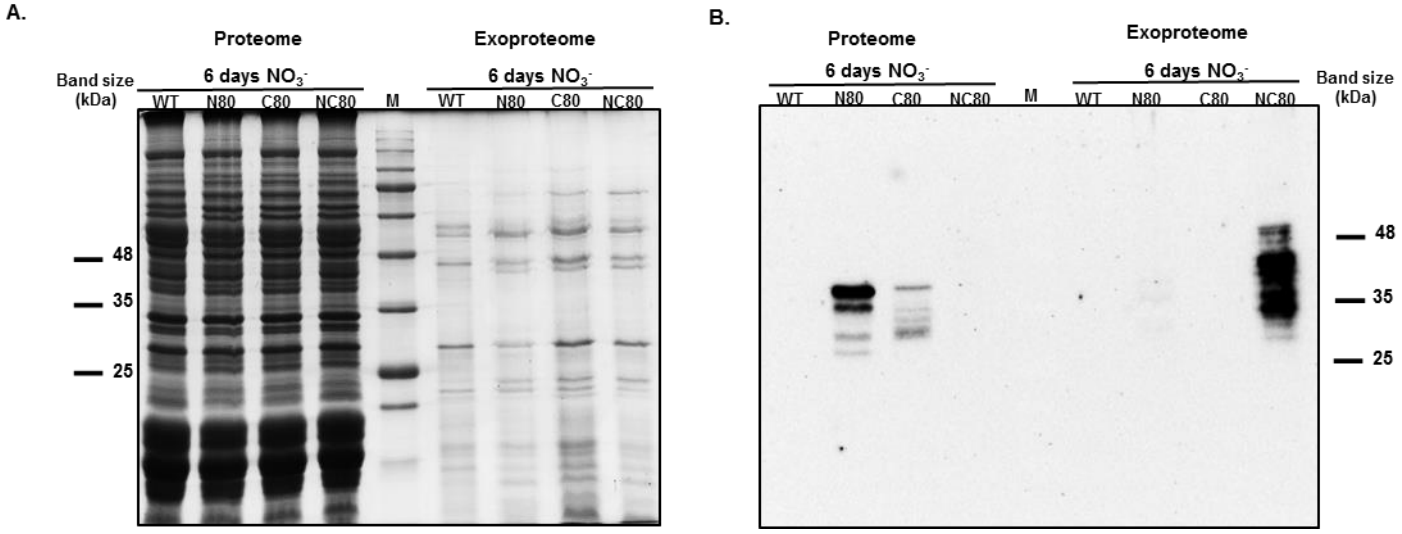

Figure 11. Proteome and exoproteome analyses concerning the transformants harboring pNir4 based constructs. **A**. Coomassie brilliant blue stained SDS-polyacrylamide gel depicting the proteome and exoproteome of *Anabaena* sp. PCC 7120 wild-type (WT) and transformants with pNir4 based constructs (N80nir, C80nir and NC80nir), collected after 6 days grown in non-nitrogen-fixing conditions (medium containing nitrate). This gel is shown to confirm that an equal protein quantity was loaded in each lane. M, NZYColour Protein Marker II (Nzytech®). The sizes in kilodaltons (kDa) of some of the ladders's protein fragments are shown for reference to the left of each panel.. **B.** Western blot result using an antibody anti-GFP depicting the proteome and exoproteome of *Anabaena* sp. PCC 7120 WT and designated N80nir, C80nir and NC80nir transformants, with abovementioned conditions. The sizes in kilodaltons (kDa) of some of the ladders's protein fragments are shown for reference to the right of each panel.

It is worthy to notice that type 1 secretion system (T1SS) consists of three components: an ABC transporter, a membrane fusion protein (MFP) and an outer membrane protein-channel or factor (OMF) of the TolC (OMF from *Escherichia coli*) family (Figure 2). Even though all three parts of T1SS are required for proper protein secretion, it is known that the ABC-transporter performs the selection of which proteins are going to be secreted [52]. Moreover, different ABC-transporters and MFP's are usually present in the same bacteria, whereas there is usually only one OMP in each species [36, 52]. The ABC-transporters work both in the recognition of the protein substrate and to energize the secretion process [93], while the apparent MFP complex role is to make a bridge between the ABC-transporter and the OMF. The OMF act as a pore, alternating from a closed position to an open position when triggered by its substrate, to secrete the substrate to the medium [94]. In *Anabaena* sp. PCC 7120 there is only one TolC homologue that acts as OMF, alr2887, known as HgdD [31, 69] . Exoprotein HesF, as already mentioned, was found to be a type 1 secretion system substrate, since an HgdD (TolC-like) mutant failed to secrete HesF [13, 31]. However, the two other T1SS protein secretion components remain to be identified in *Anabaena* sp. PCC 7120. In addition, fusing a *nirA* promoter with the HesF exoprotein encoding gene, an overexpression of the protein was achieved in vegetative cells grown in medium containing nitrate, leading to an over-accumulation of HesF in the

medium [13]. Consequently, authors concluded that HesF is secreted to the extracellular space neither using a heterocyst-specific mechanism nor a N_2-fixation-dependent complex [13]. Furthermore, the secretion system appears to be either highly active or plentiful to be responsible for exporting such high amounts of HesF to the medium [13].

Once again, others aspects that deserve close attention are the promoter strength of Pr *nirA*, the timing of expression (expression in medium supplemented with nitrate) and the cell-type specificity in *Anabaena* sp. PCC 7120 (expressed in vegetative cells). Indeed, the nir operon harbors six genes encoding specific permeases (nrtABCD) and reductases (narB) and is strongly transcribed when the nitrogen source is either nitrate or nitrite [83], as it possible to confirm observing Figure 11. On the other hand, expression of the *nir* operon in strain PCC 7120 is subjected to ammonium-promoted repression [78], as tested in this study (data not shown). These are relevant topics to bear in mind for engineering and modulating cyanobacterial protein secretion.

1.4.3 Comparison between heterocyst and vegetative specific systems

In order to compare the proteome results between the transformants with pNir4 and pRL25C based constructs, a single gel (equal abovementioned SDS-PAGE/Western blot approach) with both proteomes was performed (Figure 12, panel A and B). With the intention of allow this proteome comparison, 35 µg of total protein was loaded in pRL25C transformants's lanes (heterocyst specific system), whereas 10 µg was loaded in pNir4 transformants's lanes (vegetative cell system), as it possible to verify in the Coomassie brilliant blue stained SDS-polyacrylamide gel (Figure 12, panel A). Concerning the Western blot results (Figure 12, panel B), it was interesting to note that tranformants N80 and N80nir possess an apparent equal pattern of bands. Despite of this fact, it is relevant to notice a clear differential accumulation of each band between them. Furthermore, the pattern of bands visualized for C80nir transformant lane also appears to be dissimilar with C80 of the verified in the heterocyst-specific system. Unfortunately, with Figure 12 (panel B), it was not possible to compare the NC80nir proteome result with the respective heterocyst-specific system NC80 proteome result. Thus, these analyses appears to indicate that vegetative cells and heterocysts display a different/specific "package of maturation tools" for the post-translational cleavage process which looks to be occurring.

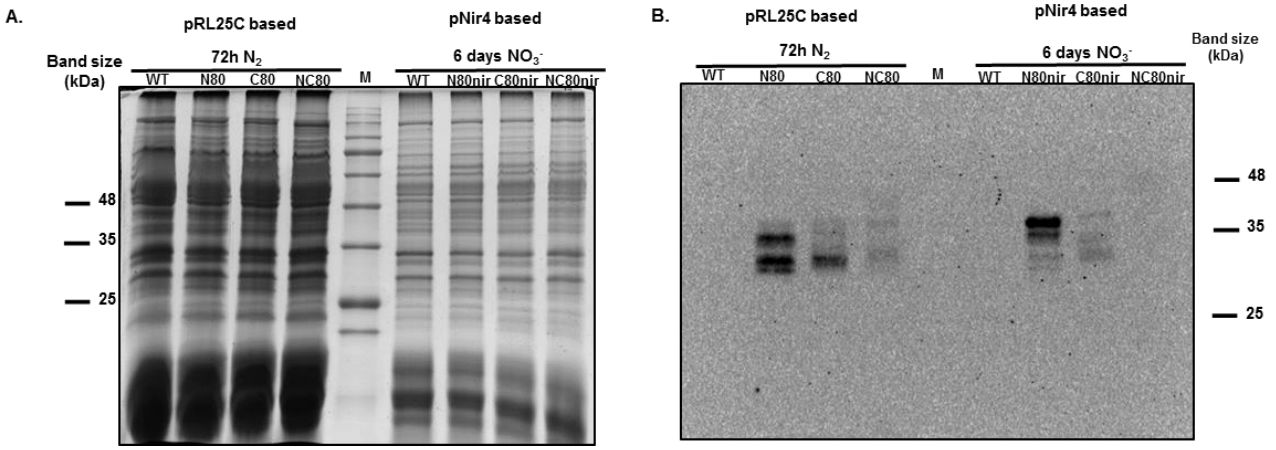

Figure 12. Proteome comparison between the transformants harboring pRL25C and pNir4 based constructs. **A**. Coomassie brilliant blue stained SDS-polyacrylamide gel depicting the proteomes of *Anabaena* sp. PCC 7120 wild-type (WT) and transformants with pRL25C (N80, C80 and NC80) and pNir4 based constructs (N80nir, C80nir and NC80nir). The transformants with pRL25C based constructs were collected in nitrogen-fixing conditions, 72 h after combined nitrogen step down, while transformants with pNir4 based constructs were collected 6 days after grown in medium containing nitrate. In order to allow this proteome comparison, 35 µg of total protein was loaded in pRL25C transformants's lanes, whereas 10 µg was loaded in pNir4 transformants's lanes. M, NZYColour Protein Marker II (Nzytech®). The sizes in kilodaltons (kDa) of some of the ladders's protein fragments are shown for reference to the left. **B.** Western blot result using an antibody anti-GFP depicting the proteomes of *Anabaena* sp. PCC 7120 WT and transformants, with abovementioned conditions. The sizes in kilodaltons (kDa) of some of the ladders's protein fragments are shown for reference to the right.

Conclusions and future perspectives

This study proposed to uncover the putative signal peptide present in exoprotein HesF of *Anabaena* sp. PCC 7120, as well as to demonstrate the applicability of this possible new finding in biotechnological applications (secretion of proteins of interest). Conclusions achieved with this study are following summarized:

- the indication that a putative signal peptide for the secretion of HesF may be present at the C-terminus of the protein;
- the evidence that the cyanobacterial protein secretion can be controlled (cell-type specificity; timing of expression; promoter strength);
- the suggestion of a specific vegetative cell/heterocyst cell post-translational cleavage process occurring against the HesF analyzed portions;
- the hypothesis that exoprotein HesF is secreted after a post-translational maturation process.

Despite the evidence presented in this work pointing to the possibility that HesF may possess a signal peptide located at the C-terminus, since the successfully secretion of a truncated protein of interest to the medium when native conditions tested, some questions remain unanswered: Does all the 80 amino acids of HesF's C-terminus really necessary to secrete a selected protein? Why the GFP-chimeric protein containing both N- and C-terminus of HesF was not secreted when native conditions tested? How to explain and/or evaluate the specific vegetative cell/heterocyst cell post-translational cleavage process occurring against the HesF analyzed portions? Why an apparently different result was obtained with the specific vegetative cell system? Some of these questions will certainly be addressed by the group where I am inserted. Consequently, in order to answer some of these topics, I propose the elaboration of a complementary study. Thus, it would be done using constructs possessing the complete *hesF* sequence, linked to the reporter *gfp* at both the N- and C-terminus, and driven by both *hesF* or *nirA* promoter (Figure 13). This experiment would be relevant to unveil the interference of the putative SP (C80 HesF) function due to the presence of a GFP located at the C-terminus of the chimeric protein. Furthermore, it would shed light about the possible post-translational cleavage process occurring against HesF. Another tremendously helpful aspect to study these interrogations would be the generation of an antibody anti-HesF.

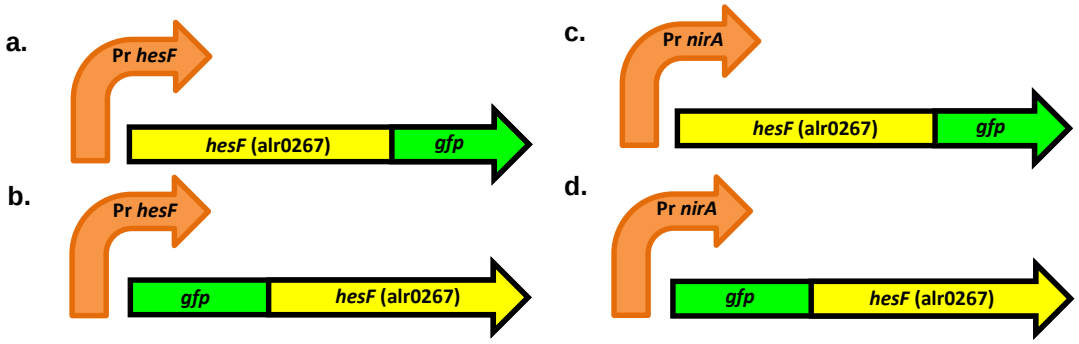

Figure 13. Schematic representation of the constructs proposed to do a complementary study, in order to shed light about some of the arisen questions with this study. Constructs **a.** and **b.** would be assembled in pRL25C plasmid, and constructs **c.** and **d.** would be built in pNir4 plasmid: **a.** Pr *hesF*::*hesF*::*gfp*, **b.** Pr *hesF*:: *gfp*::*hesF*, **c.** Pr *nirA*::*hesF*:: *gfp* and **d** Pr *nirA*:: *gfp*::h*esF*.

Still, the ability for secreting proteins to the extracellular space is not limited to cyanobacteria [12]. Moreover, certain prokaryotes are already progressively used to produce and secrete selected industrially relevant proteins [64]. Nevertheless, in order to explore protein secretion by cyanobacteria various items require to be scrutinized and finely tuned. Thus, signal peptides that can address the translocation of a newly-synthesized selected protein targets from the cyanobacterial cytoplasm to the extracellular milieu signifies one of the most critical point towards optimization of the secretion process [12]. In addition to the obvious aptitude to secrete selected proteins, the specific case of HesF here studied, with the putative identification of a possible novel signal peptide, may represent the open of an array of future applications, namely in key areas like biofuel production, bioremediation and biomass recovery.

Regarding biofuel production, generating nitrogen-fixing cyanobacteria capable of secreting large amounts of a target protein would cause the organism to be drained off of fixed nitrogen and probably relieve repression and feedback on nitrogenase activity, which has molecular hydrogen (H_2) as by-product [12]. Thus, a mutated *Anabaena* sp. PCC 7120 strain, as Δ*hypF*, which is already able to produce/release higher levels of H_2 through nitrogenase-complex activity, since deletion of *hypF* render the uptake hydrogenase inactive [95], could be used for that purpose. Thereby, a construct carrying a strong constitutive promoter linked to a “gene X” encoding a “synthetic protein X N-enriched” with nitrogen-rich amino acids (glutamine, histidine and asparagine) and to our C-terminus SP, could be assembled, and transferred into *hypF* mutant of *Anabaena* sp. PCC 7120. Consequently, from the continuous protein secretion, a higher nitrogen fixation rates and/or extended periods of nitrogenase activity to sustain growth could be expected, which would be an advantage for this nitrogenase-based technological process to produce H_2 [12, 96].

Concerning bioremediation, the putative HesF’s SP could help secreting specific heavy metal chelators in order to detoxifying heavy-metal contaminated soils, which with the theoretically combined nitrogen co-enrichment of *Anabaena* sp. PCC 7120, could contribute to that objective [12].

Biomass recovery in large volumes is an important issue to be addressed to cultivate cyanobacteria in large scale. Therefore, one of the approaches to separate growth medium from the valuable biomass has been to trigger cyanobacterial cells to adhere, aggregate and flocculate, easing its recovery [12]. In the specific case of the putative SP of HesF here found, it could be linked to others extracellular proteins in order to improve their secretion, facilitating the biomass recovery by the enhancement of adhesion and aggregation of cells [13, 28, 29]. Indeed, activation of those chimeric exoproteins expression and secretion with a precisely defined timing may aid in biomass recovery and reduce total biomass production costs. Thus, crucial aspects as the promoter strength, the timing of expression (specific growth condition) and the cell-type specificity in *Anabaena* sp. PCC 7120 (vegetative cell vs heterocyst), deserves close attention, since they are just a few of the points that need to be evaluated and harmonized for engineering a competent protein secretion *Anabaena* [12].

To sum up, these results contributed to the promise of exploiting cyanobacteria as robust protein secretion factories, as well as for the fundamental elucidation of protein secretion in cyanobacteria.

References

1. Olson, J.M., Photosynthesis in the Archean era. Photosynthesis research, 2006. **88**(2): p. 109-117.
2. Des Marais, D.J., When did photosynthesis emerge on Earth? Science(Washington), 2000. **289**(5485): p. 1703-1705.
3. Nisbet, E. and N. Sleep, The habitat and nature of early life. Nature, 2001. **409**(6823): p. 1083-1091.
4. Cavalier-Smith, T., Cell evolution and Earth history: stasis and revolution. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 2006. **361**(1470): p. 969-1006.
5. Tomitani, A., A.H. Knoll, C.M. Cavanaugh, and T. Ohno, The evolutionary diversification of cyanobacteria: molecular–phylogenetic and paleontological perspectives. Proceedings of the National Academy of Sciences, 2006. **103**(14): p. 5442-5447.
6. Delwiche, C.F. and J.D. Palmer, The origin of plastids and their spread via secondary symbiosis. Vol. 11. 1997: Springer.
7. Waterbury, J.B., The cyanobacteria—isolation, purification and identification, in The prokaryotes. 2006, Springer US. p. 1053-1073.
8. Shi, T. and P.G. Falkowski, Genome evolution in cyanobacteria: the stable core and the variable shell. Proceedings of the National Academy of Sciences, 2008. **105**(7): p. 2510-2515.
9. Valério, E., L. Chambel, S. Paulino, N. Faria, P. Pereira, and R. Tenreiro, Molecular identification, typing and traceability of cyanobacteria from freshwater reservoirs. Microbiology, 2009. **155**(2): p. 642-656.
10. Adams, D.G. and P.S. Duggan, Signalling in cyanobacteria–plant symbioses, in Signaling and communication in plant symbiosis. 2012, Springer-Verlag Berlin Heidelberg. p. 93-121.
11. Heidorn, T., D. Camsund, H.-H. Huang, P. Lindberg, P. Oliveira, K. Stensjö, and P. Lindblad, Chapter Twenty-Four - Synthetic Biology in Cyanobacteria: Engineering and Analyzing Novel Functions, in Methods in Enzymology, V. Chris, Editor. 2011, Academic Press. p. 539-579.
12. Oliveira, P., N.M. Martins, M. Santos, N.A. Couto, P.C. Wright, and P. Tamagnini, The *Anabaena* sp. PCC 7120 exoproteome: Taking a peek outside the box. Life (Basel), 2015. **5**(1): p. 130-63.
13. Oliveira, P., F. Pinto, C.C. Pacheco, R. Mota, and P. Tamagnini, HesF, an exoprotein required for filament adhesion and aggregation in *Anabaena* sp. PCC 7120. Environmental Microbiology, 2015. **17**(5): p. 1631-1648.
14. Silhavy, T.J., D. Kahne, and S. Walker, The bacterial cell envelope. Cold Spring Harbor perspectives in biology, 2010. **2**(5): p. a000414.

15. Dubey, G.P. and S. Ben-Yehuda, Intercellular nanotubes mediate bacterial communication. Cell, 2011. **144**(4): p. 590-600.
16. López, D., H. Vlamakis, and R. Kolter, Biofilms. Cold Spring Harbor perspectives in biology, 2010. **2**(7): p. a000398.
17. Hoiczyk, E. and A. Hansel, Cyanobacterial cell walls: news from an unusual prokaryotic envelope. Journal of Bacteriology, 2000. **182**(5): p. 1191-1199.
18. Jürgens, U., G. Drews, and J. Weckesser, Primary structure of the peptidoglycan from the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6714. Journal of bacteriology, 1983. **154**(1): p. 471-478.
19. Schmidt, W., G. Drews, J. Weckesser, I. Fromme, and D. Borowiak, Characterization of the lipopolysaccharides from eight strains of the cyanobacterium *Synechococcus.* Archives of Microbiology, 1980. **127**(3): p. 209-215.
20. Resch, C.M. and J. Gibson, Isolation of the carotenoid-containing cell wall of three unicellular cyanobacteria. Journal of bacteriology, 1983. **155**(1): p. 345-350.
21. Hansel, A., F. Pattus, U.J. Jürgens, and M.H. Tadros, Cloning and characterization of the genes coding for two porins in the unicellular cyanobacterium *Synechococcus* PCC 6301. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression, 1998. **1399**(1): p. 31-39.
22. Winkenbach, F., C.P. Wolk, and M. Jost, Lipids of membranes and of the cell envelope in heterocysts of a blue-green alga. Planta, 1972. **107**(1): p. 69-80.
23. Pereira, S., A. Zille, E. Micheletti, P. Moradas-Ferreira, R. De Philippis, and P. Tamagnini, Complexity of cyanobacterial exopolysaccharides: composition, structures, inducing factors and putative genes involved in their biosynthesis and assembly. FEMS microbiology reviews, 2009. **33**(5): p. 917-941.
24. Neu, T.R. and K.C. Marshall, Bacterial polymers: physicochemical aspects of their interactions at interfaces. Journal of biomaterials applications, 1990. **5**(2): p. 107-133.
25. Sergeyenko, T.V. and D.A. Los, Identification of secreted proteins of the cyanobacterium *Synechocystis* sp. strain PCC 6803. FEMS microbiology letters, 2000. **193**(2): p. 213-216.
26. Sergeyenko, T.V. and D.A. Los, Cyanobacterial leader peptides for protein secretion. FEMS microbiology letters, 2003. **218**(2): p. 351-357.
27. Moslavac, S., K. Nicolaisen, O. Mirus, F. Al Dehni, R. Pernil, E. Flores, I. Maldener, and E. Schleiff, A TolC-like protein is required for heterocyst development in *Anabaena* sp. strain PCC 7120. Journal of bacteriology, 2007. **189**(21): p. 7887-7895.
28. Zilliges, Y., J.-C. Kehr, S. Mikkat, C. Bouchier, N.T. de Marsac, T. Börner, and E. Dittmann, An extracellular glycoprotein is implicated in cell-cell contacts in the toxic cyanobacterium *Microcystis aeruginosa* PCC 7806. Journal of Bacteriology, 2008. **190**(8): p. 2871-2879.

29. Kehr, J.C., Y. Zilliges, A. Springer, M.D. Disney, D.D. Ratner, C. Bouchier, P.H. Seeberger, N.T. De Marsac, and E. Dittmann, A mannan binding lectin is involved in cell–cell attachment in a toxic strain of *Microcystis aeruginosa.* Molecular microbiology, 2006. **59**(3): p. 893-906.
30. Helm, R.F. and M. Potts, Extracellular matrix (ECM), in Ecology of Cyanobacteria II. 2012, Springer. p. 461-480.
31. Hahn, A., M. Stevanovic, E. Brouwer, D. Bublak, J. Tripp, T. Schorge, M. Karas, and E. Schleiff, Secretome analysis of *Anabaena* sp. PCC 7120 and the involvement of the TolC-homologue HgdD in protein secretion. Environmental microbiology, 2015. **17**(3): p. 767-780.
32. Vilhauer, L., J. Jervis, W.K. Ray, and R.F. Helm, The exo-proteome and exo-metabolome of *Nostoc punctiforme* (cyanobacteria) in the presence and absence of nitrate. Archives of microbiology, 2014. **196**(5): p. 357-367.
33. Christie-Oleza, J.A., J. Armengaud, P. Guerin, and D.J. Scanlan, Functional distinctness in the exoproteomes of marine *Synechococcus.* Environmental microbiology, 2015.
34. Christie-Oleza, J.A., D.J. Scanlan, and J. Armengaud, "You produce while I clean up", a strategy revealed by exoproteomics during *Synechococcus–Roseobacter* interactions. Proteomics 2015: p. 1-9.
35. Christie-Oleza, J.A., J.M. Piña-Villalonga, R. Bosch, B. Nogales, and J. Armengaud, Comparative proteogenomics of twelve *Roseobacter* exoproteomes reveals different adaptive strategies among these marine bacteria. Molecular & Cellular Proteomics, 2012. **11**(2): p. M111. 013110.
36. Kanonenberg, K., C.K. Schwarz, and L. Schmitt, Type I secretion systems–a story of appendices. Research in microbiology, 2013. **164**(6): p. 596-604.
37. Wall, D. and D. Kaiser, Type IV pili and cell motility. Molecular microbiology, 1999. **32**(1): p. 01-10.
38. Akatsuka, H., E. Kawai, K. Omori, and T. Shibatani, The three genes lipB, lipC, and lipD involved in the extracellular secretion of the *Serratia marcescens* lipase which lacks an N-terminal signal peptide. Journal of bacteriology, 1995. **177**(22): p. 6381-6389.
39. Duong, F., A. Lazdunski, B. Carni, and M. Murgier, Sequence of a cluster of genes controlling synthesis and secretion of alkaline protease in *Pseudomonas aeruginosa*: relationships to other secretory pathways. Gene, 1992. **121**(1): p. 47-54.
40. Letoffe, S., J. Ghigo, and C. Wandersman, Iron acquisition from heme and hemoglobin by a *Serratia marcescens* extracellular protein. Proceedings of the National Academy of Sciences, 1994. **91**(21): p. 9876-9880.
41. Dörr, J., T. Hurek, and B. Reinhold-Hurek, Type IV pili are involved in plant–microbe and fungus–microbe interactions. Molecular Microbiology, 1998. **30**(1): p. 7-17.

42. Bleves, S., V. Viarre, R. Salacha, G.P.F. Michel, A. Filloux, and R. Voulhoux, Protein secretion systems in *Pseudomonas aeruginosa:* A wealth of pathogenic weapons. International Journal of Medical Microbiology, 2010. **300**(8): p. 534-543.
43. Dirix, G., P. Monsieurs, B. Dombrecht, R. Daniels, K. Marchal, J. Vanderleyden, and J. Michiels, Peptide signal molecules and bacteriocins in Gram-negative bacteria: a genome-wide in silico screening for peptides containing a double-glycine leader sequence and their cognate transporters. Peptides, 2004. **25**(9): p. 1425-1440.
44. Freudl, R., Leaving home ain't easy: protein export systems in Gram-positive bacteria. Research in microbiology, 2013. **164**(6): p. 664-674.
45. Wandersman, C., Concluding remarks on the special issue dedicated to Bacterial secretion systems: function and structural biology. Research in Microbiology, 2013. **164**(6): p. 683-687.
46. Desvaux, M., M. Hébraud, R. Talon, and I.R. Henderson, Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. Trends in Microbiology, 2009. **17**(4): p. 139-145.
47. Armengaud, J., J.A. Christie-Oleza, G. Clair, V. Malard, and C. Duport, Exoproteomics: exploring the world around biological systems. Expert review of proteomics, 2012. **9**(5): p. 561-575.
48. Gennity, J.M. and M. Inouye, Protein secretion in bacteria. Current opinion in biotechnology, 1991. **2**(5): p. 661-667.
49. Costa, T.R., C. Felisberto-Rodrigues, A. Meir, M.S. Prevost, A. Redzej, M. Trokter, and G. Waksman, Secretion systems in Gram-negative bacteria: structural and mechanistic insights. Nature Reviews Microbiology, 2015. **13**(6): p. 343-359.
50. Economou, A., P.J. Christie, R.C. Fernandez, T. Palmer, G.V. Plano, and A.P. Pugsley, Secretion by numbers: protein traffic in prokaryotes. Molecular microbiology, 2006. **62**(2): p. 308-319.
51. Barnett, J.P., C. Robinson, D.J. Scanlan, and C.A. Blindauer, The Tat protein export pathway and its role in cyanobacterial metalloprotein biosynthesis. FEMS microbiology letters, 2011. **325**(1): p. 1-9.
52. Saier Jr, M.H., Protein secretion and membrane insertion systems in gram-negative bacteria. The Journal of membrane biology, 2006. **214**(1-2): p. 75-90.
53. Koster, M., W. Bitter, and J. Tommassen, Protein secretion mechanisms in Gram-negative bacteria. International journal of medical microbiology, 2000. **290**(4): p. 325-331.
54. Delepelaire, P., Type I secretion in gram-negative bacteria. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 2004. **1694**(1): p. 149-161.
55. Masi, M. and C. Wandersman, Multiple signals direct the assembly and function of a type 1 secretion system. Journal of bacteriology, 2010. **192**(15): p. 3861-3869.

56. Létoffé, S., K. Omori, and C. Wandersman, Functional characterization of the HasAPF hemophore and its truncated and chimeric variants: determination of a region involved in binding to the hemophore receptor. Journal of bacteriology, 2000. **182**(16): p. 4401-4405.
57. Omori, K., A. Idei, and H. Akatsuka, *Serratia* ATP-binding cassette protein exporter, Lip, recognizes a protein region upstream of the C terminus for specific secretion. Journal of Biological Chemistry, 2001. **276**(29): p. 27111-27119.
58. Fath, M.J., L.H. Zhang, J. Rush, and R. Kolter, Purification and characterization of colicin V from *Escherichia coli* culture supernatants. Biochemistry, 1994. **33**(22): p. 6911-6917.
59. Gilson, L., H.K. Mahanty, and R. Kolter, Four plasmid genes are required for colicin V synthesis, export, and immunity. Journal of bacteriology, 1987. **169**(6): p. 2466-2470.
60. Gilson, L., H.K. Mahanty, and R. Kolter, Genetic analysis of an MDR-like export system: the secretion of colicin V. The EMBO journal, 1990. **9**(12): p. 3875.
61. de Oliveira, J.E., C.R. Soares, C.N. Peroni, E. Gimbo, I.M. Camargo, L. Morganti, M.H. Bellini, R. Affonso, R.R. Arkaten, and P. Bartolini, High-yield purification of biosynthetic human growth hormone secreted in *Escherichia coli* periplasmic space. Journal of Chromatography A, 1999. **852**(2): p. 441-450.
62. Kamionka, M., Engineering of therapeutic proteins production in *Escherichia coli.* Current pharmaceutical biotechnology, 2011. **12**(2): p. 268.
63. Roifman, C.M., G.B. Mills, M. Chu, and E.W. Gelfand, Functional comparison of recombinant interleukin 2 (IL-2) with IL-2-containing preparations derived from cultured cells. Cellular immunology, 1985. **95**(1): p. 146-156.
64. Reed, B. and R. Chen, Biotechnological applications of bacterial protein secretion: From therapeutics to biofuel production. Research in microbiology, 2013. **164**(6): p. 675-682.
65. Berla, B.M., R. Saha, C.M. Immethun, C.D. Maranas, T.S. Moon, and H.B. Pakrasi, Synthetic biology of cyanobacteria: unique challenges and opportunities. Frontiers in microbiology, 2013. **4**.
66. Ducat, D.C., J.C. Way, and P.A. Silver, Engineering cyanobacteria to generate high-value products. Trends in biotechnology, 2011. **29**(2): p. 95-103.
67. Lindblad, P., P. Lindberg, P. Oliveira, K. Stensjö, and T. Heidorn, Design, engineering, and construction of photosynthetic microbial cell factories for renewable solar fuel production. Ambio, 2012. **41**(2): p. 163-168.
68. Murry, M.A. and C.P. Wolk, Evidence that the barrier to the penetration of oxygen into heterocysts depends upon two layers of the cell envelope. Archives of microbiology, 1989. **151**(6): p. 469-474.
69. Kaneko, T., Y. Nakamura, C.P. Wolk, T. Kuritz, S. Sasamoto, A. Watanabe, M. Iriguchi, A. Ishikawa, K. Kawashima, and T. Kimura, Complete genomic sequence of the filamentous

nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. DNA research, 2001. **8**(5): p. 205-213.

70. Kumar, K., R.A. Mella-Herrera, and J.W. Golden, Cyanobacterial heterocysts. Cold Spring Harbor perspectives in biology, 2010. **2**(4): p. a000315.
71. Cardona, T., N. Battchikova, P. Zhang, K. Stensjö, E.-M. Aro, P. Lindblad, and A. Magnuson, Electron transfer protein complexes in the thylakoid membranes of heterocysts from the cyanobacterium *Nostoc punctiforme.* Biochimica et Biophysica Acta (BBA)-Bioenergetics, 2009. **1787**(4): p. 252-263.
72. Rippka, R., J. Deruelles, J.B. Waterbury, M. Herdman, and R.Y. Stanier, Generic assignments, strain histories and properties of pure cultures of Cyanobacteria. Microbiology, 1979. **111**(1): p. 1-61.
73. Meeks, J.C. and R.W. Castenholz, Growth and photosynthesis in an extreme thermophile, *Synechococcus lividus* (Cyanophyta). Archiv für Mikrobiologie, 1971. **78**(1): p. 25-41.
74. Tamagnini, P., O. Troshina, F. Oxelfelt, R. Salema, and P. Lindblad, Hydrogenases in *Nostoc* sp. Strain PCC 73102, a Strain Lacking a Bidirectional Enzyme. Applied and Environmental Microbiology, 1997. **63**(5): p. 1801-1807.
75. Sambrook, J. and W. Russell David, Molecular cloning: a laboratory manual. Vol. 3. 1989: Cold spring harbor laboratory press.
76. Hanahan, D., Studies on transformation of *Escherichia coli* with plasmids. Journal of Molecular Biology, 1983. **166**(4): p. 557-580.
77. Wolk, C.P., Y. Cai, L. Cardemil, E. Flores, B. Hohn, M. Murry, G. Schmetterer, B. Schrautemeier, and R. Wilson, Isolation and complementation of mutants of *Anabaena* sp. strain PCC 7120 unable to grow aerobically on dinitrogen. Journal of Bacteriology, 1988. **170**(3): p. 1239-1244.
78. Frias, J.E., E. Flores, and A. Herrero, Nitrate assimilation gene cluster from the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. Journal of bacteriology, 1997. **179**(2): p. 477-486.
79. Elhai, J. and C.P. Wolk, Conjugal transfer of DNA to cyanobacteria. Methods in enzymology, 1988. **167**: p. 747.
80. Elhai, J., A. Vepritskiy, A.M. Muro-Pastor, E. Flores, and C.P. Wolk, Reduction of conjugal transfer efficiency by three restriction activities of *Anabaena* sp. strain PCC 7120. Journal of Bacteriology, 1997. **179**(6): p. 1998-2005.
81. Laemmli, U.K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. nature, 1970. **227**(5259): p. 680-685.
82. Caccia, D., M. Dugo, M. Callari, and I. Bongarzone, Bioinformatics tools for secretome analysis. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2013. **1834**(11): p. 2442-2453.

83. Desplancq, D., C. Bernard, A. Sibler, B. Kieffer, L. Miguet, N. Potier, A. Van Dorsselaer, and E. Weiss, Combining inducible protein overexpression with NMR-grade triple isotope labeling in the cyanobacterium *Anabaena* sp. PCC 7120. Biotechniques, 2005. **39**(3): p. 405.
84. Nandakumar, M., A. Cheung, and M.R. Marten, Proteomic analysis of extracellular proteins from *Escherichia coli* W3110. Journal of proteome research, 2006. **5**(5): p. 1155-1161.
85. Shin, H.D. and R.R. Chen, Extracellular recombinant protein production from an *Escherichia coli* lpp deletion mutant. Biotechnology and bioengineering, 2008. **101**(6): p. 1288-1296.
86. Flemming, H.-C. and J. Wingender, The biofilm matrix. Nature Reviews Microbiology, 2010. **8**(9): p. 623-633.
87. Branda, S.S., Å. Vik, L. Friedman, and R. Kolter, Biofilms: the matrix revisited. Trends in microbiology, 2005. **13**(1): p. 20-26.
88. Whitchurch, C.B., T. Tolker-Nielsen, P.C. Ragas, and J.S. Mattick, Extracellular DNA required for bacterial biofilm formation. Science, 2002. **295**(5559): p. 1487-1487.
89. Biller, S.J., F. Schubotz, S.E. Roggensack, A.W. Thompson, R.E. Summons, and S.W. Chisholm, Bacterial vesicles in marine ecosystems. Science, 2014. **343**(6167): p. 183-186.
90. Gasteiger, E., C. Hoogland, A. Gattiker, M.R. Wilkins, R.D. Appel, and A. Bairoch, Protein identification and analysis tools on the ExPASy server. 2005: Springer.
91. Herrero, A., A.M. Muro-Pastor, and E. Flores, Nitrogen control in cyanobacteria. Journal of Bacteriology, 2001. **183**(2): p. 411-425.
92. Camargo, S., A. Valladares, K. Forchhammer, and A. Herrero, Effects of PipX on NtcA-dependent promoters and characterization of the cox3 promoter region in the heterocyst-forming cyanobacterium *Anabaena* sp. PCC 7120. FEBS letters, 2014. **588**(9): p. 1787-1794.
93. Binet, R., S. Létoffé, J.M. Ghigo, P. Delepelaire, and C. Wandersman, Protein secretion by Gram-negative bacterial ABC exporters–a review. Gene, 1997. **192**(1): p. 7-11.
94. Andersen, C., E. Koronakis, E. Bokma, J. Eswaran, D. Humphreys, C. Hughes, and V. Koronakis, Transition to the open state of the TolC periplasmic tunnel entrance. Proceedings of the National Academy of Sciences, 2002. **99**(17): p. 11103-11108.
95. Paschos, A., A. Bauer, A. Zimmermann, E. Zehelein, and A. Böck, HypF, a carbamoyl phosphate-converting enzyme involved in [NiFe] hydrogenase maturation. Journal of Biological Chemistry, 2002. **277**(51): p. 49945-49951.
96. Tamagnini, P., E. Leitão, P. Oliveira, D. Ferreira, F. Pinto, D.J. Harris, T. Heidorn, and P. Lindblad, Cyanobacterial hydrogenases: diversity, regulation and applications. FEMS microbiology reviews, 2007. **31**(6): p. 692-720.