Rita Sofia Cebola Rebelo Manuel

Bachelor in Cellular and Molecular Biology

Influence of Ferredoxin, Iron-Sulfur Carrier and Cobalamin Importer Overexpression on the Production of High-value Products by *Escherichia coli*

Dissertation for the obtention of a Master's degree in Biotechnology for Sustainability

Supervisor: Professor Gregory Bokinsky, Technische Universiteit Delft

Internal Supervisor: Professor Inês Cardoso Pereira, Instituto de Tecnologia Química e Biológica – Universidade Nova de Lisboa

September 2019

Rita Sofia Cebola Rebelo Manuel

Bachelor in Cellular and Molecular Biology

Influence of Ferredoxin, Iron-Sulfur Carrier and Cobalamin Importer Overexpression on the Production of High-value Products by *Escherichia coli*

Dissertation for the obtention of a Master's degree in Biotechnology for Sustainability

Supervisor

Professor Gregory Bokinsky, TU Delft

Jury:

President: Professor Margarida Oliveira, ITQB-NOVA

Main Opponent: Doctor Filipe Folgosa, ITQB-NOVA

Internal Supervisor: Professor Inês Cardoso Pereira, ITQB-NOVA

ITQB-NOVA

September 2019

COPYRIGHT

Influence of Ferredoxin, Iron-Sulfur Carrier and Cobalamin Importer Overexpression on the Production of High-value Products by *Escherichia coli*

A curiosidade, instinto de complexidade infinita, leva por um lado a escutar as portas e por outro a descobrir a América [...].

Eça de Queirós

ACKNOWLEDGMENTS

A special acknowledgment to my supervisor, Professor Gregory Bokinsky, for allowing me to work in his laboratory, for his support and for his help in this project. More than that, thank you for sharing with me your knowledge, for believing in my programming and laboratorial abilities and for making me feel welcome at TU Delft.

To Elena, my day-to-day laboratory partner. After seven months, it is with the greatest affection that more than my laboratory tutor, I won a friend. Thank you for all your tolerance, for all your teachings, for showing me that making mistakes is normal and is a way for personal and professional improvement, for passing me your stunning working pace (otherwise it would be impossible for me to finish this project), for all our conversations, our lunches and for welcome me with open arms as your tutee. I cannot thank you enough for everything you have done. I hope that one day you can see the colors of all countries in the map!

To all my laboratory colleagues, Margot, Nicole and Fehrat thanks ever so much. A special thanks to Helena, for obtaining preliminary data that gave rise to this project, and to Niels, for your patience and spent hours of work to measure all our samples, it would be impossible to finish my project without your devotion to this project.

A special acknowledgment also to my internal supervisor, Professor Inês Cardoso Pereira, for your availability to give me all the support and help during all the process of research and writing of this dissertation, whenever necessary.

Thank you, LST group, for receiving me as if I was one of you! For all the indoor and outdoor lunches, for all Professors and The Kurk Friday's nights, for all botanical's bar moments, for the Macumba parties, for the Christmas party and for all international food dinners. Especially to Álvaro, Melina, Ricardo, Timmy and Venda for becoming true international best friends that I do not want to lose. During my seven months of staying in the cloudy and rainy Netherlands, you made me feel at home. You supported me, cheered me up when I was in a bad mood, made me laugh like the world was ending, allowed me to try the best brownies ever (thanks to Ricardo, for baking them for us!) and provided me the best weekend ever during those seven months. Thank you for being my family in Delft when mine was thousands of kilometers apart, you really did a great job!

Um especial obrigada também a todos os meus amigos e amigas que me acompanharam ao longo de todo ou parte do meu percurso académico e que sei que me continuarão a acompanhar. É com um enorme carinho e amizade que vos agradeço, por estarem ao meu lado em todos os momentos bons, maus e assim-assim, não só a nível académico, mas também pessoal. Seria impossível agradecer individualmente e referir o que cada um já fez por mim ao longo da nossa amizade, pelo que o faço de uma forma generalizada, sabendo que cada um de vós percebe o quanto é importante para mim. Obrigada a vós também por serem a melhor segunda família que poderia ter escolhido, ficar-vos-ei eternamente grata por tudo.

Obrigada a ti, Miguel, por me teres apoiado e teres permanecido ao meu lado durante toda esta minha jornada. Pelo amor, pela amizade, pela paciência e por toda a motivação que sempre me deste. Por acreditares em mim, mesmo quando eu não o fazia. Um enorme obrigada, com todo o carinho que sinto por ti.

Por último, e provavelmente o mais importante, um gigante e especial obrigada a toda a minha família por todo o apoio que me deram e orgulho com que encaram todos os meus sucessos do dia-a-dia. À minha avó Carolina que, apesar de já não estar presente, representa um dos grandes pilares que suportam a pessoa que sou hoje. Aos meus avós, Cruz e Dionísio, por todo o carinho, preocupação, conselhos e pelo vosso abraço que, onde quer que esteja, me faz sentir em casa. À minha tia, Carla, por todas as gargalhadas, todas as conversas e, acima de tudo, por caminhar sempre a meu lado em todas as fases da minha vida e por me ensinar que por muito mal que tudo pareça estar, serei sempre mais forte do que penso ser. À minha irmã, Beatriz, que apesar das discussões que possamos ter, não há maior ligação inexplicável do que a que temos. Obrigada por todas as recordações que ficarão para sempre guardadas em mim, pelo teu apoio, pelo teu amor, por me defenderes sempre e por ires à luta comigo seja em que situação for. Aos meus pais, Lina e Paulo, que penso que nunca existirão palavras suficientes para vos agradecer por tudo o que fizeram e fazem por mim. Um milhão de obrigadas por vos ter ao meu lado em todos os momentos da minha vida, pelo vosso apoio incondicional, pelas vossas palavras de coragem, pelo vosso espírito de sacrifício que sempre demonstraram para que nada me faltasse, por viverem comigo todas as minhas conquistas e todas as minhas derrotas, por me incutirem todos os valores e moral que hoje tenho e por me deixarem ganhar asas e voar. Pela vossa incessante e difícil luta diária para me verem feliz, um eterno obrigada.

RESUMO

Um dos problemas na produção de antibióticos é a dificuldade da sua síntese ou baixo rendimento em laboratório, no caso de moléculas biológicas complexas. Têm vias de formação complexas, onde enzimas SAM dependentes de cobalamina representam um importante papel, contendo clusters de Fe-S na sua estrutura e necessitando de transportadores secundários de eletrões para a sua atividade, tornando difícil o seu melhoramento. Exemplos são fosfomicina (enzima Fom3), antibiótico de largo espectro, e tiostrepton (enzima TsrM), antibiótico contra Gram-positivos com propriedades anticancerígenas, ambos produzidos naturalmente por *Streptomyces sp.*.

O principal objetivo desta dissertação constituiu o desenvolvimento de um método para produzi-los em *Escherichia coli* e dar importância à possibilidade de melhoramento da produção através da sobrexpressão de ferredoxinas (proteínas de transferência secundária de eletrões), de proteínas de montagem/suporte de Fe-S clusters (SufU e SufT) e importadores de cobalamina (operão *btu*). Para isso, foi criado um programa bioinformático de procura de ferredoxinas de *Streptomyces cattleya* e *Streptomyces laurentii* e dos seus genes adjacentes em um passo. Foram escolhidas três ferredoxinas de cada espécie. Seguiu-se a sobrexpressão em *Escherichia coli* das enzimas Fom e TsrM juntamente com Btu, SufUT e as ferredoxinas escolhidas. Após produção em meio MOPS com adição de cobalamina, as amostras foram preparadas para medição em LC-MS. Usando o presente método, tiostrepton foi produzido com sucesso em *E. coli,* o que não aconteceu com fosfomicina. Os resultados mostraram que a sobrexpressão de TsrM juntamente com ferredoxinas, SufUT e proteínas Btu e adição de cobalamina no meio são essenciais para melhorar a via, e as ferredoxinas de *Streptomyces laurentii* mostraram-se mais eficazes que as de *Streptomyces cattleya*.

Este estudo representa os passos iniciais para melhoramento de vias contendo enzimas SAM dependentes de cobalamina, bem como um possível início para estudos futuros cujo objetivo possa abordar o problema mundial associado aos antibióticos.

Palavras-Chave: *Enzimas SAM Dependentes de Cobalamina, Ferredoxinas, Proteínas SufUT, Proteínas Btu, Antibióticos.*

ABSTRACT

One of the problems in antibiotics production is the difficulty of their synthesis or low yield in the case of complex biological molecules. They have complex biosynthetic pathways, where cobalamin-dependent SAM enzymes play a pivotal role, with an Fe-S cluster in their structure and the necessity of secondary electron transfer proteins for their activity, making the production improvement difficult. An example of these compounds are fosfomycin (Fom3 enzyme), a broad-spectrum antibiotic, and thiostrepton (TsrM enzyme), a Gram-positive antibiotic with anticancer properties, both naturally produced by *Streptomyces sp.*.

The main objective of this dissertation was the development of a method to produce these antibiotics in *Escherichia coli* and highlight the possibility of improving production through the overexpression of ferredoxins as secondary electron transfer proteins, Fe-S cluster scaffold/assembly proteins (SufU and SufT) and cobalamin importers (*btu* operon). To address this goal, a bioinformatic program was constructed to search for *Streptomyces cattleya* and *Streptomyces laurentii* ferredoxins and their close genomic neighbors in one step. As a result, three ferredoxins of each species were chosen. This was followed by the overexpression of Fom proteins and TsrM along with Btu, SufUT and the chosen ferredoxins in *Escherichia coli*. After production using MOPS medium with cobalamin addition, samples were prepared and measured in LC-MS. Using the presented method, thiostrepton was successfully produced in *E. coli*, though this was not the case for fosfomycin. Additionally, the results showed that overexpression of TsrM together with ferredoxins, SufUT and Btu proteins along with cobalamin addition to the medium were essential to improve the synthetic pathway, and *Streptomyces laurentii* ferredoxins showed to be more effective than *Streptomyces cattleya*.

This study represents initial steps for cobalamin-dependent SAM radical enzymes pathways improvement and a possible beginning for further studies which objectives may be addressing the antibiotics problem our world is facing.

Keywords: *Cobalamin-dependent SAM enzymes, Ferredoxin, SufUT proteins, Btu proteins, Antibiotics.*

TABLE OF CONTENTS

FIGURES INDEX

TABLES INDEX

ABREVIATIONS, ACRONYMS AND SYMBOLS

2MTryp – 2-Methyl Tryptophan

5'-dA – 5'-deoxyadenosyl

ATc – Anhydrotetracycline

E° – Reduction Potential

Fom3 – 2-hydroxyethylphosphonate Methyltransferase

FNR – Ferredoxin–NADPH Reductase

Frw – Forward

GFP – Green Fluorescent Protein

HEP – 2-hydroxyethylphosphonate

HEP-CMP – cytidylyl-2-hydroxyethylphosphonate

HPP – 2-hydroxypropylphosphonate

HPP-CMP – cytidylyl-2-hydroxypropylphosphonate

IPTG – Isopropyl β-D-1-tiogalactoside

ISC – Iron-Sulfur Cluster

LC-MS – Liquid Chromatography – Mass Spectrometry

MOPS – 3-(N-morpholino)propanesulfonic Acid

$NADP^+$/ NADPH – Nicotinamide Adenine Dinucleotide Phosphate

NCBI – National Center of Biotechnology Information

NIF – Nitrogen Fixation

PEP – Phosphoenolpyruvate

PnAA – Phosphonoacetaldehyde

PnPy – Phosphonopyruvate

RAST – Rapid Annotation using Subsystem Technology

Rev – Reverse

RFP – Red Fluorescent Protein

RiPP – Ribossomally synthesized and Post-translationally modified Peptide

SAH – S-adenosyl-L-homocysteine

SAM – S-adenosylmethionine

SUF – Sulfur Assimilation

TsrM - Tryptophan 2-C-methyltransferase

1 INTRODUCTION

1.1 ANTIBIOTICS PRODUCTION: A NEW ERA PROBLEM

In the world we are living, the necessity for using antibiotics in bacterial infections is arising, increasing the appearance of antibiotic-resistant bacteria. Discovering new and more straightforward ways to produce antibiotics is becoming a research hot topic. The uncultivable bacteria are an unlimited resource of new antibiotics, but the impossibility of growing them in laboratory conditions limits the production. Thus, there is a wide range of antibiotics which production nowadays is solely possible via chemical synthesis.

A few years ago, it was believed that the production of biological compounds could only be accomplished by biological processes. Nowadays, chemical production of biological compounds has been gaining attention and impact in pharmaceutical industry. However, the chemical synthesis of biological compounds represents a massive threat to the environment, health and society in general. Furthermore, this process is characterized by very low production efficiency. The main drawbacks when it comes to this process are 1) a vast amount of noxious solvents and chemical waste that are released to the environment (causing severe health and environmental problems); 2) the feedstock used is originated from nonrenewable sources as petroleum, which abundance have been dwindling over the past decades; 3) The incredible low concentration of some complex products compared to reagents used and side products generated, making its production not economically viable; 4) The purification process due to the difficulty to select and separate the product from the mixtures; 5) The addition of several steps to the original biological pathway, increasing the energy and water expenditure and, by consequence, the economic costs (1).

On the other hand, a more ecofriendly production can be achieved importing biological pathways from non-laboratory cultivable producers into laboratory-grown bacteria. Nevertheless, this strategy is still impaired due to the low production yields. To overcome this problem it is necessary to have broad and deep knowledge of the specific pathway and its genomic context in order to identify bottlenecks, which steps may be synthetically modified (overexpression, repression, among others) or cofactors that may be necessary to add to the medium (2). Moreover, the use of well characterized or easy growing organisms as *Escherichia coli,* in which plenty of genetic tools are available, could contribute to the development or tuning of these pathways.

Taking this into account, the main objective of this study focuses on the production in laboratorial environment of antibiotics (fosfomycin and thiostrepton) produced by non-easy growing bacteria (*Streptomyces sp.)*, as well as the improvement of the production rates. Thus, the considered hypothesis is the possibility of incorporating antibiotic pathways into *Escherichia*

coli (*E. coli*) and boosting their production. Furthermore, when testing the production pathways together with proteins/compounds necessary for antibiotic production (Fe-S cluster scaffold and transport proteins, secondary electron transfer proteins and external chemical groups donors), it would be possible to increase this production.

1.2 METALLOPROTEINS: A BRIEF INTRODUCTION INTO Fe-S CLUSTER PROTEINS

1.2.1 GENERAL CHARACTERISTICS

Some of the most important energetic processes in living cells, such as respiration and photosynthesis pathways, require the reduction and oxidation of some compounds, being necessary the occurrence of electron transfers for the normal functioning of a cell. These bioprocesses (processes using living cells or specific cellular components) require specific electron transfer centers for which metalloproteins are widely used (3,4). Metalloproteins are defined as proteins which normal function directly depends on the presence of metals, either as a member of a prosthetic group or as co-factors. Although the majority of metals are known to be redox-active, living systems use a low number of these elements for electron transfer processes (4).

The three main metalloprotein classes capable of covering all reduction potential ranges in biology are cupredoxins (copper center) (4,5), cytochromes (heme/ferrous center)(3–5) and iron-sulfur (Fe-S) proteins (iron center) (3,4). The reduction potential (E°) is considered to be the tendency of a metal center to be reduced: higher E° centers are oxidants, while lower E° centers are reductants (4). It is impossible for a single protein to operate within the entire range of biologic E°. Cupredoxins, cytochromes and Fe-S proteins together are able to do it, because their E° differ from each group. Cupredoxins function at high E°, Fe-S proteins function at low E° and cytochromes have a medium E°, overlapping (almost completely) with higher E° Fe-S cluster proteins and lower E° cupredoxins (**Figure 1.1**) (3,4).

Amongst all metalloproteins, Fe-S cluster proteins are considered to be the most versatile class of electron transfer proteins, in order that they function with a wide range of E°, as seen in ***Figure 1.1***, being capable of taking part in a wide variety of bioprocesses (3,4).

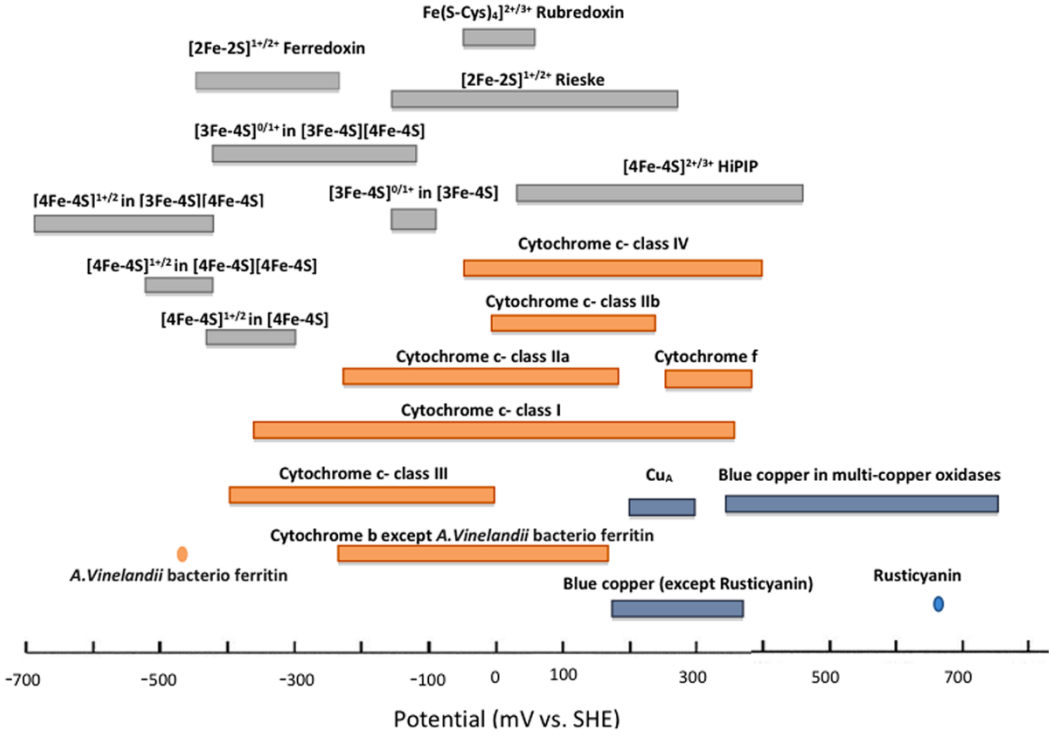

Figure 1.1 *- Reduction potential range of Fe-S cluster proteins, Cytochromes and Cupredoxins in biological electron transfer processes. Adapted from Liu, J. et al, 2014.*

The existence of Fe-S cluster proteins has been established and studied since the 1960's and, as their name indicate, contain a Fe-S cluster bound to their structure. The main types of existent Fe-S clusters are [2Fe-2S] clusters, [3Fe-4S] clusters and [4Fe-4S] (6,7). Cysteines are responsible for the coordination of the Fe-S cluster, and it is possible to identify the protein binding motif in specific C_X pattern sequences (8,9). In [4Fe-4S] cluster proteins, three close cysteines ($C_{XXX}C_{XX}C$) plus a distant one bind the cluster. This causes a tetrahedral pattern of cysteines and a distorted cubic form of the cluster whereas iron and sulfur atoms occupy alternatively vertices positions (***Figure 1.2***). Moreover, each iron is coordinated by three inorganic sulfurs and one thiol group from the closest cysteine (9,10).

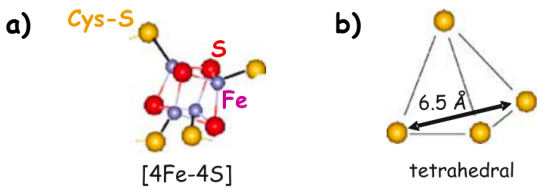

Figure 1.2 *– a) Distorted cubic form of Fe-S Cluster binding. Yellow spheres: cysteine binding residue; red spheres: sulfur atoms; purple spheres: iron atoms. b) Tetrahedral form acquired by iron binding cysteines. Adapted from Meyer, J., 2008.*

1.2.2 GENERAL FUNCTIONS AND Fe-S CLUSTERS EVOLUTION

Fe-S clusters are directly involved in important pathways in a living cell. They are involved in respiration, synthesis of amino acids, regulation of gene expression (11,12), modification/repair of DNA and RNA (11,13,14) and biosynthesis of complex products, essentially antibiotics such as fosfomycin (15) and thiostrepton (16), studied on this project. Thus, the biological interest of Fe-S cluster proteins does not just rely on their reduction potential and structure, but also on the number of functions that these clusters have in nature.

Due to the important functions Fe-S clusters have, they are active in a wide range of living species, being present in all kingdoms of life. In fact, Fe-S cluster proteins are considered to be one of the metalloproteins that first appeared on Earth (3). Due to the high abundance of iron and sulfur in Earth's early atmosphere, it was possible that spontaneous binding of the two elements into clusters occurred, essentially [4Fe-4S]. In its turn, the cluster-protein assembly is also believed to have occurred spontaneously, similarly to the cluster's formation (3,17). Thus, early living systems possibly capitalized these clusters redox features to use them as redox reactions centers. Even when the Earth atmosphere turned out into a more oxidizing medium, Fe-S proteins already established Fe-S clusters as redox centers and continued to use them in their established functions. For instance, [4Fe-4S] clusters continue to be used by anaerobic bacteria to perform electron transfers, such as the reduction of H^+ as final acceptor of electrons (3,7).

Fe-S clusters were found to be able to accomplish secondary, within or between enzymes electron transfers, substrate binding/activation and structural framework for protein folding (6,12). These features are considered essential and critical for the activation of enzymes depending on these clusters, and must be taken into account when expressing an heterologous pathway (11,12). Concretely speaking about Fe-S clusters activity in biological pathways, they are able to accomplish all the previously mentioned features in a relatively wide range of essential or important processes within a living cell.

Despite the wide variety of functions that Fe-S clusters may accomplish, they are divided into classes, depending on specific features:

- Cluster type and structure by itself;

- Catalytic activity acquired by the cluster;

- Proteins which cluster binds to (same cluster type may bind to different proteins and protein domains of diverse sequences and structures);

- Other prosthetic groups present on the bound protein (6).

1.2.3 Fe-S CLUSTERS ASSEMBLY AND PROTEINS CLASSIFICATION

In the early life on Earth, Fe-S clusters assembly occurred spontaneously. The results obtained by Malkin and Rabinowitz (18), confirmed that it is possible to activate certain Fe-S clusters apoenzymes by simply adding ferric and sulfur ions to the medium. Nevertheless, the ionic concentrations needed for protein activation in this *in vitro* study were not compatible with iron and sulfur biological toxicity *in vivo*. This suggests the existence of a specific assembly and deliver system of iron and sulfur in nontoxic forms to Fe-S clusters apoproteins (12). The existence of this system was proved in the late 1980's, when Jacobson and colleagues discovered the existence of an operon (*nif*) which function is crucial to the assembly of the Fe-S cluster in *Azotobacter vinelandii* nitrogenase (19).

Up to date, at least three different assembly pathways are known: NIF (Nitrogen Fixation), ISC (Iron-Sulfur Cluster) and SUF (SUlFur assimilation) systems, all characterized by including cysteine desulfurases to obtain sulfur from cysteine, cluster scaffold proteins to sustain the formation of the cluster and a carrier to transport and transfer the Fe-S cluster to the final protein. In *Escherichia coli*, for instance, ISC and SUF assembly systems are present (3,12,20–22). Although both systems contain proteins with similar activity, ISC proteins function in clusters assembly of a wide range of proteins while SUF proteins synthesize clusters and operate under stress conditions (23). However, under laboratory growth conditions, SUF machinery is essential for cells viability, thus, being the pathway that this study will focus on (24).

SUF machinery for Fe-S cluster assembly is the most recent discovered system and it is encoded by the *sufABCDSE* operon (25). The existence of two complementary genes were discovered, *sufU* and *sufT*, which functions were thought to rely on scaffolding and maturation/assembly of Fe-S clusters respectively. However, in the last years, studies demonstrated that *sufU* function relies essentially on the transfer of sulfur to the system, being extremely important on Fe-S cluster assembly. Its scaffolding function was not discarded, but the sulfur transfer function seems to be more viable (26–29). SufU and SufT will be used in this study to test the possibility of increasing concentration products in pathways containing Fe-S cluster proteins.

After understanding Fe-S clusters characteristics, functions, evolution and assembly, it is important to know how the proteins containing them are classified. According to the International Union of Biochemistry (IUB), Fe-S cluster proteins can be divided in two main distinct groups:

- Simple Fe-S cluster proteins. Have only Fe-S clusters and includes proteins which function is exclusively catalytic (*e.g.* transfer of electrons). Subdivided in rubredoxins, ferredoxins and other simple proteins.

- Complex Fe-S cluster proteins. Contain other prosthetic groups and generally execute enzymatic functions. Subdivided in flavoproteins, molybdenum proteins and other complex proteins, such as S-adenosylmethionine (SAM) radical enzymes (30).

1.2.4 FERREDOXINS AND SAM RADICAL ENZYMES

Ferredoxins are a group of simple Fe-S cluster proteins that are considered secondary electron transfer proteins, with no enzymatic activity. This group of proteins was discovered by Mortenson et al. in 1962 (31) and have been extensively studied due to their key role in redox reactions. The group comprises small and low molar mass proteins that contain Fe-S clusters, acting exclusively as secondary electron carriers in a wide range of essential and non essential biological pathways. This occurs due to ferredoxins capacity to be reduced by their electron transfer partners, such as NADPH, through Ferredoxin–NADPH reductase (FNR) action, and their capacity to transfer the received electron to the final acceptors (32). Other characteristics of ferredoxins are their brown color after purification and the fact that they are highly acidic and extremely stable proteins that play an essential role in the normal functioning of living cells (3,6,31,33,34). These proteins are present in a wide range of organisms in all Kingdoms of life. Furthermore, ferredoxins are ancient proteins and their emergence occurred in the early stages of life on Earth (10,33).

On the order hand, SAM radical enzymes are a very well studied Fe-S cluster superfamily characterized by using S-adenosylmethionine (SAM) in a radical mechanism. Initially, 650 SAM radical enzymes were discovered using bioinformatic techniques (13). Nowadays, this number is around 150-fold higher (13,35). SAM radical enzymes cleave SAM into a methionine and a 5'-deoxyadenosyl (5'-dA) radical, a very powerful oxidant, using the energy provided by an electron from another compatible Fe-S cluster protein (secondary electron transfer, for instance ferredoxins). They are usually composed of two or more domains (e.g. B12 vitamin dependent enzymes have in their structure an additional cobalamin binding domain to the SAM domain). Similar to ferredoxins, SAM radical enzymes are ancient proteins from early life on Earth, present in all Kingdoms of life, being abundant in bacteria and archaea (13,36–38).

In ***Figure 1.3***, a schematic perspective of most of the topics discussed previously is presented.

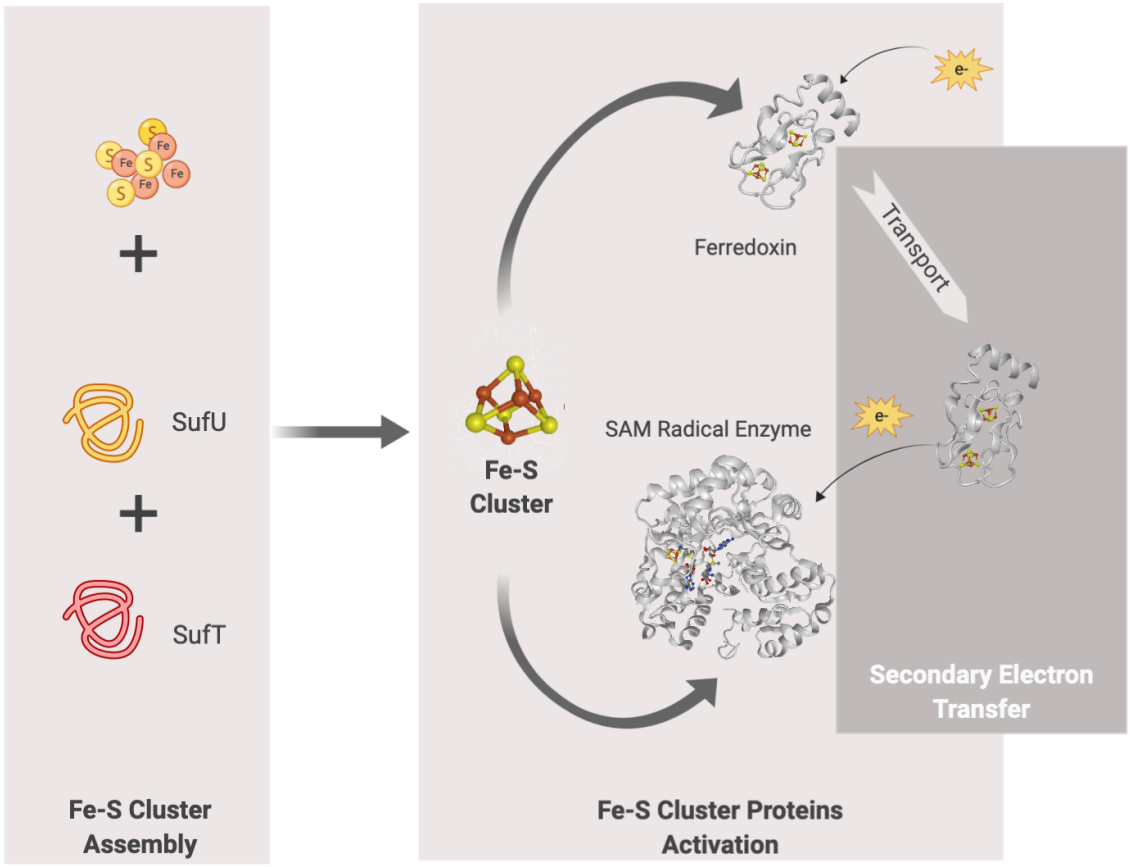

Figure 1.3 *- General scheme of Fe-S Cluster Proteins Activation. All colors used are meaningless. Proteins SufU and SufT icons are merely indicative. The structure of the general ferredoxin used represents an* E. coli *ferredoxin. The structure of the general SAM Radical Enzyme represents the* E. coli *Coproporphyrinogen III oxidase protein (also SAM Radical Enzyme). Scheme created using Biorender.*

2 BIOINFORMATICAL APPROACH TO GENOMIC CONTEXT

2.1 INTRODUCTION

Nowadays, bioinformatics is an extremely versatile area of Biology in the way that it can be used simply to analyze obtained data or to organize the original data, making discoveries. Although oftenly these discoveries need to be proven by laboratory work, bioinformatical discoveries drive us in the right direction, saving time, resources and experimental work. Some of the most used and important bioinformatical features are the annotation of genomic sequences, understanding the genomic context of certain genes, prediction of proteins structure and domains, analysis of sequencing data, construction of phylogenetic trees, and upload and organization of information in databases, being programming widely used in all of them (39). This work will be focused on the first two features.

Genomic annotation contains essential information for the researcher to use. Some important information present on annotations is data source (reference information) and some characteristics of the sequence (interpretative information), such as the function of coding sequences and respective protein names (other information). Annotation is thus a very important bioinformatics work in the sense that it represents the construction of an “identity card” to sequences. However, annotation work is also high-specialized, labor-intensive and highly dependent on computer programs. A quality annotation is accomplished by specialized personnel that spends a lot of time reanalyzing the data from sequences analysis with several computer programs (39).

An example of a fully automated annotation online service is the RAST (Rapid Annotation using Subsystem Technology) Server (National Institute of Allergy and Infectious Diseases and National Science Foundation). This program annotates partially or completely archaeal and bacterial genomes. Despite being fully automated, the program has a very complex annotation protocol. It starts by identifying rRNA and tRNA genes to exclude them from the protein-coding genes. The program will then allocate functions to the remaining genes and foretell the possible subsystems present in the genome, being subsystems considered as a group of abstract functions defined by an expert (*e.g.,* the subsystem for glycolysis is composed by the group of functional roles of the proteins that make part of the metabolic pathway). After these steps are completed, this information is used to predict the organism's metabolic network, completing the subsystem-based annotation. While subsystem-based assertions are being constructed, nonsubsystem-based annotation is made using the data analysis from several tools, not taking into account functional roles. Thus, when both annotations are done, the results are crossed to

originate the final annotation result. The complete annotation is a relatively expeditious process, being ready to download by the user after 12-24 hours after data submission (40).

As previously referred, programming is highly related to all bioinformatics features, being important to understand simple concepts related to it. First of all, it is important to understand what a program is. A program is a group of orders, written in a specific format so that they match with the computer basic operations and be possible to be executed. The written orders that compound the program are called a script (39). There are a wide range of programming languages to be used, for instance R programming language, used in this project. R is a programming language essentially used for statistics and graphical data organization/presentation. It is relatively simple and intuitive to use, and it is a free online available software inserted on the GNU Project. It is also widely used in biological researches.

Regarding bacterial genome context, this study was focused mainly on the synteny between genes. Synteny is defined as the position of several homologous genes that tend to be near each other in different genomes, implying the possible similarity in function or the necessity of acting side by side in the same pathway (41). Once working with synthetically modified metabolic pathways, it is crucial to know the individual genes and their genomic neighborhood to predict co-functions and functions complementation. The best way of analyzing the genomic context of a gene is by using bioinformatics. In this study, synteny is going to be used to predict the best ferredoxins to choose. Non-isolated ferredoxins may be considered to have an influence in specific pathways from where their neighbor genes (transcribed from the same strand) make part. However, if there are isolated ferredoxins, these may have a more general role, such as secondary electron transfer in a miscellaneous of pathways where the ones in this study may be part of (fosfomycin and thiostrepton production pathways).

In the past few years, some bioinformatics tools/algorithms have been developed and studied. Two very recent examples are the algorithm Syntenizer 3000 (2019) and the tool SynerClust (2018), both open-source. Although both are very fast, both are built for bioinformaticians or specialized personnel in programming (41,42).

Therefore, the need for a simpler, user-friendly program that can return the genomic neighborhood in a single-step arose. The main goal tackled in this chapter is the creation of an R script (simple language, relatively simple to use) that predicts the close genomic context (the genes immediately before and after) of a specific gene from a RAST annotated genome.

2.2 MATERIALS AND METHODS

The genomic context program was written via the programming language R (GNU Project, The R Foundation), using the integrated development environment RStudio Desktop Open Source Edition (RStudio). The script has two main functions. Firstly, it queries a RAST annotated

genome with user-specified keywords and sequence lengths. Then it displays the neighboring genes of the genes that match the query (close genomic context tool). The keyword used in this work was “ferredoxin” to find all ferredoxin genes present in the annotated genomes of *Streptomyces cattleya* (*S. cattleya*) and *Streptomyces laurentii* (*S. laurentii*), separately. The main packages, commands and algorithms used in the program are presented in ***Table 2.1***, as well as a short description of their direct action (consulted in the RStudio help section). After running the program, the resulting ferredoxins were compared with other ferredoxins reported in NCBI annotated genomes.

Table 2.1 *- Main Packages, Commands and Algorithms used on the program and their descriptions, divided by type and the package where they belong (if applicable).*

Name	Type	Package	Description
berryFunctions	Package		Functions essentially related to plots and hydrology.
readxl	Package		Functions to import and read .xls format files
abs	Command	*base*	Return the absolute value of an object
as.data.frame	Command	*base*	Coerce to a Data Frame
as.numeric	Command	*base*	Coerce to Numeric Objects
c	Command	*base*	Concatenate a group of elements into a vector/list.
data.frame	Command	*base*	Creates a Data Frame
for loop	Algorithm	*base*	Control Flow. For a variable in an object length, apply a specific action
grep	Command	*base*	Search for matching patterns
if loop	Algorithm	*base*	Control Flow. If the condition happens in an object, apply a specific action
insertRows	Command	*berryFunctions*	Inserts Rows to a Data Frame
install.packages	Command	*utils*	Download and install packages
is.na	Command	*base*	Select missing elements
length	Command	*base*	Set an object length
library	Command	*base*	Load installed packages
nrow	Command	*base*	Return the number of lines of an object
numeric	Command	*base*	Create numeric objects

rownames	Command	*base*	Retrieve/Set the name of the rows of a matrix
sort	Command	*base*	Order the elements of a vector
table	Command	*base*	Create a table
unique	Command	*base*	Remove duplicated elements
vector	Command	*base*	Create a vector
while loop	Command	*base*	Control Flow. While the elements of an object fit the condition, a specific action is applied
write.table	Command	*utils*	Create a file with the data on the working directory

2.3 RESULTS AND DISCUSSION

During the preparation of this work, finding specific proteins and analyzing their genomic context in order to choose the three ferredoxins from *Streptomyces cattleya* and *Streptomyces laurentii* to be used was challenging. Thanks to this difficult and time-consuming process, a program was created. This software overcomes both problems: it is able to find all proteins containing a specific keyword in their annotation, with a specific length of nucleotide sequence, and it can show the genes located on the same strand immediately before and after. The program script is presented in **Appendix I**.

After running the written program, an Excel file containing the matches was produced for each studied species. Each file was composed of several columns, such as start and stop position, function, nucleotide/amino acids sequence and nucleotide sequence length.

To test the functionality of the program, the obtained data were compared with a search of ferredoxins in *S. cattleya* and *S. laurentii* NCBI annotated genomes. Both files are shown in ***Figure 2.1*** (colors are not a direct result of the program since they were added after making the comparison between the results and NCBI search).

A – *Streptomyces cattleya*

location	strand	function.	length
NC_017585.1_253156_252959	-	Ferredoxin	197
NA	NA	NA	NA
NC_017585.1_309107_309301	+	Flavodoxin reductases (ferredoxin-NADPH reductases) family 1	194
NC_017585.1_309464_310696	+	putative lipoprotein	1232
NA	NA	NA	NA
NC_017585.1_758942_759619	+	Transcriptional regulator, AcrR family	677
NC_017585.1_759829_760053	+	Ferredoxin	224
NC_017585.1_760043_761437	+	Ferredoxin reductase	1394
NA	NA	NA	NA
NC_017585.1_840256_841017	+	hypothetical protein	761
NC_017585.1_841123_841323	+	Ferredoxin	200
NA	NA	NA	NA
NC_017585.1_1036276_1036947	+	Transcriptional regulator, AcrR family	671
NC_017585.1_1036944_1037144	+	Ferredoxin	200
NA	NA	NA	NA
NC_017585.1_1507529_1507837	+	Flavodoxin reductases (ferredoxin-NADPH reductases) family 1; Vanillate O-demethylase oxidoreductase (EC 1.14.13.-)	308
NA	NA	NA	NA
NC_017586.1_370564_369353	-	Putative cytochrome P450 hydroxylase	1211
NC_017586.1_370819_370616	-	Ferredoxin **(FERREDOXIN 3)**	203
NA	NA	NA	NA
NC_017586.1_1176600_1175836	-	Iron-sulfur cluster assembly ATPase protein SufC	764
NC_017586.1_1176925_1176608	-	Ferredoxin, 2Fe-2S	317
NC_017586.1_1178091_1176922	-	Iron-sulfur cluster assembly protein SufD	1169
NA	NA	NA	NA
NC_017586.1_4272017_4272337	+	4Fe-4S ferredoxin, iron-sulfur binding **(FERREDOXIN 1)**	320
NC_017586.1_4272487_4273587	+	N-succinyl-L,L-diaminopimelate aminotransferase (EC 2.6.1.17), type 2	1100
NA	NA	NA	NA
NC_017586.1_5098900_5098187	-	Phosphoadenylyl-sulfate reductase [thioredoxin] (EC 1.8.4.8)	713
NC_017586.1_5099079_5098897	-	Ferredoxin-like protein involved in electron transfer (NCBI: Hypothetical Protein)	182
NC_017586.1_5100773_5099076	-	Ferredoxin--sulfite reductase, actinobacterial type (EC 1.8.7.1)	1697
NA	NA	NA	NA
NC_017586.1_5581268_5580963	-	putative ferredoxin **(FERREDOXIN 2)**	305
NA	NA	NA	NA
NC_017586.1_6217582_6216338	-	Ferredoxin reductase	1244
NC_017586.1_6217773_6217579	-	Ferredoxin	194
NC_017586.1_6219050_6217833	-	Putative cytochrome P450 hydroxylase	1217

B – *Streptomyces laurentii*

location	strand	function.	length
AP017424.1_33570_32272	-	Ferredoxin reductase	1298
AP017424.1_33767_33567	-	Ferredoxin	200
AP017424.1_34687_33764	-	probable iron-sulfur binding protein YPO1417	923
NA	NA	NA	NA
AP017424.1_53331_54521	+	Putative cytochrome P450 hydroxylase	1190
AP017424.1_54536_54727	+	Ferredoxin (**FERREDOXIN 1)**	191
NA	NA	NA	NA
AP017424.1_1264058_1264363	+	putative ferredoxin (**FERREDOXIN 2)**	305
NA	NA	NA	NA
AP017424.1_1528897_1528133	-	Iron-sulfur cluster assembly ATPase protein SufC	764
AP017424.1_1529222_1528905	-	Ferredoxin, 2Fe-2S	317
AP017424.1_1530403_1529222	-	Iron-sulfur cluster assembly protein SufD	1181
NA	NA	NA	NA
AP017424.1_1903774_1904661	+	hypothetical protein	887
AP017424.1_1904661_1905014	+	Flavodoxin reductases (ferredoxin-NADPH reductases) family 1; Vanillate O-demethylase oxidoreductase (EC 1.14.13.-) (NCBI: flavodoxin reductase family 1)	353
AP017424.1_1905053_1905757	+	Flavodoxin reductases (ferredoxin-NADPH reductases) family 1; Vanillate O-demethylase oxidoreductase (EC 1.14.13.-)	704
NA	NA	NA	NA
AP017424.1_2051904_2051671	-	putative ferredoxin	233
AP017424.1_2053376_2051904	-	Aldehyde dehydrogenase (EC 1.2.1.3)	1472
NA	NA	NA	NA
AP017424.1_3122748_3122110	-	Rieske (2Fe-2S) domain protein	638
AP017424.1_3123344_3122745	-	Quad-[4Fe-4S] ferredoxin, HycB/HydN/HyfA family (NCBI: cytochrome c nitrite reductase, Fe-S protein)	599
AP017424.1_3125650_3123353	-	Assimilatory nitrate reductase large subunit (EC 1.7.99.4)	2297
NA	NA	NA	NA
AP017424.1_3263672_3263995	+	Ferredoxin, 2Fe-2S (NCBI: hypothetical protein SLA_3110)	323
AP017424.1_3263995_3265305	+	Ferredoxin reductase	1310
AP017424.1_3265298_3265939	+	hypothetical protein	641
NA	NA	NA	NA
AP017424.1_5123012_5123332	+	4Fe-4S ferredoxin, iron-sulfur binding (**FERREDOXIN 3)**	320
AP017424.1_5123454_5124548	+	N-succinyl-L,L-diaminopimelate aminotransferase (EC 2.6.1.17), type 2	1094
NA	NA	NA	NA
AP017424.1_6327442_6326717	-	Phosphoadenylyl-sulfate reductase [thioredoxin] (EC 1.8.4.8)	725
AP017424.1_6327630_6327439	-	Ferredoxin-like protein involved in electron transfer (NCBI: hypothetical protein SLA_6014)	191
AP017424.1_6327758_6327627	-	hypothetical protein	131
AP017424.1_6329455_6327755	-	Ferredoxin--sulfite reductase, actinobacterial type (EC 1.8.7.1)	1700
AP017424.1_6330425_6329856	-	Acetyltransferase	569
NA	NA	NA	NA
AP017424.1_7657391_7656609	-	Thioesterase in siderophore biosynthesis gene cluster	782
AP017424.1_7657653_7657456	-	Ferredoxin	197
AP017424.1_7658463_7657714	-	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	749

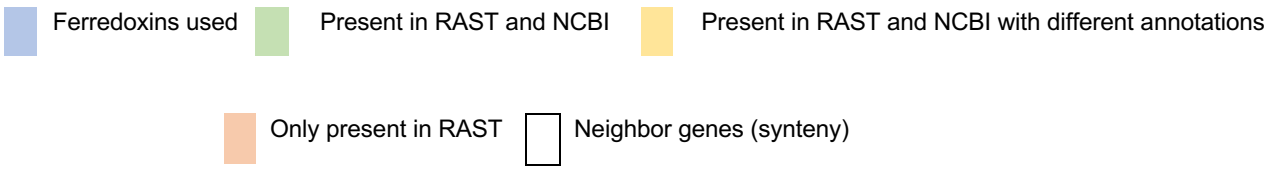

Figure 2.1 *- Final file resulting from running the program built and comparison of the obtained ferredoxins with NCBI database. Colors code was applied to both figures A and B. Nucleotide/amino acids sequences and start/stop positions columns not shown. NA lines represent empty lines for separating the synteny groups of ferredoxins.* ***A –*** *Results using* S. cattleya *RAST annotated genome.* ***B –*** *Results using* S. laurentii *RAST annotated genome.*

In both cases, the majority of identified proteins were also present in the NCBI database, which means that the proteins were similarly annotated in both databases. Despite this, there were also proteins that were not found in NCBI by direct search. They could only be found by using start and stop positions, as they were annotated differently. This is possible because each annotation program uses its own algorithms to find specific genes in a sequence. Another consequence of this effect was seen in the *S. cattleya* results, where the program found one protein annotated as ferredoxin by RAST that was not found in the NCBI database, even using the start and stop positions. Also, two proteins in NCBI search were found for *S. cattleya* annotated as ferredoxins that were found in RAST results but with a different annotation.

The results suggest that the program constitutes a simple and strong tool to find specific proteins in a genome, depending on their annotation. The fact that the program is dependent on annotation can be considered a disadvantage, because the results will be different when different annotation programs are used. On the other hand, this can also be advantageous, as users are free to choose the annotation program that best fits their needs. Since the genes' annotation is made by searching specific sequences of residues, the higher the similarity between the gene sequence and the consensus sequences, the higher the possibility of the protein to be what the program assumes it is (43). All proteins used were annotated as ferredoxins in both RAST and NCBI annotations, which suggests that the specific residues for ferredoxins (presence of a Fe-S cluster nearby cysteine residues (44)) are both strong and present in all proteins used. It is then possible to conclude and assume that the proteins used are ferredoxins, independently of the program used to annotate them.

2.4 CONCLUSIONS

The written program represents a useful, fast and user-friendly way to analyze an annotated genome regarding its close genomic context. The program searches for a specific keyword inserted by the user (*e.g.*, "ferredoxin") in an annotated genome (in this study, RAST annotated). Secondly, it returns the genes of interest containing the keyword in their function annotation, having less than the maximum of nucleotide length (user-modified feature), as well as both genes located immediately before and after in the genome. Restriction in the results upon length was important so that the genes presented could be mainly the genes of true ferredoxins, since there are several genes containing the keyword "ferredoxin" in their function annotation that are not ferredoxins (e.g. FNR genes). The length used (600 nucleotides) was based on Atkinson, et al. journal (45). Therefore, the program consists in a straightforward way of analyzing genomic neighborhoods, genes synteny (to infer the function of a protein) or even simple search for a specific gene, using a specific keyword and confined to a nucleotide sequence length.

The results show that the program was working, creating an Excel file with the several groups of three genes (ferredoxins of interest in the middle position), separated by an empty line for *S.*

cattleya and *S. laurentii* RAST annotated genomes. However, it was essential to understand the results. Both *S. cattleya* and *S. laurentii* final Excel files presented the majority of the ferredoxins annotated as “ferredoxin” in RAST results and NCBI database. On the other hand, some genes were annotated as “ferredoxin” in RAST, but not in NCBI (found by the first and last nucleotide positions), and one from *S. cattleya* was not present in NCBI, not even when searching for the start/stop positions. All ferredoxins selected and used on the study were present in both RAST and NCBI annotation, which is an indication of high probability of the protein being what the program assumes it is. The results show that the program is working correctly, in the sense that the results highly overlap with NCBI. Moreover, this program is highly dependent on the program used for genome automated annotation. This dependency does not represent a major problem, but it gives freedom for the users to use the automatic annotation program they prefer.

Thus, the main goal of this chapter consisting of the creation of a program in R capable of reading annotated genomes and returning genes of interest and their close neighbors was successfully achieved.

3 2-METHYL TRYPTOPHAN PRODUCTION IN ESCHERICHIA COLI

3.1 INTRODUCTION

Thiostrepton A is naturally produced by *Streptomyces sp.* organisms, firstly isolated from *Streptomyces azureus* in the 1950's (46) and later from *Streptomyces laurentii* in the 1970's (47). It is a potent antibiotic against Gram-positive bacteria, especially bacteria which treatment options are limited, for instance, *Staphylococcus aureus* (48). This antibiotic presents a growth-suppressive activity and functions as an inhibitor of elongation during translation by binding to the prokaryotic large ribosomal subunit (50S ribosome). In this manner, it blocks the interaction of the 50S ribosome with elongation factors and hampers translocation along mRNA (16,48–50). Furthermore, recent data suggest that thiostrepton A has anti-cancer activity by blocking the expression of a protein that causes the resistance of breast cancer cells to chemotherapy (51). Thiostrepton A is a Ribosomally synthesized and Post-translationally modified Peptide (RiPP), a very complex molecule made by ribosomes that requires extensive post-translation transformations to become active. Among the main enzymes involved in all these unusual post-translational modifications to obtain the final version of RiPPs are present cobalamin-depended radical SAM methyltransferases (52). These peptides are a class of naturally produced compounds that includes a wide range of antibiotics and bacterial toxins isolated from terrestrial and marine Gram-positive bacteria (53,54).

For the present study, the most relevant feature of the molecular structure of thiostrepton A is the presence of a tryptophan-derived quinaldic acid moiety in its structure. The further expansion of this moiety occurs due to C2 methylation of a tryptophan molecule (amino acid naturally produced by *E. coli*) on the initial steps of thiostrepton production pathway, creating 2-methyl tryptophan. The enzyme that catalyzes this methylation is Tryptophan 2-C-methyltransferase (TsrM), a cobalamin dependent SAM radical methylase (16,48,52,55–58). The molecular structure of thiostrepton is shown in ***Figure 3.1***, in which the added methyl group is highlighted in red.

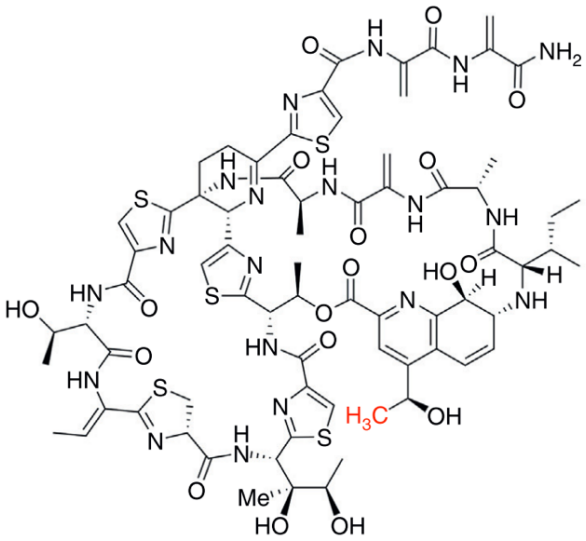

Figure 3.1 *- Molecular structure of Thiostrepton A. In red is represented the methyl group initially added to tryptophan by TsrM enzyme. Source: Fujimori, D. 2013.*

TsrM is a [4Fe-4S] enzyme that contains a cobalamin-binding domain and a SAM-binding domain, being dependent on the presence of methylcobalamin as intermediate methyl carrier. Previous studies claimed that an external electron transfer was not needed for the reaction (52). Nevertheless, Blaszczyk et al. showed that external electron transfer proteins increase TsrM activity (59).

The tryptophan methylation process is subdivided into two steps: the transfer of a methyl group from SAM to cob(I)alamin and the transfer of the same methyl group from methylcobalamin to tryptophan forming 2-methyl tryptophan. In contrast to several other SAM radical enzymes, TsrM uses SAM solely as a methyl donor and the release of a 5'-dA radical does not occur. First, cob(I)alamin is a supernucleophile that attacks the electrophilic methyl group of SAM. The methyl group from SAM is transferred to cob(I)alamin, generating methylcob(III)alamin and releasing SAH. Then, methylcob(III)alamin suffers a homolytic cleavage, creating a methyl radical that attacks C2 tryptophan, generating 2-methyl tryptophan and cob(II)alamin. This reaction arouses deprotonation, egressing an electron. Afterwards, this electron is used to replace cob(I)alamin in the cycle (59).

Nevertheless, the role of secondary electron transfer proteins in this pathway has yet to be elucidated. One hypothesis relies on the possibility of the released electron shifts to a secondary electron transfer protein that reduces the iron-sulfur cluster of TsrM and TsrM reduces cob(II)alamin to regenerate cob(I)alamin. This conjecture presents a possible function to TsrM [4Fe-4S] cluster, which remains indescribable. On the other hand, there is the possibility of occurring the electron transfer to the secondary electron transfer protein that reduces directly

cob(II)alamin to cob(I)alamin (59). These two possible functions of ferredoxins, as well as the 2-methyl tryptophan production pathway are presented in ***Figure 3.2***.

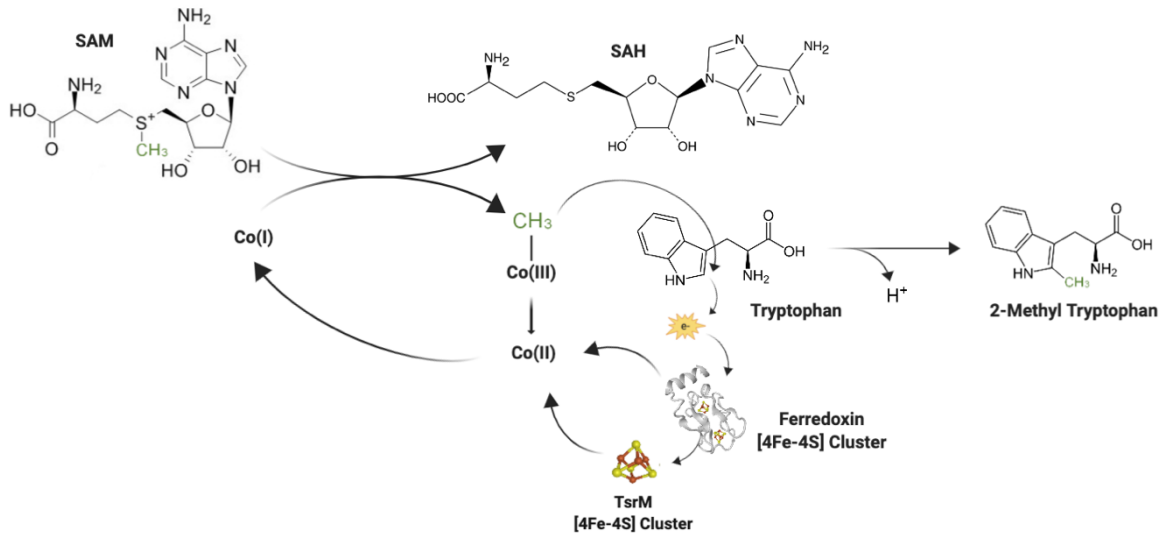

Figure 3.2 *- 2-methyl tryptophan production by TsrM enzyme. The methyl group used for tryptophan C2 methylation is presented in green. The colors of TsrM [4Fe-4S] cluster and ferredoxin [4Fe-4S] cluster are merely indicative. The structure used to represent a general ferredoxin is an* E. coli *ferredoxin. Scheme created using Biorender.*

Kuthning, A., et al. reported the *in vivo* heterologous production of several recombinant thiopeptides in *E.coli,* but not thiostrepton (60). On the other hand, TsrM has been overexpressed in *E. coli*, but its activity has been exclusively checked *in vitro*. The intracellular presence of cobalamin is essential for the activity of TsrM. For this reason, a previous research had combined the heterologous expression of TsrM and the *btu* operon (genes responsible for cobalamin uptake) in *E. coli*, even though the aim of that study was understanding the effects of *btu* operon overexpression in TsrM solubility (61). Moreover, a very recent *in vitro* study proved that the presence of secondary electron transfer proteins increases 2-methyl tryptophan yields, but the implications of TsrM and ferredoxins co-overexpression have yet to be demonstrated (59).

Thus, the existing studies can be applied to this project to test the improvement of SAM enzymes. The main goal of the present chapter is to improve 2-methyl tryptophan production by overexpressing ferredoxins, SufUT proteins, and Btu proteins (possible pathway bottlenecks) along with TsrM.

3.2 MATHERIALS AND METHODS

3.2.1 PRODUCTION VECTORS

The mature protein coding sequence of the tryptophan 2-C-methyltransferase (TsrM) from *Streptomyces laurentii* (GenBank: FJ652572.1) was codon-optimized for *E. coli* expression and synthetized by IDT. The optimized TsrM sequence was cloned into a pBbA5k BglBrick backbone to obtain the vector pBbA5K-TsrM (pTsrM). Similarly, electron transfer/Fe-S assembly vectors were constructed using the codon-optimized genes of the GenBank proteins ID AEW96347.1 (ScattFd1), AEW97532.1 (ScattFd2), AEW92689.1 (ScattFd3), CCB73802.1 (SufU) and AEW93447.1 (SufT) from *S. cattleya* NRRL 8057 = DSM 46488. The genes were assembled to create the synthetic operons ScattSufUT, Scatt3Fd and Scatt3Fd_SufUT, containing strong ribosome binding sites and cloned into the pBbS2c BglBrick backbone. To obtain the vectors pTsrM_ScattSufUT, pTsrM_Scatt3Fd and pTsrM_Scatt3Fd_SufUT, the operons were cloned between the BamHI and XhoI sites of pBbA5K-TsrM. The individual *S. cattleya* [4Fe-4S] ferredoxin genes were obtained by their amplification from the operon Scatt3Fd (see primers used in ***Table 3.1***), digested using the restriction enzymes BglII and XhoI and cloned into the pTsrM plasmid (BamHI and BglII cloning sites are compatible (62)). Similarly, vectors containing three different ferredoxins from *Streptomyces laurentii* were also constructed. The ferredoxin genes from *S. laurentii* were synthetically produced and codon-optimized for *E. coli* and used to obtain the vectors pTsrM_Slau3Fd, pTsrM_Slau3Fd_SufUT by restriction enzymes' cloning. Ferredoxins GenBank sequence ID of proteins used is BAU81007.1 (SlauFd1), BAU82127.1 (SlauFd2) and BAU85843.1 (SlauFd3). PhpK, another SAM enzyme, was used as a negative control (pPhpK).

Table 3.1 *- Primers and correspondent sequence (5'-3') used to obtain individual* S. cattleya *ferredoxins genes. Frw: Forward; Rev: Reverse*

Primer	**Sequence 5'-3'**
Frw_ScattFd1	CTAGAGCAGATCTATGACATACGTTATCGCTCAACCGT
Rev_ScattFd1	CTCGAGAAAGGATCCCTATCCATTCTGCGGTGGGAGGG
Frw_ScattFd2	CTAGAGCAGATCTATGTCAGATGCGACCGGCGAAG
Rev_ScattFd2	CTCGAGAAAGGATCCTTACGCTGCATCGGGTCCGTAAAC
Frw_ScattFd3	CTAGAGCAGATCTATGACCGTCCGGGTTTCAG
Rev_ScattFd3	CTAGAGCCTCGAGTTATGCGTCATGCACGG

The vector for cobalamin importers overexpression was already available in the laboratory. The genes for proteins btuB (AYG21241.1, GenBank), btuC (AYG19236.1, GenBank), btuD (AYG19238.1, GenBank), btuE (AYG19237.1, GenBank) and btuF (AYG20683.1, GenBank) were amplified from *E. coli* MG1655 genome and cloned into the backbone pBbS2c BglBrick backbone to create the vector pBbS2c-BtuCEDFB (pBtu).

Red Fluorescent Protein (RFP) gene was used as a negative control for Btu genes, cloned into the same backbone (pBbS2cRFP - pRFP).

3.2.2 2-METHYL TRYPTOPHAN PRODUCTION

2-Methyl Tryptophan was produced using the *E. coli* strains NCM3722 and BL21 (DE3) (the protocol was the same for both strains). *E. coli* cells were transformed with different combinations of the constructed vectors and grown overnight in Luria-Bertani Agar plates with chloramphenicol and kanamycin resistance, at 37°C.

Individual colonies were picked and inoculated for pre-cultures (triplicates) in 1 mL of SAM MOPS minimal medium (composition presented in ***Table 3.2***, all compounds used from Sigma-Aldrich), supplemented with 200 µL micronutrient stock (previously prepared in our laboratory), 1%(m/v) D-glucose (Sigma-Aldrich), 0.25%(m/v) Casamino Acids (Sigma-Aldrich, 22090) and 7 µM hydroxocobalamin, at 37°C, 250 r.p.m, O/N. For cobalamin concentration assays, were used medium variants containing cobalamin concentrations of 0 µM, 0.05 µM, 0.25 µM, 0.5 µM, 1 µM, 2 µM, 15 µM, 30 µM and 60 µM.

The cultures (triplicates) were obtained by inoculating 50 µL of pre-cultures in 3 mL of SAM MOPS minimal medium. The cultures were grown at 37°C, 250 r.p.m., until reaching an approximated OD_{600} of 0.05-0.1 for pBtu induction (83.3 ng/mL ATc) and continued to grow in the same conditions. When OD_{600} was approximately 0.5, pTsrM plasmids were induced (0.25 mM IPTG) and the medium supplemented with 0.15 mM cysteine and 32.5 µM $FeCl_3$.

Cultures were incubated at room temperature, 40 r.p.m. Samples were taken after 24h (1 ml and the equivalent of 1 ml sample with 0.5 OD_{600} samples). Samples were centrifuged, 15000 r.p.m., 2 min, and supernatants were transferred to new tubes. Pellets were quenched with NMM (methanol, acetonitrile and water, ratio of 5:3:1) + 0.1% Formic Acid and resuspended using vortex and sonication. Pellet samples were dried in speedvac at 40°C and resuspended with NMM for LC-MS measurement.

Table 3.2 *- Composition of the MOPS medium used for production assays and final concentration of each component. All reagents used from Sigma-Aldrich, Germany.*

Component	Final Concentration
MOPS	0.4 M
Tricine	0.04 M
NaCl	0.5 M
$MgCl_2 + 7H_2O$	5.23 mM
$CaCl_2$	0.005 mM
K_2SO_4	0.552 mM
$FeSO_4 + 7H_2O$	0.1 mM
$MnCl_2$	100 μM
NH_4Cl	28.5 mM
K_2HPO_4	1.32 mM

3.2.3 LC-MS MEASUREMENT/ANALYSIS

2-Methyl Tryptophan production was measured by an LC-MS system (Agilent) using an XBridge BEH Amide 2.5 μm (Waters, Bridge Columns) with a precolumn and equipped with a standard ESI source mass spectrometer (sample injection volume of 5 μL). The mobile phase was compound by two solvents, A (20 mM ammonium formate in 10% acetonitrile) and B (20 mM ammonium formate in 80% acetonitrile). After 6 min at 100% solvent B, the metabolites were separated by a gradient from 100% to 70% of solvent B for 6 min (flow rate 0.4 mL/min), followed by a gradient from 70% to 100% for 50 sec (same flow rate) and held at 100% solvent B for 3 min 10 sec. The 2-MTryp (positive polarity) precursor ion (219.1) was fragmented into product ions (144.1 and 128) using an ESI ionization in MRM mode. 2-MTryp concentration was estimated using a calibration curve constructed with standard samples. The values were corrected with the volume taken for sampling to obtain the final concentration of 2-MTryp. The average of the corrected values was calculated, and the error was considered to be the standard deviation of the triplicates.

An example of the calibration curves used is presented in **Appendix II**. A fresh calibration curve was prepared in each LC-MS run.

A general scheme of the methods used in this chapter is presented on ***Figure 3.3***.

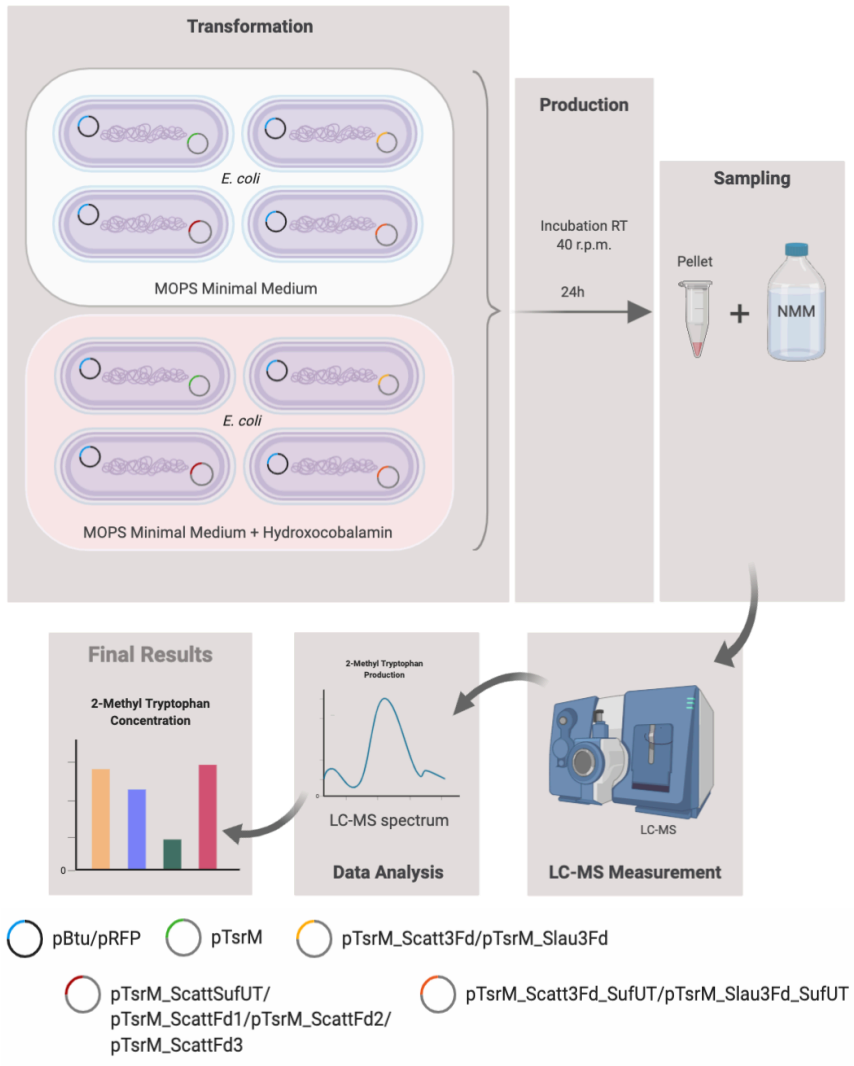

Figure 3.3 *- General scheme of the methodology used in chapter 3 of this dissertation. The spectrum and graph bars represented are merely indicative. Scheme created using Biorender.*

3.3 RESULTS AND DISCUSSION

All final concentration values obtained are presented in **Appendix III**.

3.3.1 2-METHYL TRYPTOPHAN PRODUCTION IN ESCHERICHIA COLI *NCM3722*

Production of 2-Methyl Tryptophan was performed in *E. coli* NCM3722 co-transformed with pTsrM (vector containing Tryptophan 2-C-methyltransferase) and pBtu, using pPhpK with pBtu as a negative control. The results are presented in ***Figure 3.4***.

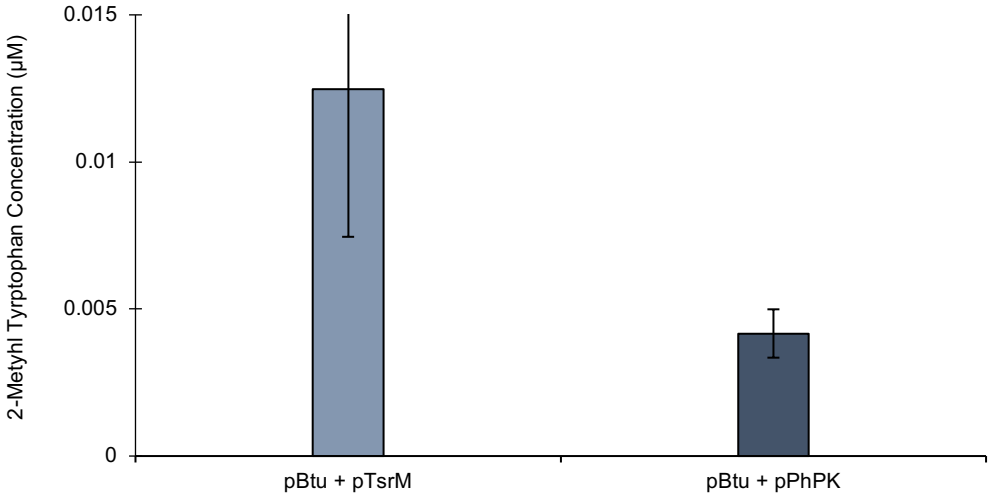

Figure 3.4 *- 2-Methyl Tryptophan obtained concentrations through* E. coli *NCM3722 production. The vector pPhpK represents the negative control. The error bars represent the standard deviation of the triplicates, while the bars represent the average of the same triplicates (μM).*

As it is shown in Figure 3.4, the strain pBtu + pTsrM yielded higher concentrations of 2-methyl tryptophan than the negative control. It is possible to conclude that the production of 2-methyl tryptophan production was achieved in *E. coli* NCM3722. The presence of a small amount of 2-methyl tryptophan on negative control samples may be due to a wrong integration of an inexistent clear LC-MS peak. Despite this, the negative control results were used in all assays.

3.3.2 ANALYSING THE IMPACT OF STREPTOMYCES CATTLEYA *FERREDOXINS AND Fe-S CLUSTER ASSEMBLY PROTEINS IN 2-METHYL TRYPTOPHAN PRODUCTION IN* ESCHERICHIA COLI NCM3722

Encouraged by the obtained results, we decided to determine the influence of *Streptomyces sp.* ferredoxins and Fe-S cluster assembly proteins in the production of 2-methyl tryptophan. The effect of the following constructs was measured: pTsrM_Scatt3Fd, pTsrM_ScattSufUT and pTsrM_Scatt3Fd_SufUT, all co-transformed with pBtu (***Figure 3.5***).

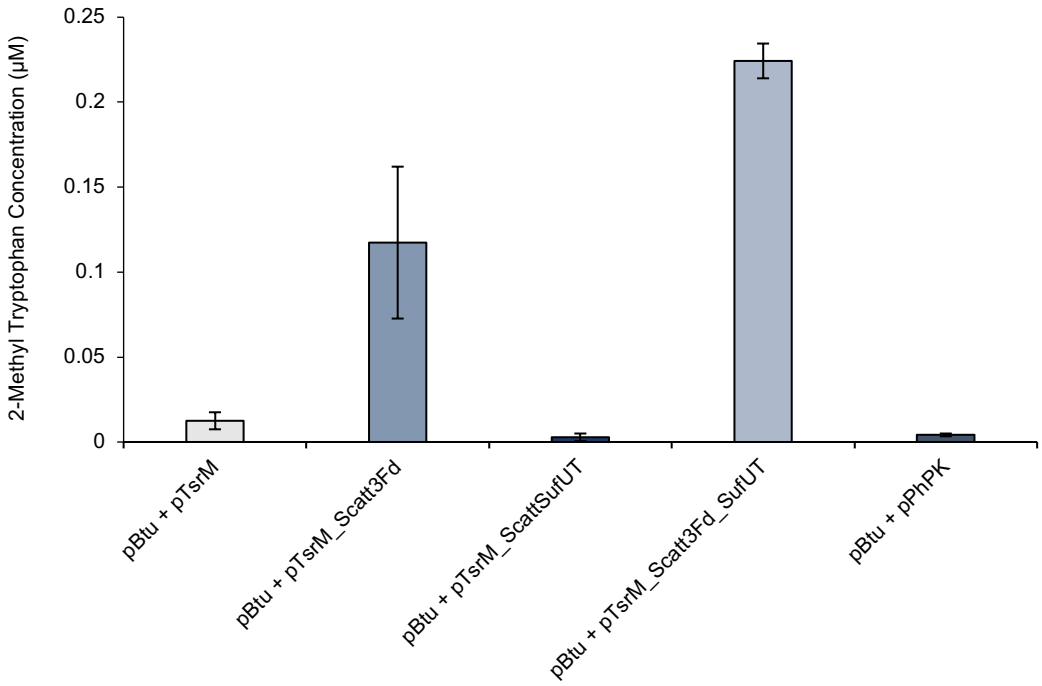

***Figure 3.5** - Increasing concentration results using* S. cattleya *ferredoxins and Fe-S cluster assembly genes (μM). Positive control: pBtu + pTsrM. Negative control: pBtu + pPhpK.*

Strains containing ferredoxins presented higher concentrations than the positive control, while the sample with only SufUT presented no production. This suggests that ferredoxins are the main cause for TsrM higher activity. Thus, ferredoxins increase the production of 2-methyl tryptophan (9.4-fold, compared to positive control). Furthermore, the strains containing ferredoxins and SufUT had higher production titers than the strains only containing ferredoxins, which implies that SufUT proteins do not increase the production by themselves. When conjugated with ferredoxins, the production is increased 18-fold compared to the positive control, and 1.9-fold in comparison to the strains with sole expression of ferredoxins. Although the reaction can be improved by

overexpressing only ferredoxins, production titers can be higher when co-overexpressing ferredoxins and SufUT proteins.

3.3.3 THE INFLUENCE OF INDIVIDUAL STREPTOMYCES CATTLEYA *FERREDOXINS IN 2-METHLY TRYPTOPHAN PRODUCTION*

Gene duplication is a common phenome among genomes, even though these genes can be corelated with specific pathways. It was then important to perceive if there was a specific ferredoxin responsible for the increase of 2-methyl tryptophan concentration or if it occurs due to the synergy between the three of them. Strains with pTsrM_Scatt3Fd were compared with strains having the three ferredoxins independently cloned (pTsrM_ScattFd1, pTsrM_ScattFd2 and pTsrM_ScattFd3). The final results obtained can be seen in ***Figure 3.6***.

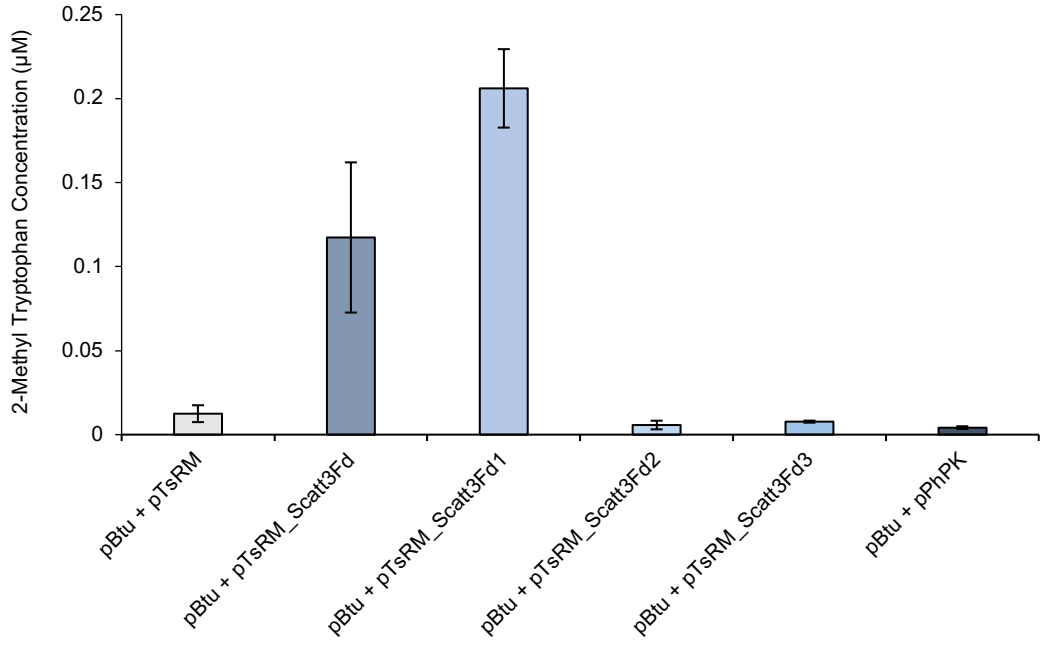

Figure 3.6 *- Comparison between 2-Methyl Tryptophan titers obtained using the three ferredoxins from* Streptomyces cattleya *cloned together and independently, using the strains containing pBtu + pPhPK as a negative control and pBtu + TsrM as positive control. Bars represent the average of triplicates (µM) and error bars the standard deviation of the same triplicates.*

As it is shown in ***Figure 3.6***, ferredoxin 2 and ferredoxin 3 produced concentrations close to the negative control, while ferredoxin 1 strain produced higher concentrations than the positive control (16.5-fold). Taking this into account, it is possible to infer that ferredoxin 1 is responsible for the ferredoxins improving effect in the reaction. Also, pTsrM_ScattFd1 strain presented higher concentration than pTsrM_Scatt3Fd strain (1.8-fold). These results may occur due to the fact that

it is expected that ferredoxin 1 carries and provides the majority of the electrons, but there is a minority that is carried by ferredoxins 2 and 3, decreasing the efficiency of the process when using three ferredoxins together. Furthermore, ferredoxin 1 might have higher affinity to receive the electrons from FNR, competing with ferredoxins 2 and 3 to be activated, which decreases the efficiency of the process when all three ferredoxins are being overexpressed. Another possibility is that ferredoxin 1 individual overexpression is more efficient than when the ferredoxin is co-overexpressed with other ferredoxins. Nevertheless, this hypothesis can only be proved after ferredoxins' quantification. It is then concluded that ferredoxin 1 from *S. cattleya*, in contrast to ferredoxins 2 and 3, is the main responsible for increasing 2-methyl tryptophan concentrations.

3.3.4 STREPTOMYCES LAURENTII *FERREDOXINS IMPACT IN 2-METHYL TRYPTOPHAN PRODUCTION*

Taking the previous results into account, it was important to understand if *S. laurentii* ferredoxins could present higher improving effect than *S. cattleya* ferredoxins in the reaction catalyzed by *S. laurentii* TsrM. Samples of *E. coli* NCM3722 transformed with pTsrM_Slau3Fd and pTsrM_Slau3Fd_SufUT were compared with a positive control (pTsrM) and with *S. cattleya* strains (pTsrM_Scatt3Fd and pTsrM_Scatt3Fd_SufUT). The resulting data is shown in ***Figure 3.7***.

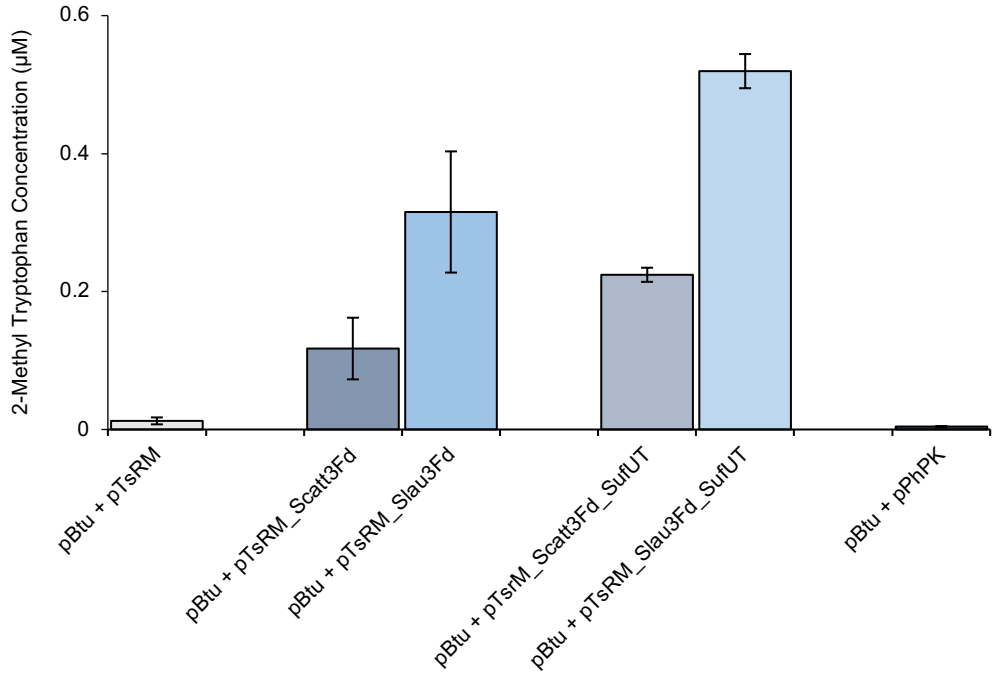

Figure 3.7 - S. laurentii *ferredoxins effect in 2-methyl tryptophan production, using the strains containing pBtu + pPhPK as a negative control and pBtu + pTsrM as positive control. Bars represent the average of triplicates (µM) and error bars the standard deviation of the same triplicates.*

Similar to *S. cattleya,* strains with ferredoxins from *S. laurentii* showed higher concentrations than the pTsrM strain, which proves that there is 2-methyl tryptophan production improvement due to the presence of electron transfer proteins. Moreover, the SufUT improving effect when combined with ferredoxins also occurred (pBtu + pTsrM_Slau3Fd_SufUT strain obtained 1.6-fold higher titers than pBtu + pTsrM_Slau3Fd strain). When comparing *S. cattleya* strains with *S. laurentii* strains, the second ones present higher titers than the first ones. This effect is 2.7-fold higher when using *S. laurentii* ferredoxins instead of *S. cattleya* ones, which might be predictable once TsrM protein is also originary from *S. laurentii*. However, the occurrence of an error in RBS optimization for *S. cattleya* ferredoxins could also be the reason why these ferredoxins presented lower titers (it is necessary to repeat this assay using new RBS-optimized *S. cattleya* ferredoxins genes). The higher concentration was obtained with pBtu + pTsrM_Slau3Fd_SufUT strain, having a concentration increase of 41.6-fold comparing to pBtu + pTsrM. As with the previous case, the question remained as to which specific gene was responsible for the higher titer. It was made a Clustal Omega (EMBL-EBI) alignment between the three proteins from *S. cattleya* and the three proteins from *S. laurentii* to search for similarities between them (***Figure 3.8***).

```
CLUSTAL O(1.2.4) multiple sequence alignment


SlauFd3         MTYVIAQPCVDVKDKACIEECPVDCIYEGQRSLYIHPDECVDCGACEPVCPVEAIFYEDD      60
ScattFd1        MTYVIAQPCVDVKDKACIEECPVDCIYEGQRSLYIHPDECVDCGACEPVCPVEAIFYEDD      60
SlauFd2         -----------------MQQEAPAVGTEEALEVWIDQDLCTGDGICAQYAPEVFELDI-D      42
ScattFd2        ----------------------MSDATGEELEVWIDQDLCTGDGICAQYAPEVFELDI-D      37
SlauFd1         ------------------------------MRISVDTGVCIGAGQCALVAPEVFTQDE-D      29
ScattFd3        ------------------------------MTVRVSVDRDRCCGAGQCVLAVPQVFDQSEED      32
                                              : :. . * . * *    *        *

SlauFd3         TPEEWKDYYKANV---EFFDDLGSPGGASKLGL--IERDHPFIAALP-------------      102
ScattFd1        TPEEWKDYYKANV---EFFDELGSPGGASKLGL--IERDHPFIAALP-------------      102
SlauFd2         GL----AYVKSA-----DDELLQAPGATTRVPLPLLQDVVDSVKECPGECIHVRRAGDGV      93
ScattFd2        GL----AYVKAPAEPGEEGELRTEPGATAPVPLPLLHDVVDSAKECPGECIHVRRVSDKV      93
SlauFd1         GF----SMVLPGREDGPVGPL-----------------VREAVRACPVQAITVNE-----      63
ScattFd3        GL----VELLTEEPAVELGPQ-----------------LHQAASFCPAGAITVHDA----      67
                                                       *

SlauFd3         ----PQNG  106
ScattFd1        ----PQNG  106
SlauFd2         EVYGPDAG  101
ScattFd2        EVYGPDAA  101
SlauFd1         --------  63
ScattFd3        --------  67
```

Figure 3.8 - *Clustal Omega multiple sequence alignment of all ferredoxin's amino acids sequences from S. cattleya and S. laurentii used in this study. (*) represents conservation of residues in all sequences, (:) represents high conservation in all sequences, (.) represents low conservation in all sequences, no symbol represents no conservation of residues. Red rectangle represents the only amino acid residues difference between the ferredoxins 1 and 3 from S. cattleya and S. laurentii respectively.*

From the alignment of ferredoxins, it is possible to observe that the ferredoxins 1 and 3 from *S. cattleya* and *S. laurentii,* respectively, are identical. The only difference between them is one residue in position number 78: a glutamic acid (E) for *S. cattleya* and aspartic acid (D) for *S. laurentii,* both of them are acid residues having identical properties (Red rectangle in ***Figure 3.5***). All other ferredoxins have low similarity between them. As concluded in the previous chapter, ferredoxin 1 from *S. cattleya* is responsible for the increase in 2-MTryp production. For this reason, we propose that ferredoxin 3 from *S. laurentii* might be the responsible for the increase in production in *S. laurentii* ferredoxins strain. However, this has yet to be confirmed experimentally.

3.3.5 2-METHYL TRYPTOPHAN PRODUCTION IN ESCHERICHIA COLI *BL21 (DE3)*

Following the previous studies, it was important to understand if it was also possible to produce the compound or if the production could be increased using a different *E. coli* strain, BL21 (DE3), a strain commonly used for the over expression of proteins. The process used was the same as described in 3.3.1 subchapter and the results are shown in ***Figure 3.9***.

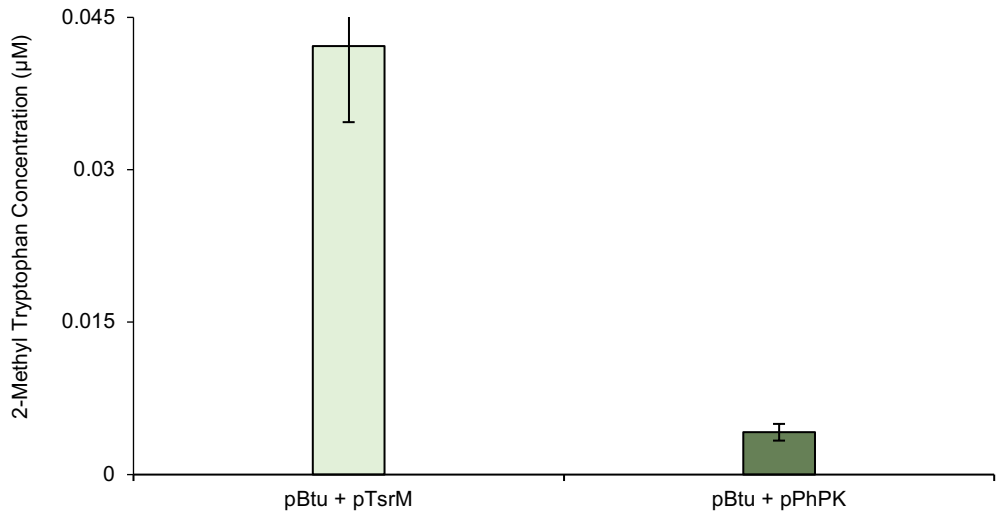

Figure 3.9 *- Production of 2-methyl tryptophan in* E. coli *BL21 (DE3). pBtu + pPhPK strains were used as negative control. Bars represent the average of triplicates' concentrations (µM) and the error bars the correspondent triplicates standard deviation. Negative control: pBtu + pPhPK.*

As it is shown in the graph, the strain containing pTsrM produced higher 2-methyl tryptophan concentrations than the negative control. Thus, it is concluded that it is possible to produce 2-methyl tryptophan using *E. coli* BL21 (DE3).

3.3.6 STREPTOMYCES LAURENTII *FERREDOXINS IMPACT IN 2-METHYL TRYPTOPHAN PRODUCTION IN* ESCHERICHIA COLI *BL21 (DE3)*

The second question addressed in this project was whether ferredoxins from *S. laurentii* also increase production when using *E. coli* BL21 (DE3). Once *S. laurentii* proteins obtained the best results, *E. coli* BL21 (DE3) cells co-transformed with pBtu and pTsrM_Slau3Fd and pBtu and pTsrM_Slau3Fd_SufUT were compared with the ones transformed with pBtu + pTsrM (positive control, negative control: pBtu + pPhPK). ***Figure 3.10*** depicts the obtained results.

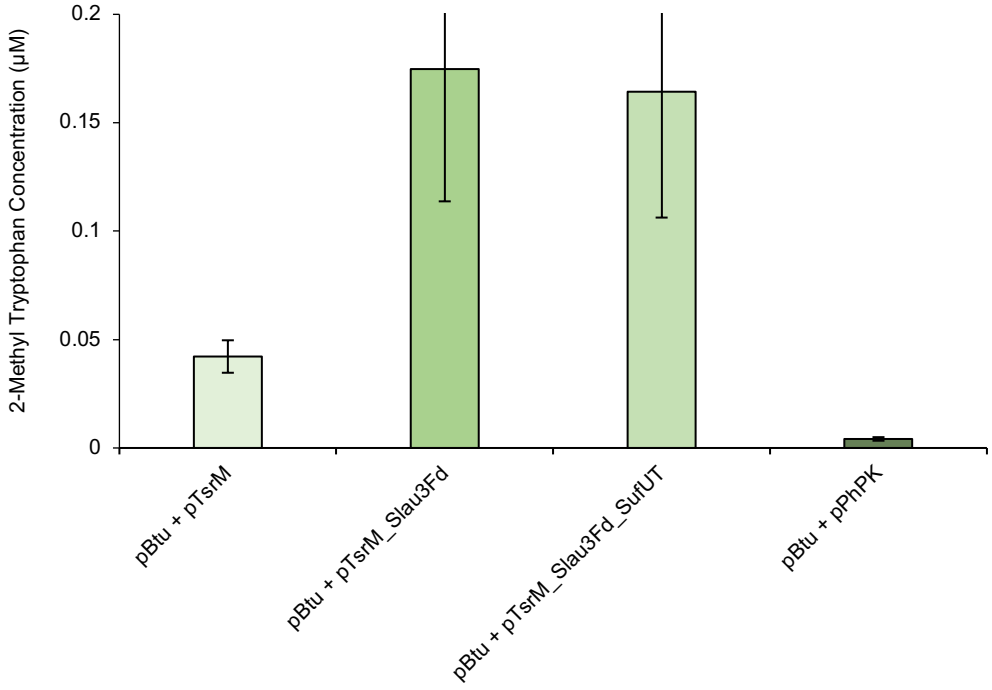

Figure 3.10 - S. laurentii *ferredoxins effect in 2-methyl tryptophan production. Positive control: pBtu + pTsrM, Negative control: pBtu + pPhPK. Bars: average of concentrations of triplicates (µM), Error Bars: standard deviation of the triplicates.*

We observe that both pBtu + pTsrM_Slau3Fd and pBtu + pTsrM_Slau3Fd_SufUT strains shown higher concentrations than pTsrM, and similar concentrations between them. The SufUT effect seen in NCM3722 is not the same as in BL21 (DE3). In NCM3722, SufUT combined with ferredoxins shown higher titers comparing to when using ferredoxins by themselves. In BL21 (DE3), the final concentrations using only ferredoxins or ferredoxins and SufUT proteins were very similar. Thus, in BL21 (DE3) strain, SufUT proteins have a very low impact in 2-methyl tryptophan production, being ferredoxins responsible for increasing the final concentration of the compound. This shows that despite the possibility of ferredoxins effect being transversal to different strains, SufUT proteins effect is not.

3.3.7 2-METHYL TRYPTOPHAN PRODUCTION IN NCM3722 vs BL21 (DE3)

To understand which *E. coli* strain is the best one to use in order to obtain the maximum improvement, pBtu + pTsrM_Slau3Fd and pBtu + pTsrM_Slau3Fd_SufUT strains were compared between NCM3722 and BL21 (DE3) (expression strain: higher final concentrations of the product were expected) *E. coli* strains. Comparing the results shown in ***Figure 3.11***, it is possible to observe that NCM3722 strains produced higher concentration of 2-methyl tryptophan than BL21 (DE3) ones, except on the positive control (pBtu + TsrM). When improving the reaction using strains containing ferredoxins, NCM3722 produced a higher amount of 2-methyl tryptophan than BL21 (DE3) (1.8-fold and 3.2-fold, respectively). This may occur once NCM3722 strain contains in its genome the gene *btuB* (hydroxocobalamin membrane importer) while BL21 (DE3) strain does not. For unknown reasons, this might result in a better functional expression of the genes, increasing the production of 2-methyl tryptophan, obtaining higher concentrations. Taking this into account, to obtain maximal production, *E. coli* NCM3722 strain is preferred over *E. coli* BL21 (DE3) strain. Since the positive control concentrations are lower in NCM3722 than in BL21 (DE3), it was important to understand if the overexpression of pBtu operon was necessary or toxic for NCM3722, decreasing the production.

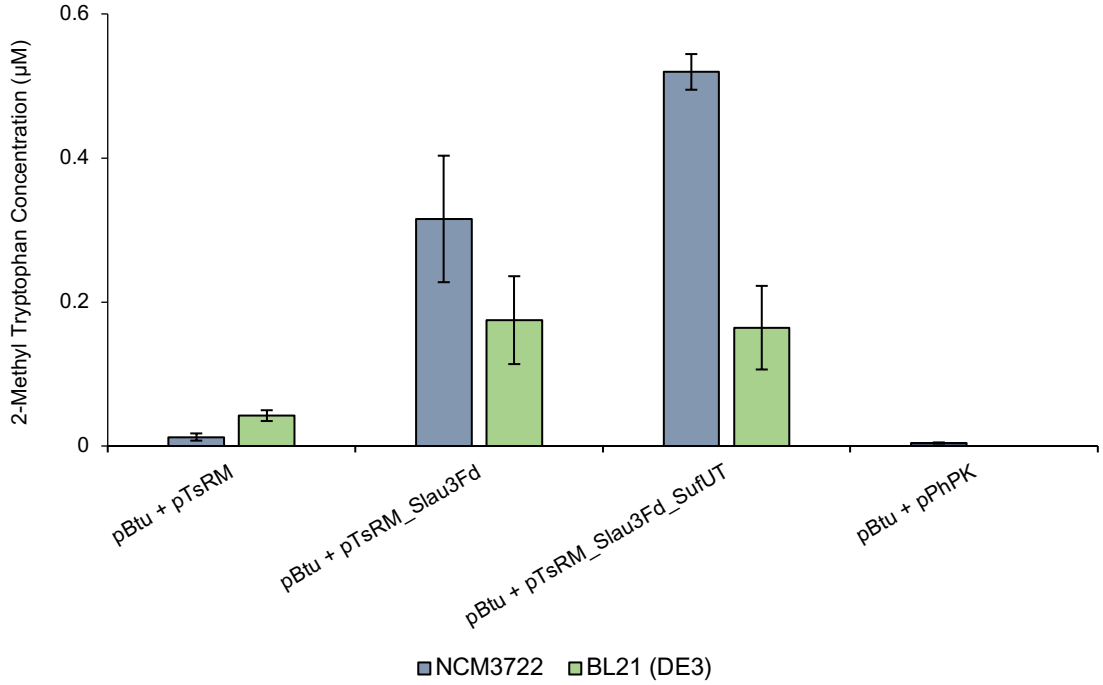

Figure 3.11 *- Product concentration comparison between* E. coli *NCM3722 and* E. coli *BL21 (DE3) strains. Blue color represents NCM3722 samples' concentration and Green color represents BL21 (DE3) samples' concentration.*

3.3.8 EFFECT OF BTU OPERON OVEREXPRESSION ON METHYLTRYPTOPHAN PRODUCTION

Strains containing pBtu were compared with strains containing pRFP (negative control to pBtu) in both *E. coli* NCM3722 and *E. coli* BL21 (DE3). The graphical results can be seen in ***Figure 3.12***.

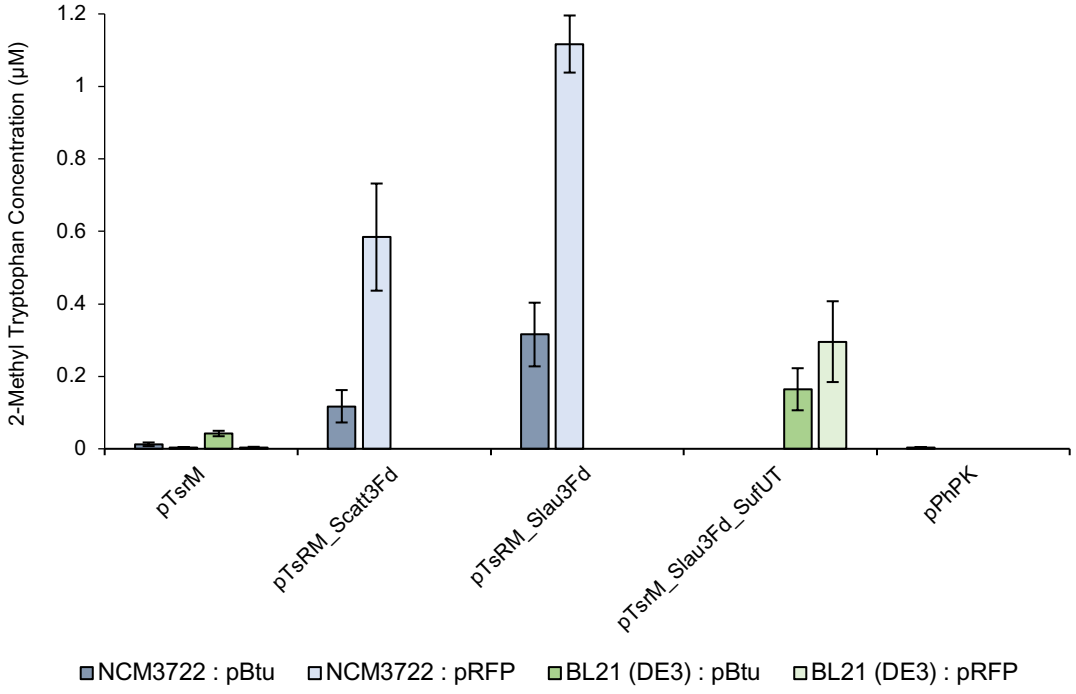

Figure 3.12 *- Comparison of 2-Methyl Tryptophan titers between pBtu expression and pRFP (negative control to pBtu). Bars represent the average of compound concentrations of the triplicates (µM) and error bars represent the standard deviation of those samples.*

Observing ***Figure 3.12***, all strains containing pRFP, except pRFP + pTsrM, present higher concentrations than the ones with pBtu. There are two possibilities to explain these results: (1) when co-expressed with ferredoxins or ferredoxins and SufUT proteins, the expression of pBtu is not strictly necessary and becomes toxic, decreasing final titers or (2) pRFP expression increases, for unknown reasons, 2-methyl tryptophan production in *E. coli*. To know which possibility was the correct one, another assay was done to compare strains transformed with pBtu and strains not transformed with pBtu (***Figure 3.13***).

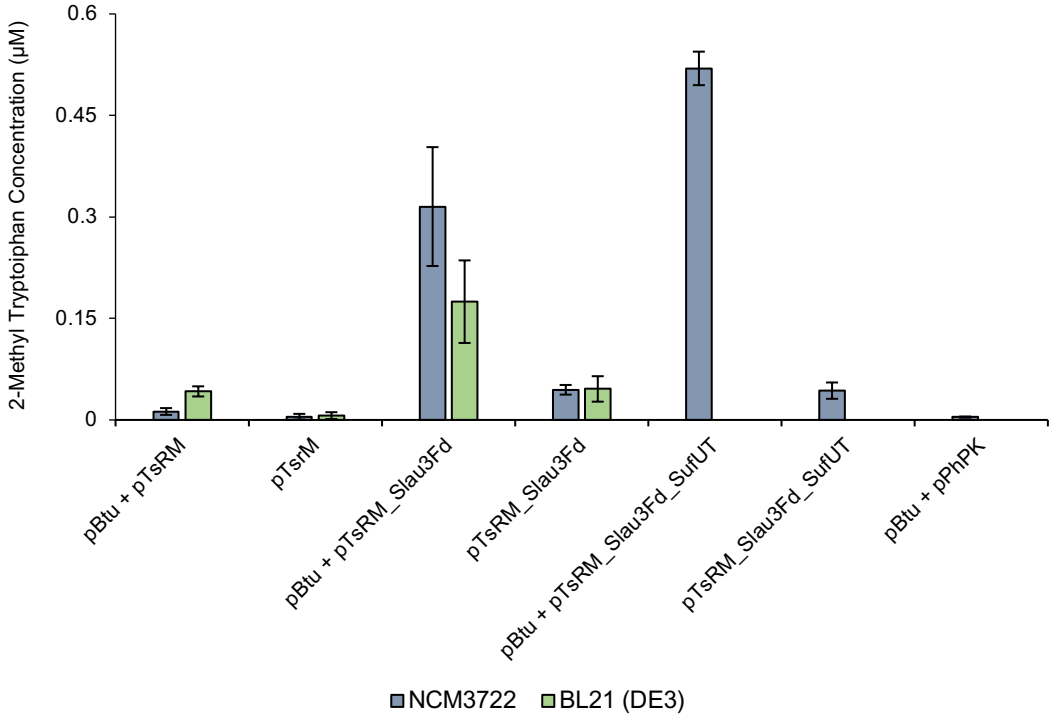

Figure 3.13 *- Testing the effect of expression/overexpression of pBtu in* E. coli *NCM3722 and* E. coli *BL21 (DE3). Error bars represent the standard deviation of the triplicates.*

All strains containing pBtu expression produced higher titers of the compound, compared to the ones just containing pTsrM vectors. This indicates that pBtu co-expression with TsrM vectors is not toxic and increases production. Moreover, in NCM3722 and even in BL21 (does not contain the BtuB gene in its genome) it is seen production when not using the vector pBtu. Due to this, it is possible to conclude that even though pBtu increases 2-methyl tryptophan production, it is not strictly necessary when co-expressed with TsrM vectors containing ferredoxins or ferredoxins and SufUT proteins. Although, when just using pTsrM, the production is very similar to the negative control. From this we draw the conclusion that the higher the quantity of Btu channels is, the higher the production by TsrM and 2-methyl tryptophan concentration increases (without pBtu, cobalamin import is also a limiting step of the reaction).

Finally, when comparing the strains pBtu + pTsrM_Slau3Fd and pTsrM_Slau3Fd in both *E. coli* strains, it is possible to observe that without pBtu the production was very similar in both, while with pBtu NCM3722 produced higher concentrations. This was an expected result, once NCM3722 already had the BtuB gene in its genome and BL21 (DE3) had not. Thus, it is expected that the overexpression of pBtu, containing BtuB, presented higher production concentrations in NCM3722 than in BL21 (DE3), where BtuB is not overexpressed.

Taking this into account, the expression/overexpression of pBtu operon is not strictly necessary, but it improves productivity. Once pBtu is not decreasing titers, the results obtained in

Figure 3.12 regarding pRFP were not expected. The results indicate that pRFP is having an enhancement impact on the production. More data is necessary to understand exactly the nature of this phenomenon.

3.3.9 HYDROXOCOBALAMIN SUPPLEMENT EFFECT IN 2-METHYL TRYPTOPHAN PRODUCTION

Following the previous results, the question if the hydroxocobalamin supplement is really necessary to improve the production had arisen. To find an answer to this question, strains which medium contained the supplement were compared with the strains containing the same vectors (pBtu + pTsrM and pBtu + pTsrM_Slau3Fd) in medium without the supplement. The *E. coli* strain used was the one that obtained the best concentrations before (*E. coli* NCM3722, subchapter 4.3.7). Graphical results are presented in ***Figure 3.14.***

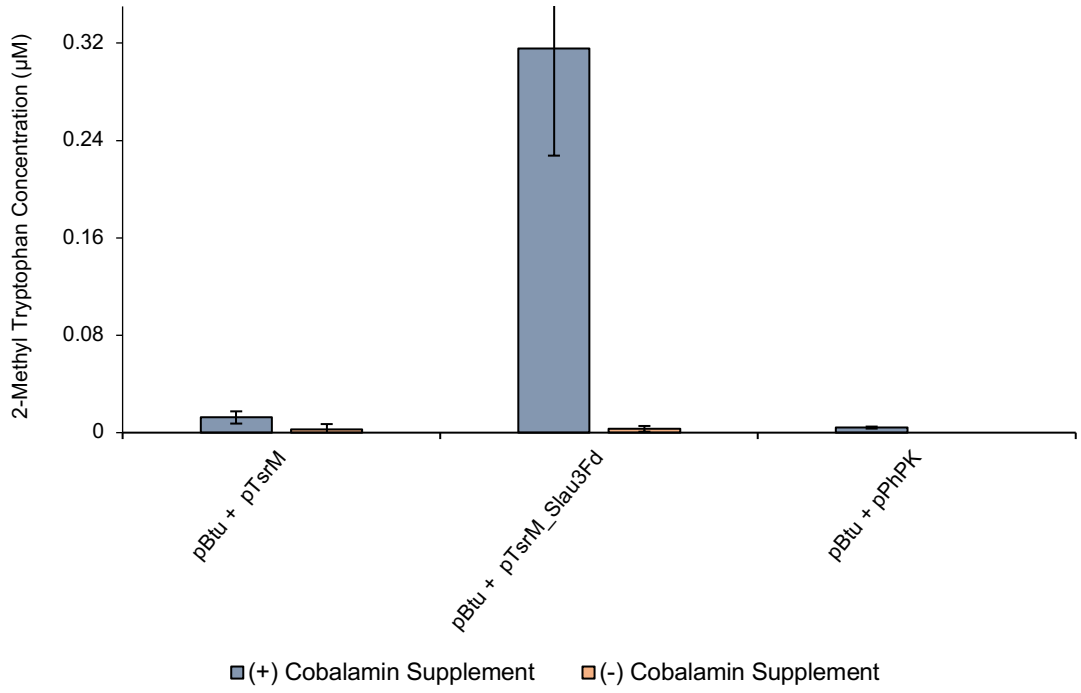

Figure 3.14 *- Measurement of 2-methyl tryptophan concentrations of* E. coli *NCM3722 samples with and without cobalamin supplement to understand its effect on the improvement of the reaction. Blue bars represent the average concentrations of triplicates which cells grew with cobalamin supplement (+); orange bars represent the average concentrations of triplicates which cells grew without adding the supplement (-). Error bars represent standard deviation of the triplicate's values.*

Observing ***Figure 3.14***, all cells grown in media without the supplement presented production rates similar to the negative control (pBtu + pPhPK). This suggests that the cobalamin supplement not only increases the production (100.6-fold increase when comparing pBtu +

pTsrM_Slau3Fd strain with and without supplement), but also that it is strictly necessary to 2-methyl tryptophan production in *E. coli*. These results corroborate the data from other previous studies (16,52,56).

Once it is concluded that cobalamin supplement is essential to increase production titers, it was important to understand if we could further increase the production by adjusting the supplement concentration in the medium. Using *E. coli* NCM3722 cells co-transformed with pBtu and pTsrM_Slau3Fd, a new assay was made using several concentrations of hydroxocobalamin in the medium (***Figure 3.15***).

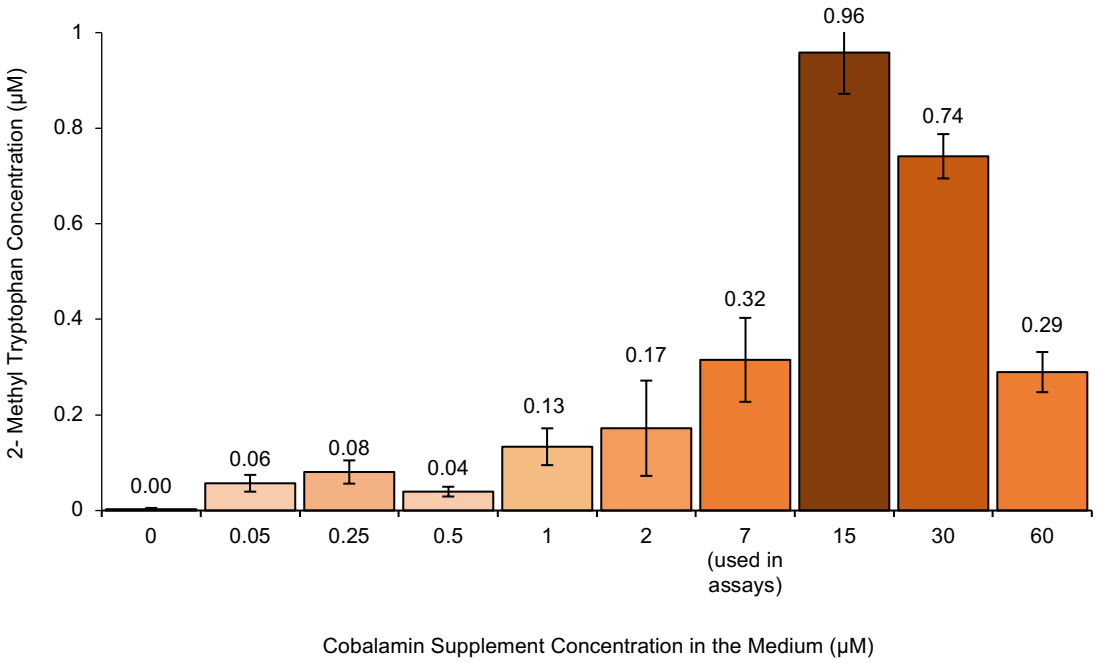

Figure 3.15 *- Measurement of 2-methyl tryptophan concentration using different cobalamin supplement concentrations in the medium (*E. coli *NCM3722 cells co-transformed with pBtu and pTsrM_Slau3Fd). Light orange bars represent lower product concentrations and dark orange bars represent higher product concentrations.*

In ***Figure 3.15*** it is observed that there is minimal production of 2-methyl tryptophan until 0.5 μM cobalamin supplement, increasing concentrations from 1 μM to 15 μM cobalamin supplement and a decrease of production from 30 μM to 60 μM. Since the maximal production obtained occurred when using 15 μM (0.96 μM), this concentration of cobalamin supplement is considered to be the best one to use in order to obtain higher titers of the compound. When using higher concentrations of supplement (30 μM and 60 μM), production decreases. This decrease may result from a negative effect that high concentrations of cobalamin might have in the TsrM pathway or in cell growth. Therefore, since highest production values are between 7 μM and 30 μM, being 15 μM the highest, the optimal cobalamin supplement concentration is either between

7 μM and 15 μM or 15 μM and 30 μM. Further studies using several intermediate concentrations of cobalamin between 7 μM and 30 μM are needed to obtain a more specific value.

3.4 CONCLUSIONS

The results presented in this chapter show that not only it is possible to import the activity of a foreign TsrM into *E. coli*, but also that the reaction catalyzed by this enzyme can be optimized using compatible ferredoxins, SufUT proteins, and Btu proteins overexpression.

This data indicates that ferredoxins' overexpression has a pivotal role in pathway improvement, while *S. cattleya* SufUT proteins do not increase 2-methyl tryptophan concentrations by themselves but are an enhancer when overexpressed together with *S. cattleya* and *S. laurentii* ferredoxins in *E. coli* NCM3772. *E. coli* NCM3772 was considered the best strain to use because it produces higher concentrations of the compound compared to BL21 (DE3). After testing the three ferredoxins from *S. cattleya* individually in *E. coli* NCM3772, it was possible to conclude that there was a specific ferredoxin (ScattFd1) responsible for the increase in titer. This implies that *E. coli* ferredoxins/flavodoxins are not sufficient to improve 2-methyl tryptophan production and heterologous ferredoxins/flavodoxins are needed to react with foreign TsrM proteins. The explanation to this lies on the discrepancy between the genus of the host (*Escherichia coli*), the heterologous overexpressed TsrM and ferredoxins genes (*Streptomyces sp.*). Similar origin of the enzyme and the secondary electron transfer proteins implies for higher compatibility between them. Even when overexpressing TsrM and ferredoxins from the same genus, individual ferredoxins have different impact on the production. This shows that there might be specific electron transfer proteins used to directly or indirectly reduce cob(II)alamin to cob(I)alamin, replacing it on the cycle, and that are ready to receive a methyl group from SAM and transfer it to tryptophan. The molecular mechanisms behind this might be related to the reduction potential of each specific ferredoxin, as well as their structural compatibility and binding-induced conformational changes that might occur (45). It has been shown that some ferredoxins have higher reduction potential than other, being more stable when reduced, hampering the reduction of cob(II)alamin (59). Moreover, it is possible that some ferredoxins have a more compatible structure to transfer the electrons to cob(II)alamin or suffer conformational changes when bound to TsrM, being easier or harder to reduce the enzyme.

S. cattleya ferredoxin 1 was the responsible for the improvement of titers and, most likely due to the homology between them, *S. laurentii* ferredoxin 3 may play the same role. Although this cannot be taken to show that there are specific ferredoxins more compatible to certain SAM enzymes than other, it can be seen as a hypothesis that demands further investigation. Moreover, the overexpression of *S. laurentii* ferredoxins presented higher product concentration than *S. cattleya* ones in *E. coli* NCM3772. This occurred probably due to an RBS optimization error of *S.*

cattleya ferredoxins genes but is also possible that is because TsrM and *S. laurentii* ferredoxins have the same origin.

For a better understanding on the influence of *btu* operon and to make sure that its overexpression was necessary or non-toxic, assays with the replacement of *btu* genes for RFP and without pBtu or pRFP plasmids were performed. Obtained results imply that *btu* overexpression is not toxic and enhances 2-methyltryptophan production. The results from RFP plasmid strains were not expected and presented the highest product concentration comparing to the other strains (pRFP + pTsrM_Slau3Fd increased 89.5-fold the production compared to pBtu + pTsrM). The possibility of the occurrence of an error was discarded once the procedure was repeated twice. Therefore, there is currently no explanation to back these observations.

Lastly, the results obtained regarding TsrM cobalamin requirement indicate the necessity of cobalamin addition to the medium for *in vivo* experiments. The maximal product concentration obtained occurred when using 15 μM cobalamin (0.96 μM), 3-fold more than when using 7 μM (0.32 μM) and 1.3-fold more than 30 μM (0.74 μM). Thus, instead of using a medium cobalamin concentration of 7 μM, should be used a higher concentration between 7 μM and 30 μM for the described method (more experiments are needed to know the specific optimal concentration). If the optimal concentration is found, all obtained concentrations may probably increase.

In conclusion, the present data represent the first steps to the improvement of thiostrepton A pathway from *Steptomyces laurentti* in *E. coli* by increasing 2-methyl tryptophan production titers by TsrM. This study shows that previous findings reported in literature do not only apply to *in vitro* conditions, but are also valid in *in vivo* experiments, making it possible to import foreign pathways containing Fe-S enzymes into *E. coli*. Taking into account all the obtained results, the best combination for improving TsrM functioning is the co-transformation of *E. coli* NCM3772 cells with the plasmids pBtu and pTsrM_Slau3Fd_SufUT (41.6-fold higher than positive control) and grown in medium containing 15 μM cobalamin supplement. Thus, the main objective of this chapter – stated as the improvement of 2-methyl tryptophan production using co-overexpression of ferredoxins, SufUT proteins and Btu proteins along with TsrM, as well as understanding the individual roles of each group of proteins – was successfully achieved.

4 FOSFOMYCIN PRODUCTION IN ESCHERICHIA COLI

4.1 INTRODUCTION

As referred in chapter 1, the search for novel and effective antibiotics, as well as the improvement of their production pathways, has been a topic of study by a large number of laboratories. In 1969 a new powerful antibiotic, fosfomycin, was discovered by Hendlin, D. et al. (63). Fosfomycin is a broad-spectrum antibiotic naturally produced by organisms from *Streptomyces sp.* and *Pseudomonas sp.* in aerated and submerged environments (63–65). It affects Gram-positive and Gram-negative bacteria, and it is an approved antibiotic drug for the treatment of lower urinary tract infections. Moreover, as a safe drug, it presents incredibly low toxicity for humans. Thus, fosfomycin is a potent and reliable antibiotic to be administered as a chemotherapeutic agent (63,66,67).

A particularly appealing characteristic of fosfomycin, despite its seemingly simple structure, is the presence of an epoxide and a C-P bond, quite a rare occurrence in bioprocesses (see ***Figure 4.1***) (15,67,68).

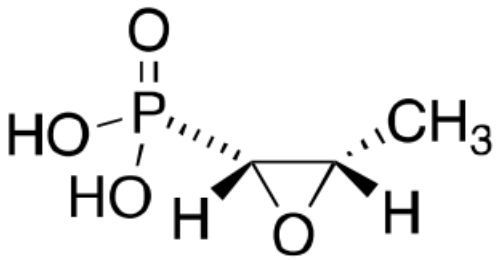

Figure 4.1 *- Fosfomycin molecular structure. Adapted from Sato, S. et al., 2017.*

Fosfomycin is considered a Phosphoenolpyruvate (PEP) analog owing to their similar chemical structures. Fosfomycin inhibits in an irreversibly way the enzyme enolpyruvyltransferase MurA, the catalyzer of the first step of peptidoglycan biosynthesis pathway, blocking cell wall development (66,69). For having a broad incidence in bacteria, fosfomycin producers have two specific genes, *fomA* and *fomB*, that confer them complete resistance to the action of this antibiotic. In addition to this, studies confirmed that when transforming *E. coli* with *fomA* and *fomB*, these bacteria present a complete resistance to fosfomycin (67,70).

The metabolic pathway for fosfomycin biosynthesis was discovered by genetic analysis and *in vitro* studies in *Streptomyces sp.* It comprises five genes (*fom1-fom4* and *fomC*) considered to be essential for the antibiotic biosynthesis (15,71). Furthermore, two different, fosfomycin

pathways have been reported. In the first pathway (PathA), the formation of 2-hydroxypropylphosphonate (HPP) from 2-hydroxyethylphosphonate (HEP) is seen as a direct process. The second pathway (PathB) was recently proposed by Sato, S. et al. (15). In contrast to PathA, PathB presents the methylation of HEP forming HPP as a three-step process, instead of the direct process presented by the first proposed pathway. Both pathways, as schematically presented by Sato, S. et al., are represented in ***Figure 4.2*** (15).

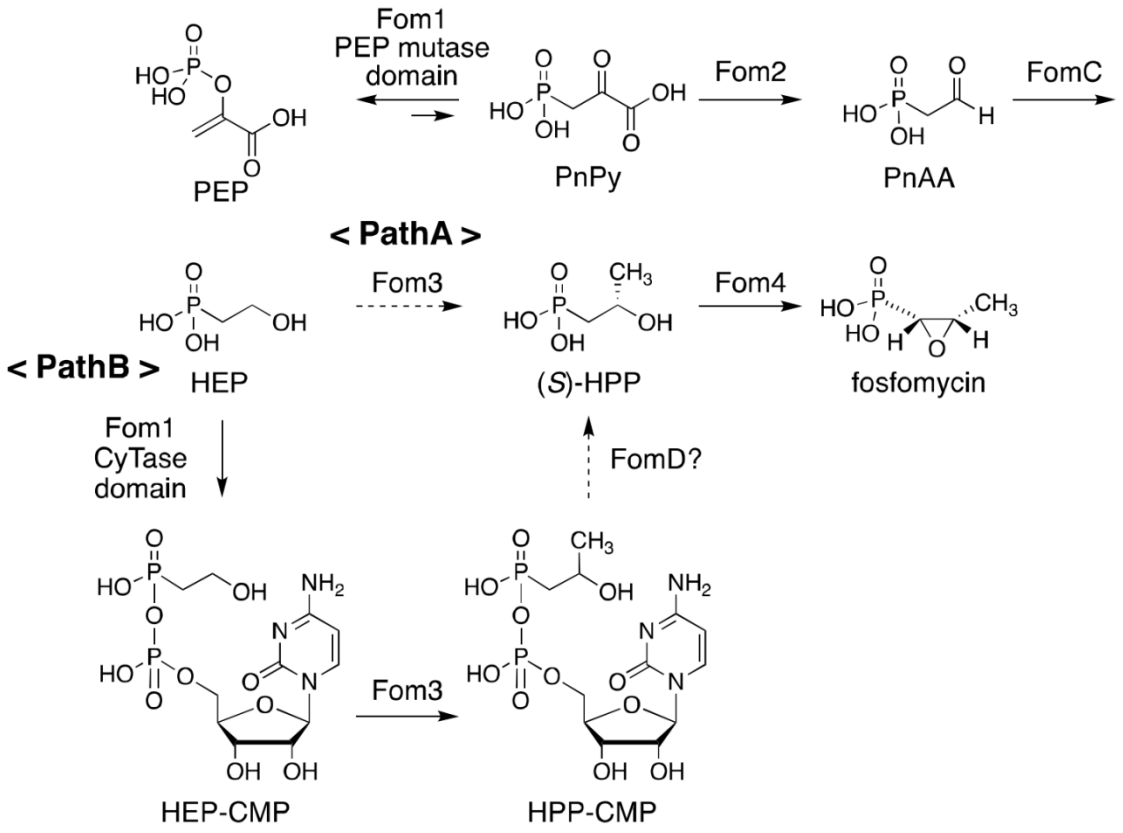

Figure 4.2 *- Two possible fosfomycin production pathways in* Streptomyces sp. *PathA: first proposed pathway. PathB: second proposed pathway. Source: Sato, S., et al., 2017.*

The three main steps on the pathway are the establishment of a C-P bound, the insertion of a CH_3 (methyl) group and the insertion of an epoxide group. As it is possible to observe in ***Figure 4.2***, both pathways consider that the enzyme Fom4 accomplishes the insertion of an epoxide group. Similarly, the C-P bound is achieved due to Fom1 catalyzation. However, the way the methyl group is added is not well established, having two possible options: direct (PathA) or indirect (PathB) addition. Pathway B is considered to be the correct one to follow as a model. In both options, 2-hydroxyethylphosphonate methyltransferase (Fom3) represents a leading role in the process, either transforming directly HEP in HPP or transforming cytidylyl-HEP (HEP-CMP)

in cytidylyl-HPP (HPP-CMP). For pathway B is believed the existence of a sixth needed enzyme, FomD, to catalyze the transformation of HPP-CMP in HPP. Moreover, the necessity for the presence of intermediary steps in pathway B is suspected to be due to Fom3 strict recognition of HEP-CMP instead of HEP by itself (15).

Fom3 is a SAM radical methyltransferase whose function relies on the catalysis of a C-methylation of HEP-CMP to obtain HPP-CMP. This enzyme is composed by a radical SAM domain and a cobalamin-binding domain, being methylcobalamin its cofactor and the donor of the methyl group necessary for the reaction (15,72,73). Similar to other SAM enzymes, SAM radical domain of Fom3 is composed by a [4Fe-4S] cluster bound to cysteines by three of the irons from the cluster. The fourth iron atom function is the binding to SAM in two sites (α-amino and α-carboxylate groups). When SAM binds to the SAM domain, the cluster is reduced, delivers an electron that cleaves the α-carboxylate bound, and an L-methionine and a 5'-dA radical are released. A hydrogen atom from Carbon 2 (C2) from HEP is received by the radical, creating a C2 radical on the HEP molecule. The HEP molecule is now ready to receive the methyl group present on methylcobalamin, donated by SAM. This group is transferred from SAM to cob(I)alamin, creating methylcob(III)alamin ready to transfer it to HEP-CMP. When this occurs, cob(III)alamin turns to cob(II)alamin. To replace cob(I)alamin in the cycle is necessary the input of an electron from a secondary electron transfer protein (66). This process is shown in ***Figure 4.3***.

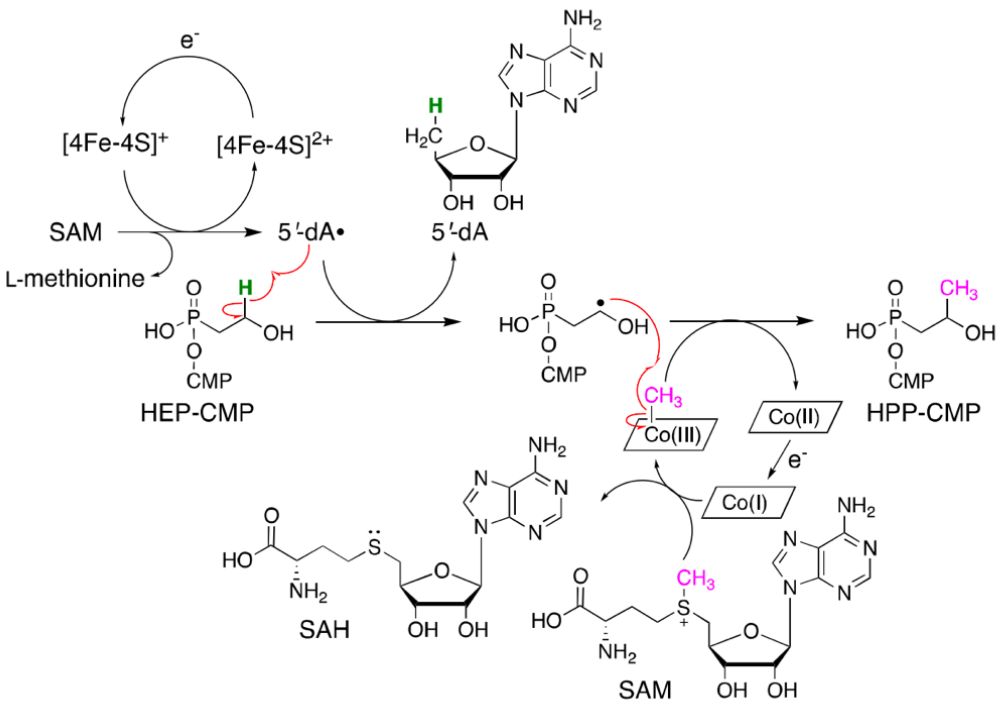

Figure 4.3 *- HEP-CMP C-methylation by Fom3 enzyme reaction. Source: Sato, S., et al., 2017.*

The fosfomycin studies accomplished until now are in their majority *in vitro* studies for enzymatic research. On the other hand, *in vivo* studies focused on the usage of different *Streptomyces, sp.* organisms for heterologous expression research (15,71,74). Also, a recent study showed that overexpressing *btu* operon increases the intracellular concentration of cobalamin (61).

Another approach can be studying the heterologous expression of fosfomycin pathway in laboratory grown bacteria species, using undemanding conditions. If these studies are well succeeded, novel and greener pharmaceutical industrial production processes can be applied in the future. The main target of this chapter is to introduce the fosfomycin pathway from *S. wedmorensis* (*S. wedmorensis*), into *Escherichia coli*. Furthermore, the possibility to obtain a maximal production by overexpressing genes for cobalamin import from *E. coli, S. cattleya* Fe-S cluster assembly/maturation proteins and *S. cattleya* ferredoxins genes is going to be tested.

A general scheme summarizing this chapter main goal is presented in ***Figure 4.4***.

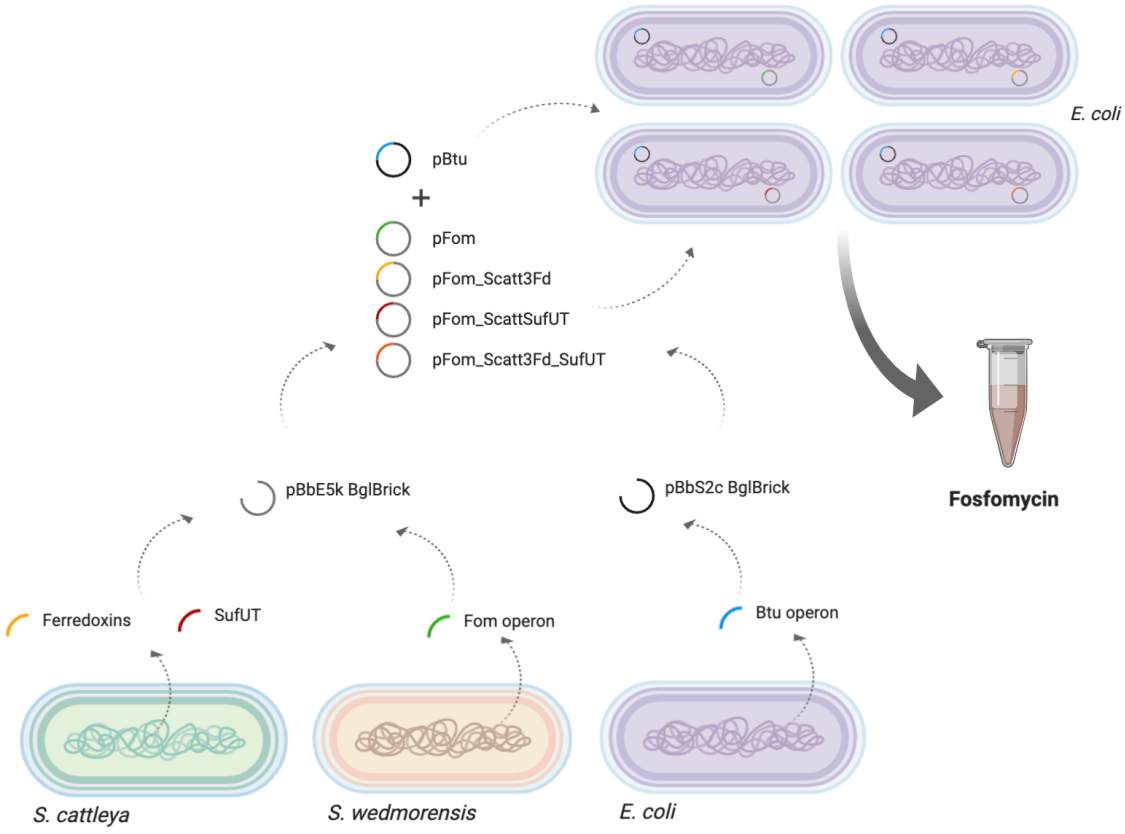

Figure 4.4 *- General scheme of fosfomycin heterologous production in* E. coli *as expected to this study. Scheme created using Biorender.*

4.2 MATHERIALS AND METHODS

4.2.1 PRODUCTION VECTORS

Fosfomycin production vectors were already available in the laboratory (work previously made). Synthetic genes (*fom1, fom2, fom3, fom4, fomC* and *fomD*, GenBank ID proteins BAA32495.1, BAA32496.1, BAA32490.1, BAA32491.1, BAA32497.1 and BAA32492.1 respectively) from *Streptomyces wedmorensis* codon-optimized for *Escherichia coli* expression were cloned in the backbone pBbE5k BglBrick to produce the Fosfomycin production vector pBbE5k_Fom34D12C (pFom). To obtain the vectors pFom_ScattSufUT, pFom_Scatt3Fd and pFom_Scatt3Fd_SufUT it was used the same operons, genes and way of cloning from the previous chapter. Furthermore, it was also produced, using the same cloning procedure, the vectors pFom_ScattFd1, pFom_ScattFd2 and pFom_ScattFd3. As negative control was used the TsrM gene (does not belong to fosfomycin production pathway) cloned into the same backbone (pTsrM).

The vector pBtu used was the one described for 2-methyl tryptophan production (section 3.2.1, chapter 3) and all the vectors were expressed using *E. coli* DH5-α.

4.2.2 FOSFOMYCIN PRODUCTION

The protocol used to produce fosfomycin was similar to the one used to produce 2-methyl tryptophan. The differences between the two protocols were the transformation in *E. coli* NCM3722 and BL21 (co-transformed with pFom and pBtu vectors) volume of the cultures (5 mL for fosfomycin), the concentration of the pBtu operon inducer (50.0 ng/mL ATc) and the sampling process (5h and 24h pellet and supernatant samples). Furthermore, supernatant samples were also measured (10 µL was taken and added to 90 µL of NMM for LC-MS measurement).

All remaining components (*E. coli* strains, medium, sampling and sampling preparation steps) were the same as described above.

All individual ferredoxins samples were taken, but not measured in LC-MS.

4.2.3 LC-MS MEASUREMENT/ANALYSIS

Fosfomycin levels were measured by an LC-MS system (Agilent) using a Merck Millipore peek ZIC-pHILIC column with a precolumn and equipped with a standard ESI source mass spectrometer (sample injection volume of 5 µL). The mobile phase was compound by two

solvents, A (20 mM ammonium carbonate) and B (20 mM ammonium carbonate in 80% acetonitrile). After 2 min at 100% solvent B, the metabolites were separated by a gradient from 100% to 55% of solvent B for 15 min (flow rate 0.2 mL/min), followed by a gradient from 55% to 100% for 1.5 min (same flow rate), increasing of flow rate to 0.4 mL/min, held at 100% solvent B for 2 min and reduction of the flow again to 0.2 mL/min for 2 min. The fosfomycin (mass 138.06, negative polarity) precursor ion (137) was fragmented into product ions (79.1 and 63.1) using an ESI ionization in MRM mode. Fosfomycin concentration was obtained using a calibration curve constructed with standard fosfomycin samples. The values were corrected with the volume taken for sampling to obtain the final concentration of fosfomycin. It was used the average of the corrected values and the error was considered as the standard deviation of the triplicates.

An example of the calibration curves used is presented in **Appendix II**. A fresh calibration curve was prepared in each LC-MS run.

A general scheme of the methods used in this chapter is presented on ***Figure 4.5***.

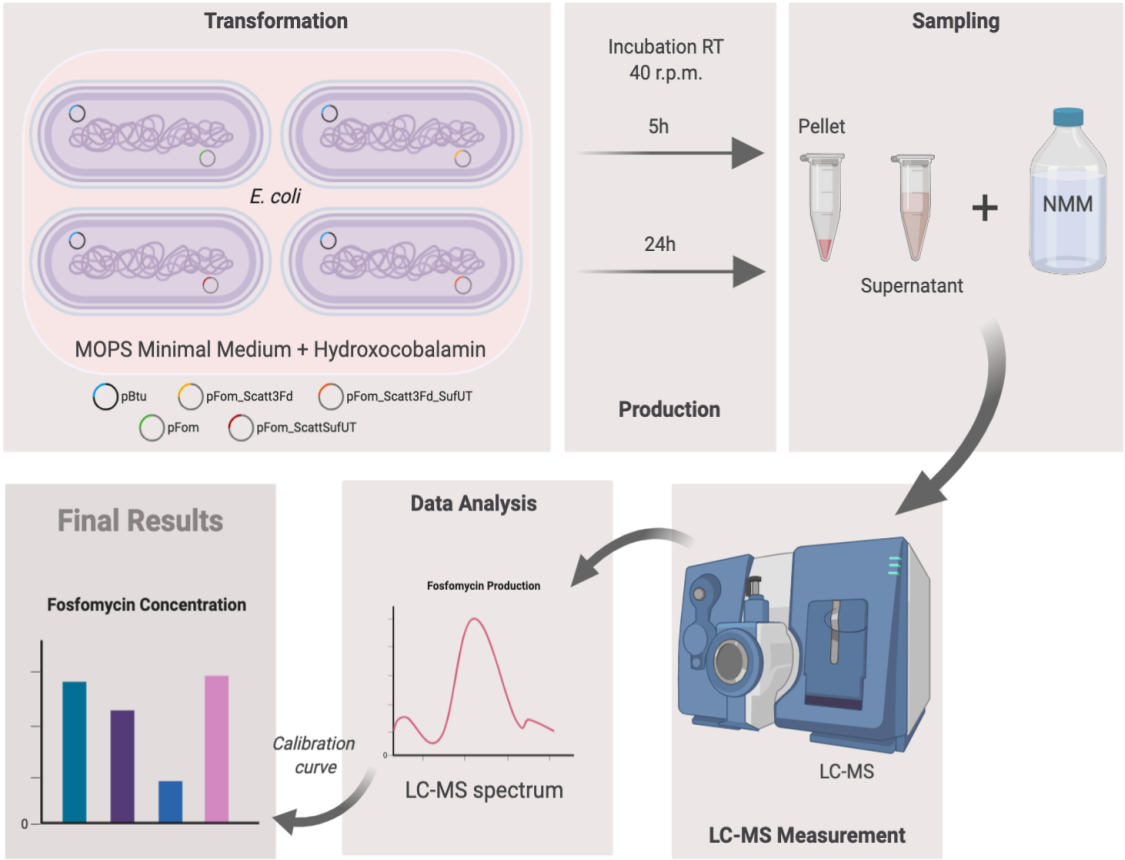

Figure 4.5 *- General scheme of the methodology used in chapter 3 of this project. The spectrum and graph bars represented are merely indicative. Scheme created using Biorender.*

4.3 RESULTS AND DISCUSSION

All final concentration values obtained are presented in **Appendix III**.

4.3.1 FOSFOMYCIN PRODUCTION IN ESCHERICHIA COLI *NCM3772*

To test whether *Streptomyces wedmorensis fom* genes and *Streptomyces cattleya* ferredoxin/SufUT genes make it possible for *E. coli* NCM3722 to produce fosfomycin, the vectors pFom, pFom_Scatt3Fd, pFom_ScattSufUT, pFom_Scatt3Fd_SufUT and pTsrM (negative control) were co-transformed with pBtu. Samples were measured in LC-MS.

The results from LC-MS data analysis suggested that the fosfomycin concentrations obtained were similar to or lower than the negative control. It was concluded that there was no production of fosfomycin. This conclusion was corroborated by the absence of LC-MS clear peaks in all strains (data not shown). Thus, all peak area values obtained were due to a wrong automatic integration of peaks. Finally, this data shows that it was not possible to produce fosfomycin using *S. wedmorensis fom* genes and *S. cattleya* ferredoxin genes in *E. coli* NCM3722, which challenges previous results obtained in our laboratory (was possible to obtain significative concentrations of fosfomycin) (75).

4.3.2 FOSFOMYCIN PRODUCTION IN ESCHERICHIA COLI *BL21 (DE3)*

Next, we moved on to prove whether the lack of fosfomycin production was due to the choice of strain. Taking BL21 (DE3) as a new chassis, we carried out experiments retaining all previous conditions.

The LC-MS analysis results are shown that the strains pBtu + pFom_ScattSufUT 24h and pBtu + pFom_Scatt3Fd_SufUT 5h had higher production than the negative control, especially the supernatant sample from the second one. This could be indicative that it is possible to produce fosfomycin in BL21 (DE3). Although, the observation of the LC-MS peaks showed the same results as the production in NCM3722. None of the strains presented clear peaks (data not shown), which stands for the values being a consequence of wrong automatic integration of peaks and therefore impossible to measure. Finally, it was concluded that, as with NCM3722, it was not possible to produce fosfomycin using *S. wedmorensis* and *S. cattleya* genes in *E. coli* BL21 (DE3). Once again, this results are in contrast with previous observations in our laboratory (75).

4.4 CONCLUSIONS

The results presented in this chapter show that in these conditions, it was not possible to produce fosfomycin, even though it was previously produced in the laboratory (75). Even when overexpressing cobalamin importers, Fe-S cluster assembly proteins and Fom3 compatible ferredoxins (main possible bottlenecks of Fom3 activity), the presence of fosfomycin was not detected on the samples. As all conditions used for cell growth, sampling and LC-MS measurement were the same as in previous studies, the results obtained here may have been caused by a possible committed error in some step during this study.

Moreover, since the used plasmids were constructed during the previous study, they were not freshly prepared plasmids when used in the present study. It is possible that their sequences had deteriorated, even though the tubes were kept at a proper temperature. Another possible explanation that may be considered is the existence of parallel pathways that may be conditioning the proper functioning of fosfomycin production. An example of this is the necessity of the presence of peroxide in the cell so that the peroxidase Fom4 can proceed to the formation of the epoxide group on fosfomycin molecule. If the concentration of peroxide within the cell is not sufficient or the peroxide is being used in other cellular functions, the last reaction for fosfomycin production is compromised (66,76). Second, it is important to consider the existence of active phosphonate (molecule containing C-P bounds) degradation pathways. The presence of genes that encode C-P cleavage proteins in *E. coli* is documented and they are known to be phosphate starvation inducible. Hence, if the phosphate concentration of the medium is low, C-P cleavage proteins are produced and fosfomycin molecules become susceptible to degradation (77,78).

Therefore, taking into account the present results, it was not possible under the presented conditions to produce fosfomycin, even when using compatible secondary electron transfer proteins. Since it is not possible to know precisely the origin of the difference between results, it is necessary, in future work, to repeat these experiences using the same conditions and freshly prepared plasmids. In addition to this, the studies should also increase, in a non-toxic way, the peroxide concentration in the cell, for instance by reducing or inhibiting catalase activity, and increase intracellular phosphate concentration by adding higher concentration of K_2HPO_4 to the medium. It can also be considered to include *fomA* and *fomB* genes in the vector to make sure that *E. coli* cells are completely resistant to fosfomycin during the production process, taking into account the previous published results (67,70).

In conclusion, the main goal of the present chapter of this study, introduced on subchapter 4.1 as the heterologous expression of *Streptomyces wedmorensis* fosfomycin pathway in *Escherichia coli* using *Streptomyces cattleya* SufU, SufT and ferredoxins overexpression was not accomplished.

5 CONCLUSIONS

As discussed at the beginning of this dissertation, the recurrent appearance of the resistance to antibiotics is a problem that humanity is facing nowadays. The production of new and powerful antibiotics is needed, but it also represents a challenge to researchers and pharmaceutical industries. Several causes are behind this, such as the impossibility of producing antibiotics biologically or the low biological production rates that can be achieved when bioproduction is possible (1,2). Thus, this dissertation presents the first steps to build a possible method to overcome these two barriers on laboratory bioproduction of antibiotics.

In this dissertation, I present an innovative method to improve reactions of cobalamin-dependent SAM radical methylases. This method includes a bioinformatics program to assist the search and choice of possible compatible ferredoxins, followed by heterologous overexpression in *Escherichia coli* of the SAM radical enzyme along with those compatible ferredoxins, SufU, SufT and Btu proteins. The process is initiated by choosing the target organism and run the program in its annotated genome. After this, it is necessary to choose the ferredoxins to be studied and clone them with SufUT and the cobalamin-dependent SAM radical enzyme of interest. This process is followed by an *E. coli* co-transformation of the previous plasmid with a vector containing Btu proteins, and the production step should occur with the addition of cobalamin to the medium. Although fosfomycin results showed no production, 2-methyl tryptophan results represent the first steps for obtaining and improving high-value compounds production pathways that include cobalamin-dependent SAM radical enzymes from bacteria impossible to grow on laboratorial environment (41.6-fold more production when co-overexpressing *S. laurentii* TsrM, *S. laurentii* ferredoxins, *S. cattleya* SufUT and *E. coli btu* operon, and adding 7 μM cobalamin to the medium). The obtained results show that it is possible to improve thiostrepton production, in the way that the range of optimal concentration of cobalamin supplement in the medium was discovered, as well as that the overexpression of *btu* operon along with compatible secondary electron transfer proteins and TsrM increases the production of 2-methyl tryptophan, an essential compound on thiostrepton synthesis. Hence, further improvement experiments might survey the impact of several different ferredoxins and TsrM from distinct species, as well as the determination of the specific optimal concentration (between 7 μM and 30 μM) of cobalamin supplement added to the medium.

Nevertheless, it is crucial in future work to apply this methodology to other compounds, as to back up its robustness. Furthermore, the existence of further studies is essential to understand the role of electron transfer partners of secondary electron transfer proteins and include them on the method if the obtained results are promising. When all these results are achieved, they can be applied to several antibiotics pathways, making it possible to enhance bioproduction. It is then necessary to do several scale-up studies to understand its capacity to be applied to industrial

production. If all these results are positive, in the last instance, these antibiotics bioproduction may be industrialized, diminishing most of the impacts from chemical pharmaceutical industry.

In summary, the hypothesis to prove – presented in the beginning of this dissertation as the possibility of producing antibiotics, naturally produced in unculturable bacteria, in synthetically modified *E. coli*, as well as the possibility of improving production rates by increasing external chemical groups donors concentration and by overexpressing Fe-S clusters scaffold/transport proteins and secondary electron transfer proteins – could be verified.

REFERENCES

1. Li C-J, Trost BM. Green chemistry for chemical synthesis. Proc Natl Acad Sci. 2008;105(36):13197–202.

2. Walsh C, Wencewicz T. Antibiotics: Challenges, Mechanisms, Opportunities, 2nd Edition. Antibiotics: Challenges, Mechanisms, Opportunities, 2nd Edition. 2015. 399–419 p.

3. Liu J, Chakraborty S, Hosseinzadeh P, Yu Y, Tian S, Petrik I, et al. Metalloproteins containing cytochrome, iron-sulfur, or copper redox centers. Chem Rev. 2014;114(8):4366–469.

4. Hosseinzadeh P, Lu Y. Design and fine-tuning redox potentials of metalloproteins involved in electron transfer in bioenergetics. Biochim Biophys Acta - Bioenerg. 2016;1857(5):557–81.

5. Ueno T, Watanabe Y. Coordination Chemistry in Protein Cages: Principles, Design, and Applications. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2013. 399 p.

6. Cammack R. Iron—sulfur clusters in enzymes: Themes and variations. Adv Inorg Chem. 1992;38(C):281–322.

7. Beinert H, Sands RH. Studies on Succinic Acid and DPNH Dehydrogenase Preparations by Paramagnetic Resonance (EPR) Spectroscopy. 1960;3(1):41–6.

8. Bruschi M, Guerlesquin F. Structure, function and evolution of bacterial ferredoxins. FEMS Microbiol Lett. 1988;54(2):155–76.

9. Fukuyama K, Nagahara Y, Tsukihara T, Katsube Y, Hase T, Matsubara H. Tertiary structure of Bacillus thermoproteolyticus [4Fe-4S] ferredoxin. Evolutionary implications for bacterial ferredoxins. J Mol Biol. 1988;199(1):183–93.

10. Meyer J. Iron-sulfur protein folds, iron-sulfur chemistry, and evolution. J Biol Inorg Chem. 2008;13(2):157–70.

11. Rocha AG, Dancis A. Life without Fe-S clusters. Mol Microbiol. 2016;99(5):821–6.

12. Johnson DC, Dean DR, Smith AD, Johnson MK. Structure, Function, and Formation of Biological Iron-Sulfur Clusters. Annu Rev Biochem [Internet]. 2005;74(1):247–81. Available from: http://www.annualreviews.org/doi/10.1146/annurev.biochem.74.082803.133518

13. Holliday GL, Akiva E, Meng EC, Brown SD, Calhoun S, Pieper U, et al. Atlas of the Radical SAM Superfamily: Divergent Evolution of Function Using a “Plug and Play” Domain [Internet]. 1st ed. Vol. 606, Methods in Enzymology. Elsevier Inc.; 2018. 1–71 p. Available from: http://dx.doi.org/10.1016/bs.mie.2018.06.004

14. Young AP, Bandarian V. Pyruvate is the source of the two carbons that are required for formation of the imidazoline ring of 4-demethylwyosine. Biochemistry. 2011;50(49):10573–5.

15. Sato S, Kudo F, Kim SY, Kuzuyama T, Eguchi T. Methylcobalamin-Dependent Radical SAM C-Methyltransferase Fom3 Recognizes Cytidylyl-2-hydroxyethylphosphonate and Catalyzes the Nonstereoselective C-Methylation in Fosfomycin Biosynthesis. Biochemistry. 2017;56(28):3519–22.

16. Mahanta N, Hudson GA, Mitchell DA. Radical S-Adenosylmethionine Enzymes Involved in RiPP Biosynthesis. Biochemistry. 2017;56(40):5229–44.

17. Imlay JA. Iron-sulphur clusters and the problem with oxygen. Mol Microbiol [Internet]. 2006;59(4):1073–82. Available from: internal-pdf://112.248.36.66/Imlay-2006-Molecular_Microbiology-clusterSchem.pdf

18. Malkin R, Rabinowitz JC. The reconstitution of clostridial ferredoxin. Biochem Biophys Res Commun. 1966;23(6):822–7.

19. Jacobson MR, Cash VL, Weiss MC, Laird NF, Newton WE, Dean DR. Biochemical and genetic analysis of the nifUSVWZM cluster from Azotobacter vinelandii. MGG Mol Gen Genet. 1989;219(1–2):49–57.

20. Tokumoto U, Kitamura S, Fukuyama K, Takahashi Y. Interchangeability and Distinct Properties of Bacterial Fe-S Cluster Assembly Systems : Functional Replacement of the isc and su f Operons in Escherichia coli with the nifSU-Like Operon from Helicobacter pylori. Japanese Biochem Soc.

2004;136(2):199–209.

21. Fontecave M, Choudens SO De, Py B, Barras F. Mechanisms of iron-sulfur cluster assembly: The SUF machinery. J Biol Inorg Chem. 2005;10:713–21.

22. Takahashi Y, Nakamura M. Functional assignment of the ORF2-iscS-iscU-iscA-hscB-hscA-fdx-ORF3 gene cluster involved in the assembly of Fe-S clusters in Escherichia coli. J Biochem. 1999;126(5):917–26.

23. Outten FW, Djaman O, Storz G. A suf operon requirement for Fe-S cluster assembly during iron starvation in Escherichia coli. Mol Microbiol. 2004;52(3):861–72.

24. Mashruwala AA, Pang YY, Rosario-Cruz Z, Chahal HK, Benson MA, Mike LA, et al. Nfu facilitates the maturation of iron-sulfur proteins and participates in virulence in Staphylococcus aureus. Mol Microbiol. 2015;95(3):383–409.

25. Takahashi Y, Tokumoto U. A third bacterial system for the assembly of iron-sulfur clusters with homologs in Archaea and plastids. J Biol Chem. 2002;277(32):28380–3.

26. Yokoyama N, Nonaka C, Ohashi Y, Shioda M, Terahata T, Chen W, et al. Distinct roles for U-type proteins in iron–sulfur cluster biosynthesis revealed by genetic analysis of the Bacillus subtilis sufCDSUB operon. Mol Microbiol. 2018;107(6):688–703.

27. Selbach BP, Chung AH, Scott AD, George SJ, Cramer SP, Dos Santos PC. Fe-S cluster biogenesis in gram-positive bacteria: SufU Is a zinc-dependent sulfur transfer protein. Biochemistry. 2014;53(1):152–60.

28. Mashruwala AA, Bhatt S, Poudel S, Boyd ES, Boyd JM. The DUF59 Containing Protein SufT Is Involved in the Maturation of Iron-Sulfur (FeS) Proteins during Conditions of High FeS Cofactor Demand in Staphylococcus aureus. PLoS Genet. 2016;12(8):1–39.

29. Johnson MK, Frazzon J, Dean DR, Agar JN, Krebs C, Huynh BH. IscU as a scaffold for iron-sulfur cluster biosynthesis: Sequential assembly of [2Fe-2S] and [4Fe-4S] clusters in IscU. Biochemistry. 2000;39(27):7856–62.

30. Palmer G, Reedijk J. Nomenclature Committee of the International Union of Biochemistry (NC-IUB). Nomenclature of electron-transfer proteins. Recommendations 1989. J Biol Chem. 1992;267(1):665–77.

31. Mortenson LE, Valentine RC, Carnahan JE. An Electron Transport Factor from Clostridium Pasteurianum. Biochem Biophys Res Commun [Internet]. 1962;7(6). Available from: http://www.ncbi.nlm.nih.gov/pubmed/14244728%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC441251

32. Mulo P, Medina M. Interaction and electron transfer between ferredoxin–NADP+ oxidoreductase and its partners: structural, functional, and physiological implications. Photosynth Res. 2017;134(3):265–80.

33. Eck R V., Dayhoff MO. Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences. Science (80-). 1966;152(3720):363–6.

34. Arnon DI. Ferredoxin and Photosynthesis. Science (80-). 2006;149(3691):1460–70.

35. Sofia HJ. Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. Nucleic Acids Res [Internet]. 2001;29(5):1097–106. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/29.5.1097

36. Lanz ND, Booker SJ. Auxiliary iron-sulfur cofactors in radical SAM enzymes. Biochim Biophys Acta - Mol Cell Res [Internet]. 2015;1853(6):1316–34. Available from: http://dx.doi.org/10.1016/j.bbamcr.2015.01.002

37. Moss L, Frey PA. Activation of Lysine 2,3-Aminomutase by S-Adenosylmethionine. J Biol Chem. 1990;265(30):0–4.

38. Baraniak J, Moss ML, Frey PA. Lysine 2,3-aminomutase. Support for a mechanism of hydrogen transfer involving S-adenosylmethionine. J Biol Chem. 1989;264(3):1357–60.

39. Lesk AM. Introduction to Bioinformatics. 4th ed. Oxford University Press; 2014.

40. Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: Rapid annotations using subsystems technology. BMC Genomics. 2008;9(75):1–15.

41. Moslemi C, Skovbjerg CK, Moeskjaer S, Andersen SU. Syntenizer 3000: Synteny-based analysis of orthologous gene groups. bioRxiv [Internet]. 2019;618678. Available from: https://www.biorxiv.org/content/10.1101/618678v1?rss=1

42. Georgescu CH, Manson AL, Griggs AD, Desjardins CA, Pironti A, Wapinski I, et al. SynerClust: a highly scalable, synteny-aware orthologue clustering tool. Microb genomics. 2018;4(11):0–6.

43. Médigue C, Moszer I. Annotation, comparison and databases for hundreds of bacterial genomes. Res Microbiol. 2007;158(10):724–36.

44. Rothery RA, Workun GJ, Weiner JH. The prokaryotic complex iron-sulfur molybdoenzyme family. Biochim Biophys Acta - Biomembr. 2008;1778(9):1897–929.

45. Atkinson JT, Campbell I, Bennett GN, Silberg JJ. Cellular Assays for Ferredoxins: A Strategy for Understanding Electron Flow through Protein Carriers That Link Metabolic Pathways. Biochemistry. 2016;55(51):7047–64.

46. Dutcher D, Vandeputte J. Thiostrepton, a new antibiotic. II. Isolation and chemical characterization. Antibiot Annu [Internet]. 1955;3:560—561. Available from: http://europepmc.org/abstract/MED/13355326

47. Trejo WH, Dean LD, Pluscec J, Meyers E, Brown WE. Streptomyces laurentii, a new species producing thiostrepton. J Antibiot (Tokyo). 1977;30(8):639–43.

48. Kelly WL, Pan L, Li C. Thiostrepton biosynthesis: Prototype for a new family of bacteriocins. J Am Chem Soc. 2009;131(12):4327–34.

49. Bash E. Thiostrepton: A Ribossomal Inhibitor Of Translocation. Biochem Biophys Res Commun. 1970;40(3):667–74.

50. Porse BT, Leviev I, Mankin AS, Garrett RA. The antibiotic thiostrepton inhibits a functional transition within protein L11 at the ribosomal GTPase centre. J Mol Biol. 1998;276(2):391–404.

51. Kongsema M, Wongkhieo S, Khongkow M, Lam E, Boonnoy P, Vongsangnak W, et al. Molecular mechanism of Forkhead box M1 inhibition by thiostrepton in breast cancer cells. Oncol Rep. 2019;1(2):953–62.

52. Benjdia A, Pierre S, Gherasim C, Guillot A, Carmona M, Amara P, et al. The thiostrepton A tryptophan methyltransferase TsrM catalyses a cob(II)alamin-dependent methyl transfer reaction. Nat Commun. 2015;6:1–10.

53. Suzumura K-I, Yokoi T, Funatsu M, Koji N, Koichi T, Huiping Z, et al. YM-266183 and YM-266184, Novel Thiopeptide Antibiotics Produced by Bacillus cereus Isolated from a Marine Sponge. J Antibiot (Tokyo). 2003;56(2):129–34.

54. Smith T, Jiang Y-F, Floss H. Thiopeptide Antibiotics. Chem Rev. 2005;105(2):685–714.

55. Chen L, Zeng AP. Rational design and metabolic analysis of Escherichia coli for effective production of L-tryptophan at high concentration. Appl Microbiol Biotechnol [Internet]. 2017;101(2):559–68. Available from: http://dx.doi.org/10.1007/s00253-016-7772-5

56. Fujimori DG. Radical SAM-mediated methylation reactions. Curr Opin Chem Biol. 2013;17(4):597–604.

57. Zhou P, O'Hagan D, Yuen LD, Mocek U, Zeng Z, Frenzel T, et al. Biosynthesis of the antibiotic thiostrepton. methylation of tryptophan in the formation of the quinaldic acid moiety by transfer of the methionine methyl group with net retention of configuration. J Am Chem Soc. 1989;111(18):7274–6.

58. Frenzel T, Zhou P, Floss HG. Formation of 2-methyltryptophan in the biosynthesis of thiostrepton: Isolation of S-adenosylmethionine: Tryptophan 2-methyltransferase. Arch Biochem Biophys. 1990;278(1):35–40.

59. Blaszczyk AJ, Knox HL, Booker SJ. Understanding the role of electron donors in the reaction catalyzed by Tsrm, a cobalamin-dependent radical S-adenosylmethionine methylase. JBIC J Biol Inorg Chem [Internet]. 2019; Available from: http://link.springer.com/10.1007/s00775-019-01689-8

60. Kuthning A, Mösker E, Süssmuth RD. Engineering the heterologous expression of lanthipeptides in Escherichia coli by multigene assembly. Appl Microbiol Biotechnol. 2015;99(15):6351–61.

61. Lanz ND, Blaszczyk AJ, McCarthy EL, Wang B, Wang RX, Jones BS, et al. Enhanced Solubilization of Class B Radical S-Adenosylmethionine Methylases by Improved Cobalamin Uptake in

Escherichia coli. Biochemistry. 2018;57(9):1475–90.

62. Wu GC, Goler JA, Anderson JC, Keasling JD, Leguia M, Arkin AP, et al. BglBricks: A flexible standard for biological part assembly. J Biol Eng. 2010;1–12.

63. Hendlin D, Stapley EO, Jackson M, Wallick H, Miller AK, Wolf FJ, et al. Phosphonomycin, a New Antibiotic Produced by Strains of Streptomyces. Science (80-). 1969;166(3901):122–3.

64. Shoji J, Kato T, Hinoo H, Hattori T, Hirooka K, Matsumoto K, et al. Production of fosfomycin (phosphonomycin) by Pseudomonas syringae. J Antibiot (Tokyo). 1986;39(7):1011–2.

65. Katayama N, Tsubotani S, Nozaki Y, Harada S, Ono H. Fosfadecin and fosfocytocin, new nucleotide antibiotics produced by bacteria. J Antibiot (Tokyo). 1990;43(3):238–46.

66. Horsman GP, Zechel DL. Phosphonate Biochemistry. Chem Rev. 2017;117(8):5704–83.

67. Seto H, Kuzuyama T. Bioactive natural products with carbon-phosphorus bonds and their biosynthesis. Nat Prod Rep. 1999;16(5):589–96.

68. Christensen BG, Leanza WJ, Beattie TR, Patchett AA, Arison BH, Ormond RE, et al. Phosphonomycin: Structure and Synthesis. Science (80-). 1969;166(3901):123–5.

69. Kahan FM, Kahan JS, Cassidy PJ, Kropp H. the Mechanism of Action of Fosfomycin (Phosphonomycin). Ann N Y Acad Sci. 1974;235(1):364–86.

70. Kuzuyama T, Kobayashi S, O'Hara K, Hidaka T, Seto H. Fosfomycin Monophosphate and Fosfomycin Diphosphate, Two Inactivated Fosfomycin Derivatives Formed by Gene Products of fomA and fomB from a Fosfomycin Producing Organism Streptomyces wedmorensis. J Antibiot (Tokyo). 1996;49(5):502–4.

71. Hidaka T, Goda M, Kuzuyama T, Takei N, Hidaka M, Seto H. Cloning and nucleotide sequence of fosfomycin biosynthetic genes of Streptomyces wedmorensis. MGG Mol Gen Genet. 1995;249(3):274–80.

72. Kuzuyama T, Hidaka T, Kamigiri K, Seto H, Imai S. Studies on the biosynthesis of fosfomycin 4. The biosynthetic origin of the methyl group of fosfomycin. J Antibiot (Tokyo). 1992;45(11):1812–4.

73. Bauerle MR, Schwalm EL, Booker SJ. Mechanistic diversity of radical S-adenosylmethionine (SAM)-dependent methylation. J Biol Chem. 2015;290(7):3995–4002.

74. Woodyer RD, Shao Z, Thomas PM, Kelleher NL, Blodgett JA V, Metcalf WW, et al. Heterologous Production of Fosfomycin and Identification of the Minimal Biosynthetic Gene Cluster. Chem Biol. 2006;13(11):1171–82.

75. Shomar Monges H. Producing high-value chemicals in Escherichia coli through synthetic biology and metabolic Engineering. TU Delft University. 2019.

76. Olivares P, Ulrich EC, Chekan JR, Van Der Donk WA, Nair SK. Characterization of Two Late-Stage Enzymes Involved in Fosfomycin Biosynthesis in Pseudomonads. ACS Chem Biol. 2017;12(2):456–63.

77. Mcgrath JW, Hammerschmidt F, Quinn JP. Biodegradation of phosphonomycin by Rhizobium huakuii PMY1. Appl Environ Microbiol. 1998;64(1):356–8.

78. Fox EM, Mendz GL. Phosphonate degradation in microorganisms. Enzyme Microb Technol. 2006;40(1):145–50.

APPENDIX I

The written script that constitute the program used in this study (Chapter 2: Bioinformatical Approach to Genomic Context) is presented below:

```
#Install the packages required
install.packages("berryFunctions")
install.packages("readxl")

#Use the packages
library(berryFunctions)
library(readxl)

#First step: Import the genome Dataset to Rstudio

#to select the lines of ferredoxins with less than 200 aa:
genome = data.frame(#GENOME DATASET NAME#)

#select the genes with less than 600 nucleotides (less than 200 aa):
aalenght = data.frame(genome[abs(genome$stop – genome$start ) <= 600,])

#select the lines from the previous one that have the word ferredoxin/flavodoxin (just change the word from "ferredoxin" to "flavodoxin", also change on script line 102):
truefds = data.frame (grep("ferredoxin", aalenght$function., ignore.case = TRUE))

#Problem: the lines that it returns are the lines from aalenght dataset and not from E.coli.K12.MG1665 dataset

#Discover the desired E.coli.K12.MG1665 dataset line numbers:
rw = numeric(0)

for(I in 1:nrow(truefds)){
  val = truefds[I,]
  rw[i] = rownames(aalenght[val,])
  #print(rw)
}

rw = as.numeric(rw)
```

```
   #to show all the lines before and after the desired protein that are
transcribed in the same strand (Genomic Context):
   rows_to_save = vector(mode = 'numeric')


   for (I in 1:length(rw)) {
     my_row=as.numeric(rw[i])
     current_strand = genome[my_row,7]
     if (current_strand == "+") {
       other_strand = "-"
     }
     else if (current_strand == "-") {
       other_strand ="+"
     }


     strand = current_strand


     while (strand != other_strand) {
       my_row = my_row-1
       strand = genome[my_row, 7]
       different = my_row
       start = different+1
     }
     strand = current_strand
     while (strand != other_strand) {
       my_row = my_row+1
       strand = genome[my_row, 7]
       different = my_row
       stop = c(different-1)


     }
     #save the desirable table lines
     rows_to_save = c(rows_to_save,start:stop)
     #print(genome[start:stop, 9])


   }
   #Create the dataset with the desired rows
   table = data.frame(genome[rows_to_save,], stringsAsFactors = FALSE)


   #Add a column with the gene length
```

```
   table$length = abs(table[,"stop"] – table[,"start"])


   #If the script gives errors on the for cycle starting on line 166, that
means that the protein is already shown in the desired format or there is no
genes correspondent to the search.

   #In either one or another case, take of the "#" from the next line and
finish the run here

   #write.table(as.data.frame(table[,1:14]),file="#FILE NAME#.csv",
quote=F,sep="\t",row.names=F)


   #Add a column with the row's names

   table$rows = rownames(table)


   #Change the row's names to the row's numbers

   rownames(table) = 1:nrow(table)


   #Add a columns with the row's numbers

   table$rowsnbrs = rownames(table)


   #Change the row's names to the initial ones

   rownames(table) = table[,"rows"]


   #Select from dataset "table" the rows numbers of the genes annotated as
"ferredoxin"/"flavodoxin"

   #(just change the word from "ferredoxin" to "flavodoxin"):

   fds_selected = grep("ferredoxin", table$function., ignore.case = TRUE)

   #print(fds_selected)


   #To only show the lines before and after if those lines are on the same
strand as the selected genes


   #Discover the lines' numbers to be presented on the final result

   this_row = as.numeric(0)

   row_before = as.numeric(0)

   row_after = as.numeric(0)

   prev_this_row= as.numeric(0)

   after_this_row = as.numeric(0)


   for (I in 1:length(fds_selected)) {

     this_row [i] = fds_selected[i]

     row_before[i] = fds_selected[i]-1

     row_after[i] = fds_selected[i]+1
```

```
  if(row_before[i] != 0){
    if(table[row_before[i], "strand"] == table[this_row[i], "strand"]){
      prev_this_row [i] = row_before[i]
    }
    if(table[row_after[i], "strand"] == table[this_row[i], "strand"]){
      after_this_row [i] = row_after[i]
    }
  }
}
prev_this_row = prev_this_row[!is.na(prev_this_row)]
after_this_row = after_this_row[!is.na(after_this_row)]

#Group all the lines' numbers
all_rows = c(this_row, prev_this_row, after_this_row)
#print(all_rows)

#Sort all the lines' numbers in decreasing order
all_rows = sort(all_rows)
#print(all_rows)

#Make sure that there are no repeated numbers
all_rows = unique(all_rows)
#print(all_rows)

#Create the dataset with all desired lines
table2 = data.frame(table[all_rows,])

#Add blank lines between each group of ferredoxin/neighbor genes

#Add a column with the row's names
table2$rows = rownames(table2)

#Change the row's names to the row's numbers
rownames(table2) = 1:nrow(table2)

#Add a column with the row's numbers
table2$rowsnbrs = rownames(table2)
```

```
   #Change the row's names to the initial ones
   rownames(table2) = table2[,"rows"]

   #Add the blank lines when the difference between rows' number is higher
than 1
   the_row = numeric(0)
   next_row = numeric(0)
   rows_numbers = numeric(0)

   for(I in 1:nrow(table2)){
     the_row[i]=as.numeric(table2[I,"rows"])
     next_row[i]=as.numeric(table2[i+1,"rows"])
     #print(the_row)
     #print(next_row)
     if(next_row[i]-the_row[i] > 1){
       rows_numbers[i] = table2[i+1,"rowsnbrs"]
     }
     #break(next_row[i]==the_row[i])
   }

   rows_numbers = rows_numbers[!is.na(rows_numbers)]
   rows_numbers = as.numeric(rows_numbers)

   #Everytime a blank line is added, it is added another line to the dataset.
The next for loop is used so that all the blank lines are added to the right
place
   for (I in 1:length(rows_numbers)) {
     rows_numbers[i] = rows_numbers[i]+(i-1)

   }
   #Create the dataset with the desired rows and blank lines
   table2 = insertRows(table2, rows_numbers)

   #Create a file to save the results
write.table(as.data.frame(table2[,1:14]),file="#FILE NAME#.csv",
quote=F,sep="\t",row.names=F)
```

APPENDIX II

An example of the calibration curves used in chapter 3 (2-methyl Tryptophan Production in *Escherichia coli*) and chapter 4 (Fosfomycin Production in *Escherichia coli*) of this dissertation are presented below in ***Figures II.1*** and ***II.2***. The necessity of a calibration curve relies on the fact that each LC-MS run has slight differences on the machine accuracy. Thus, once all samples were not run at the same time and there are some samples that run several times (due to errors), it is necessary to run calibration curve samples in each run to obtain the concentration of fosfomycin present in the samples and to be possible to compare results from different runs.

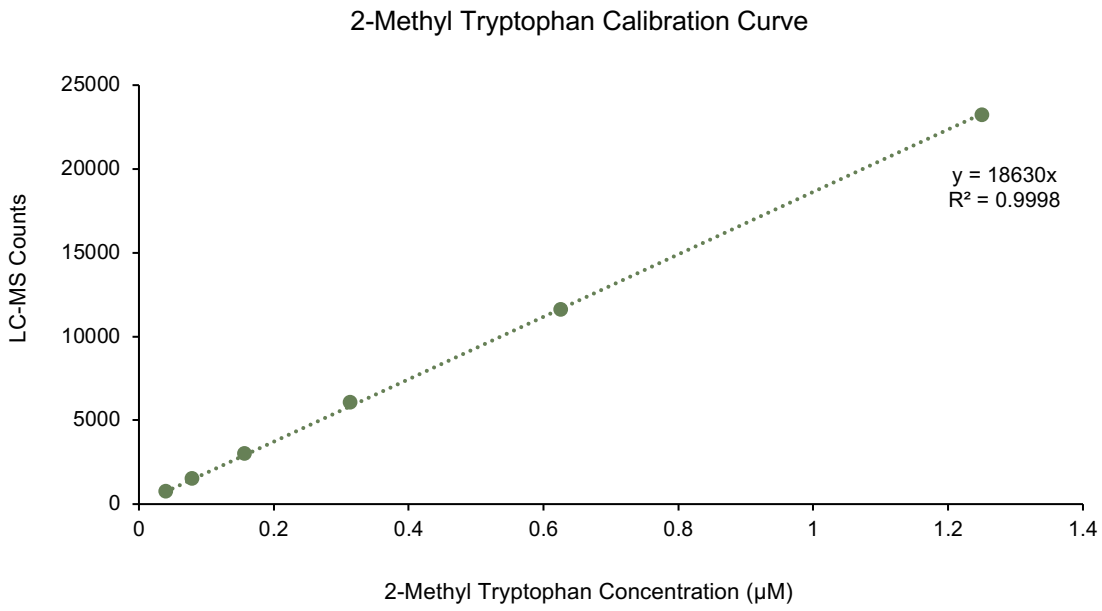

Figure II.1 *- 2-Methyl Tryptophan Calibration Curve. The calibration curve was constructed by measuring 2-Methyl Tryptophan samples with different known concentrations (39 nM, 78 nM, 156 nM, 313 nM, 625 nM and 1.25 µM).*

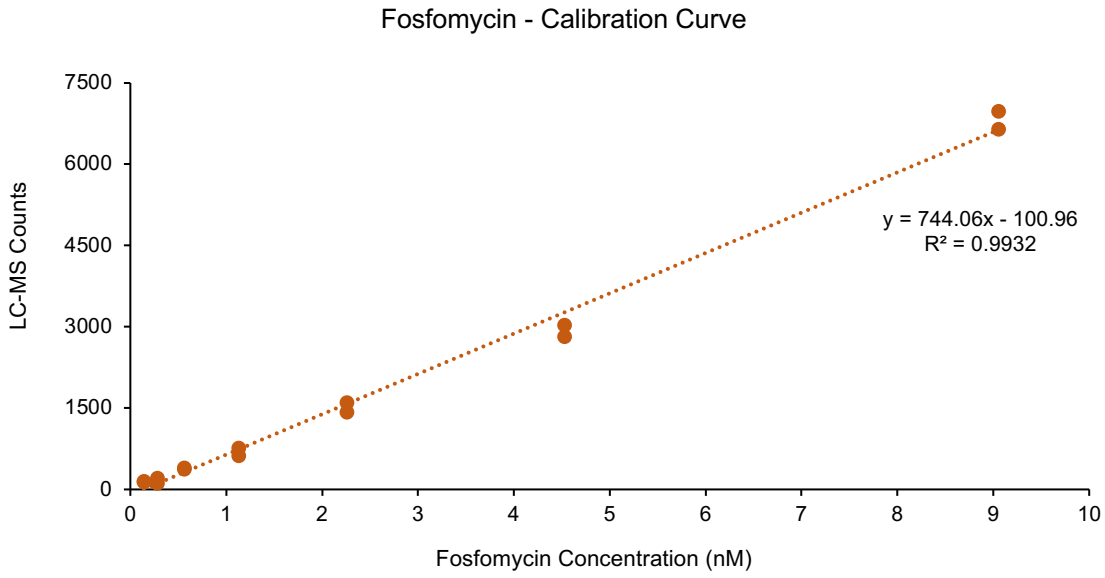

Figure II.2 *- Fosfomycin Calibration Curve. The calibration curve was constructed by measuring fosfomycin samples with different known concentrations (0.141 nM, 0.283 nM, 0.566 nM, 1.132 nM, 2.264 nM, 4.527 nM and 9.054 nM).*

APPENDIX III

The final concentrations obtained by LC-MS measurement and using a model of the calibration curves previously shown (**Appendix II**) are presented in ***Table III.1*** (2-methyl tryptophan results, chapter 3) and in ***Table III.2*** (Fosfomycin results, chapter 4).

Table III.1 *- Final 2-Methyl Tryptophan Concentrations (μM) obtained by calibration curve application to LC-MS measurements.*

SAMPLE	2-METHYL TRYPTOPHAN CONCENTRATION (μM)	STANDARD DEVIATION
NCM3772 pTsRM	0.0047	0.0042
NCM3772 pBtu + pTsrM	0.0125	0.0050
NCM3772 pRFP + pTsrM	0.0041	0.0006
NCM3772 pBtu + pTsrM (-)Cob	0.0029	0.0041
NCM3772 pBtu + pTsrM_Scatt3Fd	0.1173	0.0447
NCM3772 pRFP + pTsrM_Scatt3Fd	0.5842	0.1478
NCM3772 pBtu + pTsrM_ScattFd1	0.2061	0.0233
NCM3772 pBtu + pTsrM_ScattFd2	0.0057	0.0026
NCM3772 pBtu + pTsrM_ScattFd3	0.0077	0.0006
NCM3772 pBtu + pTsrM_ScattSufUT	0.0028	0.0022
NCM3772 pBtu + pTsrM_Scatt3Fd_SufUT	0.2243	0.0102
NCM3772 pTsRM_Slau3Fd	0.0445	0.0071
NCM3772 pBtu + pTsrM_Slau3Fd	0.3153	0.0879
NCM3772 pRFP + pTsrM_Slau3Fd	1.1169	0.0788
NCM3772 pBtu + pTsrM_Slau3Fd (-)Cob	0.0031	0.0024
NCM3772 pTsRM_Slau3Fd_SufUT	0.0433	0.0121
NCM3772 pBtu + pTsrM_Slau3Fd_SufUT	0.5195	0.0248
BL21 pTsRM	0.0064	0.0051
BL21 pBtu + pTsrM	0.0422	0.0075
BL21 pRFP + pTsrM	0.0036	0.0019
BL21 pTsRM_Slau3Fd	0.0458	0.0188
BL21 pBtu + pTsrM_Slau3Fd	0.1748	0.0610
BL21 pBtu + pTsrM_Slau3Fd_SufUT	0.1643	0.0580

BL21 pRFP + pTsrM_Slau3Fd_SufUT	0.2957	0.1115
pBtu + pPhPK	0.0042	0.0008
Cobalamin Concentration Assay (NCM3772 pBtu + pTsrM_Slau3Fd)		
0 μM	0.0031	0.0023
0.05μM	0.0570	0.0176
0.25μM	0.0806	0.0243
0.5μM	0.0396	0.0102
1μM	0.1335	0.0385
2μM	0.1722	0.0997
7μM	0.3153	0.0879
15μM	0.9583	0.0864
30μM	0.7412	0.0464
60μM	0.2897	0.0420

Table III.2 *- Final Fosfomycin Concentrations (nM) obtained by calibration curve application to LC-MS measurements.*

SAMPLE	FOSFOMYCIN CONCENTRATION (nM)	STANDARD DEVIATION
NCM3772 pBtu + pFom 5h Pellet	0.3006	0.0543
NCM3772 pBtu + pFom 24h Pellet	0.2476	0.0108
NCM3772 pBtu + pFom_Scatt3Fd 5h Pellet	0.4485	0.0495
NCM3772 pBtu + pFom_Scatt3Fd 5h Supernatant	0.2150	0.0032
NCM3772 pBtu + pFom_Scatt3Fd 24h Pellet	0.2273	0.0308
NCM3772 pBtu + pFom_Scatt3Fd 24h Supernatant	0.2288	0.0451
NCM3772 pBtu + pFom_ScattSufUT 5h Pellet	0.2202	0.0255
NCM3772 pBtu + pFom_ScattSufUT 5h Supernatant	0.2213	0.0148
NCM3772 pBtu + pFom_ScattSufUT 24h Pellet	0.2207	0.0012
NCM3772 pBtu + pFom_ScattSufUT 24h Supernatant	0.1941	0.0313
NCM3772 pBtu + pFom_Scatt3Fd_SufUT 5h Pellet	0.3844	0.0860
NCM3772 pBtu + pFom_Scatt3Fd_SufUT 5h Supernatant	0.2238	0.0121
NCM3772 pBtu + pFom_Scatt3Fd_SufUT 24h Pellet	0.2230	0.0145
NCM3772 pBtu + pFom_Scatt3Fd_SufUT 24h Supernatant	0.2358	0.0144
BL21 pBtu + pFom 24h Pellet	0.0708	0.0197
BL21 pBtu + pFom_ScattSufUT 24h Pellet	0.7976	0.1698
BL21 pBtu + pFom_Scatt3Fd_SufUT 5h Pellet	1.2423	1.2292
BL21 pBtu + pFom_Scatt3Fd_SufUT 5h Supernatant	5.5947	2.0396
BL21 pBtu + pFom_Scatt3Fd_SufUT 24h Pellet	0.2478	0.0090
pBtu + pTsrM	0.4404	0.2502