UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
DEPARTAMENTO DE BIOLOGIA CELULAR E MOLECULAR

MARIA JULIANA ROLON ROJAS

Development of synthetic biology tools by methods *in silico* and *in vivo* applied to bacteria of biotechnology importance

Ribeirão Preto
Brasil
2019

MARIA JULIANA ROLON ROJAS

Versão Corrigida

Development of synthetic biology tools by methods *in silico* and *in vivo* applied to bacteria of biotechnology importance

Dissertação apresentada à faculdade de Medicina de Ribeirão Preto na Universidade de São Paulo para a obtenção do título de Mestre em Ciências – Área de concentração: Biologia Celular e Molecular

Orientador: Prof. Dr. Rafael Silva Rocha

Ribeirão Preto
Brasil
2019

ROLON ROJAS, María Juliana
Development of synthetic biology tools by methods *in silico* and *in vivo* applied to bacteria of biotechnology importance, 2019
89f. p. : il. ; 30 cm

Dissertação de Mestrado, apresentada à Faculdade de Medicina de Ribeirão Preto/USP. Área de concentração: Biologia Celular e Molecular.
Orientador: Prof. Dr. Rafael Silva-Rocha.
1. Biología Sintética 2. Plasmídeos 3. Tags de Degradação 4. Biossensores 5. *Pseudomonas putida KT2440* 6. Bioinformatica. 7. Modelagem de Proteínas 8. Docking Molecular

Aluna: María Juliana Rolon Rojas

Título: Development of synthetic biology tools by methods *in silico* and *in vivo* applied to bacteria of biotechnology importance.

Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do título de Mestre em Ciências.
Área de concentração: Biologia Celular e Molecular.
Orientador: Prof. Dr. Rafael Silva Rocha

Aprovado em: ___/___/___

Banca Examinadora

Prof(a).Dr(a).:_____

Instituição:_____

Assinatura:_____

Prof(a).Dr(a).:_____

Instituição:_____

Assinatura:_____

Prof(a).Dr(a).:_____

Instituição:_____

Assinatura:_____

AGRADECIMENTOS

Primero quisiera agradecer a mis padres Elsa y Nelson, a ellos están dedicados cada uno de mis logros; como al resto de mi familia, en especial a mi Tía Hilda que es una mirada al pasado de como el esfuerzo forma grandes personas y para mi ahijado Santiago en el que veo el futuro y a quien amo con toda mi alma. A mi gran amigo Jorge Silva por introducirme en el mundo de la ciencia y en gran parte formar mi pensamiento crítico. A mi casi hermano Fenner por sus palabras que siempre me abrazaron en los momentos en los que los kilómetros de distancia abalaban mi corazón y a las personas que al final de este camino no puedo dar la mano, pero de las que me siento infinitamente grata mi abuelo Ernesto y mi nona Rosa.

Agradezco a mi orientador, el Profesor Rafael Silva, por ser un líder inspirador, por introducirme en el mundo de la Biología sintética, recibirme en su laboratorio e infundir junto con la Profesora Eugenia la búsqueda de un “alem” en mi. A mis segundos orientadores durante estos dos años Leonardo, Lummy y Caua que siempre estuvieron dispuesto a regalarme su tiempo y conocimiento con la mayor paciencia. Por su puesto, a todos mis compañeros del laboratorio de Biología Sintética y del Laboratorio de Metagenómica funcional que hicieron del día a día una experiencia increíble. Además, a mi galera latina que siempre con unas palabras en español consiguieron aliviar la distancia de casa en especial a Alejandro y David, amigos que llevo para el resto de mi vida y a Adriano que fue mi gran compañero, con el que además del camino académico compartí mi vida.

Finalmente agradezco por el apoyo financiero a la Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) y a la Coordenação de aperfeiçoamento de pessoal de nivel superior (CAPES) sin las que no hubiese sido posible realizar esta disertación junto con el Departamento de Biología Celular y Molecular de la FMRP-USP, que me formo con profesores extraordinarios que me enseñaron realmente a pensar de forma critica y a todos los alumnos y funcionarios que de alguna forma me ayudaron en este proceso.

“Los seres humanos no nacen para siempre el día en que sus madres los alumbran, sino que la vida los obliga a parirse a sí mismos una y otra vez”

Gabriel García Márquez.

GENERAL INDEX

RESUMO

ROLON, Rojas Maria Juliana. Desenvolvimento de ferramentas de biologia sintética por métodos *in silico* e *in vivo* aplicados a bactérias de importância biotecnológica. 2019. 80F. Dissertação (Mestrado) do Departamento de Biologia Celular e Molecular de Bioagentes Patogênicos, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo USP.

Apesar dos enormes avanços científicos e biotecnológicos alcançados pela Biologia Sintética nas últimas décadas, uma das grandes limitações em seu desenvolvimento consiste na baixa diversidade de partes biológicas padrão (caracterizadas qualitativa e quantitativamente) que permitem a construção de circuitos sintéticos complexos. Nesse contexto, o crescente número de genomas bacterianos sequenciados, as ferramentas bioinformáticas e o avanço das técnicas moleculares permitem a exploração desses sistemas. Essa abordagem permite a exploração dos diversos elementos genéticos para criar ferramentas sintéticas ortogonais bio-inspiradas em microorganismos que evoluíram para conter em seus genomas elementos genéticos que poderiam ser úteis para fins biotecnológicos. Um dos microorganismos mais notáveis que se encaixam nessa descrição é *Pseudomonas putida* KT2440, devido à sua grande plasticidade e controle transcricional, o que permite degradar mais de 100 compostos aromáticos derivados da lignina. Assim, no presente trabalho, foram projetadas novas ferramentas para a análise de fatores transcricionais com seus elementos reguladores *cis*, o que resultou em quatro novos vetores com capacidade de validação em um grande número de espécies de bactérias diferentes, com base na arquitetura da família dos plasmídeos pSEVA; com sistemas repórter contendo as proteínas sfGFP (super folding Green Fluorescence protein), mCherry e etiquetas de degradação ajustáveis à temperatura de crescimento de cada microorganismo. Além disso, estudamos uma abordagem de modelagem de proteínas *in silico*, modelando fatores transcricionais do genoma de *Pseudomonas putida* e, em seguida, realizamos um acoplamento molecular e virtual screening usando os principais compostos aromáticos que foram relatados na literatura como degradados por esta bactéria e um banco de dados com drogas aprovadas pela FDA. Isso resulta na seleção de proteínas candidatas da família MarR PP_3359 e VanR, que respondem a moléculas da via de degradação do ácido ferúlico; E o regulador de proteínas GalR da familia LysR, que responde ao ácido gálico, aos homólogos e algums medicamentos, a fim de criar ferramentas para a biologia sintética inspiradas em sistemas naturais para aplicações biotecnológicas.

Palavras chave: Biologia sintética, biossensores, bioinformática.

ABSTRACT

ROLON, Rojas Maria Juliana Development of synthetic biology tools by methods *in silico* and *in vivo* applied to bacteria of biotechnology importance. 2019. 89F. dissertation (Master) of the Department of Cellular and Molecular Biology of Pathogenic Bioagents, Faculty of Medicine of Ribeirão Preto, University of São Paulo USP.

In spite of the enormous scientific and biotechnological advances achieved by Synthetic Biology in the last decades, one of the great limitations in its development consists of the low diversity of standard biological parts (qualitatively and quantitatively characterized) that allow the construction of complex synthetic circuits. In this context, the increasing number of sequenced bacterial genomes, the bioinformatic tools and the advancement of molecular techniques allow the exploitation of these systems. Such an approach allows the exploration of the diverse genetic elements to create orthogonal synthetic tools bio-inspired in microorganisms that have evolved for containing in their genomes genetic elements that could be useful for biotechnological purposes. One of the most remarkable microorganisms that fit with this description is *Pseudomonas putida*, due to its great plasticity and tightly transcriptional control which allow it to degrade more than 100 aromatic compounds derived from lignin. Thus, in the present work, new tools for the analysis of transcriptional factors (TFs) with their *cis*-regulatory elements were designed, which resulted in four new vectors with validation capacity in a large number of different bacteria species based on pSEVA plasmids family architecture; with reporter systems containing the proteins sGFP (super folding green fluorescent protein), mCherry and degradation tags adjustable with the growth temperature of each microorganism. Also, we studied an *in silico* approach modelling proteins by threading in TF from the genome of *Pseudomonas putida* and then, we perform a molecular docking using the major aromatic compounds that have been reported in the literature as degraded by this bacteria. This result in the selection of candidates proteins from the MarR family PP_3359, VanR which response to molecules of the degradation pathway to Ferulic acid; And the protein GalR, LysR type regulator which responds Galic Acid and homologues in order to create tools for synthetic biology-inspired by natural systems for biotechnological applications.

Keywords: Synthetic biology, biosensors, bioinformatics.

I. INTRODUCTION

I INTRODUCTION

1. Synthetic Biology

It unveiled the discovery of elegance in DNA structure was unveiled in the ‘80s by the biologist James Watson, the physicist Francis Crick and the chemist Rosalind Franklin[1]¹¹¹¹ They laid the first rules of the molecule of life that most living beings Bluntly and simplistically: a double helix structure formed by Adenine-Thymine and Guanine-Cytosine linked by hydrogen bonds, which are supported by a phosphate and deoxyribose sugar skeleton forming a helical structure of grooves separated by standard distances. These first brushes on the molecule of life promoted molecular biologists to experiment with this configuration until they reach methods of sequencing and synthesis in the laboratory.

This knowledge led to a natural step, the manipulation of DNA. It was under this premise it created synthetic biology, covering interdisciplinary areas such as biology, computing, physics, and electrical engineering to achieve its primary objective; Extract parts of DNA of interest (genes, promoters, terminators, TFs among others) from living systems to test, validate and reassemble them in other living organisms for program them to perform new functions. [2–4]

Validating the DNA molecules of interest in foreign guests by recreating unnatural chemical systems can reveal emerging properties of the original systems and give information to predict the behavior of the cells in which DNA was synthetically introduced[5]. Because of this workflow, we get sufficient information to reprogram organisms to perform specific cellular functions. It is a concept that can study genetic material and information in many organisms, synthetic biology has covered applications in different areas ranging from basic research, through applications for human health and mainly biotechnology uses.[6,7]

The incredible expansion of this field also meant the finding of bottlenecks of this nascent area, because of the demand of these biological parts that worked for original purposes with predictable functionality. Thus, recent research trends in synthetic biology have focused their efforts on expanding tools that can design increasingly complex and different biological systems.[8]

1.1 Tools in Synthetic Biology

Looking at for additional parts of DNA for the development of synthetic biology is important because It is way how this area can serve other fields. These tools have been more widely explored in prokaryotic organisms because they have smaller and minimalist genomes that can be exploited more easily, besides, bacteria are accessible to manipulation in the laboratory and it optimizes most molecular biology techniques in these organisms[9,10]. For this reason, it described the tools available to bacteria in this section.

1.1.1 Plasmids

Plasmids are non-essential non-chromosomal genomic components found in bacteria, they replicate themselves autonomously in a controlled manner, these are mainly circular and in nature in sizes from 1 to 1000 Kbs[11] containing genes that confer adaptive advantages and increase survival under specific conditions; we have classified these according to the characteristics that confer:

-Resistance: Give bacteria the ability to oppose antibiotics, heavy metals, enzymes, UV light, toxic agents, bacteriophages, and/or bacteriocins.

-Energy metabolism: Which are the ones that carry genes that allow the use of alternative sources of nutrients for catabolism and anabolism.

-Pathogenicity and symbiosis: These plasmids carry genes that produce endo/exotoxins, enzyme transport, colonization, serum resistance, capsule production, or iron sequestration.

-Dissemination and perpetuation: They load the genes that allow conjugation to pass the plasmids to other microorganisms and replicate them.

Besides, the vectors to loading these resistances genes have their replication mechanisms that can be of the theta type, which mimics the chromosomal replication or by rolling circle. Both kinds of replication start at a specific sequence, the origin of replication namely OriV (vegetative origin) this DNA sequence determines properties such as host amplitude and the number of copies it has inside the cell (which can be low, medium or high).

Another important sequence in vectors is the systems of partition and maintenance, used for guaranteeing the segregation of plasmids in daughter cells when the bacteria replicates. It varies these mechanisms; for example, high copy number vectors perform a stochastic partition because they are highly distributed in cells. Vectors with low copy numbers have genes that encode proteins for these purposes such as ParA and ParB that bind to a DNA sequence called ParS to address the vectors in the median region of the cell. Futhermore, another interesting mechanism is one that guarantees cells possesses at least one copy of the vector to survive in an antitoxin-toxin system, so the plasmid contains a gene that encodes a protein that kills the cell if it is expressed (among these targets can be DNA gyrase, translation, and replication) and also has a gene whose product an antitoxin that binds to the promoter region of the gene to prevent it from occurring; This antitoxin product is very unstable and has to be produced continuously causing death in daughter cells that do not possess this plasmid[1,12].

The ability to self-replicate, maintained for generations, be easily synthesized, manipulated in the laboratory and function of transport elements to deliver foreign DNA made by vectors the milestone for genetic manipulation and from we base them the classical cloning strategies. Thanks to vectors, molecular mechanisms have been understood and cellular expression patterns were written, providing important information used for optimized and improvement over the years the main tools that boosted biotechnology to produce numerous of molecules of interest (production of insulin, antibiotics, vaccines among others)[13,14,15,16,17]

This manipulation of vectors began with a minimization of the natural size in which were found in bacteria, because like in genomes it possesses DNA product of evolutionary traces that are not useful for the laboratory goals. Besides, the use of crucial components such as selection markers, which are genes that give a characteristic that allows cells that have genes of interest to be separated from those that do not; these can be of auxotrophy, of drug resistance or alpha complementation and finally to facilitate their manipulation and as a last part to introduce DNA of interest a multiple cloning site (MCS)

that contains many restriction enzyme recognition sites. It classifies plasmids in molecular biology into:

-Cloning vectors: The most used, its aim is to load exogenous DNA to produce an outrageous amount of it, for this reason, most of them charge a high number of copies, these vectors are usually small to be easier to purify. Some of these are hybrid and contain replication sites that can serve in several organisms or even contain two replication sites for each organism (even eukaryotes) and are called shuttle vectors. Although most of the cloning vectors can be easily transformed, some of them contain besides to their OriV a transfer origin (OriT), this allows it to be mobilized to other microorganisms by conjugative transfer.

-Expression vectors: Mainly used for the production of recombinant proteins of interest, these have a strong promoter next to their MCS responsible for controlling the production of a protein that can respond to different incubation conditions and culture media. (The most commonly used in the classical methods are the Lac operon that can be induced by the addition of a lactose-like substance called IPTG and sometimes a modification of these promoters that makes a tighter control depending on the concentration that is added of this substance that is the *tac* or *trc* promoters) these promoters are important because they control the amount of protein. In some cases, these can be cytotoxic to organisms, so a powerful control of protein production is very strict. It focuses the principal purpose of these vectors on the biosynthesis of products. Thus, the proteins after being produced by the bacteria need to be purified, for these cases, they have a peptide fusion system that allows affinity to separate them into a solid matrix. The most widely used strategy is the fusion of proteins with a peptide-rich in histidines with a link to the protein that facilitates their separation by affinity chromatography and subsequent removal, so it does not interfere with the tertiary structure of the protein.

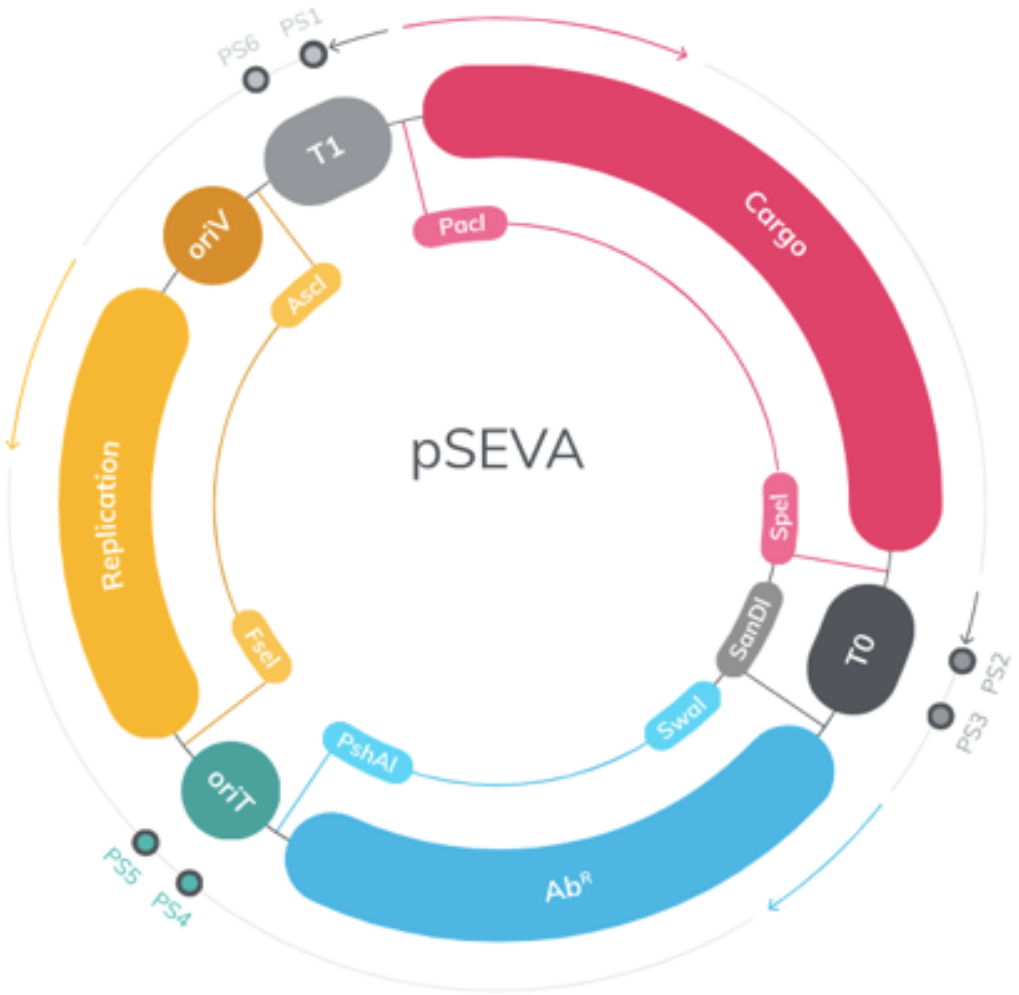
-Reporter vectors: Used to study the kinetics of transcriptional regulation *in vivo* with distinct types of conditions or promotors sequences. This is possible because contain indicator proteins, which can be fluorescent, luminescent, or enzymatic that can be measured over time to discover how genes, promoters, or conditions print strength in transcription. Studies of specific gene expression in bacteria have been greatly facilitated by the use of reporter genes. And these vectors have allowed a deep understanding of genetic regulation[18,19].

Thanks to these advances and knowledge, a family of modular vectors was designed in 2013, these have the particularity of being modular and can be broken down into a series of components that can be mixed and combined into a variety of configurations that could be replaced as needed; The Standard European Vector Architecture pSEVA is a plasmid system that represents minimalism and orthogonality in a very pragmatic and standardized way, these plasmids have a simple and elegant nomenclature that allows their recognition easily and quickly, with a platform where all the sequences are available (192 canonical vectors that have been verified and curated by the database managers, and 178 siblings (Sib) vectors allow some standardization criteria collection SEVA code standardized by other research groups)[20,21,22,23], has an on line platform (http://seva.cnb.csic.es/) which helps to choose the best vector according to different research goals and can be ordered for use. Therefore it has been widely used by researchers around the world (in a quick Google search it published 212 articles since 2014).

It makes these vectors up of six parts (Figure 1.A). Three parts of them are constantly in the architecture of all these vectors. Two strong transcriptional terminators (T0 and T1) dependent on the Rho protein, an OriT that allows it to be contained in several hosts and an MCS to put the charges. And the other three parts that compose them are variable and are flanked by enzymes that facilitate being exchanged at will. These are, antibiotic resistance genes (with sizes between 0.8 to 1.3 kbs), an OriV to define in which organisms this microorganism can be contained and the number of copies and finally unique types of charges. These interchangeable parts define the nomenclature of these vectors. Thus, the first number marks the antibiotic marker, the second the OriV, and the third the cargo that contains them. (Figure 1.B).

Figure 1. pSEVA Vector Family. A. The general structure of the SEVA plasmids family. the cargo (pink), two origin one of replication, one vegetative (dark yellow circle) and one of transfer OriT (green circle), the terminators T0 and T1 (dark gray, light gray respectively), the protein in charge of replication of the vector (yellow) and the gene that codes for antibiotic resistance that serves to make a positive selection (light blue). It indicates the directionality of the strongest transcription flow in the designed genetic device and the enzymes flanking between each of the spaces of each part. B. Expanded nomenclature of pSEVA. These vectors have 4 modules, organized in each of their names, the first corresponds to the resistance antibiotic marker (Listed 1-9 in the blue-delineated table), the second to the origin of replication (1-9 in the yellow table), the third position corresponding to the cargo (pink table listed from 1-13).

A.



B.

pSEVA XXX

No.	Ab*
1	Ap
2	Km
3	Cm
4	Sm/Sp
5	Te
6	Gm

No.	OriV
1	R6K
2	RK2
3	pBBR1
4	pRO1600/ColE1
5	RSF1010
6	p15A
7	pSC101
8	pUC
9	pBBR322/ROP

No.	Cargo
1	MCS-default
2	LacZ a-pUC19
3	LacZ a-pUC18
4	aclq-Ptrc
5	LacZ
6	LuxCDABE
7	GFP
8	xy/S-Pm
9	alkS-Palk
10	araC-Pbad
11	chnR-PchnB
12	cprK1-PDB3
13	PEM7

Taken and adapted from Silva Rocha, Martinez (2014)

1.1.2 Degradation Tags

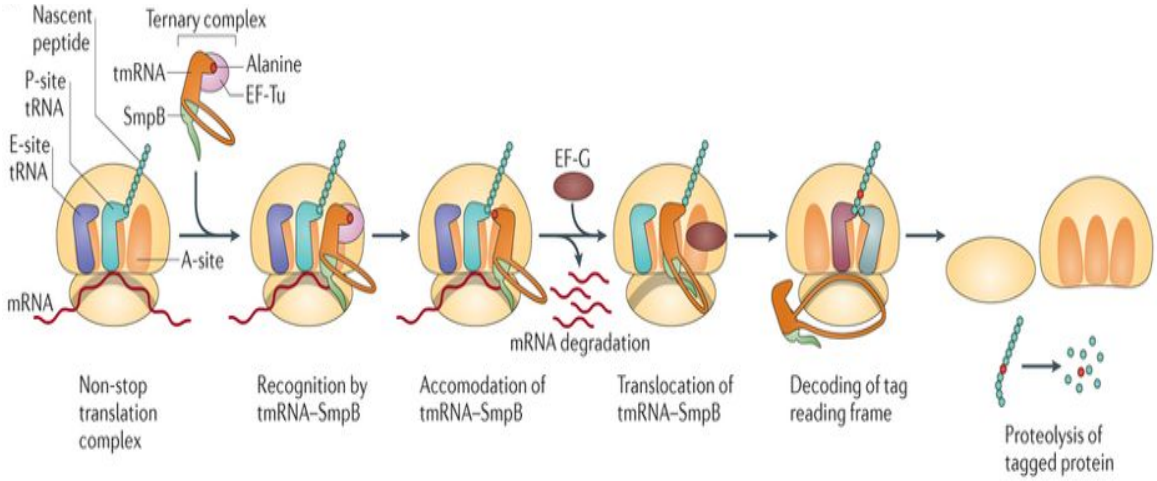
The degradation of synthetic proteins is important sometimes, especially when transient behaviors are being analyzed and real-time dynamics of gene expression are required to be evaluated. Reporter proteins that are more widely used in indicator vectors (Green fluorescent protein, galactosidase, and luciferase) have the disadvantage of being very stable (they could even persist for hours which would be reflected even in daughter generations when bacteria that replicate are studied quickly) and did not reflect a photo of the moment to have a kinetic transcript. [24]

This is how degradation tags was used in proteins produced heterologously in synthetic biology, the degradation tags are formed by sequences that encode specific short peptides in the n-terminal or c-thermal extremities of proteins so when they are translated, they will be recognized by the protease machinery of the prokaryotic cell and consequently decrease it is half-life.[25,26]

The solution to this problem is bioinspired in the endogenous process that bacteria activate to rescue ribosomes. This occurs during protein synthesis when a defective mRNA gets stuck in the complex formed by the ribosome and RNAt (because these junctions are very strong because evolutionarily the nascent peptide-RNAt and ribosome interactions are strongly anchored so as not to affect the processivity until beginning the termination process and not generate errors in the transcription); The bacterium activates a process called trans-transduction to reuse these ribosomes because the accumulation of these complexes decreases the ability of protein synthesis and can be toxic to the cell because this process spends the valuable machinery (ribosomes and amino acids) that can be used in other rounds of protein synthesis.

In trans-transcription a ribonucleoprotein system that contains a 183 nucleotide RNA that contains RNAm and RNAt (tmRNA) domains recognizes the stagnant ribosome, binds as an RNAt to the ribosome and adds an alanine to the defective nascent peptide, then a portion of this molecule acts as an RNAm replacing the rest of the original message of the peptide with a degradation tag that consists in 11 residues (AANDENYALAA) sequence and when they finish this process, they release the ribosome and the sequence of these peptides are recognized by the ClpXP and ClpAP proteases that form a complex that degrades it[25,27,28] (Figure 2).

Figure 2. Graphical representation of trans transcription. It shows how the ribosome is stagnant and this is recognized and by the tmRNA that together with the SspB protein articulates the tail of amino acids that are the target of proteolysis.



The figure is taken from E, Jabri. (2003)

Thus, in the synthetic systems where this mechanism is mimicked, the sequence encoding the signal amino acids is added to the protein sequence, so they are recognized by proteases and degraded. The standardization of this process has led to it being played with the white sequence determining that if the last 3 amino acids (called *ssrA* label) of this peptide are modified by AAV, ASV, LVA mainly to get differents half-lives of proteins.

1.1.3. Biosensors

It defines biosensors in synthetic biology as cellular sensors made by host cells that produce signals when they recognize a certain molecule. 35 years ago the first bases were laid for them to be explored when operons and cellular regulatory mechanisms were discovered in prokaryotes[29] that have been widely exploited to where today there are three main classes of sensors that are commonly designed of the synthetic form:

-Cytoplasmic regulatory proteins

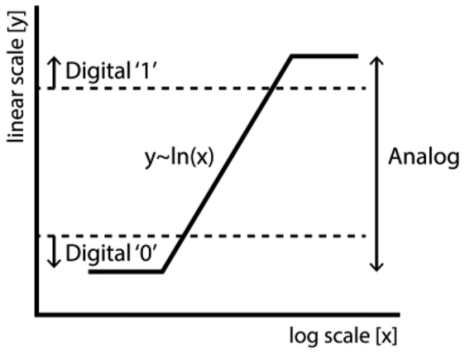
Comprise an inducible system that allows specific genes to be activated by adding a small molecule to the growth media. The inducing molecule passes through the membrane binds to the regulatory protein called TFs (See 2.2.2). This binding causes

conformational changes that pass these active or inactive proteins, this changes their affinity for a certain DNA sequence and causes a stronger or lower expression of the promoter to which it responds.

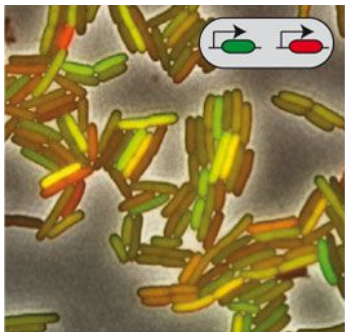
These biosensors can respond to different concentrations of inducing molecule and have a dynamic range that depends on each of them, sometimes, they have an all-or-nothing response (digital) or a response in which a dynamic range of concentration of the inductor can be adjusted to see a predictable linear regression behavior (analog)[30,31] (Figure 3.A). These inducible systems may have atypical behaviors at the single-cell level experimental studies revealed stochastic fluctuations because of the heterogeneity of the environment called noise. (Figure 3.B)[31]

Figure 3. A. The Dynamic range of digital and analog systems. B. *E. coli* expressing two identical promoters driving two different fluorescent proteins, in red and green, respectively leaving in evidence the noise in the intensity from cell to cell.

A.



B.



Taken from Woo, Sung, Sik Sarpeshkar (2015) and Eldar, Avigdor, Elowitz, Michael B. (2010)

-Two-component systems

Two-component systems are the most frequent in prokaryotes. These correspond to inter-membrane proteins with a part that goes to the extracellular surface and responds specifically to light, temperature, touch, metals, metabolites, and chemicals. When this occurs, the protein formed, and this can activate or repress the transcription.

-Environmentally sensitive promoters

The bacteria have systems for detecting environmental conditions such as pH, temperature, oxygen concentration, or UV light that can modulate the transcriptional response and have sensory inputs to manipulate cells.

Through these key parts and the grand world of synthetic biology tools, they grow and are optimized almost every day. This is possible because this area grows along with the development of molecular biology techniques to make a faster design of genetic circuits such as cloning methods that are becoming faster than are Golden Gate, Gibson, yeast-mediated Cloning, and Oligonucleotide Stitching. Genetic manipulation is more accessible by techniques such as CRISPR-CAS9 (short palindromic repeats grouped regularly spaced) that can make targeted mutagenesis in microorganisms and maps on bacterial regulation are constructed under many amounts of conditions to define along with the cellular behavior that they give the large-scale analysis of RNA-seq. Thus this area is constantly growing, planning and seeking to find more answers to the questions and challenges of the different areas it can reach. [1,32]

2. *Pseudomonas putida KT2440*

2.1 Generalities

Pseudomonas putida mt-2 is a gram-negative proteobacterium isolated from the soil that identified as degrading aromatic substrates such as toluene, *x-xylene*, and *p-xylene* for use as carbon and energy sources. It soon discovered that the metabolic nature of this bacterium was a product of the presence of the TOL pWWo a catabolic plasmid that possesses degradation genes by a meta-cleavage pathway for the use of toluene, *m-xylene*, *3-Ethyltoluene* and, *1,2,4-trimethylbenzene* and its derivatives of alcohol, aldehyde, and carboxylic acid and has a size of 116.5 kbs. And when this plasmid removed from strain mt-2, this variant of the bacterium called *Pseudomonas putida KT2440* which resulted in the state's assignment recognized as safe (GRAS) and a new tool for studying the degradation of aromatic compounds.[33–35]

The first analyzes made to *Pseudomonas putida KT2440* by classical culture methods revealed that it could grow using at least 100 different carbon fluxes including 50 aromatic compounds among which *benzoate, p-hydroxybenzoate, benzylamine, phenylacetate, phenylalanine, tyrosine, phenylethylamine, phenylhexanoate, phenylheptanoate, phenyloctanoate, coniferyl alcohol, p-coumarate, ferulate, caffeate, vanillate, nicotinate* and *quinate* (hydroaromatic compound). And other aromatic compounds, such as *2-hydroxybenzoate* (salicylate), *3-hydroxybenzoate*, *2,3-dihydroxybenzoate, 2-aminobenzoate* (anthranilate), *p-hydroxyphenylacetate, tyramine, aniline, atropine, 2-phenylethanol, phenol, mandelate, phenylglyo, p-methoxybenzoate*

(p-anisate), *3,4-dimethoxybenzoate* (veratrate), *p-hydroxy-3,5-dimethoxybenzoate* (syringate), *cinnamate, phenylpropionate, 3-hydroxyphenylpropionate, vanillylmandelate, pyridoxal, pinateinexal, pinateinexal, isinate quinoline, isoquinoline, gallate,* and *resorcinol*. It is the first clues of the great metabolic capacity that this microorganism possessed.[36]

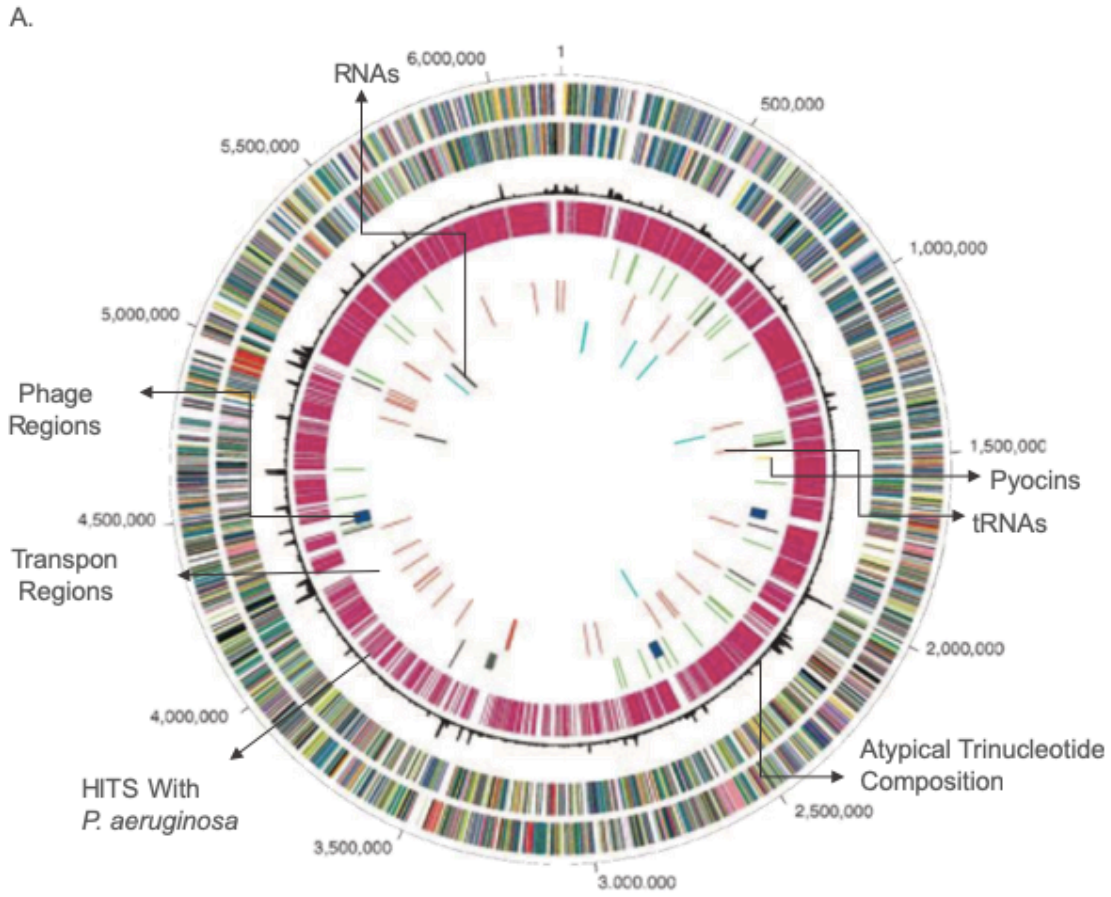
The complete analysis of the genome of P. putida KT2440 carried out for the first time in 2002 and reviewed in 2016[33,37]. A chromosome composed of 6,181,873 bp were found, with many of mobile elements, repetitive extragenic palindromic sequences of 35-40 bp, genome mosaic structure and 105 chromosomal islands (± 4000 kb) that exhibited a different pattern of the Guanine-Cytosine relationship where genes and regulatory regions involved in their metabolic capacity. (Figure 4)[38]

Besides revealing these characteristics, the genes of this bacterium written and described the metabolic maps of *P. putida KT2440*. This resulted in the discovery of peripheral metabolic pathways that this bacterium used for degrading aromatic compounds and the finding of their five main central pathways. Each one has the machinery to perform the aromatic ring cleavage by epoxidase, dioxygenases and reductases enzymes. These metabolic pathways are catechol (*cat*), protocatecatechuate (*pca*), phenylacetate (*pha*), homogentisate (*hmg*) and the catabolic pathway of N-heterocyclic aromatic compounds (*nic*) that are distributed throughout the entire genome[39–41] (Figure 5).

Because of many genes destined to degrade distinct types of carbon sources and because this microorganism developed in a highly dynamic environment where nutritional conditions can change quickly this bacterium has a very versatile transcriptional regulation. It can handle stress conditions or changes in the food source quick way through the repression or activation of one or several genes (local or global regulation) without involving a large energy expenditure for it. It achieves this through the transduction of signals made by regulatory proteins that have modular domains capable of interacting with environmental signals and gene transcription of the TFs[42].

Figure 4. Circular representation of the *P. putida KT2440* genome. The outer circle, coding regions predicted in positive and lower band color correspond to negative line genes concentrated by function category: salmon, amino acid biosynthesis; light blue, biosynthesis of cofactors, prosthetic groups and carriers; light green, cellular envelope; red cell processes; Brown, central intermediate metabolism; yellow, DNA metabolism; green, energy metabolism; purple, fatty acid, and phospholipid metabolism; pink, fate protein/synthesis;

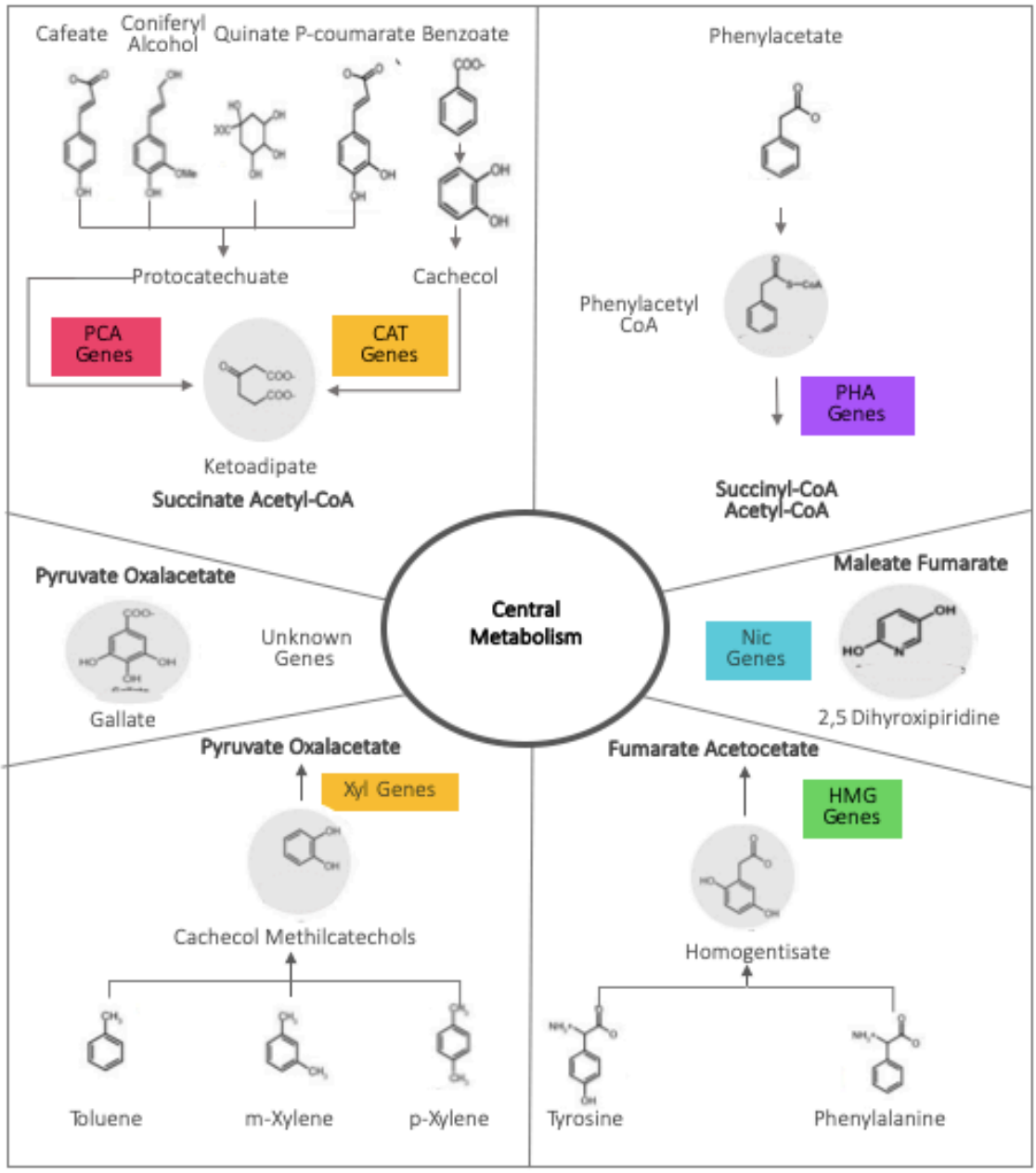
orange, purines, pyrimidines, nucleotides, nucleotides; blue, regulatory functions; gray, transcription; teal, transport and binding proteins; hypothetical, conserved black hypothetical proteins. The inner circles sample RNAs, trinucleotide composition, and a comparison with the *Pseudomonas aeruginosa* genome.



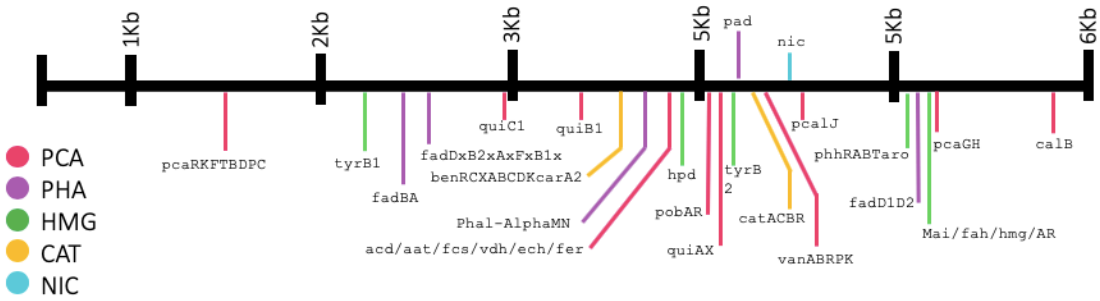
Taken and adapted from K.E Neslon (2002).

Figure 5. General representation of the degradation pathways of aromatic compounds present in *P. putida KT2440*. A. Schematic representation of the central metabolism of metabolic *P. putida*. Representing in yellow the cat genes and the *xyl* genes next to them the degradation pathway compounds in which it is involved, the protocatechuate branch with the PCA genes in pink, *pha* is represented in purple, the HMG homogenate pathway Green color and the *nic* genes of light blue. The main compounds of each path are indicated by grey circles for each path. B. Linear representation of the *Pseudomonas putida KT2440* chromosome showing the ubiquity of the genes involved in the catabolic pathways of degradation of different aromatic compounds with the same color code in Figure A.

A



B



Taken and adapted from Weinel, Christian, Nelson, Karen E. (2002).

2.2 Transcription of *Pseudomonas putida KT2440.*

2.2.1 Sigma Factors and Promotor Regions

One of the main regulatory steps in bacteria is the transcription control, the enzyme RNA polymerase (RNAp) being the key protein. The RNAp cannot start transcription alone, it must bind to a polypeptide, the sigma factor (σ). This factor interacts reversibly with the RNAp forming the called holoenzyme of RNAp. The RNAp bound to the σ factor can recognize a promoter sequence that leads to the formation of the closed complex to open complex with the DNA. Then, the holoenzyme RNAp synthesizes several fragments of 2 to 12 bp (pair base) until it releases the σ factor, allowing the chain to lengthen until the RNAp finds an ending region. The frequency in which RNAp starts transcription is determined by the strength of the promoter and is directly related to the sequence of the promoter region and the conformation of the DNA. The factors recognize the promoter region, which always joins two regions called consensus regions that, with exceptions, are located at -35 and -10 base pairs from the beginning of transcription.[43,44] Sigma factors have 4 regions conserved and according to their structure can be divided into 3 major groups:

-Primary factors: These are the factors that regulate the transcription of the basic survival genes of the organism and are responsible for the expression of exponential growth genes.

-Alternative factors: These factors differ significantly from the primary ones and are responsible for the transcription of specific regulons. In this group belong the factors σ 28, responsible for the expression of flagella proteins and σ32, responsible for the response to denaturing agents (ethanol, temperature increase, hydrogen peroxide)[45–47], and several other sigma factors that regulate the response to iron deficiency or even control the producing of pigments for light protection.

-Factors of the σ54 family: The σ54 family is structurally and functionally different from the proteins belonging to σ70 we can recognize no homologous structure between these two families. The σ54 factor has a constitutive transcription and its cell concentration corresponds to 15-20%, remaining constantly in different stages of growth of *E. coli* and *P. putida*. σ54 responds to a variety of protomers that have an architecture with the key sequence in positions −12 and −24. RNA polymerase can recognize these promoters

but cannot form transcriptionally competent open complexes without the intervention of a specialized transcription activator that contains an activator protein that binds σ54.

These proteins are TFs and are normally activated with the activation of cellular stress response genes (such as the aromatic compounds of *Pseudomonas putida*, the transport of dicarboxylates in *Rhizobium meliloti* and the anaerobic metabolism genes of *E. coli*), and the activation of genes of catabolic interest of atypical compounds.[4849]

2.2.2 Transcriptional Factors TFs

As stated earlier, there are additional promoter elements that contact RNAp and influence the start of transcription. What are the prokaryotic TFs that typically act as homodimers that bind to palindromic or pseudopalindromic DNA sites, among which the helix-turn-helix (HTH) domain (Figure 6[50]). These proteins control transcription through different mechanisms, such as modification of DNA binding affinity for small ligands (such as aromatic compounds), and covalent modifications (such as phosphorylation of regulatory proteins) or by varying protein concentrations. Which are regulated by their expression, location, degradation or a combinatorial effect of these factors.[47,51–54].

Figure 6. Crystallographic representation of TF GRDBD and smGRE complex in the *E. coli* expression system.

Different factors have emerged independently as regulators that are less conserved than their target genes and evolve independently of them. Therefore, there is a wide repertoire of TFs that respond to specifically promiscuous signals, although there is an expert knowledge of genotype-phenotype relationships in bacteria under relevant conditions.

Pseudomonas putida KT2440 has 423 TFs[55,56]. These proteins together with the sequences in the genome to which they can bind are responsible for the result of their tightly controlled metabolic pathways and TFs that respond to environmental conditions quickly and effectively [56] (Figure 7). These types of compounds can influence that have been structurally grouped. The first family and one of the most described is the AraC/XylS family that interacts with dimer-shaped regions of DNA. These proteins have been described by recruiting different sigma factors depending on environmental conditions (for example, σ32 in the first phase of exponential growth and σ38). Another group of regulators is those of the IclR family, in which proteins can self-regulate and

control the initiation of transcription in other regions of the DNA by optimizing the space between -35 and -10 boxes to recruit RNAp. The LysR family, where we find members that control the expression of the degeneracy pathways of catechol and chlorocatechol, helps RNA polymerase interact with the promoter and activate transcription; the members of the FNR and MarR families can perform various functions by activating or deactivating gene expression in response to different aromatic compounds. Besides, the GntR repressor family has members capable of interacting with the aryl-CoA compound. TFs, which are proteins sensitive to environmental variations with one or more DNA binding motifs. One of these proteins acts as an environmental stimulus sensor while the other interacts with the nucleic acid of specific sequences. to activate the activation and repression of transcription either. In *P. putida KT2440*, several TFs of the IclR (PcaR, HmgR), LysR (CatR), NtrC (PhhR) and MarR (FerR) families were identified. However, not all have been widely described, and the regulatory mechanisms that control gene expression are diverse and the extensive metabolic activity of this microorganism, there may still be many regulatory mechanisms not described [43,55,55–57].

2.2.3 Transcriptional control and the degradation pathway of Ferulic Acid and Gallic Acid

The catabolic route that uses as carbon sources degradation products of aromatic lignocellulose, such as fermentation, vanillin, ferulic acid, and protocatechuate in *P. putida KT2440* is found in two clusters with genes encoding enzymes that mediate its biotransformation. One of these clusters is the VanABK operon. In which VanK takes part in vanillin transport, and VanA and VanB form a vanillin degrading enzyme complex, vanillin-O-demethylase, which converts vanillate into protocatechuate. The expression of this operon is controlled by the VanR, a protein from the GntR family, which were previously described and crystallographer in the *Bacillus* spice in which the functional number of dominoes were revealed from bind to palindromic regions (together with pseudoboxes near of these regions17–19) of DNA that is only at the end of the repression, then, when two molecules of vanillin between the intervening cavities VanR and released from the DNA of the operator allowing RNA to transcribe two vanAB genes that modulate the expression of enzymes and proteins transport of the metabolic pathway by the affinity of binding to DNA in response to covalent modifications of effector molecules. These proteins are VanA and VanB that code for the vanillate O-demethylase oxygenase subunit and vanillate O-demethylase oxidoreductase genes respectively in the metabolic pathway of ferulic acid bioconversion[60,62–67] (Figure 8), and the enzyme Vdh vanillin dehydrogenase (PP_3358), a key to the bioconversion of vanillin to protocateuate in many of other bacterial species together with a hydroxycinnamoyl-CoA hydratase-lyase hydratase (PP_3357) and feruloyl-CoA- synthetase (PP_3356) that have strong evidence of being under the control of the TF PP_3359[61]. (Figure 8).

Figure 7. Circular representation of the *P. putida KT2440* genome. Blue Outer circle, coding regions predicted in positive and lower band color correspond to negative line genes. In the two circles that have interiors that have a gray shadow, the 423 TFs of *Pseudomonas putida KT2440* that are found throughout their genome positively and negatively respectively are shown. These TFs are divided by color the main families of TFs as follows: salmon color is the Cro / CI family transcriptional regulatory family, black for LuxR, orange LysR, ArsR / Asnc families are represented in red, GntR proteins Magenta, MarR of bluish green, AraC of Blue and TetR of dark green. In the other orange circle shows the distribution of GC and the circles inside the graph represent the PP_3359, PP_3738, PP_3946 and PP_2516 genes.

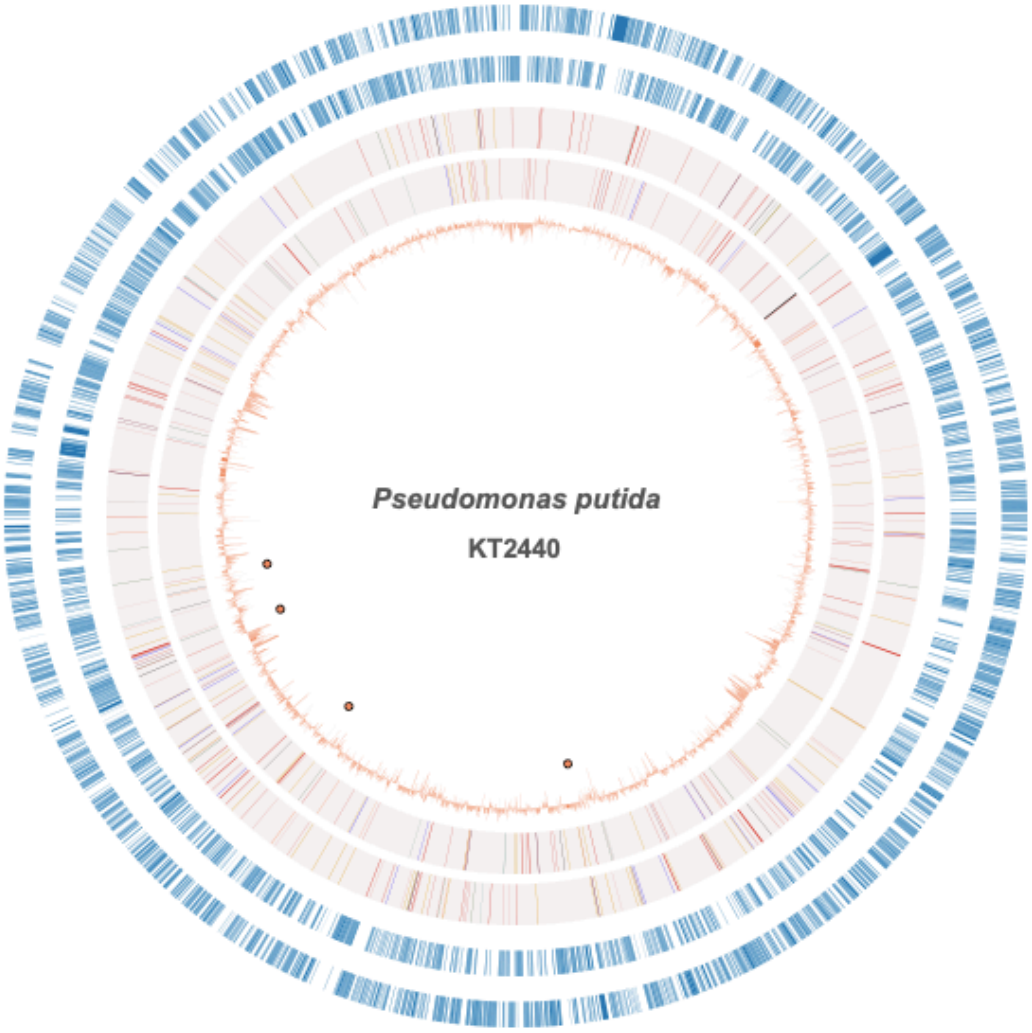


Figure made in online platform for cIVI[67] genome design with information on TFs obtained in Pseudomonas Genome DB[68]

This molecule contrasts with the characteristic marker of the MarR regulatory family in which were typecast according to its proteomic analysis. Most of these proteins act as genetic repressors and are divergently oriented to the proteins under their control. This design allows the TFs to specifically bind the site to the intergenic regions between genes

transcribed divergently to repress the transcription of both. In the binding of a small molecule ligand or response to the oxidation of specific cysteines, DNA binding is attenuated, resulting in the expression of the activated gene which in turn negatively self-regulates itself (Figure 9).

Carrying out a more in-depth study of these proteins, it was found that the VanR protein is very conserved in Pseudomonas sp. along with the genes it controls (in which VanA / B was found in the surrounding area without exception in all species only sometimes in the positive or negative strands). In a phylogenetic analysis, it was revealed that the protogenes and fluorescence species share 87% to 88% of the amino acid sequence with high coverage, unlike the *Pseudomonas aeruginosa PAO* that does not contain this gene so conserved because it is a very environmental gene and the use of compounds derived from lignin and the genotype of the latter is mainly pathogenic. We have *Bacillus subtilis* which they share a 17% Identity percentage and it is the closest protein found in the PDB database with a tertiary structure that acts as a repressor of the GntR Family that regulates NagR , genes involved in the absorption and metabolism of N-acetylglucosamine (GlcNAc)[54,69] (Figure 10.A) (Alignment data for this protein are shown in Appendix 1)

Figure 8. A. Metabolic pathway of degradation of Ferulic Acid. B. Schematic representation of the VanR and PP_3359 clusters with the genes they control.

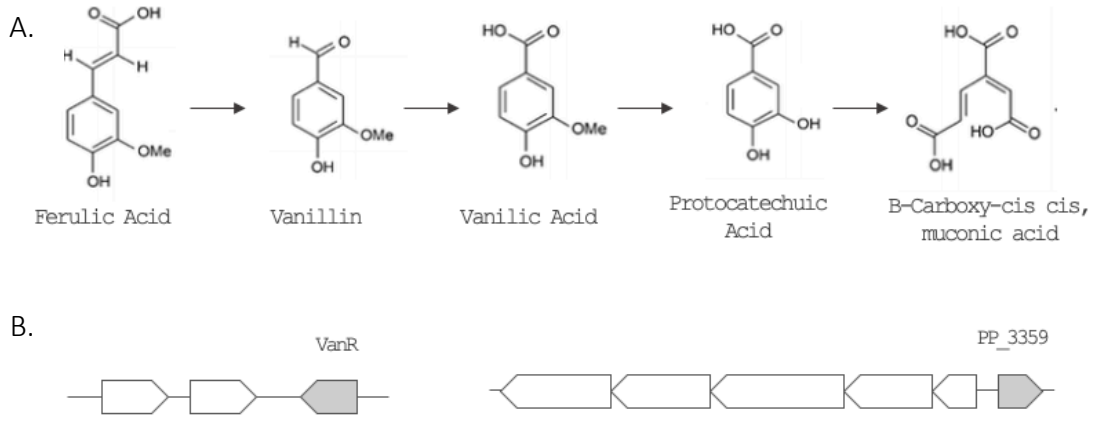


Figure 9. Representative homologs of MarR and its mode of gene regulation. In the one shown in the box above that when the TFs encounters ligand, it does not bind to DNA and allows the transcription machinery to begin the protein synthesis process.

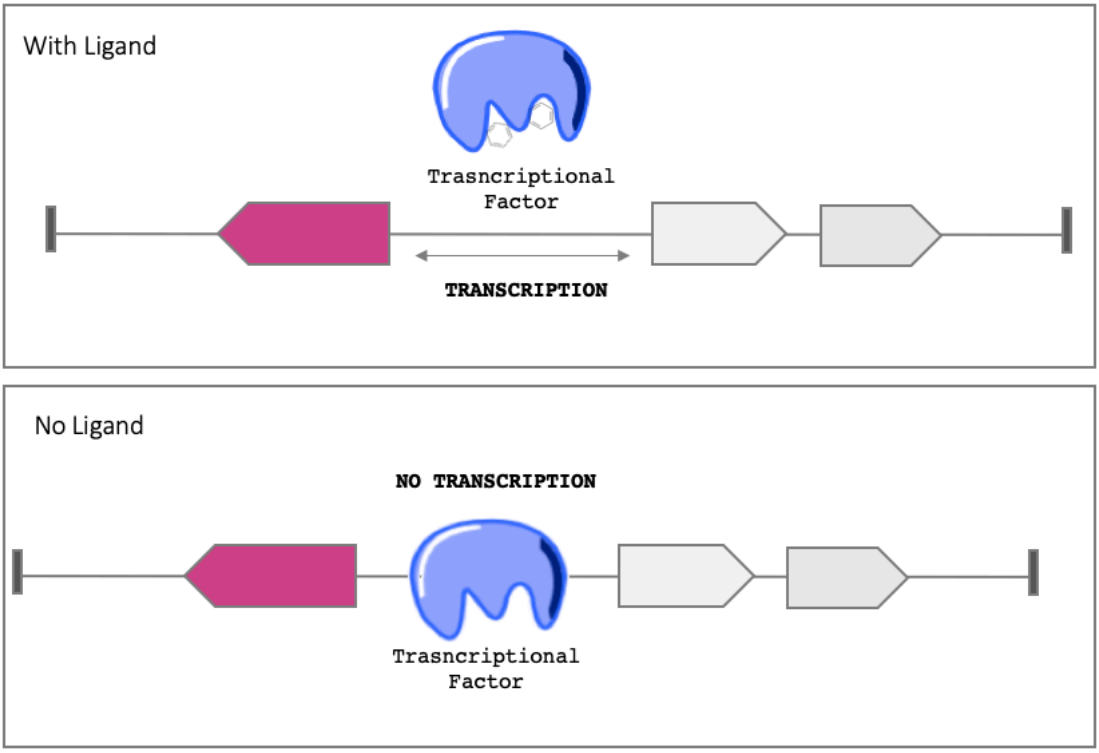
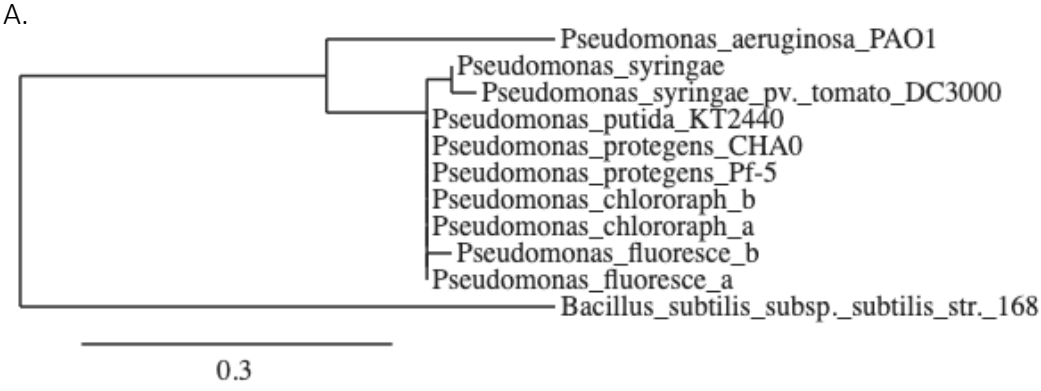
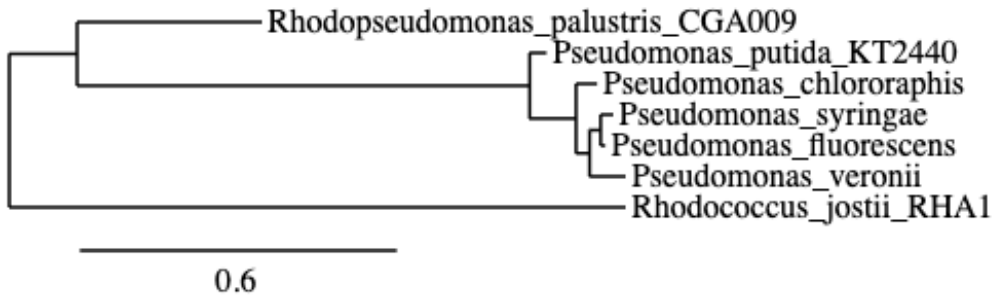


Figure. 10. A. Phylogenetic analysis of the amino acid sequence of the A. VanR protein. B. PP_3359 gene.

A.



B.

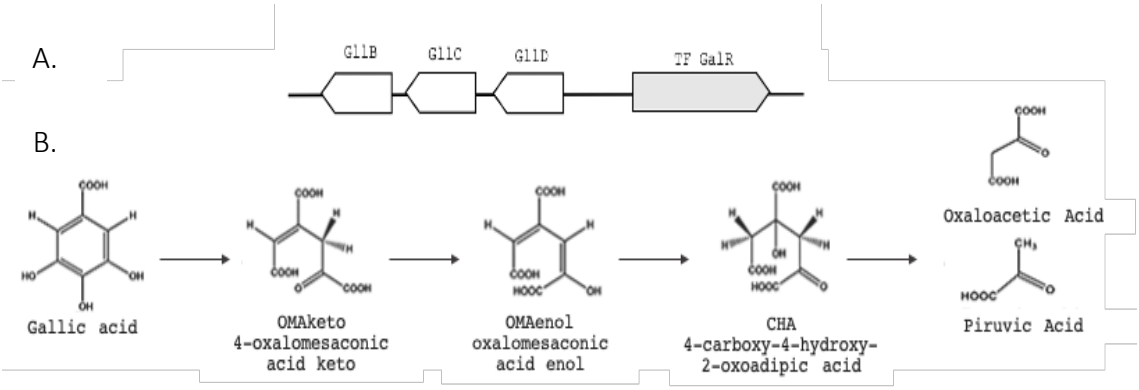


Obtained by http://www.phylogeny.fr/[70]

Analyzing the second of the proteins involved in the degradation of ferulic acid, we found it that PP_3359 gene is less conserved (we found even no match for genes that have a phylogenetic kinship in the PAO1 aeruginosa species). This protein shares with the fluorescence species 83.10% of identity with a coverage of 91%, cloraphis 80.14%, syringae 73%. And the closest knots that have a 3D structure by X-ray diffraction method are the A chains of transcriptional regulators of the MarR family that share a percentage of identity of 38.46% and 30.77% for *Rhodopseudomonas laustris* in which acts as a Structural basis of transcriptional regulation by CouR, a repressor of coumarate catabolism and *Rhodococcus Jostii* (chain A, TF) in which is not found annotated functions (Alignment data for this protein are shown in Appendix 1)

Another of the interesting degradation pathways of Pseudomonas that can be explored as a regulator is that of Galic Acid, in which GalR, a LysR type regulator that controls a degradation pathway that begins with a direct ring cleavage reaction and formation of the central intermediate *4- Oxalomesaconic acid* formed by GalD gene, the prototype of a new isomerase subfamily that catalyzes the tautomerization of the OMAketo generated by the GalA dioxygenase to OMAenol; followed by its hydration to *4-carboxy-4-hydroxy-2-oxoadipic acid* by the GalB hydratase that uses Zinc as an adjuvant, then a ring excision that results in pyruvic and oxaloacetic acids by GalC to go to the central metabolism of this microorganism(Figure 11)[71]

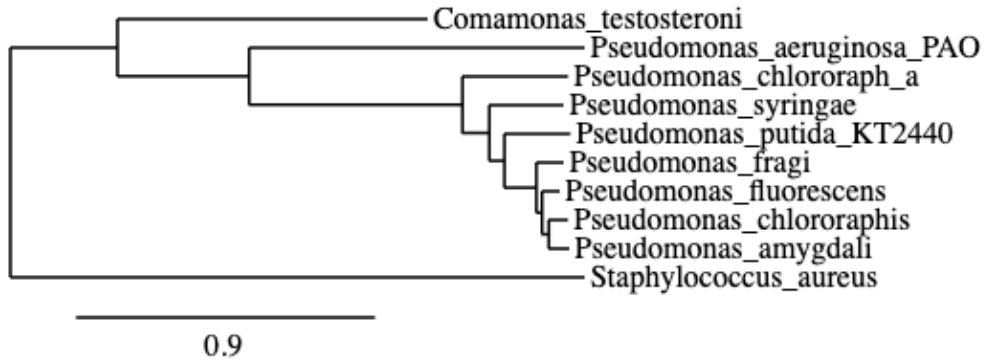
Figure 11. A. Schematic representation of the GalR clusters with the genes they control.
B. Metabolic pathway of degradation, of Gallic acid.



A systematic observation found a similar behavior with the other two regulators in which the species that moves farther away is aeruginosa PAO with an identity of 72.35% in the sequence and a sequence got by nuclear magnetic resonance corresponding to the A chain of a TF of the LysR family of *Staphylococcus aureus* wich share a 30.09% identity Citrate-responsive mechanism of catabolite control protein and also with the A chain of a regulator

of the Lys family that response p-toluenesulfonate operon tsaMBCD in the microorganism *Comamonas testosteroni* with which shares a 27% identity (The data of the alignments of this protein are shown in Appendix 1).

Figure 12. Phylogenetic analysis of the amino acid sequence of the GalR gene.



Obtained by http://www.phylogeny.fr/

2.3 *Pseudomonas putida* synthetic biology and biotechnology

The industry an enormous demand for processes, based on the needs, it chooses the chassis used,, but in recent years *P. putida* has been proposed as a definitive chassis for synthetic biology[6,72]. And although *Escherichia coli* and *Saccharomyces cerevisiae* have represented success stories genetic and genomic manipulations that lead to the overproduction of metabolites synthesized natively. These organisms have been displaced because *E. coli* which has the challenge of suffering when removed from laboratory conditions because it grows under very strict conditions or *S. cerevisiae* difficult to manipulate due to its eukaryotic nature. Also, Pseudomonas has and competes with these last two because they have remarkable characteristics:

-The remarkable metabolic and physiological capacity to resist stress (against extreme and fluctuating environmental conditions of temperature, pH, presence of toxins or solvent) that explodes rapidly with low nutrient demands.

-It is a GRAS and HV1 organism that can released into the environment that can live in rhizospheres using its versatile metabolism to degrade lignin-derived compounds without establishing pathogenic relationships with plants.

-Ideal to accommodate recombinant DNA because all its protocols widely developed, besides to all its genomic and metabolic knowledge allows a modeler and precision behavior.

These characteristics have boosted its use in biotechnological systems, have integrated approaches to metabolic engineering, bioprocess engineering including designs that have reflected in the production of renewable raw materials because of its evolution are becoming efficient cellular factories for various industrially relevant products, including a full range of unnatural chemicals[72–75].

P. putida KT2440 was the first bacteria approved like GMO (genetically modified organism), for an approach engineered to introduced into the environment with a family of plasmids that by conjugation could passed to other organisms and grant to other communities feature degradation genes for aromatic compounds to combat oil spills more efficiently and quickly[42,61,76–78].

After this success story has not stopped its second most notable feature has been its use for the biosynthesis of polymers based on polyhydroxyalkanoates (PHA) that are a biocompatible and biodegradable alternative to petroleum plastics and even has applications in the tissue engineering area because this bioprocess has optimized by systematic analysis of metabolic networks to get PHA material with the characteristics of elasticities, crystallinity and stiffness from the manipulation of their metabolism. [59,79,80] besides other important applications in the production of drugs or important proteins for the industry[15] (Table 1).

Another of the applications that been found in *P. putida KT244* that also gives it its wide metabolic capacity is the finding of TFs in it for the engineering of biosensors in other organisms. Because as already described above it has a very large reservoir of these proteins (Figure 7) and they all respond to aromatic molecules and could help the growing demand for parts for synthetic biology to create genetic circuits that encode new functions in biological machines.[72]

Table 1. Overview of biotechnology products derived by natural and recombinant strains of *Pseudomonas putida.*

Product	Strain	Comment
PHA hetero-polymer	KT2442	Bioplastics
PHA hetero-polymer	KT2442	Altered β-oxidation with modified monomer composition

PHA hetero-polymer	KT2442	Altered with high content of 3-hydroxytetradecanoate
PHA homo-polymers	KT2442	Modified β-oxidation; feeding of hexanoate (c6) and heptanoate (c7)
Myxochromide S	KT2440	Heterologous expression of myxochromide S biosynthetic gene cluster from *Stigmatella aurantiaca*
4-Valerolactone	KT2440	Introduction of *tesB* (*E. coli* G1655) to secrete 4-hydroxyvalerate and pon1 (human paraoxonase I), expressed extra-cytosolic to catalyze the utilization the intermediate 4-hydroxyvalerate
p-Hydroxybenzoate	KT2440	Additional copy of *aroF-1* gene, introduction of *xylAB_FGH* genes (*E. coli*) and the *pal* gene (*Rhodosporidium toruloides*)
2-Alkyl-4(1H)-quinolones	KT2440	Insertion of *pqsABCD* from *Pseudomonas aeruginosa*
4-Hydroxyquinaldine	KT2440	Inserted genes *qoxLMS* (*Arthrobacter nitroguajacolicus* Rü61a)

Taken and adapted from Bernd (2010)

These components have the advantage of being orthogonal, because they do not significantly interfere with the endogenous machinery of evolutionarily separated organisms that do not contain metabolic pathways or transcriptional control natively. Without a doubt, the most used chassis for this purpose is the *E. coli* bacteria, a laboratory model since the development of classical molecular biology techniques. In this organism, there is a deep knowledge of the metabolic networks that operate molecular and biochemical dynamics, which arise from decades of genetic manipulation, and a wide resource of computational tools such as RegulonDB[81]. In this sense, *E. coli* is a model ideal prokaryotic to possess a robust arsenal of information, paving the way for the engineering of endogenous metabolic networks, predicting information flow processes (transcription, translation) and cellular resources (metabolites, ATP, amino acids). This is how the promoters and TFs of several species of pseudomonas have provided to produce heterologous and to design various biosensors that have standardized in other species of Pseudomonas or in *E. coli* (Table 2) that have widely standardized and modeled to respond to different concentrations of the compound.

Based on the expansion and robustness of the *P. putida KT2440* metabolic networks, our group has a significant interest interest in industrially important molecules that not processed by endogenous *E. coli* machinery. For example, xenobiotic compounds such as chlorinated solvents, herbicides and pesticides, with great biotechnological relevance in the Brazilian and Latin America industry, can apply as inputs of synthetic genetic circuits that perform their detection (construction biosensors) and their bioremediation. Similarly, aromatic compounds such as gallic

acid and vanilla (which recognized by P. putida machinery can use for various biomedical applications).

Table 2. Examples of gene expression systems and promoters designed using parts harnessed from Pseudomonas for regulated gene expression, along with some applications of these systems in metabolic engineering.

Expression system	Source	Characterization and uses in biotechnology
XylS/*Pm* (3-methylbenzoic acid)	*P. putida* mt-2	Cha. in *E. coli* and *P. putida*; *p*-coumaric production in engineered *P. putida* strains; heterologous expression of genes for fermentation under low oxygen availability in *P. putida*
XylR/*Pu* (3-methylbenzyl alcohol)	*P. putida* mt-2	Cha. in *E. coli* and *P. putida*; expression of phosphoglucomutase (*celB*) from *Acetobacter xylinum* in recombinant *E. coli* and *Xanthomonas campestris* strains
AlkS/P_{alkB} (short-chain alkanes)	*P. putida* GPo1	Cha. in *E. coli* and *P. putida*; production of (*S*)-styrene oxide and heterologous proteins in recombinant *E. coli* strains
NahR/P_{sal} (Salicylic acid)	*P. putida* NCIB 9816-4	Cha. in *E. coli* and *P. putida*; surface reporter of catabolic promoter activity; gene expression cascade with XylS variants
TodST/P_{todX} (4-chloroaniline)	*P. putida* T-57	Cha. in *E. coli*
CymR/P_{cym} (4 Isopropylbenzoic acid)	*P. putida* F1	Cha. in *E. coli*
ClcR/P_{clcA} (3/4-chlorocatechol)	*P. putida* PRS2015	Cha. in *P. putida*
HpdR/P_{hpdH} (3-hydroxypropionic acid)	*P. putida* KT2440	Cha. in *E. coli* and *Cupriavidus necator*
MmsR/P_{mmsA} (3-hydroxypropionic acid)	*P. putida* KT2440	Cha. in *E. coli* and *Cupriavidus necator*
NagR/P_{nagA} (salicylic acid)	*Comamonas testosteroni*	Production of 3-methylcatechol by *P. putida*
ChnR/P_{chnB} (cyclohexanone)	*Acinetobacter johnsonii*	Cha. in *E. coli*; regulation of biofilm formation in engineered *P. putida* strains

CprK/P_{DB3} (3-chloro-4-hydroxyphenylacetic acid)	*Desulfitobacterium hafniense*	Cha. in *E. coli* and *P. putida*
MekR/P_{mekA} (methyl-ethyl ketone)	*P. veronii*	Cha. in *E. coli* and *P. putida*; expression of *cre* for recombineering in *P. putida*
MtlR/P_{mtlE} (mannitol)	*P. fluorescens*	Cha. in *E. coli* and *P. putida*
AraC/P_{araB} (arabinose)	*E. coli*	Cha. in *E. coli* and *P. putida*; *p*-coumaric production in engineered *P. putida* strains
RhaRS/P_{rhaB} (arabinose)	*E. coli*	Cha. in *E. coli* and *P. putida*; *p*-coumaric production in engineered *P. putida* strains
TetR/P_{tetA} (anhydrotetracycline)	*E. coli*	Synthesis of myxochromide, epothilone and tubulysin from *Cystobacter* sp. in engineered *P. putida* strains
LacIQ/P_{lac}, P_{tac}, P_{trc} (isopropyl β-D-1-thiogalactopyranoside)	*E. coli*	Cha. in *E. coli* and *P. putida*; biological containment of *P. putida*; production of long-chain rhamnolipids and vanillin in engineered *P. putida* strains
P_{T5}	Bacteriophage T5	Expression of 6×His-tagged proteins (*pfrA* siderophore regulatory gene) in *P. putida*
P_{T7}	Bacteriophage T7	Cha. in *P. putida*; containment of *P. putida* strains by conditional expression of the streptavidin gene
P_{EM7}	Synthetic, constitutive	Cha. in *P. putida*; CRISPR-Cas9—based counter-selection for genome manipulations in *P. putida*
P_{oprL}	Synthetic, constitutive	Cha. in *P. putida*; laccase gene from *Shigella dysenteriae*
P_{gro}, P_{tuf}, P_{tufs}	Synthetic, constitutive	calibrated gene expression in *P. putida*; production of *cis*,*cis*-muconic acid
P_{JE}	Synthetic, constitutive	Calibrated gene expression in *P. putida*
$P_{SynP\text{-}BG}$	Synthetic, constitutive	Calibrated gene expression in *P. putida*

Taken and adapted from Bernd (2010)

In the Laboratory of Synthetic and Systemic Biology of the University of São Paulo USP for the first time generated a system of promoter biosensors with their respective TFs with improved response to acetylsalicylic acid (Aspirin) from a natural TF of *P. putida*, responsive to xylose (an aromatic compound found in wood)[82,83]. The above highlights the evolutionary path that bacteria have traveled for the use of these aromatic compounds as carbon sources and raise an interesting field of research for the development of new therapies because most of the active compounds discovered At present, against various diseases they have a plant origin and therefore the bacteria have in their regulatory networks proteins that could link to drugs of conventional use through the structural homology present in these compounds, which will allow the use of these valuable proteins to the use and treatment of various diseases, especially tumor type in animal or human hosts[16,17,84,85].

However, regardless of the field of application, synthetic biology characterized by a constructive approach to understanding and manipulating biological systems. This aim represents an important challenge because of the complexity of the organisms and given that the principles that determine their functioning not well characterized to achieve orthogonal and robust functions. Also, in recent years there has been an increase in the design of several biological circuits, it has reduced the number of regulatory biological parts that have described and characterized. Besides, the compounds that widely used in synthetic circuits are expensive and cytotoxic, which limits their actual use in the pharmaceutical industry (such as IPTG, which is one of the commercially used but commonly used expression systems). Therefore, it is necessary to characterize the new regulators responsible for the most accessible compounds, thus increasing our ability to manipulate microorganisms.

3. Bioinformatics

The fields of bioinformatics and computational biology arose for the need to solve the problems generated by the enormous amount of biological information, got by the different methods developed in the laboratory. Bioinformatics presents many areas of research that include sequence alignment, genome assembly, gene search, gene regulation and expression, comparative genomics, evolutionary biology, prediction, and alignment of protein structures, determination of metabolic networks, and experimental proteomics. Each of these areas goes hand in hand with the techniques developed in computer science, statistics, computer science, and applied mathematics [86].

The main goals of bioinformatics are three[87]. The first is to organize the data so that researchers can access existing information and provide recent entries as it gives new products; Protein Data Bank PDB[88], GenBank[89] databases are examples that affirm the fulfillment. The second aim is the development of tools and resources that help in the analysis of data; for example, the implementing programs based on punctuation matrices and alignments such as BLAST or PSI-BLAST[90,91] and the third aim is to use these tools to analyze the data and interpret the results biologically meaningfully.

Traditionally, biological studies examine individual systems and it compared to some that related.. In bioinformatics, global analyzes of all available data can performed, discover common principles that occur in many biological systems that have supported synthetic biology mainly to automate the genetic design and improve the cycle of design with the prediction by bioinformatics, these efforts have mainly linked to the design of circuits and to analyze models that mimic metabolic networks, predict how these can behave in cells on platforms such as iBioSim[92] or OptCircuit[93]. The first steps to programming high-level decision making in synthetic networks show would produce more robust and dynamic organisms, including those that can perform many tasks simultaneously. Besides, given that adaptive and predictive behaviors. However, all these efforts still focused on DNA sequences that introduced or on RNA fragments characterized by ignoring that proteins are molecules that have a tertiary structure that can also predict and used in favor of the design of genetic circuits. This is how the constant search for TFs that respond to different molecules has converged in using computational models based on structural biology, because they represent savings in terms of resources and research time, added to their great robustness that allows analyzing hundreds of data at the same time and get valuable conclusions by establishing relationships between them in a pipeline that includes protein modeling, finding surfaces that can receive aromatic protein activating compounds and docking tests in which an *in silico* test of these unions and interactions[94].

3.1 Protein structures

The getting of protein structure by experimental methods mainly done by 3 methods:

-X-ray crystallography, which solves the structure of a protein from the X-ray diffraction pattern, product of the rebound of the atoms that make up a macromolecule and interpreted by the electron density map[95].

-Nuclear Magnetic Resonance (NMR) in which the three-dimensional structure of a protein revealed because of the resulting absorption spectrum emitted by a protein when it passed through an intense and static magnetic field that relieves chirality, Van der Waals radii, distances of union and angles[96].

-Cryogenic electron microscopy, which has become one of the most powerful tools for solving biomolecular structures. It is electron microscopy operated at temperatures of liquid nitrogen or mediums in which they study the sample at cryogenic temperatures, it can do this in its native environment, unlike crystallization[97].

The knowledge got by these techniques has been very extensive for the study of macromolecules that have been available in Repositories as the Database The Protein Data Bank (PDB) (http://www.rcsb.org/pdb/), a database that contains information on biological macromolecules and located available to the public. (it has registered 157145 structures until October 2019) that can classify as protein structures, nucleic acid oligonucleotide structures, carbohydrate structures, and hypothetical models of protein structures. There are websites supported by academic entities that help classify proteins taking into account their structure and functionality: the SCOP (structural classification of proteins)[98], the CATH (class, architecture, topology, homologous superfamily) [99] are the main sites and each presents its classification scheme. Thanks to these tools, it has been possible to detect patterns of variation and conservation that speak of the evolutionary history in proteins, an important biological feature in the study of life with subsequent studies, was possible to show that the patterns of secondary structures (motifs) correlate with the physical-chemical characteristics of the amino acids that make up them by limiting the number of conformations. [100]

In this way, the direction of these arguments leads us to understand how bioinformatics in this wide range of scientific knowledge has served in the theoretical resolution of three-dimensional protein structures. X-ray crystallography, NMR and CrioEm methods are not always affordable and take a long time, bioinformatics tools speed up the understanding of structural studies in a rather *a priori* context, and generate alternative routes that can optimize research crucial [101]. However, it is important to clarify that the structural bioinformatics of proteins, depends on databases such as PDB, and thus, inherently is subject to experimental science.

Thus, the structural bioinformatics based on renaturation, the limited number of folds, structures already determined, sequence alignments and evolutionary relationships,

have developed programs with algorithms capable of generating three-dimensional models of proteins unknown in terms of their structure but known as to its sequence. It write the highly robust algorithms under mathematical tools such as hidden Markov models (HMM), Monte Carlo methods, or neural networks, which speed up the search for information in databases that gives structural patterns producing optimal results for later analysis. The tertiary structure prediction methods developed experimentally are three (in Table 3 shows some the programs that implement this method):

-Homology Modeling: Model construction by the homology technique includes the prediction of the three-dimensional structure of a target protein of known sequence, from another protein of known sequence and structure (template). The modeling can given as long as the two proteins (target and template) closely related together, that is when they are homologous. It has defined the following steps in homology modeling. However, this grinding is only possible when you have a structure with a nearby related, that is when they are homologous (greater than 30% homology). The steps to perform homology modeling are the alignment of amino acid sequences between the target and template proteins where regions containing inserts or deletions determined in parallel; followed by a replacement of side chains of waste that have mutated. Waste that has not mutated, keep the conformation. The waste they have mutated maintain the same conformational chain angles lateral, and can model on this basis and generation and evaluation of the three-dimensional model of the target protein[98,102,103].

-The threading method: It is based on the principle that in nature there is a limited number of basic folds in the "nuclei of the protein domains", thus, threading predicts tertiary structures of target proteins using folds of other proteins of known structure. The way this method operates has given it alternative names such as folding recognition or remote homology. The threading method takes into account that the construction can base on non-homologous proteins that registered in the PDB, FSSP, SCOP, or CATH databases. The basic idea of threading is to build many hypothetical models of the target protein, based on each known structure using different alignments and energy equations (which got empirically). The folding recognition implements statistical methods of assigning scores to the generated models and selecting the best structure and also calibrating these scores to determine if the best-estimated model is probably correct. This model represents the advantage of predicting protein structures with a low percentage of identity[104–106].

-Modeling by Ab Initio: The ab initio (or de novo) method for modeling, initially breaks the target sequence into short segments (from 3 to 9 residues), then predicts the structure of these divided segments using statistics (such as HMM). The Bioinformatics of Protein Structures 23 next step of the method is to join each of the divisions, to carry out the three-dimensional configuration exploring all possible combinations. It choose the conformation with the lowest global energy for this modeling to take place, ab initio takes into account three elements: a representation of protein geometry, a force field, and a surface energy search technique. The ab initio method is far from being efficient, the quality of the predictions made by the people is very low, but in recent years new strategies have implemented to improve this technique and bring it closer to the reality of the protein structural nature. Ab initio, unlike those described above (homology and threading), is a non-comparative method.[107,108]

Table 3. Some programs for modeling structures. Each program uses a certain modeling method and is available on a website.

PROGRAM	METHOD	WEBSITE	BIBLIOGRAPHY
LOMETS	Threading	http://zhanglab.ccmb.med.umich.edu/LOMETS/	(Wu & Zhang, 2007)[109]
MODELLER	Homology	https://modbase.compbio.ucsf.edu/scgi/modweb.cgi	(Fiser & Sali, 2003)[110]
PHYRE PHYRE 2	Homology/ threading	http://www.imperial.ac.uk/phyre/	(Kelley & Sternberg, 2009)[111,112]
ROBETTA	Threading and ab initio	http://robetta.bakerlab.org/	(Chivian et al., 2005)[113]
SWISS-MODEL	Homology	http://swissmodel.expasy.org/	(Arnold et al., 2006)[114]
I-TASSER	Threading and ab initio	http://zhanglab.ccmb.med.umich.edu/I-TASSER/	(Zhang, 2008)[104]
RAPTORX	Threading	http://raptorx.uchicago.edu/	(Peng & Xu, 2011)[115]
QUARK	Ab initio	http://zhanglab.ccmb.med.umich.edu/QUARK/	(Xu, 2012)[108]

.2 Docking

Docking or molecular docking is a computational technique that couples a molecule small (ligand) at the binding site of a macromolecular target (receptor) and that estimates the affinity of the link, that predicts the energy interaction between two molecules[116,117]. Build a structure *in silico* receptor-ligand involves two key components: a search algorithm and a punctuation function. The search algorithm comprises a sampling of the conformation and orientation of the ligand on the limited binding site of the receiver. The punctuation function selects the best "posture" (conformation, orientation, and translation) of the ligand on a molecule. For being successful, the coupling must accurately predict two things concerning the experimental information available: ligand structure (pose prediction) and the binding trend (affinity prediction), performed docking experiments using theoretical models and showed that ligand coupling predictions about structures do not keep reliable accuracy, however, they found that these trials can improve by employing multiple models instead of one alone, even taking into account the template proteins so that the enrichment of the calculations and coupling consensus can establish.

Rigid docking It is one docking performed and based on the “lock and key” hypothesis, where the receiver treated as a body rigid, and the ligand quickly and efficiently explores the conformations which can opt for the receptor cavity. However, it accepted that ligand binding is not a static fact, but a dynamic process where both (the ligand and the protein) may undergo conformational changes. To this change in form working where the protein has flexible residues called: flexible docking or induced change by incorporating flexibility in the protein the search space is increasing exponentially, therefore, the computational cost also increases compared to rigid docking these algorithms determine all possible optimal conformations for a complex (protein-protein, protein-ligand) in an environment, the position and orientation of one molecule concerning the other. They can also calculate the energy resulting from the complex and each interaction[118].

The different algorithms that can use by rigid docking analyzes are Monte Carlo methods, genetic algorithms, methods based on fragments, complementary methods of points, methods of distance geometry, and systematic search (in table 4 enlisted some most used programs for performing docking). To allow the search for the large conformational space available from a linking to a receiver, programs like AutoDock[119] use a method based in a grid, this allows a rapid evaluation of the link energy of the conformations that generated.In this method, it embeds the receiver in a grid. Then, each atomic probe

place sequentially at each point of the grid, the energetic interaction between the probe and the receiver calculated, and it stores the resulting energy value in the grid. They can then use this energy grid can then as a search table during docking simulation.[120].

Table 4. Some programs used for docking experiments. The address of Internet to download each program. Some have free licenses, another does not.

PROGRAM	WEBSITE	REFERENCE
AUTODOCK	http://www.scriPPs.edu/mb/olson/doc/autodock/	(Oleg Trot et al., 2010)[119,121]
GLIDE	http://www.schrodinger.com/productpage/14/5/	(Friesner et al., 2004)[122]
GOLD	http://www.ccdc.cam.ac.uk/Solutions/GoldSuite/Pages/GOLD.aspx	(Verdonk et al., 2003)[123]

Docking has been used for various applications, one of them is the Virtual screening (VS) a high-throughput methodology in which algorithm docking is used to find the way to join a database of compounds (Databases can be public and commercial chemicals that are commonly analyzed in real practice or drugs)[124]. Compounds that show better scoring function based on the best binding conformations with the target protein and lowest free energy data are considered as potential candidates for protein pocket inhibition or activation[125].

Docking is used for drug discovery from small-molecule libraries, thus, this methodology does a docking with each one of the molecules that are going to be studied in a protein cavity in a fast automated docking method. This method makes the first filter of possible compounds that can bind to a protein that could usually be potential drugs *in silico* for *in vivo* tests. Thus, during this work, we proposed to analyze drugs that could theoretically bind aromatic compounds and activate transcription from the pocket sites of TFs[126].

II OBJECTIVES

1. General objective

This work seeks to develop synthetic biology tools by *in silico* and *in vivo* methods for applications in biotechnology from the genome of *Pseudomonas putida* KT2440.

2. Specific objectives

Expand the number of reporter vectors with fluorescent proteins to create a promoter/protein hunting system useful for biosensor kinetic analysis or finding new genes by genomics or metagenomics.

Conduct a study of the genome of *Pseudomonas putida KT244*0 and choose TFs that are candidates for biosensors

Model and do a docking of the possible molecules that can respond to these families of TFs to predict their activity.

III CHAPTER 1

Design and construction of hunting vectors for search of biosensors.

1. Material and Methods

1.1 Bacterial Strains and Growth Conditions

Bacterial strains, plasmids, and primers used in this study listed in Table 5. E. coli DH5α cells used for cloning procedures and for testing the natural by *Pseudomonas putida* KT2440 strain for testing the circuits *E. coli* strains were grown at 37 ° C in media solid and liquid LB (Lysogeny broth composed of 10 g/L triptone, 5 g/L yeast extract, 5 g/L NaCl with final pH ~ 6.8 - 7.2) for cloning process. *Pseudomonas putida KT2440* were grow at 30 ° C in citrate medium for Pseudomonas (composed of ammonium dihydrogen phosphate, 0.2 g / L, magnesium sulfate heptahydrate, 0.2 g / L, sodium chloride, 5 g / L, trisodium citrate, 2 g / L with final pH ~ 7.0) and minimal media M9 (6.4 g / L Na2HPO4·7H2O, 1.5 g / L KH2PO4, 0.25 g / L NaCl, and 0.5 g / L NH4Cl) were used for growth expression analysis of both strains. The antibiotics streptomycin (Sm, 34 µg / mL) and chloramphenicol (50 µg / mL) were added to the media to ensure plasmid retention for pSEVA and pMR1, respectively.

Table 5. Bacterial strains, plasmids, and primers

STRAINS	
Pseudomonas putida KT2440	mt-2 KT2440
Escherichia coli DH5A	F– ϕ80ΔlacZΔM15Δ(lacZYAA-argF); U169 recA1 end A1 hsdR17 R–M+supE4 thi gyrA relA
VECTORS	
pMR1	CmR; orip15a
PMR1 XYLS	CmR; orip15a; promoter and regulatory region as native arrangement from Pseudomonas putida KT2440
pMR1 P100	
pSEVA 431	CmR; orip15a; Strong promotor p100
pSEVA 421	SmR; MCS default pSEVA; Ori RK2
pRJ1	SmR; MCS default pSEVA; Ori pBBR1
pRJ2	SmR; MCS default pSEVA; Ori RK2; sfGFP LVA
pRJ3	SmR; MCS default pSEVA; Ori RK2; sfGFP AAV
pRJ4	SmR; MCS default pSEVA; Ori RK2; sfGFP LVA; mCherry LVA
pJ3 P100	SmR; MCS default pSEVA; Ori RK2; sfGFPAAV; mCherry AAV
pRJ3 XYLS	PRJ3; Strong promotor p100
PRIMERS	pRJ3; Promoter and regulatory region as native arrangement from Pseudomonas putida KT2440
VECTOR CONSTRUCTION PRIMERS	
5'FOR_MCHERRY	GCGCTTAATTAAGCTTAAGCTACTAGAGCGTAGTTTTCGTCGTT TGCTGCCTTGTACAGCTCGTCCATGC
3'REV_MCHERRY_LVA	GCGCTTAATTAAGCTTAAACAGCAGCAGCGTAGTTTTCGTCGTT TGCTGCCTTGTACAGCTCGTCCATGC

3’REV_MCHERRY_AAV	GCGCAAGCTTAAAGAGGGGACAATGCGTAAAGGCGAAGAGC
5’FOR_SFGFP	GCGCACTAGTGCTTAAGCTACTAGAGCGTAGTTTTCGTCGTTTG CTGCTTTGTACAGTTCATCCATACC
3’REV_SFGFP_AAV	GCGCCCTAGGAAAGAGGGGACAATGGTGAGCAAGGGCGA
PRIMERS SEQUENCING pSEVA	
5 T1 SEQ	GCCCAGCTGTCTAGGG
3 MCS SEQ	GGTCGACTCTAGAGGATCC
5 MCS SEQ	GCGAATTCGAGCTCGG
3 T0 SEQ	CCCCTGGATTCTCACC
PRIMERS SEQUENCING pMR1	
MCHERRY_PMR1	GGTCACCTTCAGCTTGG

1.2 Construction of Expression vectors

We constructed new expression vectors unidirectional and bidirectional expression systems comprising four vectors with the fluorescent proteins sfGFP and mCherry, each with a bacterial protease degradation tag of LVA (sequence ATTAAGCTACTAAAGCGTAGTTTTCGTCGTTTGCTGC) and AAV (sequence GCAGCAAACGACGAAAACTACGCTGCTG.). These vectors based on pSEVA vectors 431 and constructed using 5'For_mCherry, 3'Rev_mCherry_LVA, 3'Rev_mCherry_AAV, 5'For_sfGFP, 3'Rev_sfGFP_LVA, 3'Rev_sfGFP_AAV primers (referred on Table 5), a strong RBS sequence at the 5′ region of the gene and on the reverse the sequence for LVA and AAV tags, to have half-life protein variants that allow a better analysis of protein expression. The fragments were obtained using Phusion High-Fidelity DNA polymerase (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's protocol and PCR performed on an Eppendorf thermocycler (Eppendorf AG, Hamburg, Germany). sfGFP protein got by PCR were amplified from the pSB1C3 vector with the part BBa_I746916 by Igem parts repository and mCherry cloned from pMR1 vector. It flanked each of these parts with the enzymes AscI, PacI for mCherry, and SpeI and SacI for sfGFP protein. We named the resulting plasmids pRJ1 to pRJ4.

Also, during the construction of these vectors, the MiraPrep protocol with the Miniprep Kit Wizard (Promega Corporation, Madison, Wisconsin, EUA) standardized in the laboratory of synthetic and systematic Biology that used for the rest of the construction for the study of induction systems and detailed below.[127]

1.2.1 Miraprep protocol

A 50ml inoculum of bacteria of interest enriched with streptomycin or chloramphenicol for pSEVA or pMR1 respectively were incubated at 37 ° C for 12-16 hours in a shaker at

220 rpm. Then the resulting culture were transferred to 50 ml falcon and centrifuged 4000xg at 4 ° C for 10 minutes, then, the leftover discarded and resuspended in 1.5 ml of Resuspend buffer and 80 µg /ml RNaseA. 1.5 ml of lysis buffer and 50 µl of alkaline solution added to pellet solution and incubated for 5 minutes at room temperature. 2 ml of neutralization buffer added, and it inverted the tube 3–4 times. It distributed the bacterial lysate into 4 tubes by pouring and spun at 13,200xg at room temperature for 10 minutes. Next, the supernatants collected in a 15 ml tube, and the pellets discarded. 5ml volume of 96% ethanol added to the supernatant and mixed thoroughly for 5 seconds. The sample-ethanol mix loaded onto 3 spin-columns for 30 seconds at 13,200xg after the addition of each aliquot.

After each spin, the flow-through discarded. These steps were repeated until it passed the entire sample through the spin-columns. The columns were washed two times with 500 µl wash buffer, spun after each at 13,200xg at room temperature for 30 seconds, and the flow-through discarded. It then spun the empty columns were then one last time at 13,200xg at room temperature for 1.5 minutes to remove any residual wash buffer. After this, the old collection tube discarded and each column put into a new tube. 40 µl of distilled water or elution buffer were added to the column which incubated for 2 minutes at room temperature and spun at 13,200xg for 2 minutes to elute the DNA from columns. The eluted DNA from all 3 columns combined in one tube. (Figure 13)

Figure 13. MiraPrep protocol overflow.

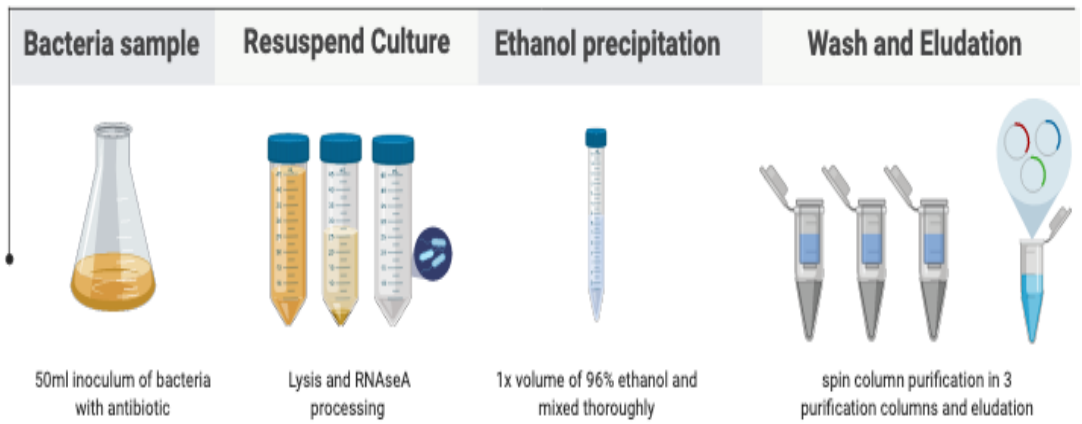


Image made on the Biorender platform (https://biorender.com)

1.2.2 Linearization of vectors and cutting fluorescence proteins by restriction enzymes.

The pSEVA 431/421 LVA / AAV vectors obtained by Miraprep protocol prepared with restriction enzyme digestion to receive each of the fragments of interest with the

respective enzymes. 1 µg of the vectors got by Miraprep protocol, it incubated with 1 µL of each enzyme (Biolabs - Ipswich, USA) and the buffer recommended for each of the enzymes by the commercial house and treatred with water to have a final volume of 50 µL. We incubated these reactions for 2 hours at 37 ° C. Each reaction containing vector dephosphorylated with the addition of 1U of CIP (Calf Intestinal Alkaline Phosphatase) (Invitrogen - Carlsbad, USA) for 30 minutes at 37 ° C and then inactivated at 65 ° C for 20 minutes. To confirm the digestion of vectors, a 0.8% agarose gel electrophoresis ran together with a molecular weight marker of GeneRuler 1kb Plus DNA (Thermo Fisher - Waltham, USA) and gel documentation performed in ChemiDoc XRS + (Bio -Rad Laboratories, Inc - California USA). Each of the vectors purified using the Wizard® SV Gel and PCR Clean-Up System Kit (Promega - Madison, USA) and all fragments were quantified in NanoDrop (Thermo Scientific NanoDrop 2000c Spectrophotometer). We performed the same procedure for each of the genomic fragments got by PCR except for CIP treatment.

1.2.3 Ligation, Transformation, and Confirmation of Constructions.

100µg of the linearized pSEVA 431 vectors incubated with their respective fragments previously cut in the equimolar ratio in a 1:3 vector insert ratio incubated with 1.0 U of T4 DNA Ligase enzyme (Biolabs - Ipswich, USA) in buffer 1X ligament (50 mM Tris-HCl, pH 7.6, 10 mM MgCl2, 1 mM ATP, 10 mM DTT) in a final volume of 10µL; As control of each of these reactions, each cut insert replaced with water to ensure that it digests the vector and it does not relegate this and another unlinked to know if there are traces of undigested vector. We incubated the reaction overnight at 10 ° C and then inactivated at 65 ° C for 15 minutes.

It transformed the resulting ligaments into electrocompetent bacteria *E. coli* DH5α in the MicroPulser electroporator (Bio-Rad - Hercules, USA) in voltage of 2.5kV in Ec1. Then, these cells resuspended in 1mL of LB sterile liquid medium and incubated for 1 hour at 37 ° C. After 500 µL of cells plated on solid LB medium supplemented with streptomycin (Sm, 34 µg / mL) and chloramphenicol (34 µg / mL) for pSEVA and incubated overnight at 37 ° C.

Confirmation of vectors carried the fragment confirmed by colony PCR, resuspending each colony in 10 µL of sterile water and using 1 µL as a template using the respective primers for each fragment of interest of the colonies chosen on each plate and following PCR conditions were used to amplify each fragment (94 °C for 7 min, followed by 30 cycles of, 45 seconds at 94 °C, 1 min at 56 °C, and 1 min, at 72 °C with a final extension

of 7 min at 72 ° C) with Gotaq polymerase (Promega GoTaq® DNA Polymerase). As a negative reaction control, it replaced the template with water.

Transformations in *E. coli* also confirmed by growth of inoculums from each overnight colony with their respective antibiotic and a miniprep plasmid DNA extraction (Promega Wizard® Plus SV Minipreps) followed by restriction digestion by the enzymes for each fragment and analyzed by agarose gel electrophoresis 0.8% run together with a GeneRuler 1kb Plus DNA molecular weight marker (Thermo Fisher, USA) and gel documentation performed on ChemiDoc XRS+ (Bio-Rad Laboratories, USA).

1.2.4 Growth Curves, Fluorescence Tests (sfGFP and mCherry) and data processing.

To validate and measure the capacity of the vectors constructed, promoters widely used in the Laboratory of Synthetic and Systematic Biology used, which is a promoter-TFa system that responds to different concentrations of 3-methylbenzoate in *E. coli* and these constructs and empty vectors also treated with different concentrations of candidate inductors, because these are phenolic compounds and in sometimes, they can inhibit bacterial growth their oxidizing activity.

In that way. Individual colonies cultivated during the night in LB medium, centrifuged, and resuspended at minimum (M9) fresh. A culture (10 µL) analyzed in 96-well microplates in biological triplicates with 1 µL of M9 medium supplemented with antibiotics in streptomycin 25 µg / ml as appropriate for each vector and each compound in final concentrations of 0.1mM, 0.01mM, 0.001mM and as a negative control it replaced with MilQ water and each construction tested with the compounds ferulic acid, vanillic acid, and vanillin. Cell growth and fluorescence emitted by sfGFP and mCherry were quantified using the Victor X3 plate reader (Perkin Elmer). In response to it calculated the promoter as arbitrary units by dividing the fluorescence levels of optical density at 600 nm (GFP / OD600) after normalization with a blank of each compound. It carried fluorescence and absorbency measurements out at intervals of 30 minutes for 8 hours (until reaching the maximum log growth phase of the bacteria) in technical and biological triplicates in all experiments. The data processed and analyzed using R and R Studio software (https://www.r-project.org/) in a script made for this purpose (Appendix 1) This code was made available on the GitHub[128] platform (https://github.com/MariaJulianaRolonRojas/Victor-3X-Analysis)

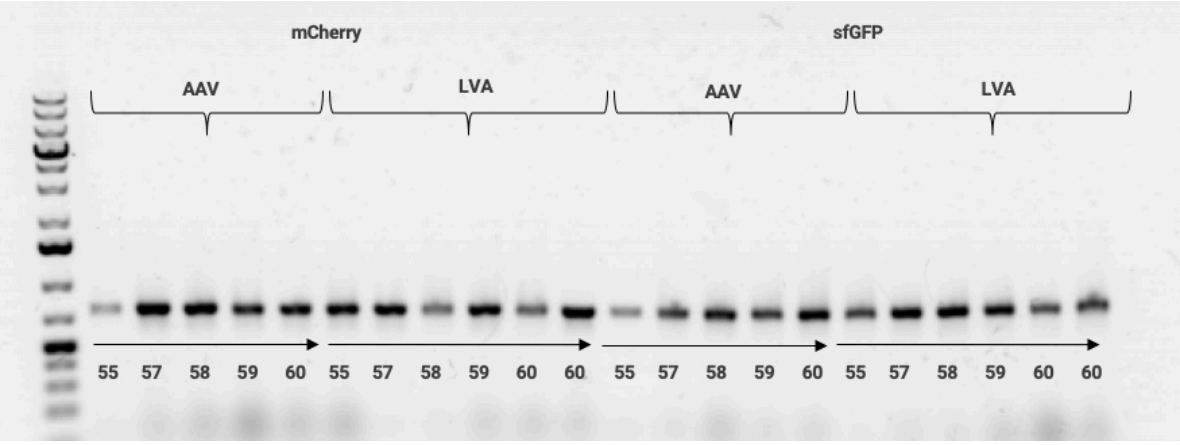
2. Results

2.1 Construction of expression vectors

To construct the vectors, a temperature gradient PCR reaction of the mCherry fragments started from the PMRI and sfGFP vector of pSB1C3 in which each of the protein added to the LVA tail and another AAV tags resulting in two variants of each one with the primers indicated in the methodology. These fragments were running in an electrophoretic experiment colored with Syber Safe to be the size of each of the expected DNA sequences. 744 bp for mCherry and 772 for sfGFP (Figure 14) and purified. Then, these fragments but by restriction enzymes SpeI / HindIII for sfGFP fragments and PacI / AvrII for mCherry for two hours according to the protocol suggested by the manufacturer and purified.

To have more quantity and quality of plasmidid DNA to make each of the constructions a standardization of the MiraPrep process carried out. In a first screening, better plasmid quantities were obtained than with the conventional MiniPrep process but with top levels of RNA contamination (Figure 15.A), thus, the concentration of RNAse A

Figure 14. Agarose electrophoresis gel (0.8%) of PCR gradient for sfGFP and mCherry fragments with LVA and AAV labels from 55 to 60 ° C. 1 Kb DNA ladder.

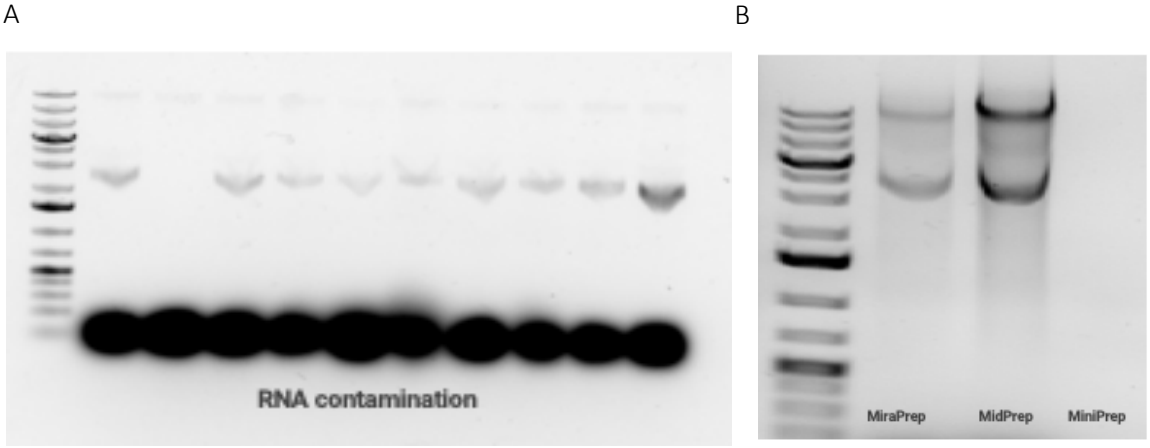


increased to 80 µg / ml instead of 50 µg / ml as it did In the described protocol, in this way it got comparable vector DNA quantities with a MidPrep (Figure 15.B) and it compared with a conventional Miniprep process in which it cannot observed after applying 5uL colored with Saber Safe. Then, each of the 1 µg of the pSEVA431LVA / AAV vectors got by Miraprep protocol vectors incubated with 1 µL of each enzyme following the corresponding protocol (10X NEBuffer 1X, enzyme A 1.0 µl 10 units

enzyme B 10 units and nuclease free water to 50 µl) with the enzymes referred to in table 6 and an electrophoretic run performed to confirm the cut of each of them (Figure 15).

Thus, a binding reaction with the enzyme T4 DNA Ligase performed, and the protocol suggested by NEB (10X T4 DNA Ligase Buffer, Vector DNA 3 kb 50 ng Insert DNA 50 ng, T4 DNA Ligase 1 µl Nuclease-free water to 20 µl) incubated overnight and then subjected to inactivation. It transformed this ligation with 5 µl of the reaction into 50 µl competent cells along with its negative controls. The cells that grew from the transformation analytically digested with the SpeI and HindIII enzymes, in which we observed the 740 bp fragment to leave for two carries (Figure 9). From these vectors with the AAV and LVA variants of the sfGFP protein, the same process performed for the cloning of the mCherry fragments into their AAV and LVA variants flanked by the PacI and AvrII enzymes and confirmed by analytical digestion and PCR of confirmation where the clones observed in the electrophoretic run 3,4,6,7 positive for the charge of the AAV variant and in the positive lane the 6,7,8 positive clones for the LVA tag.

Figure 15. Agarose electrophoresis gel (0.8%) A. MiraPreps of vector pSEVA 421 observing RNA contamination below. B. Gel in lane 2 a sample of Miraprep treated with 80 um / ml RNAseA, lane 3 in MidPrep and lane 4 in a miniprep of pSEVA 421. 1 Kb DNA ladder.



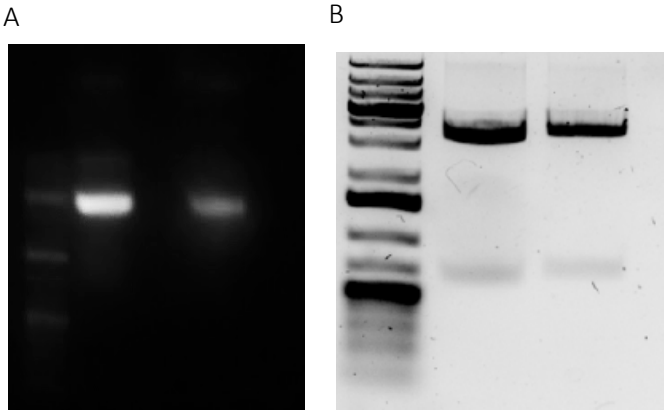


Figure 16. *Agarose electrophoresis gel (0.8%) A*. Linearization of pSEVA 431LVA vectors, in lane 2 the cut with SpeI and HindIII is observed to receive the sgGFP fragment and in lane 4 the same cut to receive the LVA variant. B. Confirmation digestion of PSEVA 431 to receive the sFgfp fragment with the same enzymes. *1 Kb DNA ladder.*

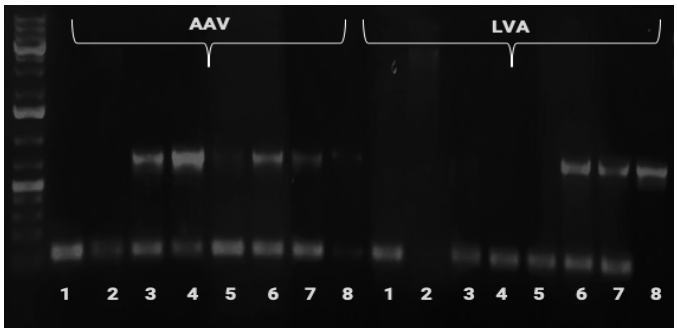
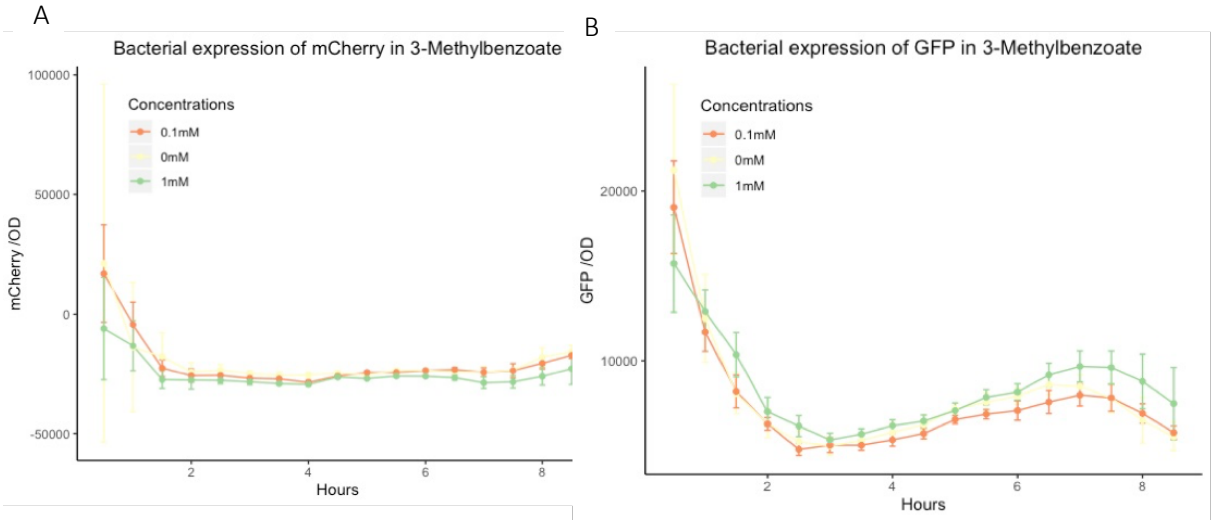


Figure 17. *Agarose electrophoresis gel (0.8%)*. Confirmation PCR results for the mCherry fragment in pSEVA 431 with the AAV and LVA variants. *1 Kb DNA ladder.*

2.2 Growth Curves and Fluorescence Tests (GFP and mCherry) and data processing.

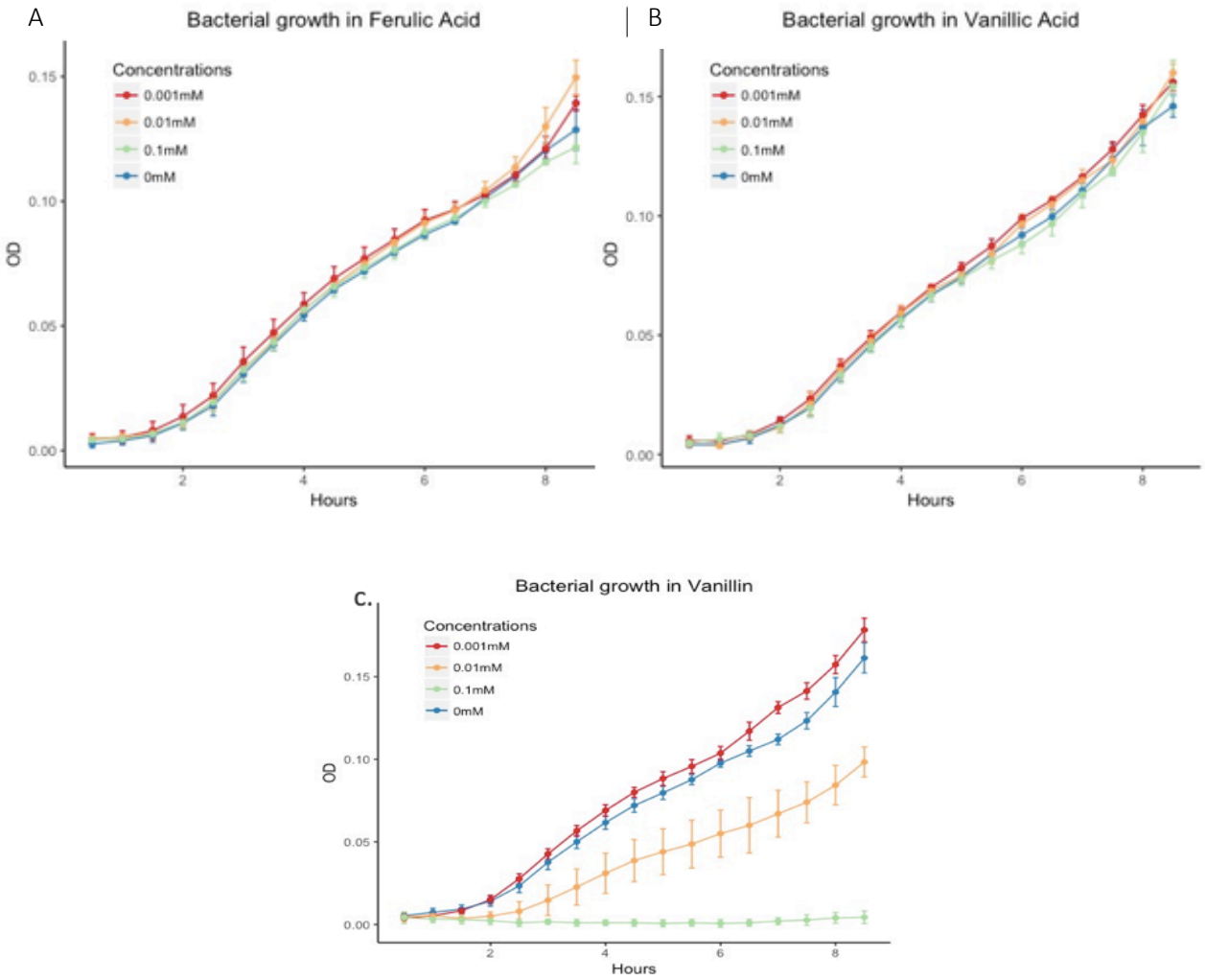
During the kinetics in which mCherry and sfGFP measured for 8 hours, we observed no expression levels of either protein before induction with *3-methylbenzoate*. This may occur because sfGFP is a fast-induction protein proteases because targeted that its degradation tags optimized for this and cannot be measured over time and mCherry could occurring late and degraded in the same way. (Figure 18). The screening to analyze the bacterial growth of each of the constructions with the aromatic compounds used, because some may exhibit cellular cytotoxicity because *E. coli* does not have the metabolic machinery for their degradation. Thus, during an analysis, absorbency measurements carried out at intervals of 30 minutes for 8 hours (until reaching the maximum log growth phase of the bacteria). It observed that for the Vanillic Acid and Ferulic Acid compounds compared to the control they show a bacterial growth dynamics, on the contrary, the Vanillin compound showed bacterial cytotoxicity at 0.1 0.01 molar concentrations (Figure 19), so for future experiments, this compound would not is ideal to treated as an inductor to treated in the *E. coli* chassis.

Figure 18. Analysis of *E. coli* DH5α in M9 liquid media for 8 hours exposed to increasing concentrations of *3 methyl benzoate*. A. Expression pattern of mCherry B. Expression pattern of sfGFP.



Graphics obtained by R studio software

Figure 19. Analysis of *E. coli* DH5α in M9 liquid media for 8 hours exposed to increasing concentrations of aromatic inducing compounds (a) Growth curve of *E. coli* in concentrations of Vanyllic Acid (B) Growth curve of *E. coli* in concentrations of Acid Ferulic (c) Growth curve o*f E. coli* in Vanillin concentrations.



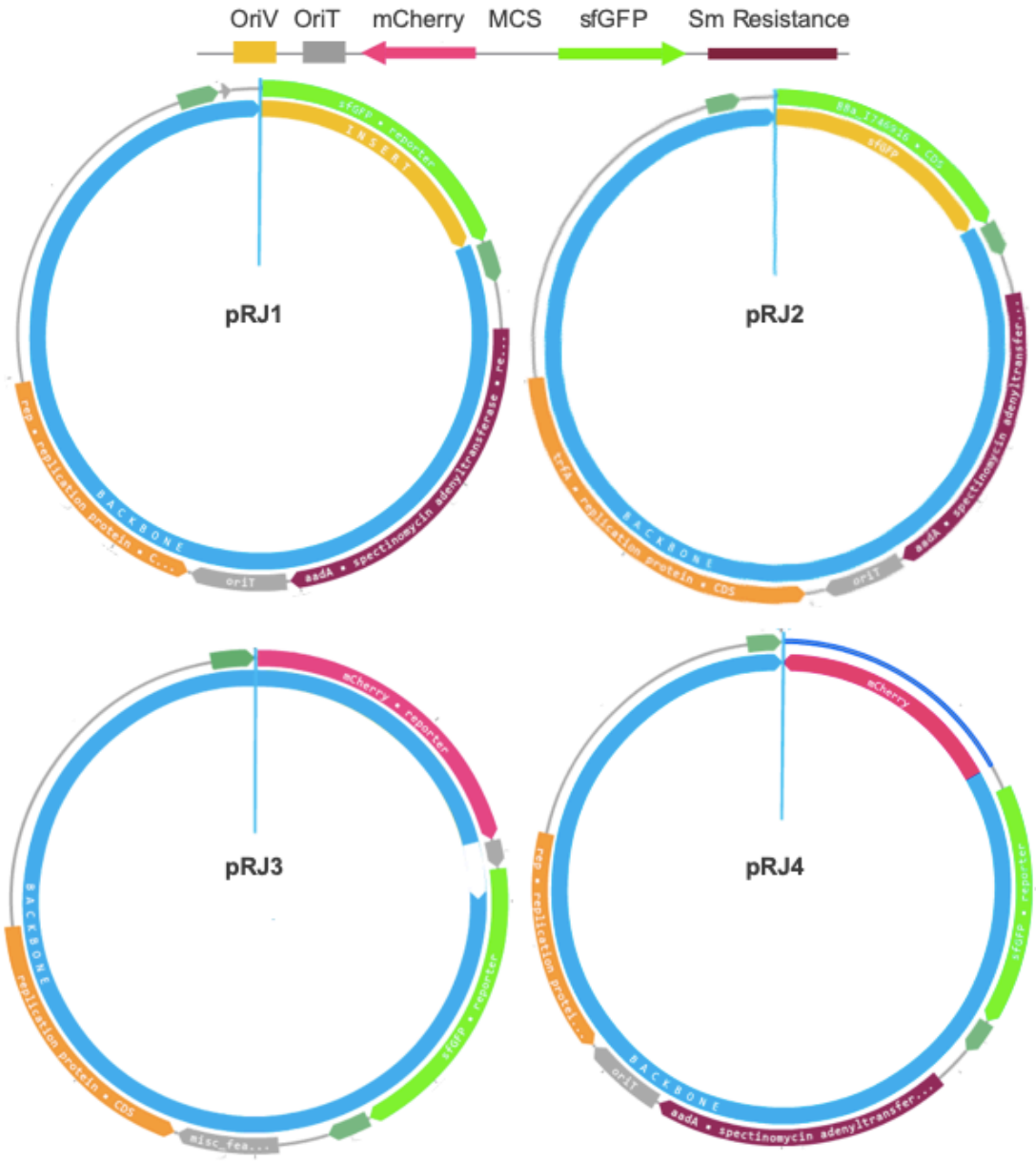
Graphics obtained by R studio software

This process resulted in a new family of vectors that have an architecture that gives them helpful characteristics for the analysis of TF-promoter expression systems because they are bidirectional systems that monitor the amount of TFa and analyze the strength that these unions print in the transcription. All of them have an origin of replication ori-pBBR1 of low/medium number of copies (~ 15 copies per cell) that prevents excessive cellular expression of the unnatural genes of the microorganism that overload their metabolism and cellular machinery, but allow a level of a suitable expression for the analysis of dynamics with fluorescent proteins, this characteristic avoids large differences of

expression at the single-cell level (as it could happen with systems of a high number of copies that are more variable in their quantity of plasmids).

One system reported is a variant of GFP, sfGFP (super folding GFP) protein with improved tolerance to circular permutation, greater resistance to chemical denaturants and an improved and divergent folding kinetics found mCherry that have a wavelength of excitation and emission different from that of sfGFP to avoid noise between them; each of these proteins has a degradation tag of LVA and AAV that allows the degradation of the protein in a range of 15 minutes which gives the security of protein production between measurement intervals. As another advantage is that because it bases these plasmids on pSEVA vectors, however, it needs more validations, in other microorganisms or in other conditions to test its expression of fluorescent proteins.t contain a Transfer OriT that narrow-host range and broad-host-range. Thus, the vectors that possess sfGFP with the degradation tags LVA and AAV called pRJ1 and pRJ2 respectively. They called the plasmid containing the two reporter proteins with AAV degradation tail pRJ3 and the one with the LVA pRJ4 degradation tail (Figure 20). However, in the first characterization experiments, these did not present good kinetic, they suggest it to study other types of proteins and, like the synthetic biology statements, they dictate to play with each of the parts that contain them for future characterization.

Figure 20. A. Schematic representation of vectors constructed with the sfGFP reporter protein and in a divergent manner mCherry in the middle of the MCS for cloning and Vegetative Origin (Ori V) of pBBR1 together with the Transfer origin and the streptomycin resistance gene Sm AadA resistance of plasmid pVLT35. B Schematic representation of the vectors constructed with their variants that only contain the sfGFP (772 pbs) gene useful for future standardizations and the PJR3 and PRJ4 vectors that contain mCherry (744pbs) and that were constructed from PRJ1 and PRJ2.



Graphics obtained by Genome Compiler software.

IV CHAPTER 2

Modeling of TFs, *in silico* analysis and docking with potential ligands.

1. *In silico* methods

1.1 Threading modeling and 3D ligand models.

The modeling of the genes PP_3359, VanR and GalR implemented using the threading method to model proteins, this methodology chosen because after conducting a template search, no nearby proteins could find that had a crystal structure (See instruction point 2.1.2, Figure 10 / Figure 12), however, these families of TFs, although not highly conserved, have typical structures, which could be a starting point for using the threading method and getting good protein models.

The sequences of the proteins GalR (Gene ID 1045766) PP_3359 (Gene ID 046535) and VanR (Gene ID 1046176) got in the NCBI repository (https://www.ncbi.nlm.nih.gov/) from the genome of *P. putida KT2440*. For this, it developed a 3D model of the target protein using the sequence alignment by suitable homologous templates.Thus, performed by I-TASSER platform (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) protein models in 4 steps: threading template identification, iterative structure assembly simulation, model selection and refinement, and structure-based function annotation.

It carried the evaluation of the structure of the models out with the QMEAN program from the Swiss Institute of Bioinformatics (SIB). QMEAN is a software that estimates the quality of three-dimensional structures making an exhaustive comparison with the structures determined by crystallography or NMR deposited in the PDB, using two scoring functions QMEAN and QMEANclust.: (1) the interactions potentials of Cβ and (2) of all atoms, (3) the local geometry of the structure (i.e. torsion angles that form three consecutive residues), (4) the solvation potential (to describe the state of the most internal waste), and finally, (5) the predicted secondary structure and (6) solvent accessibility as for the last two descriptors. This score given locally and globally for the protein and given from 0 to 1 (1 being the highest and best) in terms of protein structure quality findings.

We got the structure files of ligands used for molecular docking studies against proteins models in the ZINC ligands repository (http://zinc.docking.org/), where phenolic compounds were: ferulic acid (ZINC accession 58258), vanillin (ZINC accession 2567933), vanillic acid (ZINC accession 338275) which are found naturally and form part of their degradation pathway. And, in a test, compounds with similarity to gallic acid were found to evaluate how homologs could bind to these proteins and measure these

interactions *in silico* form got from the platform PubChem (because not all were available in ZINC): Gallic acid (Pubchem CID 370), P-cresol (Pubchem CID 2879) and Phyrogallol (Pubchem CID 1057) (Figure 23). On the other hand, the compounds used to perform the virtual screening were taken from a database of 13571 commercial molecules of drugs approved by the FDA (Food and Drug Administration) obtained from the database of DrugBank (https : //www.drugbank.ca/releases/latest13) include 2,630 approved small molecule drugs, 1,377 approved biologics (proteins, peptides, vaccines, and allergenics), 131 nutraceuticals and over 6,374 experimental (discovery-phase) drugs.

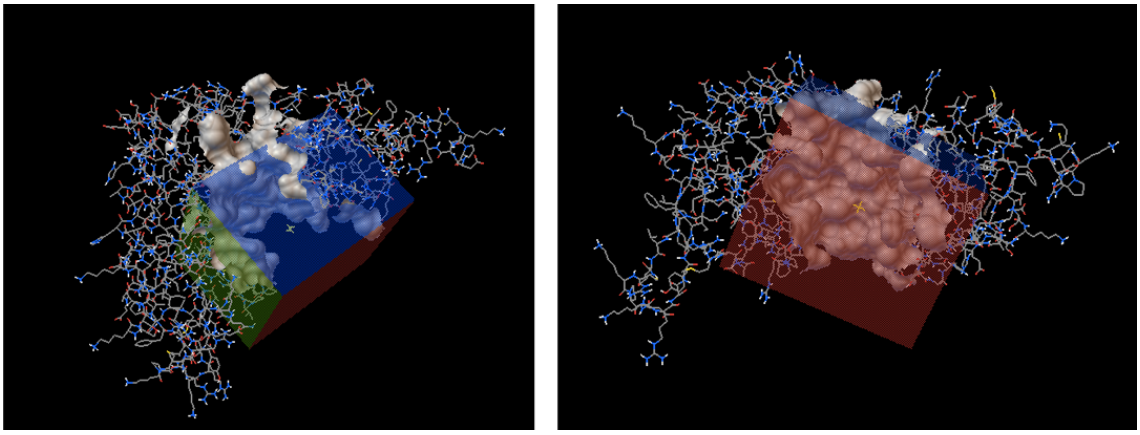
1.2 Pocket recognition and completion of Docking

To find pocket structures using CASTP 3.0 software (The Computed Atlas of Protein Surface Topography) that provides a quantitative characterization of protein cavities that could function as pockets for biologically important ligand binding. This program shows the cavities in a graphic representation of the structure of the proteins and also demarcates in the amino acid sequence the places where these pockets are located along with the residues of the peptide chains. These results were used to make a grid, which is the place where the ligands do the docking and measure the interactions that can occur in this place to bind (Figure 21). The motifs HTH that have it discarded these proteins to perform these dockings, since these are DNA binding sites that do not bind ligands. (we knew These data from the results got by BLAST and I-TASSER, which showed interactions of specific residues along the amino acid chain in the peptide chain with DNA).

Docking experiments performed using the AutoDock 4.2 program, which uses a combination of an empirical force field with a genetic algorithm Lamarckian and used the protocol proposed by the same program followed and it modified some processes. Polar hydrogens were added to each receptor, and then the charges were added to these and ligands by the Gasteiger method. Copper atom charges were supplied by modifying the coordinate file that AutoDock provides in the receiver's preparation.Thus, the docks that were performedmade with the compounds of the degradation pathway of the ferulic acid for PP_3359 and VanR along with those of the degradation pathway of Galic acid for GalR. 1000 energy evaluations were performed for each aromatic compound, and 5 models were obtained each time in search of different conformations of the ligands. Finally, the molecular interactions given in Autodock were determined by binding free energy of each ligand with each model and the binding energies given in kilo-heat per

mol (kcal = mol) were recorded. The smaller this number is, the greater the probability that the union will occur. We use as a control a molecule of salsalate (Pubchem CID DB01399) which is a non-steroidal anti-inflammatory agent with structures divergent to the compounds evaluated and from which no activity is expected.

Figure 21. Pocket adjustment in Pymol in VanR protein according to data obtained by CASTP 3.0



1.3 Virtual screening on commercial molecules database - FDA

VS experiments performed using the Gold Docking, this program uses set points on the protein hydrogen levels and assigns a donor point on the protein and vice versa. In addition to this, the CH groups of the ligand are exposed to the hydrophobic points of the receptor protein throughout the entire structure that is analyzed.

The search algorithm to explore possible link modes; GOLD uses a genetic algorithm in which the dihedral parameters of rotary ligand bonds, ligand ring geometries, dihedrals of OH protein groups and NH groups and the mappings of the set points (i.e. the position of the ligand) are considered. at the binding site). Of course, at the start of a coupling run, all of these variables are random.

We perform simulations of molecular coupling between the structure of PP_3359, VanR and GalR and a database of 3000 commercial molecules of drugs approved by the FDA (Food and Drug Administration) obtained from the database of DrugBank (https: //www.drugbank.ca/releases/latest13) using the Gold Docking software14 performing 10 repetitions for each of the molecules. The objective was to select higher hits the criterion of the best score value for coupling in the coordinates of the region of the protein structure function site.

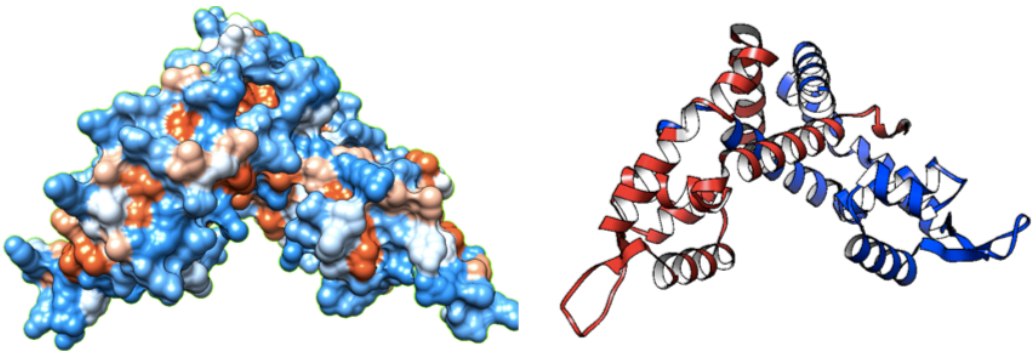
2. *In silico* Results

2.1. Protein and models

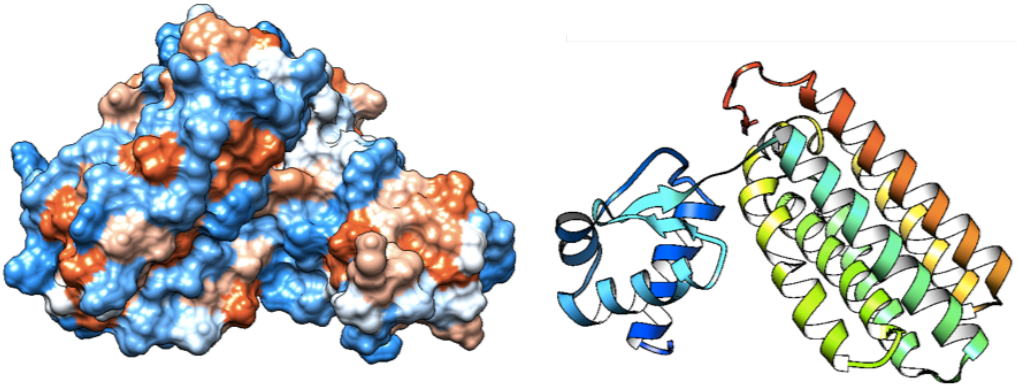
The PP_3359 protein modeled as a heterodimer and even I-TASSER were modeled by MarR protein family, their prediction resulted in the modeling obtained a quaternary structure in the form of homodimer and a QMEAN value of 0.71 ±0.5 (Figure 22.A). The VanR gene was predicted as a homotetramer from the template of a TF of the GntR family from *Pseudomonas sp.* with a QMEAN of 0.6 ±0.5 (Figure 22.B). Finally, the GalR protein modeling was performed with the NMR structure template TFs of LysR type regulator from different bacterial with a value of 0.67 ± 0.07. (Models with a global score greater than 0.5 chosen because above this cut-off point, It is acceptable for coupling and dynamic studies). And we observe similar behavior in all protein structures because they all have DNA-binding H-T-H domains that give this structure.

Figure 22. Tertiary structure obtained by threading of TFs from *P. putida KT2440*. On the right, a surface structure that leaves in colors of heteroatoms that leaves the protein hydrophobicity map in interview (under the color code indicated by White for hydrogen, Black for carbon Sky blue for nitrogen and Red for oxygen) and to the Left a model in a tape diagram where you can see the protein chains. A. PP_3359 Protein model B. VanR Protein model. C. GalR protein model.

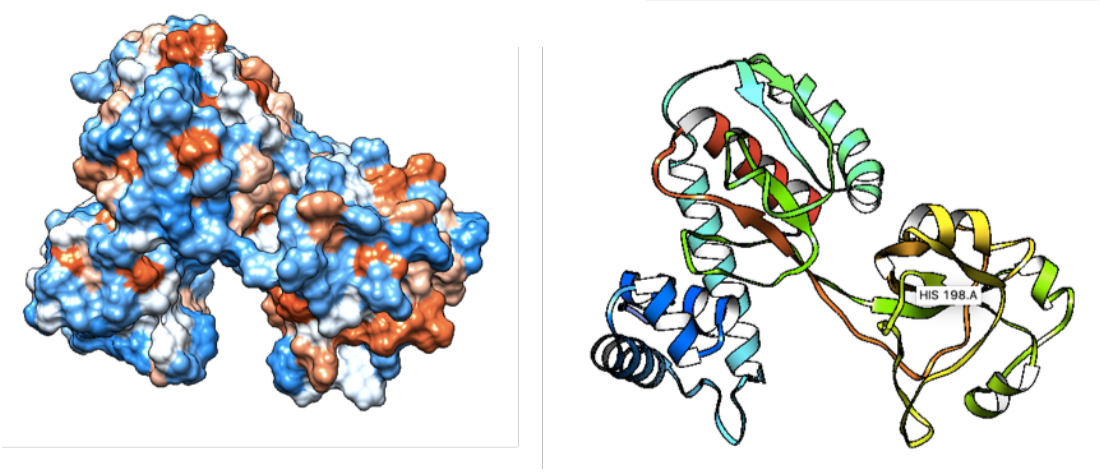
A.



B.

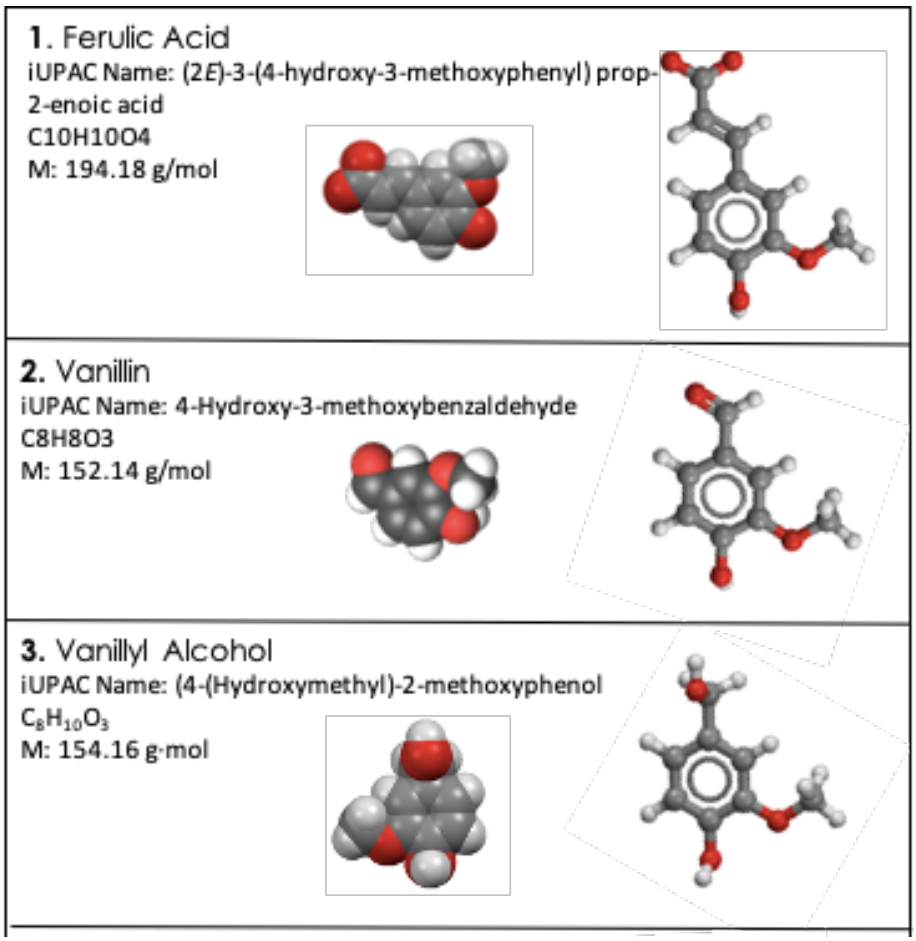


C.

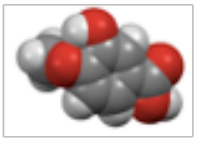


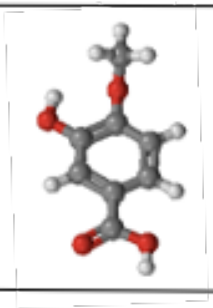
Obtained by UCSF Chimera.[129]

Figure 23. Graphical representation in sticks and balls of Aromatic compounds of the degradation life of ferulic acid (1-6) and Gallic acid (7-10). Molecular Formula, Molecular Weight, Ball and Stick, Space-Filling and. IUPAC name

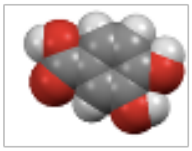


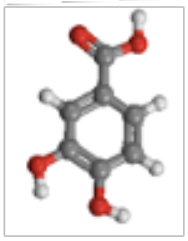
4. Isovanilic Acid
iUPAC Name: 3-hydroxy-4-methoxybenzoic acid
C8H8O4
M: 168.15 g·mol



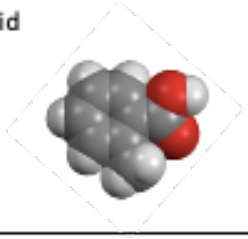


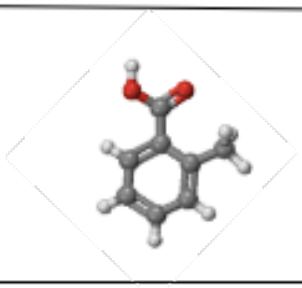
5. Pyrocatechuic acid
iUPAC Name: 2,3-Dihydroxybenzoic acid
C7H6O4
M: 154.12 g/mol



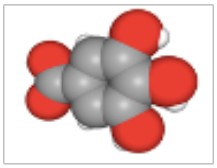


6. Benzoic acid
iUPAC Name: benzoic acid
C7H6O2
M: 122.12 g/mol

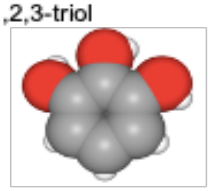


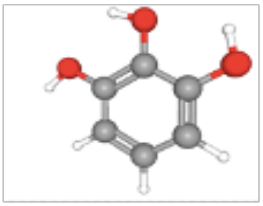


7. Gallic Acid
iUPAC Name: Zinc;4-carboxy-2,6-dihydroxyphenolate
C14H10O10Zn
M: 403.6 g/mol

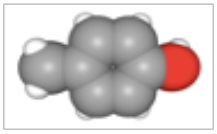


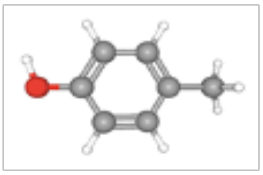
8. Pyrogallol
iUPAC Name: benzene-1,2,3-triol
Chem. Formula:C6H6O3
M: 126.11 g/mol



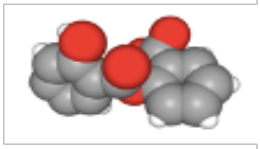


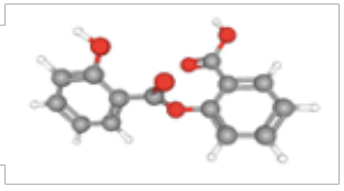
9. Pcresol
iUPAC Name: 4-methylphenol
C7H8O
M: 108.14 g/mol





10. Salsate
iUPAC Name: 2-(2-hydroxybenzoyl)oxybenzoic acid
C14H10O5
M: 258.23 g/mol





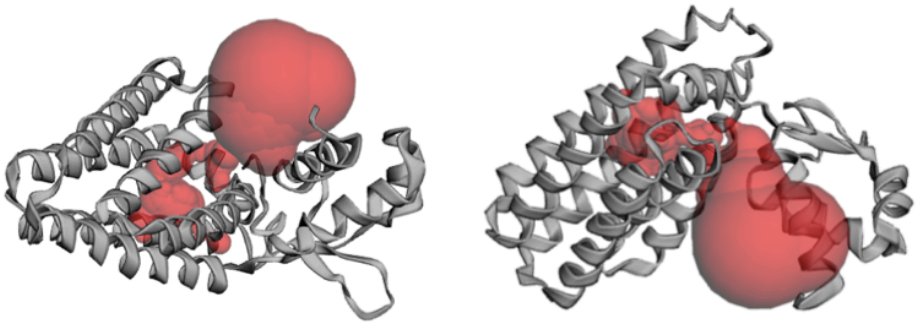
2.2. Pocket recognition, performing of Docking and VS

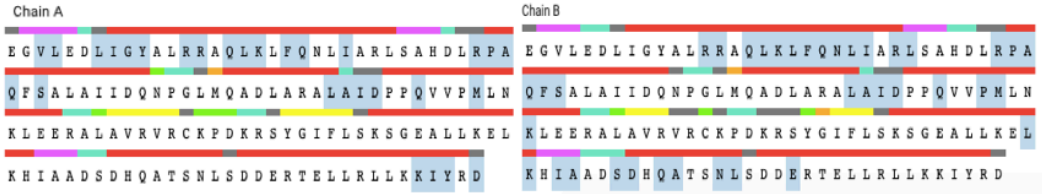
CASTP 3.0 recognized the sites where TFs might join by having anatomy and pocket. And the amino acid clusters that make up the pocket of each structure that are mostly hydrophobic. It shows the outputs of this program in Figure 24. And these were the ones chosen to do the grid formation during the docking process to have contact with the ligands in which a cube formed within these places of not over 30 x 30 x 30 Angstroms (data recommended by the software, AutoDock search space sizes are given in Angstroms, a unit of measurement used to represent molecular distances in which 100 million Anstroms fit in a centimeter) to direct docking and ligands to that location.

Regarding the studies carried out the configurations with the lowest energy that each of the aromatic compounds adopted chosen, based on the principle of minimal entropy. It represents these associations in Figure 25 and 26. Where it always observed that the negative sauce control is the one with the highest score. For the couplings made with VanR and PP_3359, the same behavior is observed, being the ferulic acid that is the most abundant molecule of all in nature were the one that presented a greater bond strength, and this is evolutionarily understandable because as it progresses the degradation the force with which the other compounds can bound is weaker because less amount of the proteins produced by the operons is needed to degrade the aromatic compounds and use them as carbon sources These results can be interesting if these interactions could modify the strength of the promoters because they could make biosensors that have different forces that could be controlled by changing only the inductor and not having to exchange each of these parts to have different forces (as is the case and the names of different promoters responsible for IPTG).

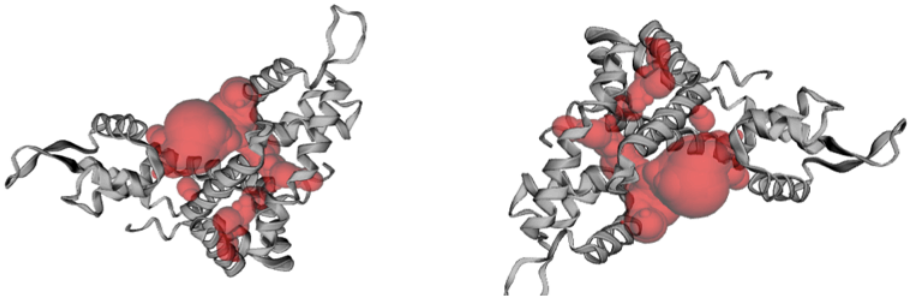
Figure 24. CASTP 3.0 outputs. In the upper part, the protein structure is located showing in red the cavities found in two different angles, on the other hand, in the lower part is the peptide chain showing each one of the residues that have contact in shading With the structures it is pocket shape and showing in the α -helix in red, β-bridge in orange, 3_{10} helix in purple and in pink π helix representation, and green and grey colors for bend al coil structures in the protein. A. VanR output B. PP_3359 output and C. GalR output.

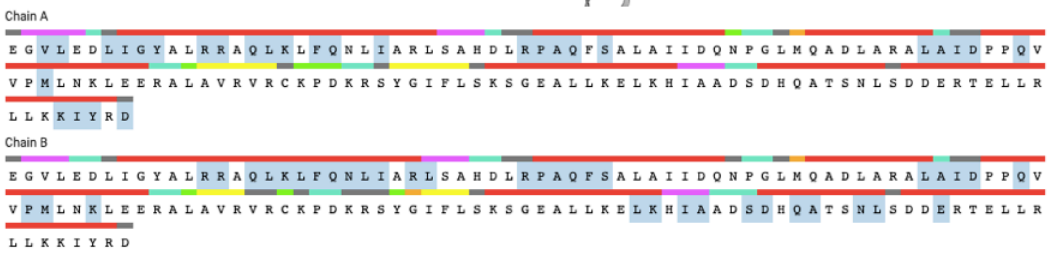
A.



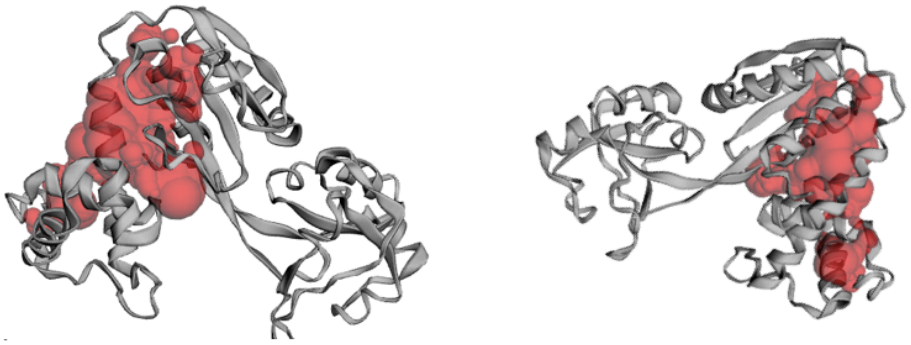

Chain A
EGVLEDLIGYALRRAQLKLFQNLIARLSAHDLRPA
QFSALAIIDQNPGLMQADLARALAIDPPQVVPMLN
KLEERALAVRVRCKPDKRSYGIFLSKSGEALLKEL
KHIAADSDHQATSNLSDDERTELLRLLKKIYRD
Chain B
EGVLEDLIGYALRRAQLKLFQNLIARLSAHDLRPA
QFSALAIIDQNPGLMQADLARALAIDPPQVVPMLN
KLEERALAVRVRCKPDKRSYGIFLSKSGEALLKEL
KHIAADSDHQATSNLSDDERTELLRLLKKIYRD

B.




Chain A
EGVLEDLIGYALRRAQLKLFQNLIARLSAHDLRPAQFSALAIIDQNPGLMQADLARALAIDPPQV
VPMLNKLEERALAVRVRCKPDKRSYGIFLSKSGEALLKELKHIAADSDHQATSNLSDDERTELLR
LLKKIYRD
Chain B
EGVLEDLIGYALRRAQLKLFQNLIARLSAHDLRPAQFSALAIIDQNPGLMQADLARALAIDPPQV
VPMLNKLEERALAVRVRCKPDKRSYGIFLSKSGEALLKELKHIAADSDHQATSNLSDDERTELLR
LLKKIYRD

C.



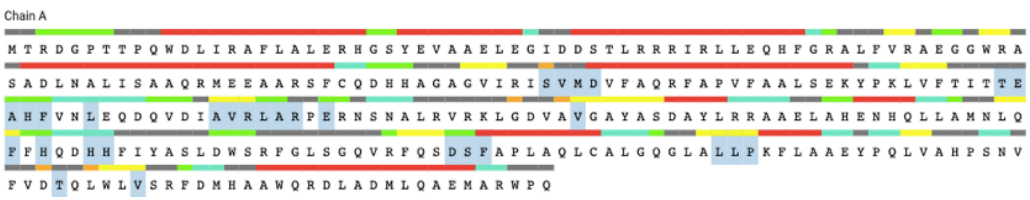

Chain A
MTRDGPTTPQWDLIRAFLALERHGSYEVAAELEGIDDSTLRRRIRLLEQHFGRALFVRAEGGWRA
SADLNALISAAQRMEEAARSFCQDHHAGAGVIRISVMDVFAQRFAPVFAALSEKYPKLVFTITTE
AHFVNLEQDQVDIAVRLARPERNSNALRVRKLGDVAVGAYASDAYLRRAAELAHENHQLLAMNLQ
FFHQDHHFIYASLDWSRFGLSGQVRFQSDSFAPLAQLCALGQGLALLPKFLAAEYPQLVAHPSNV
FVDTQLWLVSRFDMHAAWQRDLADMLQAEMARWPQ

Figure 25. Docking results for VanR and PP_3359.

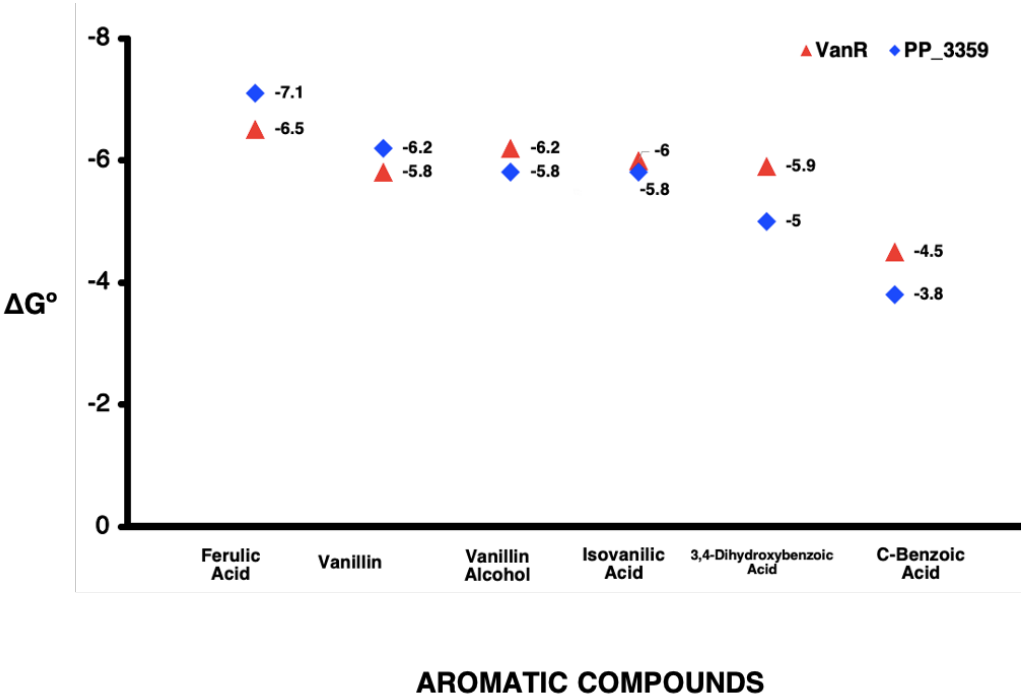
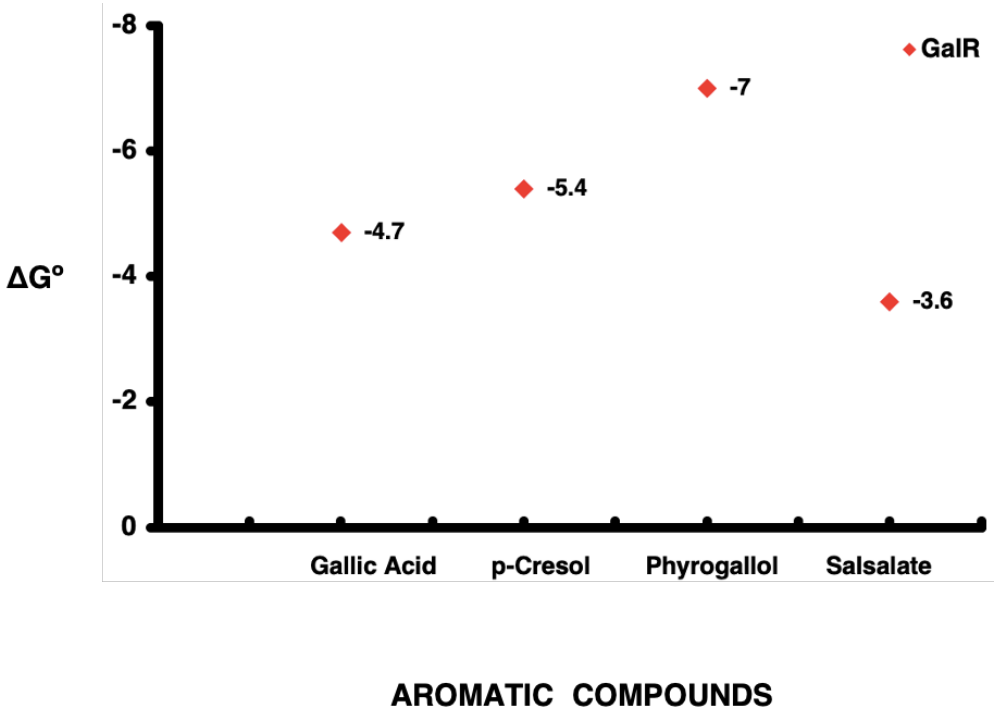


Figure 26. Docking results for GalR.



With the GalR protein and the results of its dockings, and the previous two, the control shows a fairly positive behavior concerning the molecules evaluated, as it used homologous molecules, and they presented structural similarity here the behavior was different.The P-cresol, which is a smaller molecule, has a lower score and it was seen in one observation that it is because they do not have molecules that can have an esthetic impediment in the performance of molecular docking. On the other hand, Phyrogallol has an even higher score than gallic acid, this molecule is derived from gallic acid (it has the same number of donors and receptors) and is obtained by heating gallic acid and its conformation presents more fit to the pocket than the gallic acid itself and could inducer for the design of this system.

After analyzing compounds that belong to the metabolic pathways of each of the proteins studied. The VS was perform. The results obtained were graphed showing the 220 compounds with the highest performance during the SV. The score> 60 was taken as a limit to consider a molecule as a candidate for inducing TF.

Thus, during a first screening and a triplicate of the protein PP_3359 with the drug library (Figure 27), no compound was found with an outstanding score (> 60) with these proteins. The second analysis of the TFs belonging to the degradation of phenolic compounds with the VanR protein (Figure 28). It showed 5 outstanding molecules. Two of them were ruled out because if antibiotics were used as inductors, the bacteria could inhibit their bacterial growth. Thus, the candidate molecules were Pemetrexed, which is an anticancer agent and 2 antifungals from the azole family (miconazole and ketoconazole) (Table 6) Which could be interesting, because in the structure of the azoles a pharmacophore could be found to activate these proteins. Added to that they are attractive candidates because these molecules are accessible and cheap.

Likewise, in the analysis of the protein GalR (Figure 29), affinities with these compounds were observed with azole groups in which the butoconazole had the highest score during the trial. On the other hand, we observed the antibiotic Cephatothin that was discarded and naloxone and letrozole (Table 7), which are molecules that could be analyzed for cytotoxicity and pharmacokinetics if it were to delve into their study for potential use in circuits for the health area.

Figure 27. PP_3359 Results VS.

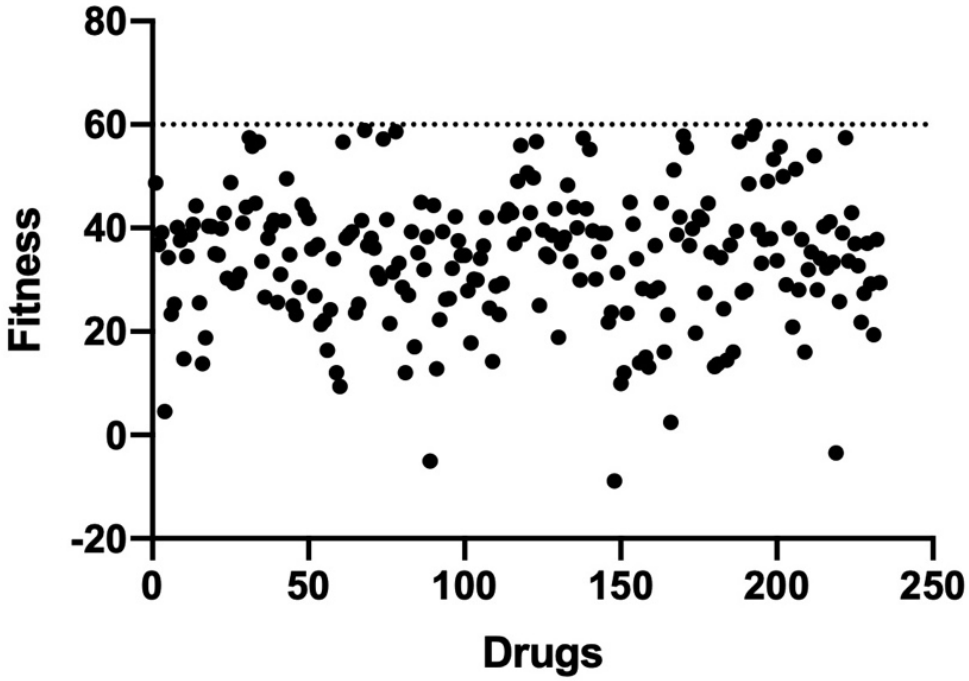


Figure 28. VanR Results VS drugs.

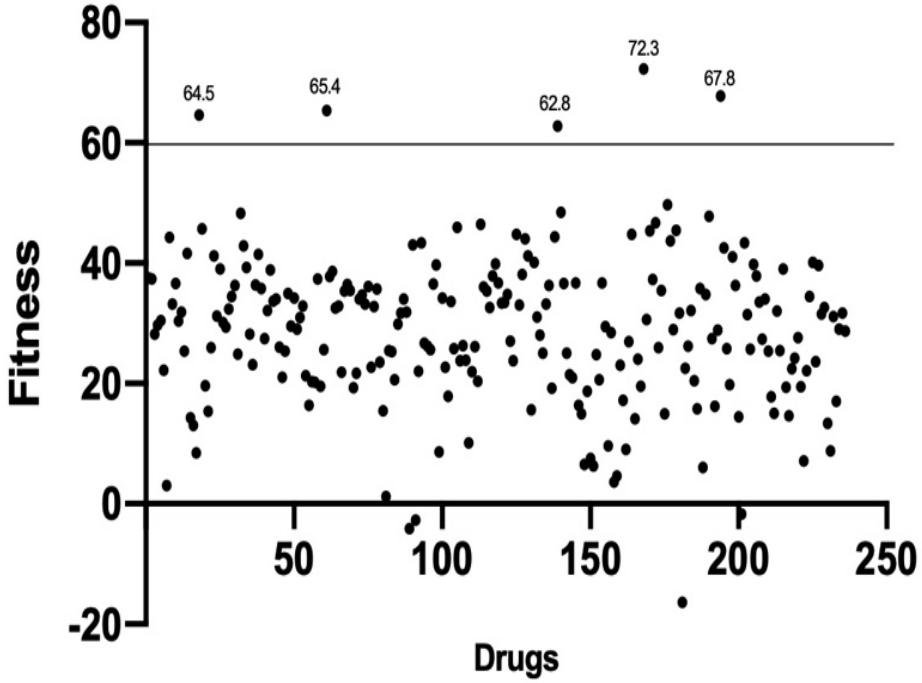


Table 6. Output VS with SCORE of VanR.

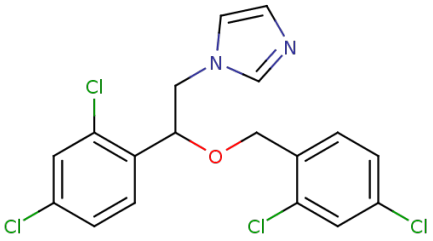
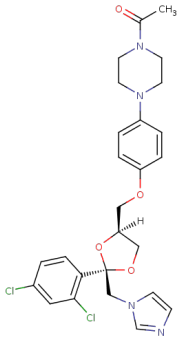
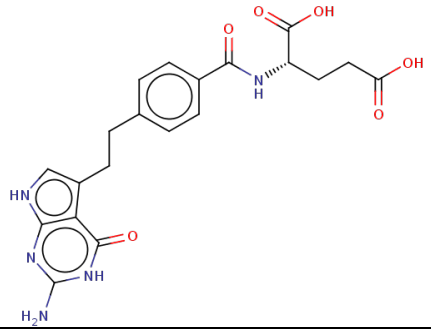
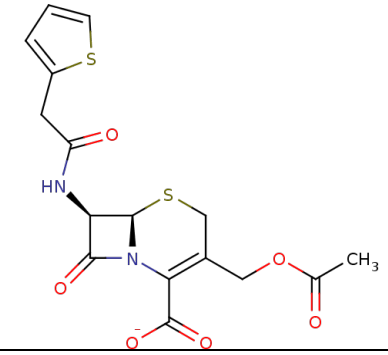
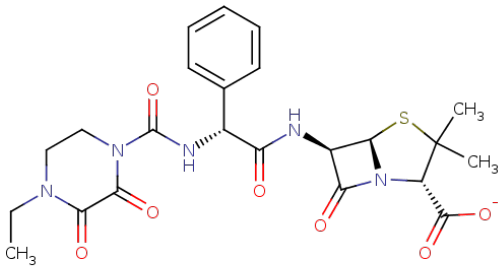
Fitness VanR	Ligant	Estructure
62.8	'BindingDB_31772_1' Miconazole An imidazole antifungal agent.	
64.59	'BindingDB_8610_1' Ketoconazole Is an antifungal medication used to treat a number of fungal infections.	
65.4	'BindingDB_18796_1' Pemetrexed Is a chemotherapy drug manufactured. Its indications are the treatment of pleural mesothelioma and non-small cell lung cancer.	
67.8	'BindingDB_53513_1' Cephalothin/Cefalexina Antibiotic of the group of cephalosporins known as first generation. It is used to treat bacterial infections in the respiratory tract (pneumonia, pharyngitis).	
72.3	'BindingDB_39348_1' Piperacillin It is a broad-spectrum β-lactam antibiotic of the ureidopenicillin class.	

Table 7. Output VS with SCORE of GalR.

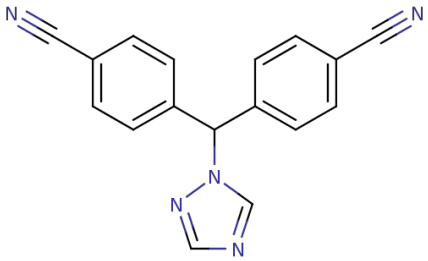
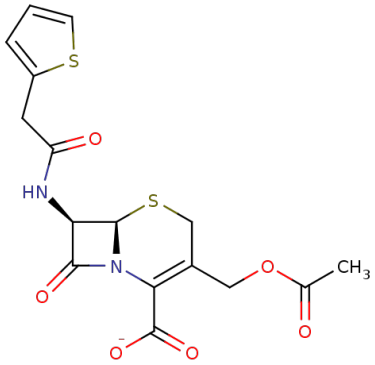
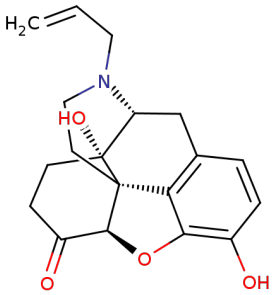
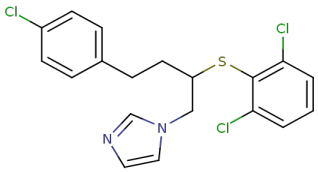
Fitness GalR	Ligand	Structure
65.77	BindingDB_13061_1' Letrozole Nonsteroidal aromatase inhibitor that has been introduced for the adjuvant treatment of hormonally sensitive breast cancer.	
68.16	'BindingDB_53513_1' Cephalothin Antibiotic most impermeable to attack by staphylococcal betalactamase, it is very effective in severe staphylococcal infections, such as endocarditis.	
75.71	'BindingDB_54795_1' Naloxone Opioid antagonist medication that is used to block or reverse the effects of opioid medications.	
67.49	'BindingDB_79206_1' Butoconazole Imidazole antifungal used in gynecology.	

Figure 29. GalR Results virtual screening drugs

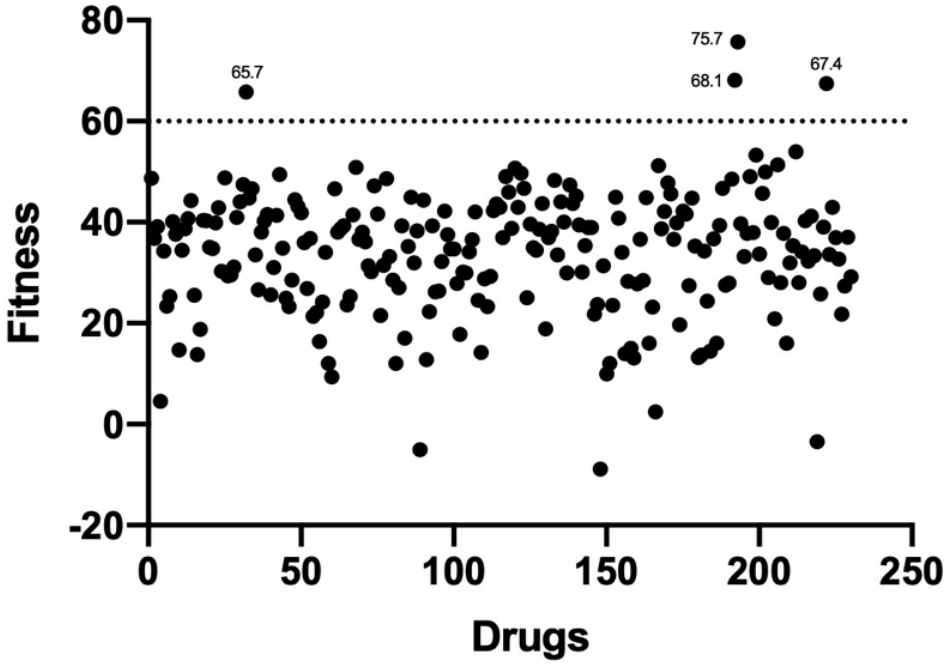
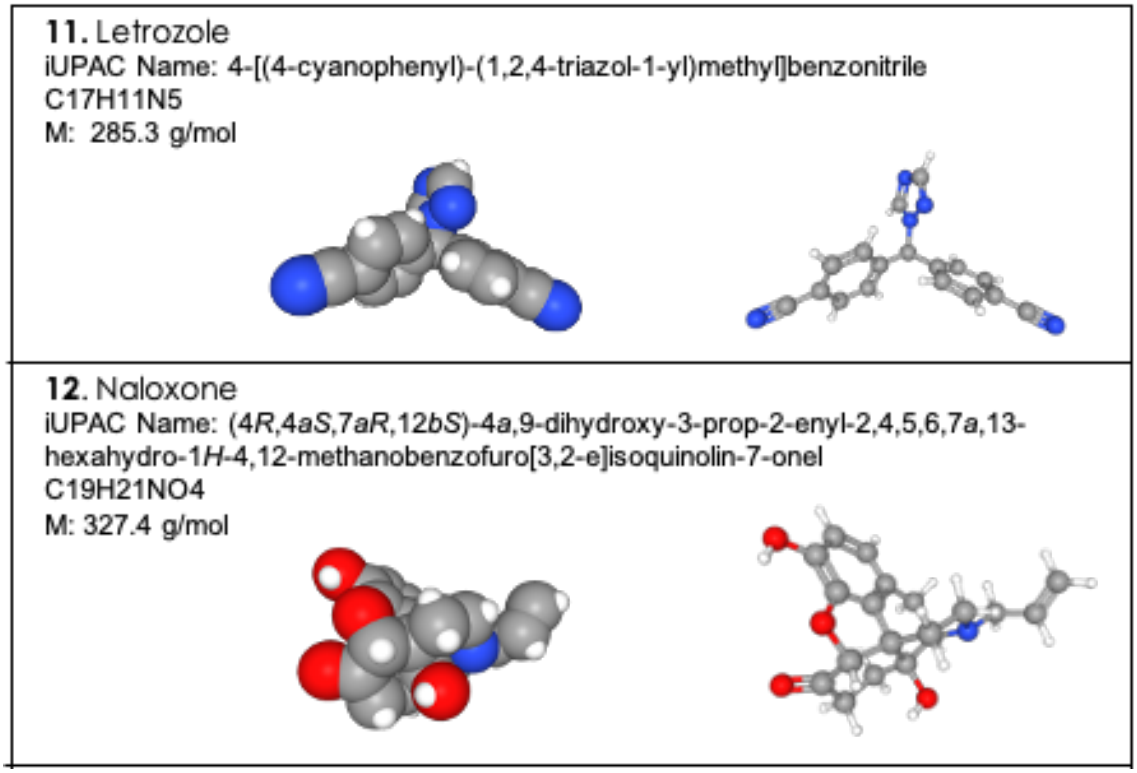


Figure 30. Graphical representation in sticks and balls of Drugs out put in VS. Molecular Formula, Molecular Weight, Ball and Stick, Space-Filling and. IUPAC name

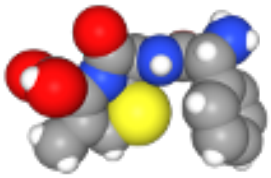


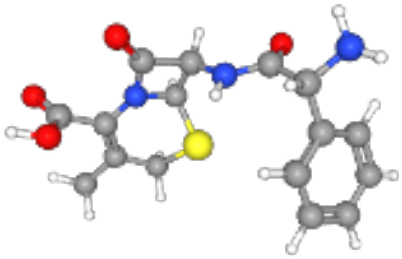
13. Cephalexin

iUPAC Name: (6*R*,7*R*)-7-[[(2*R*)-2-amino-2-phenylacetyl]amino]-3-methyl-8-oxo-5-thia-1-azabicyclo[4.2.0]oct-2-ene-2-carboxylic acid

C16H17N3O4S

M: 347.4 g/mol



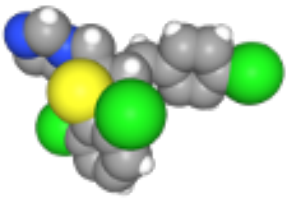


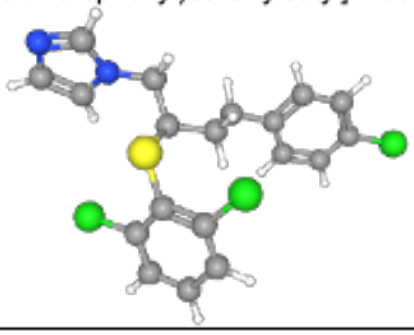
14. Butoconazole

iUPAC Name: 1-[4-(4-chlorophenyl)-2-(2,6-dichlorophenyl)sulfanylbutyl]imidazole

C19H17Cl3N2S

M: 2411.8 g/mol



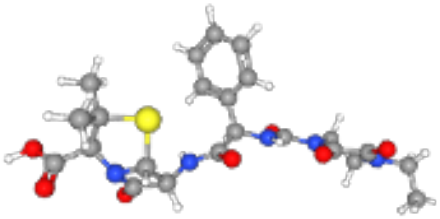


15. Piperacillin

iUPAC Name: (2*E*)-3-(4-hydroxy-3-methoxyphenyl)prop-2-enoic acid

C23H27N5O7S

M: 517.6 g/mol

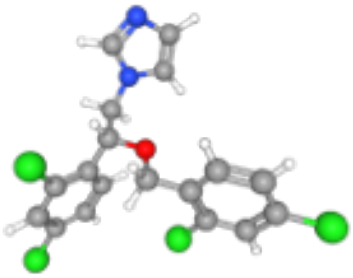


16. Miconaxole

iUPAC Name: 1-[2-(2,4-dichlorophenyl)-2-[(2,4-dichlorophenyl) methoxy] ethyl]imidazole

C18H14Cl4N2O

M: 416.1 g/mol

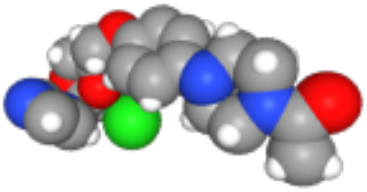


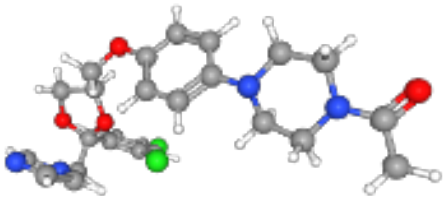
17. Ketoconazole

iUPAC Name: 1-[4-[4-[[(2*S*,4*R*)-2-(2,4-dichlorophenyl)-2-(imidazol-1-ylmethyl)-1,3-dioxolan-4-yl]methoxy]phenyl]piperazin-1-yl]ethanone

$C_{26}H_{28}Cl_2N_4O_4$

M: 531.4 g/mol



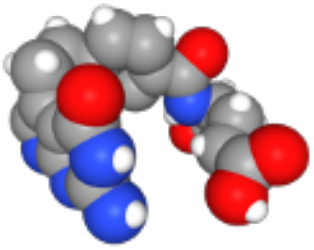


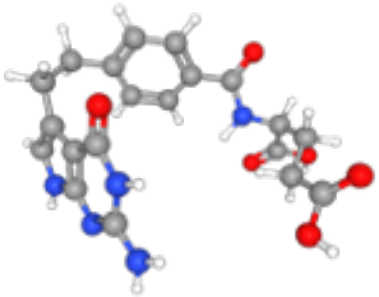
18. Pemetrexed

iUPAC Name: (2*S*)-2-[[4-[2-(2-amino-4-oxo-3,7-dihydro pyrrolo[2,3-d]pyrimidin-5-yl) ethyl] benzoyl]amino] pentanedioic acid

$C_{20}H_{21}N_5O_6$

M: 427.4 g/mol





V. CONCLUSIONS

As we have highlighted it throughout this document, the creation of new tools for synthetic biology is the key point for the development of itself, this is how throughout the realization of this master we designed these tools using experimental and computational approaches.

In an experimental approach, a family of hunting vectors of TF detection systems and promoters with fluorescent proteins was designed for applications in genomics, metagenomic or synthetic approaches in which it can see how the production of a protein imprints force on the transcription of an easy way in various microorganisms. Unfortunately, the standardization of these types of vectors requires more in-depth analysis to use and standardized. We suggest that this system may not work due to two theories. The first of them is that when proteins optimized to be smaller, such as sfGFP and a degradation tag are added that will interfere with 11 amino acids (33 bps) the formation of the tertiary structure that formed the amino acids responsible for emitting fluorescence may not work. And the second is that the level like this protein is rapidly doubled, produced and degraded very quickly and not be detected at high levels during kinetic tests (which are every 30 minutes).

In this work, an analysis was performed in silico by molecular docking (Autodock Vina 4.2 Software) to analyze the possible bind that TFs have with ligands that belong to the metabolic pathway that they control. This was done with the metabolic pathway of degradation of phenolic compounds and gallic acid. Thus, the compounds with highest scores in the Docking carried out were the first in the metabolic pathway of degradation, while the aromatic compounds that were intermediate and final in the metabolic pathway had a less binding affinity in our tests made for the VanR, GalR, and PP_3359 proteins. This could suggest that the groups that are lost in the first reactions could be important for the binding of the ligands to the transcription factor. Data obtained could be analyzed in-depth and contrasted using other tools of medicinal chemistry.

In the results of the VS analyzes (God Docking Software), very interesting molecules were found as compounds derived from azoles. These are affordable and are not cytotoxic to animal cells. Therefore, they are accessible to be used in the industry and in the health area as promising compounds. However, as with canonical compounds, additional studies are proposed from these results such as pharmacophore prediction, molecular dynamics, toxicological predictions, and even coupling to measure the

distances in angstroms of these junctions to propose these molecules for the construction of new genetic circuits.

VI. REFERENCES

1. Madigan, M., Martinko, J., Dunlap, P. & Clark, D. Brock Biología de los microorganismos, 12th Edition. *Pearson* (2009). doi:10.1007/s13398-014-0173-7.2
2. Church, G. M., Elowitz, M. B., Smolke, C. D., Voigt, C. A. & Weiss, R. Realizing the potential of synthetic biology. *Nat. Publ. Gr.* **15**, (2014).
3. de Lorenzo, V. & Danchin, A. Synthetic biology: Discovering new worlds and new words. The new and not so new aspects of this emerging research field. *EMBO Rep.* (2008). doi:10.1038/embor.2008.159
4. Cheng, A. A. & Lu, T. K. Synthetic Biology: An Emerging Engineering Discipline. *Annu. Rev. Biomed. Eng.* (2012). doi:10.1146/annurev-bioeng-071811-150118
5. Medema, M. H. & Zhao, H. Editorial: Synthetic biology and bioinformatics. *Natural Product Reports* **33**, 913–914 (2016).
6. Keasling, J. D. Synthetic Biology for Synthetic Chemistry. *ACS Chem. Biol.* **3**, 64–76 (2008).
7. Isaacs, F. J., Dwyer, D. J. & Collins, J. J. RNA synthetic biology. *Nature Biotechnology* (2006). doi:10.1038/nbt1208
8. Kwok, R. Five hard truths for synthetic biology. *Nature* (2010). doi:10.1038/463288a
9. Adams, B. L. The Next Generation of Synthetic Biology Chassis: Moving Synthetic Biology from the Laboratory to the Field. *ACS Synth. Biol.* **5**, 1328–1330 (2016).
10. Keasling, J. D. Synthetic biology and the development of tools for metabolic engineering. *Metab. Eng.* **14**, 189–195 (2012).
11. Edition, E. Brock Biology of Microorganisms. *Cell* (2006).
12. Biologia Molecular e Genética Bacteriana | Sociedade Brasileira de Genética. Available at: https://www.sbg.org.br/pt-br/livros/biologia-molecular-e-genetica-bacteriana. (Accessed: 29th October 2019)
13. Anirban, R. O. Y. & Parker, R. S. Dynamic modeling of free fatty acid, glucose, and insulin: An extended 'minimal model'. *Diabetes Technol. Ther.* **8**, 617–626 (2006).
14. Abil, Z., Xiong, X. & Zhao, H. Synthetic Biology for Therapeutic Applications. *Mol. Pharm.* **12**, 322–331 (2015).
15. Rehm, B. H. A. Bacterial polymers: Biosynthesis, modifications and applications. *Nature Reviews Microbiology* **8**, 578–592 (2010).
16. Kramer, M. G., Masner, M., Ferreira, F. A. & Hoffman, R. M. Bacterial Therapy of Cancer: Promises, Limitations, and Insights for Future Directions. *Front. Microbiol.* **9**, 16 (2018).
17. Chien, T., Doshi, A. & Danino, T. Advances in bacterial cancer therapies using synthetic biology. *Curr. Opin. Syst. Biol.* **5**, 1–8 (2017).
18. Nora, L. C. *et al.* The art of vector engineering: towards the construction of next-generation genetic tools. *Microbial Biotechnology* **12**, 125–147 (2019).
19. de Lorenzo, V. Designing microbial systems for gene expression in the field. *Trends in Biotechnology* **12**, 365–371 (1994).
20. Martínez-Garćía, E., Aparicio, T., Goñi-Moreno, A., Fraile, S. & De Lorenzo, V. SEVA 2.0: An update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. *Nucleic Acids Res.* **43**, D1183–D1189 (2015).
21. Silva-Rocha, R. *et al.* The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res.* **41**, D666-75 (2013).
22. seva. Available at: http://seva.cnb.csic.es/. (Accessed: 26th October 2019)
23. Jahn, M., Vorpahl, C., Hübschmann, T., Harms, H. & Müller, S. Copy number variability of expression plasmids determined by cell sorting and droplet digital

PCR. *Microb. Cell Fact.* **15**, (2016).
24. Jabri, E. Tag, you're degraded. *Nature Structural Biology* **10**, 676 (2003).
25. Levchenko, I., Grant, R. A., Wah, D. A., Sauer, R. T. & Baker, T. A. Structure of a delivery protein for an AAA+ protease in complex with a peptide degradation tag. *Mol. Cell* **12**, 365–372 (2003).
26. Purcell, O., Grierson, C. S., Bernardo, M. D. & Savery, N. J. Temperature dependence of ssrA-tag mediated protein degradation. *Journal of Biological Engineering* **6**, (2012).
27. Karzai, A. W., Roche, E. D. & Sauer, R. T. The SsrA–SmpB system for protein tagging, directed degradation and ribosome rescue. *Nat. Struct. Biol. 2000 76* **7**, 449–455 (2000).
28. Baker, T. A. & Sauer, R. T. ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochimica et Biophysica Acta - Molecular Cell Research* **1823**, 15–28 (2012).
29. Juárez, J. F., Lecube-Azpeitia, B., Brown, S. L., Johnston, C. D. & Church, G. M. Biosensor libraries harness large classes of binding domains for construction of allosteric transcriptional regulators. *Nat. Commun.* **9**, 3101 (2018).
30. Teo, J. J. Y., Woo, S. S. & Sarpeshkar, R. Synthetic Biology: A Unifying View and Review Using Analog Circuits. *IEEE Transactions on Biomedical Circuits and Systems* **9**, 453–474 (2015).
31. Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. *Nature* **467**, 167–173 (2010).
32. Martínez-García, E. & de Lorenzo, V. Pseudomonas putida in the quest of programmable chemistry. *Curr. Opin. Biotechnol.* **59**, 111–121 (2019).
33. Belda, E. *et al.* The revisited genome of *Pseudomonas putida* KT2440 enlightens its value as a robust metabolic *chassis*. *Environ. Microbiol.* **18**, 3403–3424 (2016).
34. Nelson, K. E. *et al.* Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440. *Environ. Microbiol.* **4**, 799–808 (2002).
35. Nakazawa, T. Travels of a Pseudomonas, from Japan around the world. *Environmental Microbiology* **4**, 782–786 (2002).
36. Regenhardt, D. *et al.* Pedigree and taxonomic credentials of Pseudomonas putida strain KT2440. *Environ. Microbiol.* (2002). doi:10.1046/j.1462-2920.2002.00368.x
37. Nelson, K. E. *et al.* Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440. *Environ. Microbiol.* (2002). doi:10.1046/j.1462-2920.2002.00366.x
38. Kim, Y. H. *et al.* Analysis of aromatic catabolic pathways in Pseudomonas putida KT 2440 using a combined proteomic approach: 2-DE/MS and cleavable isotope-coded affinity tag analysis. *Proteomics* **6**, 1301–1318 (2006).
39. Heim, S. *et al.* Proteome reference map of Pseudomonas putida strain KT2440 for genome expression profiling: Distinct responses of KT2440 and Pseudomonas aeruginosa strain PAO1 to iron deprivation and a new form of superoxide dismutase. *Environ. Microbiol.* **5**, 1257–1269 (2003).
40. Weinel, C., Nelson, K. E. & Tümmler, B. Global features of the Pseudomonas putida KT2440 genome sequence. *Environ. Microbiol.* **4**, 809–818 (2002).
41. Stjepandić, D. *et al.* The genome structure of Pseudomonas putida: High-resolution mapping and microarray analysis. *Environ. Microbiol.* **4**, 819–823 (2002).
42. Halan, B., Schmid, A. & Buehler, K. Maximizing the productivity of catalytic biofilms on solid supports in membrane aerated reactors. *Biotechnol. Bioeng.* **106**, 516–527 (2010).
43. Eduardo Díaz* and María A Prieto. Bacterial promoters triggering biodegradation of aromatic pollutants. *Curr. Opin. Biotechnol.* **11**, 467–475 (2000).

44. Browning, D. F. & Busby, S. J. W. The regulation of bacterial transcription initiation. *Nature Reviews Microbiology* **2**, 57–65 (2004).
45. Moretfl, E. & Segovia2t, L. MINIREVIEW The r54 Bacterial Enhancer-Binding Protein Family: Mechanism of Action and Phylogenetic Relationship of Their Functional Domains. *J. Bacteriol.* (1993).
46. Schulz, S. *et al.* Elucidation of Sigma Factor-Associated Networks in Pseudomonas aeruginosa Reveals a Modular Architecture with Limited and Function-Specific Crosstalk. *PLoS Pathog.* (2015). doi:10.1371/journal.ppat.1004744
47. Gross, C. A. *et al.* The functional and regulatory roles of sigma factors in transcription. in *Cold Spring Harbor Symposia on Quantitative Biology* (1998). doi:10.1101/sqb.1998.63.141
48. Morett, E. & Segovia, L. The σ54 bacterial enhancer-binding protein family: Mechanism of action and phylogenetic relationship of their functional domains. *Journal of Bacteriology* (1993). doi:10.1128/jb.175.19.6067-6074.1993
49. Barabási, A. L. & Oltvai, Z. N. Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics* (2004). doi:10.1038/nrg1272
50. RCSB PDB - 5EMQ: Transcription factor GRDBD and GRE complex. Available at: https://www.rcsb.org/structure/5EMQ. (Accessed: 12th May 2020)
51. Reva, O. N. *et al.* Functional genomics of stress response in Pseudomonas putida KT2440. *J. Bacteriol.* **188**, 4079–4092 (2006).
52. Keller, A. D. Model genetic circuits encoding autoregulatory transcription factors. *J. Theor. Biol.* **172**, 169–185 (1995).
53. Renzi, F., Rescalli, E., Galli, E. & Bertoni, G. Identification of genes regulated by the MvaT-like paralogues TurA and TurB of Pseudomonas putida KT2440. *Environ. Microbiol.* **12**, 254–263 (2010).
54. Fillenberg, S. B., Grau, F. C., Seidel, G. & Muller, Y. A. Structural insight into operator dre-sites recognition and effector binding in the GntR/HutC transcription regulator NagR. *Nucleic Acids Res.* **43**, 1283–96 (2015).
55. Tropel, D. & van der Meer, J. R. Bacterial transcriptional regulators for degradation pathways of aromatic compounds. *Microbiol. Mol. Biol. Rev.* **68**, 474–500, table of contents (2004).
56. Díaz, E., Jiménez, J. I. & Nogales, J. Aerobic degradation of aromatic compounds. *Current Opinion in Biotechnology* (2013). doi:10.1016/j.copbio.2012.10.010
57. Fukumori, F. & Saint, C. P. Nucleotide sequences and regulational analysis of genes involved in conversion of aniline to catechol in Pseudomonas putida UCC22(pTDN1). *J. Bacteriol.* **179**, 399–408 (1997).
58. Nogales, J., Palsson, B. & Thiele, I. A genome-scale metabolic reconstruction of Pseudomonas putida KT2440: iJN746 as a cell factory. *BMC Syst. Biol.* **2**, (2008).
59. Nikodinovic-Runic, J., Flanagan, M., Hume, A. R., Cagney, G. & O'Connor, K. E. Analysis of the Pseudomonas putida CA-3 proteome during growth on styrene under nitrogen-limiting and non-limiting conditions. *Microbiology* **155**, 3348–3361 (2009).
60. Ravi, K., García-Hidalgo, J., Gorwa-Grauslund, M. F. & Lidén, G. Conversion of lignin model compounds by Pseudomonas putida KT2440 and isolates from compost. *Appl. Microbiol. Biotechnol.* (2017). doi:10.1007/s00253-017-8211-y
61. Nishimura, M., Kawakami, S. & Otsuka, H. Molecular cloning and characterization of vanillin dehydrogenase from Streptomyces sp. NL15-2K. *BMC Microbiol.* **18**, 154 (2018).
62. Di Gioia, D. *et al.* Metabolic engineering of Pseudomonas fluorescens for the production of vanillin from ferulic acid. *J. Biotechnol.* **156**, 309–316 (2011).
63. Graf, N. & Altenbuchner, J. Genetic engineering of Pseudomonas putida KT2440 for rapid and high-yield production of vanillin from ferulic acid. *Appl.*

Microbiol. Biotechnol. **98**, 137–149 (2014).
64. Simon, O., Klaiber, I., Huber, A. & Pfannstiel, J. Comprehensive proteome analysis of the response of Pseudomonas putida KT2440 to the flavor compound vanillin. *J. Proteomics* **109**, 212–227 (2014).
65. Morabbi Heravi, K., Lange, J., Watzlawick, H., Kalinowski, J. & Altenbuchner, J. Transcriptional regulation of the vanillate utilization genes (vanABK Operon) of Corynebacterium glutamicum by VanR, a PadR-like repressor. *J. Bacteriol.* **197**, 959–72 (2015).
66. Andersen, J. B. *et al. New Unstable Variants of Green Fluorescent Protein for Studies of Transient Gene Expression in Bacteria*. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY* **64**, (1998).
67. Overmars, L., Van Hijum, S. A. F. T., Siezen, R. J. & Francke, C. CiVi: Circular genome visualization with unique features to analyze sequence elements. *Bioinformatics* **31**, 2867–2869 (2015).
68. Winsor, G. L. *et al.* Enhanced annotations and features for comparing thousands of Pseudomonasgenomes in the Pseudomonas genome database. *Nucleic Acids Res.* **44**, D646–D653 (2016).
69. Lönneborg, R. & Brzezinski, P. Factors that influence the response of the LysR type transcriptional regulators to aromatic compounds. doi:10.1186/1471-2091-12-49
70. Phylogeny.fr: 'One Click' Mode. Available at: http://www.phylogeny.fr/simple_phylogeny.cgi?workflow_id=669522b57c96e489e3b031209351f438&tab_index=6. (Accessed: 31st October 2019)
71. Nogales, J. *et al.* Unravelling the gallic acid degradation pathway in bacteria: the gal cluster from Pseudomonas putida. *Mol. Microbiol.* **79**, 359–374 (2011).
72. Nikel, P. I., Martínez-García, E. & De Lorenzo, V. Biotechnological domestication of pseudomonads using synthetic biology. *Nature Reviews Microbiology* **12**, 368–379 (2014).
73. Puchałka, J. *et al.* Genome-scale reconstruction and analysis of the Pseudomonas putida KT2440 metabolic network facilitates applications in biotechnology. *PLoS Comput. Biol.* **4**, (2008).
74. Biotecnología industrial de Pseudomonas putida y especies relacionadas | SpringerLink. Available at: https://link.springer.com/article/10.1007/s00253-012-3928-0. (Accessed: 23rd October 2019)
75. Kim, J. *et al.* Properties of alternative microbial hosts used in synthetic biology: towards the design of a modular chassis. *Essays Biochem.* **60**, 303–313 (2016).
76. Stephan, S. *et al.* Metabolic physiology of Pseudomonas putida for heterologous production of myxochromide. *Process Biochem.* **41**, 2146–2152 (2006).
77. Domínguez-Cuevas, P., González-Pastor, J. E., Marqués, S., Ramos, J. L. & De Lorenzo, V. Transcriptional tradeoff between metabolic and stress-response programs in Pseudomonas putida KT2440 cells exposed to toluene. *J. Biol. Chem.* **281**, 11981–11991 (2006).
78. Moreno, R., Martínez-Gomariz, M., Yuste, L., Gil, C. & Rojo, F. The Pseudomonas putida Crc global regulator controls the hierarchical assimilation of amino acids in a complete medium: Evidence from proteomic and genomic analyses. *Proteomics* **9**, 2910–2928 (2009).
79. Faizal, I. *et al.* Isolation and characterization of solvent-tolerant Pseudomonas putida strain T-57, and its application to biotransformation of toluene to cresol in a two-phase (organic-aqueous) system. in *Journal of Industrial Microbiology and Biotechnology* **32**, 542–547 (2005).
80. Albuquerque, M. G. E., Martino, V., Pollet, E., Avérous, L. & Reis, M. A. M. Mixed culture polyhydroxyalkanoate (PHA) production from volatile fatty acid (VFA)-rich streams: Effect of substrate composition and feeding regime on PHA productivity, composition and properties. *J. Biotechnol.* **151**, 66–76 (2011).
81. Salgado, H. *et al.* RegulonDB (version 4.0): transcriptional regulation, operon

organization and growth conditions in Escherichia coli K-12. *Nucleic Acids Res.* **32**, 303D – 306 (2004).
82. Monteiro, L. M. O. *et al.* Reverse engineering of an aspirin-responsive regulator in bacteria. *bioRxiv* 400788 (2018). doi:10.1101/400788
83. Arruda, L. M., Monteiro, L. M. O. & Silva-Rocha, R. The Chromobacterium violaceum ArsR arsenite repressor exerts tighter control on its cognate promoter than the Escherichia coli system. *Front. Microbiol.* (2016). doi:10.3389/fmicb.2016.01851
84. Ferlay, J. *et al.* Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *Eur. J. Cancer* **103**, 356–387 (2018).
85. Anderson, J. C., Clarke, E. J., Arkin, A. P. & Voigt, C. A. Environmentally Controlled Invasion of Cancer Cells by Engineered Bacteria. *J. Mol. Biol.* **355**, 619–627 (2006).
86. Advances in Computers: Computational Biology and Bioinformatics - Google Libros. Available at: https://books.google.com.br/books?hl=es&lr=&id=SbDcnpj4sSEC&oi=fnd&pg=PP2&dq=advances+in+computers+computational+biology+and+bioinformatics+tseng&ots=2z8dLC3S-s&sig=5M5FjM5ra3y2c9avxB45BlzyAC0#v=onepage&q=advances in computers computational biology and bioinformatics tseng&f=false. (Accessed: 31st October 2019)
87. Luscombe, N. M., Greenbaum, D. & Gerstein, M. What is bioinformatics? An introduction and overview. *Yearb. Med. Inform.* (2001). doi:10.1055/s-0038-1638103
88. Berman, H. M. *et al.* The protein data bank. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2002). doi:10.1107/S0907444902003451
89. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. GenBank. *Nucleic Acids Research* (2003). doi:10.1093/nar/gkg057
90. BLAST. Nucleotide BLAST: Search nucleotide databases using a nucleotide query. *Basic Local Alignment Search Tool* (2009).
91. Adetunji, M. O., Lamont, S. J. & Schmidt, C. J. TransAtlasDB: An integrated database connecting expression data, metadata and variants. *Database* (2018). doi:10.1093/database/bay014
92. Myers, C. J. *et al.* iBioSim: A tool for the analysis and design of genetic circuits. *Bioinformatics* (2009). doi:10.1093/bioinformatics/btp457
93. Dasika, M. S. & Maranas, C. D. OptCircuit: An optimization based method for computational design of genetic circuits. *BMC Syst. Biol.* **2**, (2008).
94. Bergmann, D. C. & Fleming, A. J. From molecule to model, from environment to evolution: an integrated view of growth and development. *Current Opinion in Plant Biology* (2010). doi:10.1016/j.pbi.2009.12.001
95. Smyth, M. S. & Martin, J. H. J. Review x Ray crystallography. *Mol Pathol* (2000).
96. Kalbitzer, H. R. Protein NMR Techniques. *Zeitschrift für Phys. Chemie* (1999). doi:10.1524/zpch.1999.208.part_1_2.293
97. Murata, K. & Wolf, M. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochimica et Biophysica Acta - General Subjects* **1862**, 324–334 (2018).
98. Dalton, J. A. R. & Jackson, R. M. An evaluation of automated homology modelling methods at low target-template sequence similarity. *Bioinformatics* (2007). doi:10.1093/bioinformatics/btm262
99. Sillitoe, I. *et al.* CATH: Comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res.* (2015). doi:10.1093/nar/gku947
100. Albert, L. L. & Lehninger, A. L. *Lehninger principles of biochemistry / David L. Nelson, Michael M. Cox*. *Principles of biochemistry* (2005).
101. Chakraborty, H. J., Gangopadhyay, A., Ganguli, S. & Datta, A. Protein structure prediction. in *Applying Big Data Analytics in Bioinformatics and Medicine* (2017).

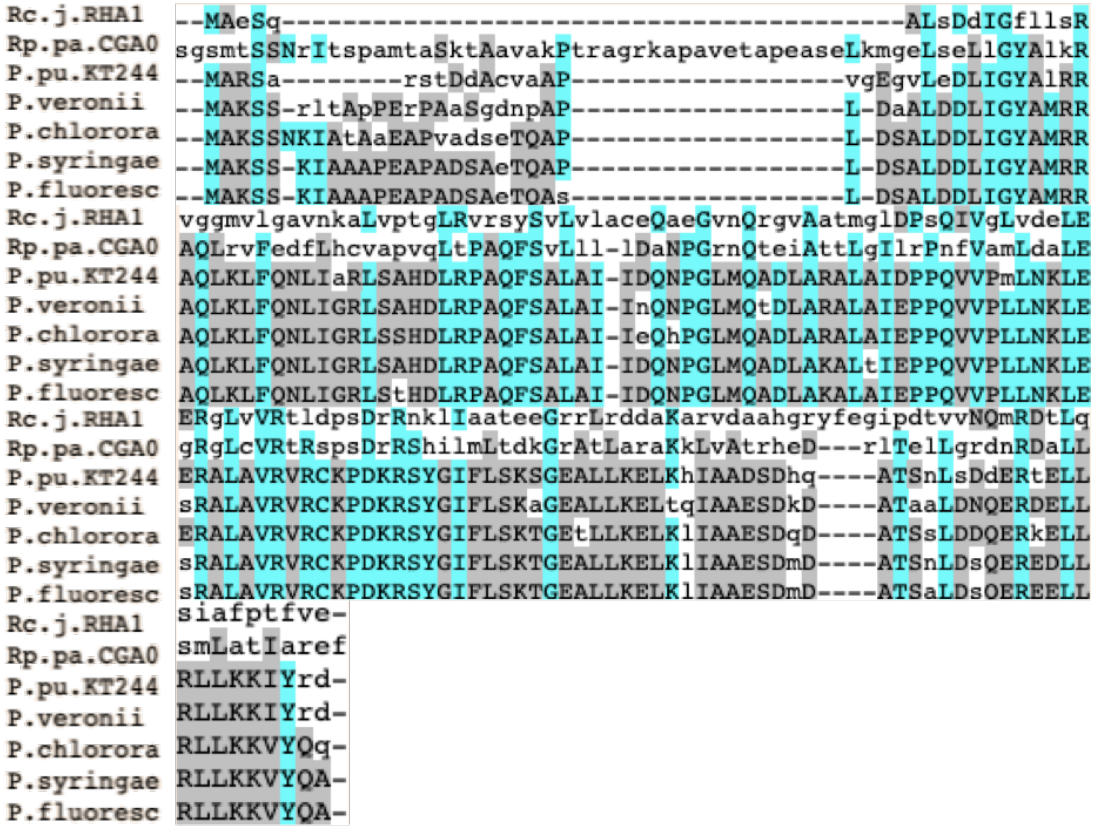
doi:10.4018/978-1-5225-2607-0.ch003
102. Krieger, E., Nabuurs, S. B. & Vriend, G. Homology Modeling. in *Structural Bioinformatics* (2005). doi:10.1002/0471721204.ch25
103. Forster, M. J. Molecular modelling in structural biology. *Micron* (2002). doi:10.1016/S0968-4328(01)00035-X
104. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* (2008). doi:10.1186/1471-2105-9-40
105. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* (2010). doi:10.1038/nprot.2010.5
106. Zhang, Y. Template-based modeling and free modeling by I-TASSER in CASP7. in *Proteins: Structure, Function and Genetics* (2007). doi:10.1002/prot.21702
107. Pliego Jr., J. R., Resende, S. M. & Almeida, W. B. De. Ab initio. *Theor. Chim. Acta* (1996). doi:10.1007/s002140050157
108. Xu, D. & Zhang, Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins Struct. Funct. Bioinforma.* (2012). doi:10.1002/prot.24065
109. Wu, S. & Zhang, Y. LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Res.* (2007). doi:10.1093/nar/gkm251
110. Pieper, U. *et al.* MODBASE, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res.* (2004).
111. Kelley, L. A. & Sternberg, M. J. E. Protein structure prediction on the web: A case study using the phyre server. *Nat. Protoc.* (2009). doi:10.1038/nprot.2009.2
112. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* (2015). doi:10.1038/nprot.2015.053
113. Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* (2004). doi:10.1093/nar/gkh468
114. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997).
115. Peng, J. & Xu, J. Raptorx: Exploiting structure information for protein alignment by statistical inference. *Proteins Struct. Funct. Bioinforma.* (2011). doi:10.1002/prot.23175
116. Taylor, R. D., Jewsbury, P. J. & Essex, J. W. A review of protein-small molecule docking methods. *J. Comput. Aided. Mol. Des.* (2002). doi:10.1023/A:1020155510718
117. Mohan, V., Gibbs, A., Cummings, M., Jaeger, E. & DesJarlais, R. Docking: Successes and Challenges. *Curr. Pharm. Des.* (2005). doi:10.2174/1381612053382106
118. Yuriev, E., Agostino, M. & Ramsland, P. A. Challenges and advances in computational docking: 2009 in review. *Journal of Molecular Recognition* (2011). doi:10.1002/jmr.1077
119. Trott, O. & Olson, A. J. Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* (2010). doi:10.1002/jcc.21334
120. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Res.* **46**, W363–W367 (2018).
121. Forli, W., Halliday, S., Belew, R. & Olson, A. AutoDock Version 4.2. *Citeseer* (2012).
122. Friesner, R. A. *et al.* Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* (2004). doi:10.1021/jm0306430
123. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D.

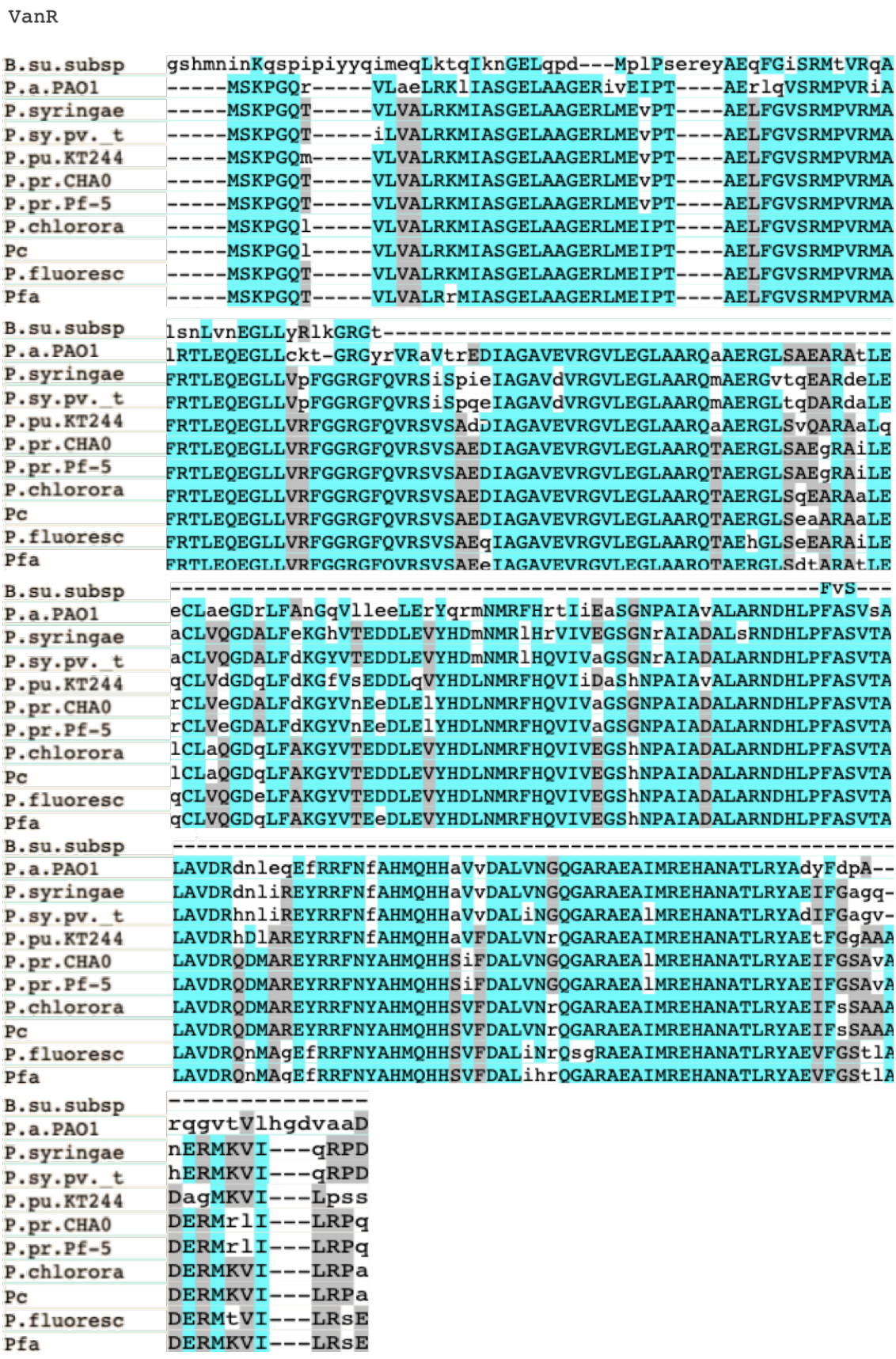
Improved protein-ligand docking using GOLD. *Proteins Struct. Funct. Genet.* (2003). doi:10.1002/prot.10465
124. McInnes, C. Virtual screening strategies in drug discovery. *Curr. Opin. Chem. Biol.* **11**, 494–502 (2007).
125. Alonso, H., Bliznyuk, A. A. & Gready, J. E. Combining docking and molecular dynamic simulations in drug design. *Medicinal Research Reviews* (2006). doi:10.1002/med.20067
126. Dias, R. & de Azevedo Jr., W. Molecular Docking Algorithms. *Curr. Drug Targets* (2008). doi:10.2174/138945008786949432
127. Pronobis, M. I., Deuitch, N. & Peifer, M. The Miraprep: A protocol that uses a Miniprep kit and provides Maxiprep yields. *PLoS One* (2016). doi:10.1371/journal.pone.0160509
128. Dabbish, L., Stuart, C., Tsay, J. & Herbsleb, J. Social coding in GitHub: Transparency and collaboration in an open software repository. in *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW* 1277–1286 (ACM Press, 2012). doi:10.1145/2145204.2145396
129. Pettersen, E. F. *et al.* UCSF Chimera?A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

VII. APPENDIX

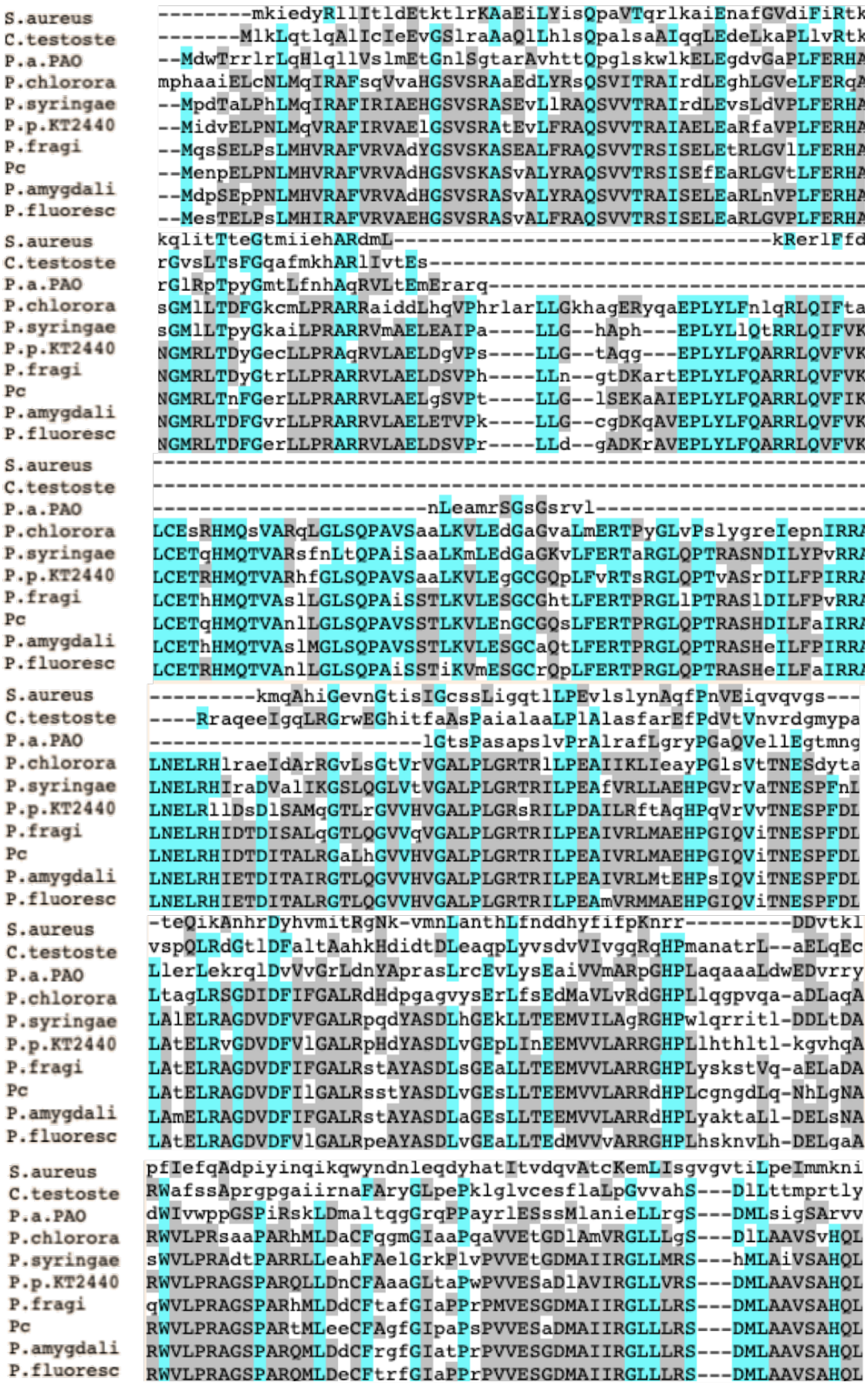
1. Apendix I. BLAST of VanR, GalR and PP_3359 proteins with their systematic neighbors and with bacteria with which it has a 3D structure.

PP_3359




VanR
B.su.subsp
P.a.PAO1
P.syringae
P.sy.pv._t
P.pu.KT244
P.pr.CHA0
P.pr.Pf-5
P.chlorora
Pc
P.fluoresc
Pfa

GalR


S.aureus
C.testoste
P.a.PAO
P.chlorora
P.syringae
P.p.KT2440
P.fragi
Pc
P.amygdali
P.fluoresc

2. Appendix II: R script used to analyze the kinetics and growth of the bacteria analyzed.

```
mediaydesvest <- function(filename,columnas) { #Numero de columnas que
el vittor analiza
  print(filename)
  lines=scan(filename, what="character",sep="\n")
  print(lines[1])
  divisiones=grep("Plate      Repeat      End time    Start temp. End
temp. BarCode",lines)
  print(divisiones)
  dat=read.table(textConnection(lines[1:(divisiones[1]-1)]),
header=TRUE,sep="\t")
  x=tail(dat$Well,n=1)
  letra=utf8ToInt(substr(x, start = 1, stop = 1))-utf8ToInt("A")+1

  numero=strtoi(substr(x, start = 2, stop = 3))

  posos=(letra-1)*columnas+numero

  repeticion=tail(dat$Repeat,n=1)


  dims = (length(dat)-4)/2

  result=array(dim=c(dims,repeticion,posos))
  resultName=c()

  for(j in 1:dims){
    pos = 4+j*2  #posicion de la columna

    for (i in 1:repeticion) {
      result[j,i,]=dat[((i-1)*posos+1):(i*posos),pos]
    }

    colName = strsplit(names(dat)[pos],"\\.")[[1]][1]
    resultName = c(resultName, colName)
  }
  print(result[1,,])

  background=array(dim=c(dims,repeticion,posos-3))
  for(j in 1:dims){

    for (i in 1:repeticion) {
      background[j,i,]=result[j,i,1:(posos-3)]-mean(result[j,i,(posos-
2):posos])
    }
  }
  print(background[1,,])

  ByOD=array(dim=c(dims-1,repeticion,posos-3))
  for(j in 2:dims){
    ByOD[j-1,,]=background[j,,]/background[1,,]
  }
  print(ByOD[1,,])

  media=array(dim=c(dims,repeticion,(posos-3)/3))
  desvest=array(dim=c(dims,repeticion,(posos-3)/3))
  #Loop para abs,GFP,mCherry
  for(k in 1:dims){
```

```
    for (i in 1:((posos-3)/3)){

      for (j in 1:repeticion){
        if(k==1){

          media[1,j,i] = mean(background[1,j,((i-1)*3+1):(i*3)])
          desvest[1,j,i] = sd(background[1,j,((i-1)*3+1):(i*3)])
        } else {

          media[k,j,i] = mean(ByOD[(k-1),j,((i-1)*3+1):(i*3)])
          desvest[k,j,i] = sd(ByOD[(k-1),j,((i-1)*3+1):(i*3)])
        }
      }
    }
  }
  return(list(media=media,desvest=desvest,tipos=resultName))
}

toggplot <- function(triplicata, concentraciones, compuestos,
lectura=0.5){

  df1 = as.data.frame.table(triplicata$media, responseName = "media")
  df2 = as.data.frame.table(triplicata$desvest, responseName = "sd")

  df1$sd = df2$sd
  #Vuelve numericos los valores de las repeticiones y de las muestras
  df1$Var2 = as.numeric(as.factor(df1$Var2))*lectura
  df1$Var3 = as.numeric(as.factor(df1$Var3))

  df1$comp = NA
  df1$ctrl = NA
  a = 1
  b = 1
  ctrl=FALSE
  for (i in seq(1,dim(triplicata$media)[3])){
    df1[df1$Var3==i,]$conc = concentraciones[a]
    df1[df1$Var3==i,]$comp = compuestos[b]
    df1[df1$Var3==i,]$ctrl = ctrl
    a=a+1
    if(a>length(concentraciones)){
      b=b+1
      a=1
    }
    if(b>length(compuestos)){
      b=1
      ctrl=TRUE
    }
  }

  df1$Var1 = as.numeric(as.factor(df1$Var1))
  for (i in 1:length(triplicata$tipos)){
    df1$Var1 = replace(df1$Var1, df1$Var1==i, triplicata$tipos[i])
  }
  return(df1)
}

plotConcentracion = function(plot, tipo, compound, ctrl){
  plotOD= plots[plots$Var1==tipo & plots$comp==compound &
plots$ctrl==ctrl,]

  if(tipo=="Absorbance"){
    titulo=paste("Bacterial growth in",compound)
    ytext="OD"
```

```
  }else{
    titulo=paste("Bacterial expression of", tipo, "in", compound)
    ytext=paste(tipo,"/OD")
  }

  if(ctrl){
    titulo=paste(titulo,"(control)")
  }

  p = ggplot(plotOD, aes(x=Var2, y=media, group=conc, color=conc)) +
    geom_errorbar(aes(ymin=media-sd, ymax=media+sd), width=.1) +
    geom_line() + geom_point()+
    scale_color_brewer(palette="Spectral")+
    theme(panel.grid.major = element_blank(), panel.grid.minor =
element_blank(),
          panel.background = element_blank(), axis.line =
element_line(colour = "black"))
  p+labs(title=titulo, x="Hours", y=ytext, color="Concentrations")+ #
for the main title, axis labels and legend titles
    theme(plot.title = element_text(hjust = 0.5))+
    theme(legend.position=c(0.2,0.8))
}


filename1=file.choose()
#calcula media, desvio, quita el ruido de fondo del blanco de todos los
posos en todas las lecturas
triplicata = mediaydesvest(filename1,12)

concentraciones = c("0mM","0.001mM","0.01mM","0.1mM")
compuestos = c("Ferulic Acid","Vanillic Acid", "Vanillin")
plots = toggplot(triplicata, concentraciones, compuestos)


plotConcentracion(plots, tipo = "Absorbance", compound = "Vanillin",
ctrl = FALSE)

for(ctrl in c(FALSE,TRUE)){
  for(t in 1:length(triplicata$tipos)){
    for (c in 1:length(compuestos)) {
      print(plotConcentracion(plots,     tipo     =triplicata$tipos[t],
compound = compuestos[c], ctrl))
      #print(paste(triplicata$tipos[t], compuestos[c], ctrl))
    }
  }
}

folder = ""
for(k in 1:dims){
  dirout = paste0(folder,resultName[k],".txt")
  write.table(triplicata[k,,],file=dirout, sep="\t",row.names=FALSE)
```