Protein engineering and synthetic pathways in *Escherichia coli* for effective production of 5-hydroxytryptophan and serotonin

Vom Promotionsausschuss der
Technischen Universität Hamburg

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

M.Sc Jose Aníbal Mora Villalobos

aus

Alajuela, Costa Rica

2018

Gutachter:

Prof. Dr. An-Ping Zeng

Prof. Dr. Andreas Liese

Prüfungsausschussvorsitzender:

Prof. Dr. Rainmund Horn

Tag der mündlichen Prüfung:

23. Februar 2018

Publications

Mora-Villalobos, J.A. and A.P. Zeng. 2017. Synthetic pathways and processes for the effective production of 5-hydroxytryptophan and serotonin from glucose in *Escherichia coli*. *J. Biol. Eng.* Accepted

Mora-Villalobos, J.A. and A.P. Zeng. 2017. Protein and pathway engineering for the biosynthesis of 5-hydroxytryptophan in *Escherichia coli*. *Eng. Life Sci.* 17, 892-899.

Acknowledgment

I am very grateful to Prof. Dr. An-Ping Zeng for the opportunity he gave me at the Institute of Bioprocess and Biosystems Engineering and for his supervision. The experience that I have gained during the past four years is extremely valuable for my career. I am very thankful for his support, guidance, and trust.

I would like to express my gratitude to Prof. Dr. Andeas Liese for having an open office and being willing to discuss new ideas and projects. I also thank him for being a member of my thesis committee; I would also like to extend my gratitude to the Chair of the committee, Prof. Raimund Horn.

I would like to thank Dr. Uwe Jandt and Dr. Samira Hezaveh for their support, advice, and discussions during this years. I would also like to express my gratitude to other IBB members that were always willing to share their scientific experience with me: Dr. Sugima Rappert, Dr. Chengwei Ma, Dr. Jana Tatur, Dr. Matthias Wurm, Dr. Wael Sabra and Dr. Wei Wang. I would also like to thank Mrs. Cornelia Hoffmann for her support and help with administrative issues. I want to thank Jan Sens, Olaf Schmidt, Jan Bomnüter, Anna Gorte and Birgit Stacks for their technical support in many ways. I am sincerely grateful to Prof. Dr.-Ing. Ralf Pörtner and all the members of his group.

I would like to thank the Deutscher Akademischer Austausch Dienst (DAAD), the Ministry of Science, Technology, and Telecommunications of Costa Rica (MICITT) and the National Center for Biotechnological Innovations (CENIBiot) for the support and funding. I would like to thank Dr. Marta Valdez, Dr. Sergio Carballo, Dr. Randall Loaiza, and Francisco Aguilar for extending me institutional support from the CENIBiot.

I am really grateful to all the friends and colleagues for the very nice moments we spend together. I would like to thank the generation of Ph.D. candidates that received me: Enrico, Grischa, Christin, Julian, Johannes, Doinita, Jin, Lin, Libang, Lifu, and Feng. I also extend my gratitude to those that are still at the institute –and I wish them my best–: Sibel, Ludwig, Yujun, Rebekka, Sukanya, Eduardo, Tyll, Minliang, Cornelius, and Philipp.

I am very grateful to my family for their love and support, especially María Clara for her never-failing support, immense love and constant motivation.

Hamburg 2017

Abstract

Metabolic engineering has improved the production of tryptophan in *Escherichia coli* during the last two decades, opening itself a plethora of opportunities for the production of tryptophan derivatives. 5-Hydroxytryptophan (5HTP) and serotonin are two important derivatives not only important for their pharmaceutical values but also because they can serve as precursors of other molecules which include sleep cycle regulator, anti-migraine medications, sedatives, anticonvulsants, antitumors, antimicrobials, antivirals, among others.

To the present, 5HTP is mainly obtained by extraction from the plant *Griffonia simplicifolia* and serotonin is produced by chemical synthesis. In both cases, the processes involve the use of organic solvents and energy-intensive conditions during the procedure. Moreover, in the case of serotonin, its chemical synthesis starts from a complex molecule (5-benzyloxyindole). Therefore, biotechnological processes for the production of these compounds from a simple renewable organic source are desired.

This dissertation presents work done to extend tryptophan metabolism for the production of 5HTP and serotonin. For this purpose, serotonin production via tryptamine and via 5HTP were compared and analyzed. In both cases, the hydroxylation step appeared to be the bottleneck due to the low activity of the enzymes when expressed in *E. coli* and the requirement of a cofactor, plus its regeneration pathway. The serotonin production pathway via 5HTP was chosen, and for this purpose, an aromatic amino acid hydroxylase from *Cupriavidus taiwanensis* (*Ct*AAAH) was selected using an *in silico* structure-based approach. Several substrate-determining residues were predicted and selected using sequence, phylogenetic and functional divergence analyses. Whole cells analysis with the wild-type and variants were done to study the shift of the enzyme preference from phenylalanine to tryptophan. All the variants increased the tryptophan hydroxylation activity in detriment to phenylalanine. The best performer, *Ct*AAAH-W192F, was transformed into a strain that had the *tryptophanase A* gene disrupted and carried a human tetrahydrobiopterin (BH_4) regeneration pathway. The resulting strain was capable of synthesizing 2.5 mM 5HTP after 24 hours in medium supplied with tryptophan.

After this first rational design round, a second semi-rational approach was selected to improve the efficiency of the enzyme. A tryptophan intracellular

concentration sensor was used to screen two independent libraries, and the variants found in the best performer of each library were combined to create *Ct*AAAH-LC. This double mutant showed higher activity and reaction speed than its predecessor. *Ct*AAAH-LC was transformed into a tryptophan producer strain (S028), which was modified by the addition of a pterin (cofactor consumed during hydroxylation) regeneration pathway. In this case, 5HTP was synthesis from glucose.

Tryptophan decarboxylase (TDC) was incorporated in the 5HTP producer strain to produce serotonin from glucose. However, the serotonin production was low and undesired side reactions were identified. To circumvent this problem, a two-step process was constructed in which the 5HTP production and the serotonin conversion are separated.

In this work, results of the highest concentration of microbial 5HTP production reported so far are presented. Afterwards, 5HTP was decarboxylated to produce serotonin in a second fermentation. This is the first report for microbial serotonin production from glucose. The process can be further optimized by more efficiently streamilining the hydroxylation and decarboxylation reactions in one strain. TDC selectivity could be engineered to shift the preference toward 5HTP in detriment of tryptophan. To this end, the development of a novel biosensor sensitive to 5HTP is critical for screening of variants.

Zusammenfassung

Durch Metabolic Engineering wurde in den letzten zwei Jahrzehnten die Produktion von Tryptophan und Tryptophan-Derivaten, wie zum Beispiel 5-Hydroxytryptophan (5HTP) und Serotonin, in *Escherichia coli* gesteigert. Diese sind nicht nur aufgrund ihres pharmazeutischen Werts wichtig, sondern auch, weil sie als Vorstufen anderer Pharmazeutika dienen können, wie zum Beispiel Schlafzyklusregulatoren, Anti-Migräne Medikamente, Sedativa, Antikonvulsiva, Anti-Tumor Medikamente, Antimikrobielle Wirkstoffe, und Antivirale Wirkstoffe. Gegenwärtig wird 5HTP hauptsächlich durch Extraktion aus der Pflanze *Griffonia simplicifolia* gewonnen und Serotonin wird durch chemische Synthese hergestellt. In beiden Fällen beinhalten die Verfahren die Verwendung von organischen Lösungsmitteln und energieintensiven Reaktionsbedingungen. Im Falle von Serotonin beginnt seine chemische Synthese zusätzlich noch mit einem komplexen Molekül (5-Benzyloxyindol). Daher ist ein biotechnologisches Verfahren ausgehend von erneuerbaren organischen Stoffen zur Herstellung dieser Verbindungen wünschenwert.

In dieser Arbeit wurde die Erweiterungdes Tryptophan-Metabolismus zur Produktion von 5HTP und Serotonin angestrebt. Zu diesem Zweck wurde die biologische Serotoninproduktion durch Tryptamin oder 5HTP verglichen und analysiert. In beiden Fällen war der Hydroxylierungsschritt reaktionslimitierend. Dies lässt sich auf die geringe Aktivität in *E. coli* exprimierten Enzyme zurückführen, sowie der zusätzlichen Notwendigkeit eines Cofaktors und der Abwesenheit eines Regenerationswegs des Cofaktors. Der Serotonin-Produktionsweg via 5HTP wurde ausgewählt und weiter optimiert. Zu diesem Zweck wurde eine aromatische Aminosäure-Hydroxylase aus *Cupriavidus taiwanensis* unter Verwendung eines strukturierten *in silico* Ansatzes ausgewählt. Mehrere substratbestimmende Aminosäuren wurden unter Verwendung von sequentiellen, phylogenetischen und funktionellen Divergenzanalysen vorhergesagt und ausgewählt. Ganzzellanalysen wurden mit dem Wildtyp und seinen Varianten durchgeführt, um die Verschiebung der Enzympräferenz von Phenylalanin zu Tryptophan zu untersuchen. Alle Varianten erhöhten die Hydroxylierungsaktivität von Tryptophan zum Nachteil von Phenylalanin. Der beste Performer, *Ct*AAAH-W192F, wurde in einen *Tryptophanase* A Gen defizienten Stamm eingebracht, welcher aber den menschlichen Tetrahydrobiopterin (BH_4) Regenerationsweg beinhaltet. Der resul-

tierende Stamm war in der Lage, 2,5 mM 5HTP nach 24 Stunden in Medium mit Tryptophan zu synthetisieren.

Nach diesem ersten Erfolg mit rationalem Design wurde ein zweiter semirationaler Ansatz ausgewählt, um die Effizienz des Enzyms weiter zu verbessern. Ein intrazellulärer konzentrationsabhängiger Tryptophan-Sensor wurde verwendet, um zwei unabhängige Bibliotheken zu prüfen. Die besten Performer aus jeder Bibliothek wurden kombiniert, dabei ein Doppelmutant *Ct*AAAH-LC entstand. Dieser Doppelmutant zeigte eine höhere Aktivität und Reaktionsgeschwindigkeit als ihr Vorgänger. *Ct*AAAH-LC wurde in einen Tryptophan produzierenden *E. coli* Stamm(S028) eingebracht, der Pterin beinhaltet, welcher als Cofaktor während der Hydroxylierung verbraucht wird. Dadurch wurde 5HTP erfolgreich aus Glukose synthetisiert.

Desweiteren wurde Tryptophan-Decarboxylase (TDC) in den 5HTP-Produzentenstamm eingebaut um Serotonin aus Glucose zu erzeugen. Allerdings war die Serotoninproduktion niedrig und es wurden unerwünschte Nebenreaktionen identifiziert. Ein zweistufiges System wurde entwickelt, um dieses Problem zu überwinden. Hierfür wurde die 5HTP-Produktion von der Serotonin-Synthese entkoppelt.

In dieser Arbeit wurde die bisher höchste publizierte Konzentration von 5HTP gezeigt. Zu diesem Zweck wurde Protein-Engineering an *Ct*AAAH durchgeführt und einem synthetischer Stoffwechelweg in *E. coli* implementiert. Anschließend wurde 5HTP decarboxyliert, um Serotonin fermentativ zu produzieren. Dies ist der erste Bericht der Serotoninproduktion aus Glucose. Das Verfahren kann durch Kombination der Hydroxylierungs- und Decarboxylierungsreaktion in einem Stamm weiter optimiert werden. Die TDC-Selektivität kann verschoben werden, um die Präferenz gegenüber 5HTP auf Kosten von Tryptophan zu verschieben. In diesem Fall ist die Entwicklung eines neuen 5HTP-detektierenden Biosensors entscheidend für das Screening von effizienten Produktionsstämmen.

Contents

Abbreviations

5HTP	5-Hydroxytryptophan
AAAH	Aromatic amino acid hydroxylases
CASP	Critical Assessment of techniques for protein Structure Prediction
CrTDC	Tryptophan decarboxylase from *Catharanthus roseus*
*Ct*AAAH	AAAH from *Cupriavidus taiwanensis*
DAHP	3-deoxy-D-arabino-heptulosonate-7-phosphate
DHPR	Human dihydropteridine reductase
DNA	Deoxyribonucleic acid
E4P	Erythrose 4-phosphate
EtOH	Ethanol
FD-I	Functional diverse analysis type - I
FD-II	Functional diverse analysis type - II
FRT	Flippase recognition target
GFP	Green fluorescent protein
GST	Glutathione S transferase
IAA	Indole acetic acid
IPA	Indole pyruvic acid
IPTG	Isopropyl-ß-D-thiogalactopyranoside
ISM	Iterative Site Mutagenesis
LB	Lysogenic Broth
ML	Maximum likelihood
MSA	Multiple sequence analysis
NJ	Neighbor joining
PAH	Phenylalanine hydroxylase
PCD	Human pterin-4 alpha-carbinolamine dehydratase
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEP	Phosphoenolpyruvate
RF2	Release Factor 2
RT	Room temperature
SOB	Super optimal broth
SOC	SOB with catabolite repression
T5H	Tryptamine 5-hydroxylase
TCA	Trichloro acetic acid
TDC	Tryptophan decarboxylase

TH	Tyrosine hydroxylase
tnaA	Tryptophanase A
TPH	Tryptophan hydroxylase
trpR	Tryptophan repressor

List of Figures

List of Tables

1 Introduction

Tryptophan is an essential amino acid with medical, industrial and pharmaceutical importance. Potential therapeutic agents have stimulated the interest in the design and synthesis of tryptophan-related structures, which could have direct health benefits or may work as key biosynthetic precursors for other molecules. 5-Hydroxytryptophan (5HTP) and serotonin are two important tryptophan derivatives (Fig. 1).

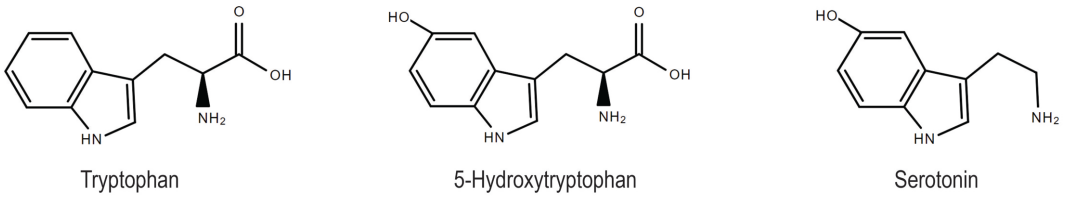

Figure 1: Tryptophan, 5-hydroxytryptophan and serotonin structures

5HTP is a natural non-canonical amino acid, and it is the precursor of the neurotransmitter serotonin. Over the last 30 years, it has been used to treat a wide variety of conditions related to serotonin imbalance, such as depression, insomnia, fibromyalgia, chronic headaches and binge eating associated with obesity. Unlike serotonin, 5-HTP is well absorbed from oral doses and can easily cross the blood-brain barrier (Birdsall, 1998). Regarding world volume demand, 5HTP stood at 136.4 tons in 2012, which corresponds to a value of 31.7 million US dollars and is anticipated to reach over 50 million US dollars by 2019 (www.transparencymarketresearch.com). For many years there have been chemical synthesis methods reported (Frangatos and Chubb, 1959; Gong et al., 2013). Still, its production is not economically feasible in large scale. A key challenge for its manufacture stands in the fact that the production depends on the extraction from seeds of the African plant *Griffonia simplicifolia*, and therefore the supply of the raw material is linked to seasonal and regional conditions.

Serotonin is naturally present in animals and plants. It is produced by the decarboxylation of 5HTP or the hydroxylation of tryptamine, respectively. In both cases, it is implicated in fundamental physiological roles (Kang et

al., 2007; Turner et al., 2006). Analogs that resemble serotonin structure act on a wide range of therapeutic targets, such as phosphodiesterase, 5-hydroxytryptamine 5-hydroxytryptamine receptors, cannabinoid receptors and HMG-CoA reductases. Many of these targets contain a binding pocket that recognizes the indole scaffold (de Sá Alves et al., 2009).

The indole aromatic heterocyclic backbone, present in 5HTP and serotonin, is a valuable molecular framework that provides a plethora of opportunities for medical chemistry and drug discovery. It is not unusual that the structure of drugs, or their precursors, resemble bioactive molecules with the elimination, addition or modification of functional groups. Therefore, 5HTP and serotonin, could serve as building blocks for active ingredients that may be used as pharmaceuticals for different conditions, such as sleep cycle regulator (Roth et al., 2015), anti-migraine medication (Cameron et al., 2015), sedatives, anticonvulsants, antitumors, antimicrobials (Cao et al., 2007) and antivirals (Chadha and Silakari, 2017) (Fig. 2). Furthermore, 5HTP and serotonin *per se* possess free radical scavenging and antioxidant activity (Christen et al., 1990; Gülçin, 2008). Biotechnology itself presents as a tempting promise for the production of these molecules using microorganisms with high yields, in short time with low costs (Chen and Zeng, 2013; Jullesson et al., 2015; Picataggio, 2009), especially because microorganisms have a relatively simpler genetic background and better reprogrammable metabolic network than the native producers.

To date, *Escherichia coli* remains as the dominant industrial microorganism producer of many complex compounds, and as the prime prokaryotic genetic model (Burk and Van Dien, 2016; Chen et al., 2013; Tang and Zhao, 2009). Different approaches have been adopted during the last four decades to channel the carbon flux towards the production of aromatic amino acids. The first attempt to rationally improve a strain for the production of tryptophan in *E. coli* was performed by Tribe and Pittard (1979). They increased tryptophan production by amplification of the *trp* operon with a deregulated *trpE* gene. In their work, the strain NST100 was able to produce 1 g/L. Since then, many groups have created different strains. To the date, the strain S028 developed by Chen and Zeng (2017) is one of the most efficient rationally designed strain reported. S028 can produce 34-40 g/L of tryptophan with a yield of 0.15 g/g and a productivity of 0.6 g/L/h. The construction of these producers has driven the deep research and engineering of its pathways,

a condition that entails the emergence of new and effective strategies for the production of tryptophan derivatives.

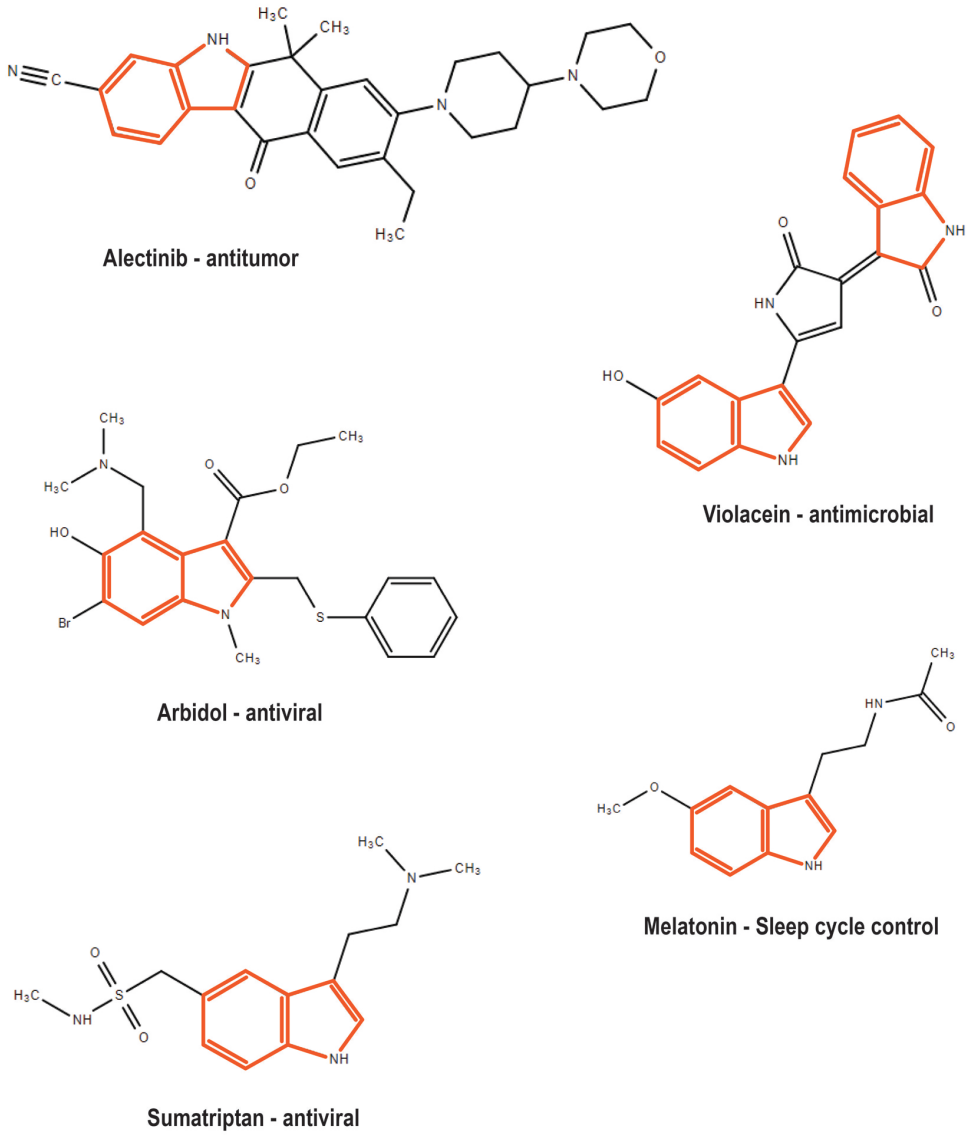

Figure 2: Tryptophan derivative molecules relevant for the pharmaceutical industry. The indole scaffold in each molecule is highlighted in orange color.

E. coli has been engineered to produce 5HTP (Hara and Kino, 2013; Lin et al., 2014). Unlike natural amino acids, the biosynthesis of non-canonical amino acids from simple sugars often requires the incorporation of artificial or synthetic metabolic pathways to expand the cell´s capabilities. The optimization of these synthetic metabolic pathways usually requires extensive protein

engineering (Arnold, 2001; Chen and Zeng, 2016). Mammalian tryptophan 5-hydroxylase is capable of synthesizing 5HTP via tryptophan hydroxylation, but it has low activity and poor stability when expressed in prokaryotes (Wang et al., 2002). Therefore, the use of native enzyme is not suitable for production purposes. Often, metabolic engineering uses enzymes with a non-native substrate to perform transformations within the novel synthetic pathways. Pioneering studies have used bacterial aromatic amino acid hydroxylases (AAAH), which has a strong preference toward phenylalanine, and shifted the substrate preference for tryptophan, based on sequence comparison and modification (Kino et al., 2009; Lin et al., 2014).

In the case of serotonin, one extra enzyme is required to decarboxylate 5HTP and convert it into serotonin. Noé et al. (1984) reported that both, tryptophan and 5HTP, are natural substrates of tryptophan decarboxylase (TDC) from *Catharanthus roseus.* This same enzyme has been used for the production of serotonin in *E. coli.* However 5HTP was supplied as a substrate, and the yield was quite low (35 mg/mL) (Park et al., 2008).

A combination of rational design and directed evolution approaches is normally used to tailor enzymes. Engineered enzymes can be incorporated into the design synthetic pathways for the production of 5HTP and further conversion to serotonin. The combination of crystallographic studies and the continuous progress in molecular modeling methods have opened new perspectives for structure-based protein engineering (rational design). Furthermore, the comparison of the sequences and functions of related proteins can also be used to identify important residues for substrate specificity (Chen et al., 2010; Chen and Zeng, 2013). The development of these areas allows, for example, the design and optimization of the binding pocket conformation guided by the features of the ligands. In the same way, point mutations could also be rationally designed, and the preference of different substrates could be evaluated *in silico*, reducing the number of mutants significantly to be tested in the laboratory. On the other hand, directed evolution by saturation mutagenesis (SM) has proven to be a useful method for protein engineering in a variety of different applications.

Currently, a mixture of both methods, rational design and directed evolution, is the most common approach for protein engineering. The access of tertiary structures, the development of *in silico* screening and prediction methods are

helpful tools in this endeavor. Protein evolution methods often suffer from bottlenecks in the design of the library and screening process due to the high number of theoretical combinations of mutants. Iterative saturation mutagenesis (ISM) is an option to overcome some of the problems, especially when it is coupled with smart-libraries that integrate structural and evolutionary data (Acevedo-Rocha et al., 2015; Reetz and Carballeira, 2007a). The combination of both reduces drastically the screening efforts required to select novel enzymes with the desired activity (Arnold, 2001; Kille et al., 2013; Parra et al., 2013).

1.1 Objectives

The main goal of this thesis was to develop biotechnological routes for the production of 5HTP and serotonin using *E. coli* as cells factories and applying a synthetic biology approach.

For this purpose, a specific route from tryptophan to 5HTP and serotonin was designed considering possible bottlenecks such as the activity and stability of the involved enzymes and the cofactors consumed during the production of the intermediate molecules. Alternative routes for serotonin production were also evaluated and discarded (Section 4). Sequence, phylogenetic and functional analyses were performed to identify hotspots in a selected hydroxylase enzyme. Structural and docking analysis restricted the number of candidates, and these were generated and characterized *in vivo* in a strain harboring a cofactor regeneration pathway (Section 5). After this first round of protein engineering, the hydroxylase activity was improved using a semi-rational approach. Two smart libraries were designed, and the colonies were selected using an intracellular tryptophan biosensor. The best performers were combined and introduced into a tryptophan producer strain to create a 5HTP producer (Section 6). Finally, the decarboxylase enzyme was incorporated into the strain. Single strain and two-steps production strategies were compared for the production of serotonin (Section 7).

2 Theoretical and Technological Background

This section is a short review of the current state-of-the-art of technologies and methods available in system biology and metabolic engineering for the development of novel strains. Strategies used for the creation of rationally designed strains for the production of tryptophan derivatives are described, as well as the status of the 5HTP and serotonin biosynthesis in *E. coli*.

2.1 Phylogenetic analysis

2.1.1 Evolutionary basis of sequence alignment

In nature, evolution acts conservatively, i.e., it does not develop a new machinery for every life form, but continuously changes and adapts from previously existing forms. At a molecular level, the same principle could be applied, different protein structures or functions are preceded by previous existing ones with slight alterations. These changes mean mutations, which are changes in the DNA and protein sequences and could bring alone changes in the protein function, activity, preference, among others. As mutations accumulate, sequences derived from the original template, yet the final product will still have enough similarity to allow the identification of a common ancestor. Evolutionary changes in a sequence do not always have to be large, small changes in a certain crucial section of a sequence can have profound functional consequences (Choudhuri, 2014).

As expected, the comparison of multiple sequences based on alignments is a fundamental analysis of most bioinformatics studies. It is the first step towards understanding the evolutionary relationship and the pattern of divergence between sequences. The relation between nucleotides or amino acids also helps to predict the potential function of unknown sequences, since normally sequences are related to proteins within the same family.

2.1.2 Multiple sequence alignment

The simultaneous alignment of nucleotide or amino acid sequences is one of the essential tasks in applied bioinformatics. It is a useful tool for the identification of regions with structural, functional and evolutionary value. Multiple

Sequence Alignments (MSA) are an essential prerequisite to many further protein analysis such as identification of conserved and variables sites within a family, phylogenetic reconstructions, or homology modeling (Davidson, 2006; Hogeweg and Hesper, 1984; Phillips et al., 2000; Procter et al., 2010).

Most of the MSAs are carried out using a progressive approach. In this method, the alignment of sequences and the construction of phylogenetic trees cannot be treated separately. An alignment always refers to a phylogenetic tree, and the construction of the phylogenetic trees requires an alignment. Hogeweg and Hesper (1984) proposed this integrated method that generates both. It uses a putative tree to align the sequences and the alignment obtained is used to adjust the tree. The sensitivity of the alignment is greatly improved without sacrificing, neither the speed nor efficiency of the analysis, which makes this approach so practical.

Many different programs use the progressive approach to carry out MSA. The most popular ones include ClustalW (Thompson et al., 1994), MAFFT (Katoh et al., 2002), MUSCLE (Edgar, 2004), T-coffee (Notredame et al., 2000) and ProbCons (Do et al., 2005). They all have different consideration regarding the biological background of the sequences. Therefore the algorithms give different results with variations in the biological accuracy, execution time and memory usage. Among of these, ClustalW is widely used for phylogenetic tree construction, and it is provided by most web portals. ClustalW bases its algorithm in the fact that similar sequences are homologous. The program first generates a phylogeny that represents the relationships between the sequences. Then pairwise alignments are carried out, beginning with the most similar sequences. Once all the pairwise alignments scores relative to all the other sequences have been calculated, they are used to group sequences. Finally, the groups are presented as multiple alignments (Thompson et al., 1994).

2.1.3 Functional diversification of protein families

Proteins are proficient, accurate, and specific. These characteristics are correlated with lack of versatility. However, proteins also exhibit a marked ability to acquire new functions and structures. The evidence for evolutionary adaptability of proteins is compelling, manifest in the vast range of proteins that have presumably derived from a common ancestor (Tokuriki and Tawfik,

2009). Comparative analysis of enzymes, encoded in a variety of prokaryotic and eukaryotic genes, reveals that evolution entails divergence in the sequence and structure of the proteins which leads to new functions of the enzymes. As a consequence, many structurally similar enzymes can act on distinct substrates and catalyze similar biochemical reactions (Galperin and Koonin, 2012).

The previously mentioned basic principle of evolution supports the historical grouping of protein families based on sequence similarity. Families are further combined into superfamilies based on similar catalytic activities, sequence motifs, and other conserved features (Todd et al., 2001). Moreover, even if enzymes have dramatically different enzymatic activities, proteins within the same (super-) family can be confidently inferred to have evolved from a common ancestor (Glasner et al., 2006). Functional promiscuity seems to be the starting point for the emergence of new features. Mutations can promote alternative reactions, change the preferences of substrates, shift the equilibrium of the reaction, and so on. Therefore, mutations create the raw material on which selection acts (Tokuriki and Tawfik, 2009).

Chemical aspects of the catalysis constrain the evolution of the enzymes, and generally, only a subset of the catalytic residues is conserved. Mechanistic diversity arises from the differential placement of other catalytic residues, and substrate diversity often is related to variation in loops and accessory binding domains (Glasner et al., 2006). Molecular evolution also referred as sequence divergence, is a function that relates the rate of neutral, deleterious and advantageous mutations, their selection coefficients, and the effective population size. The neutral theory states that functionally important sites will remain constant over time with high probability, whereas neutral sites will evolve at a much faster speed determined by the mutation rate (Kimura, 1968).

The analysis and comparison of individual sites could give hints about the relative importance of different residues. Furthermore, the combination of these results with structure and mechanistic information promotes the elucidation of functions, substrate preference, activity, stability, among others. Therefore, it is critical to combine different kinds of approaches to analyze the relationship between protein sequence and function.

2.2 Protein structure-function relationship and modeling

The relationship between sequence, structure and biochemical (biological) activity are tightly entangled. The bedrock behind this idea was established by Anfinsen (1973) when he demonstrated that bovine pancreatic ribonuclease activity could be recovered after regaining its native tertiary structure. He established, this way, a clear line between protein structure and function. From an evolutionary point of view, this also has its consequences: a strong structural similarity is an indication of divergence, and contrary to the intuitive idea, the conservation of the structure is not entirely due to the maintenance of the function (Sadowski and Jones, 2009). This contrasting idea opens a gap in Anfinsen‘s dogma where protein function, therefore, cannot be straightforward predicted from the structure. For example, there are 27 different homologous superfamilies that adopt the TIM barrel fold (eight alternated -helices and -strands), covering over 60 different EC classifications (Greene et al., 2007), meaning that the simple identification of this fold in a novel structure would do little in the prediction of its function (Lee et al., 2007).

Nevertheless, structural data can be used to detect and compare proteins with similar functions whose sequences have diverged beyond a level in which amino acid similarity cannot detect filial relations. With 3D structure information, it would be possible to predict protein function at a high level (e.g., “hydroxylases”). Furthermore, if the structure analyses are combined with additional information such as the identification of which part(s) of the proteins are more relevant for its function and the comparison with the range of possible functions that the protein adopts, then refinements in protein function can be done.

At last, protein structure is required for many functional prediction analyses. Unfortunately, the determination of protein structure is both experimentally expensive and time-consuming, and in most of the cases, this information is not available. Although protein structures deposited in public databases are increasing at an accelerated rate, at the same time the number of known protein sequences is growing even faster. Computational protein 3D structure predictions provide a potential solution to bridge this sequence-structure gap. These methods are not as accurate as experimental methods, but they

often offer molecular insight from the predicted structure which is useful for the generation of the hypothesis and to complement the experimental work. Therefore, if experimentally determined structures are unavailable, predicted structures might serve as a starting point for functional studies (Khor et al., 2015).

The biological usefulness of the predicted structure relies on the accuracy of the new protein model. Structural biology divides available algorithms into four different groups. The first two methods aim to predict the new structures only through simulation processes. The difference between them is that the first one uses general rules from protein databases to create and compare small fragments (threading method) (discussed in detail in Section 2.2.1). The second method is founded on thermodynamics principles and is based on the idea that the global minimum of free energy of a model should correspond to the native structure of the protein (*ab initio* method). The third algorithm uses the target sequence of amino acids and aligns them against the sequence of a known structure with similar sequence (homology modeling method). Finally, the last method creates fragments by the threading method and then compares the fragments against a known structural model (combined method) (Dorn et al., 2014).

2.2.1 Threading assembly method

Threading method of 3D structures based on amino acid sequences. I-TASSER (Roy et al., 2010) and Rosetta (Bonneau et al., 2001) are two commonly used methods based on this approach. These are the most successful/accurate structure prediction software to date, according to the "Critical Assessment of techniques for protein Structure Prediction" (CASP) experiments, which is a bi-annual evaluation of the state-of-the-art within the field of computational biology (Ovchinnikov et al., 2016; Zhang et al., 2015).

Threading methods fragment the target sequence and create small fold subunits. The fold fragments are compared against a known structure to predict the structure of the target. This sequence-structure comparison is based on the principle that when a new fold is discovered this is composed of common structural motifs (or fragments) with known structure (Tramontano and Büssow, 2006). The fold fragments can then be used to construct the 3D model through an assembly procedure with the purpose of finding the structure with

the lowest energy potential. The critical step of threading is to identify correct template(s) with similar folds to the target protein and make correct sequence-structure alignment (Dorn et al., 2014) (Fig. 3).

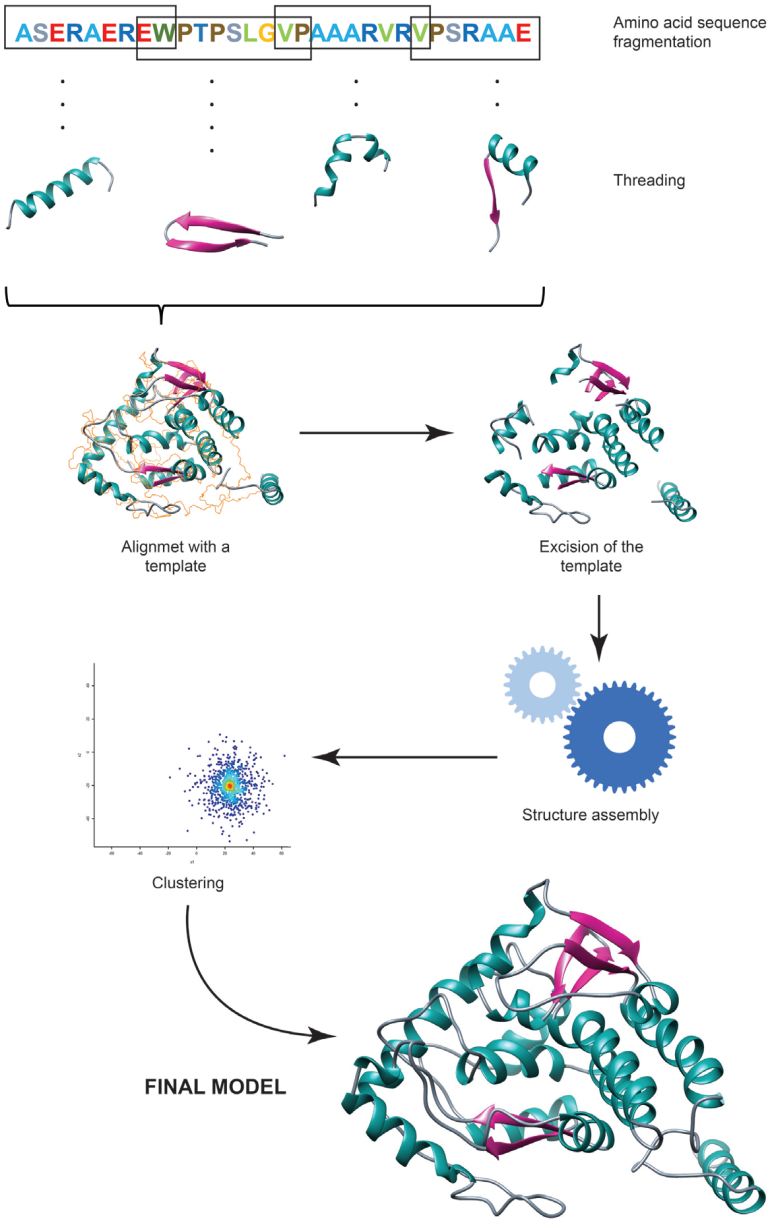

Figure 3: Flowchart of the threading assembly method. The target sequence is divided into small fragments which are fold and compared with a reference structure. The template structure is then eliminated, and several structures are assembled. After clustering and scoring a final model is generated.

This method offers advantages over the other prediction methods mentioned above. Threading can be capable of predicting new folds which are not present in the Protein Data Bank (PDB). The generation of this new folds cannot be achieved by the homology modeling method. The fragment-based approach reduces the search space, which saves a significant amount of computational power, without sacrificing accuracy if compared to an *ab initio* method. This special feature, fragmentation, brings along intrinsic limitations to the method: low accuracy in large conformational searches caused by a different combination of such fragments, and the discrimination of different combination of fragments in regions where the potential energy is relatively low (Dorn et al., 2014).

2.2.2 Modeling assembly method

Homology modeling (or comparative modeling) is considered to be a very successful and reliable method for the generation of new models. However, the accuracy highly depends on the existence of a previously determined and closely related structure. Homology modeling is grounded on the fact that all members of a protein family present similar fold. The structure of related proteins will remain similar, despite the accumulation of variations during evolution. The method uses experimentally validated structures from homologous proteins as templates and aligns the sequence of the target protein over the structure. High homology between proteins generates accurate models. However, suitable models can be obtained even with low sequence similarity, between 20 and 30 % (Mariani et al., 2011). Lower sequence identity decreases the probability of identifying a correct template and due to the misalignment of the sequences errors as side-chain packing, distortion and shifts in the correctly aligned regions or errors in regions without a template might occur.

A scheme of the homology modeling pipeline is present (Fig. 4). In general, it comprises the following steps which can be performed in a linear way or repeated until a suitable model is obtained. (i) The process starts with the identification and selection of a homologous related structure which is used as a template, (ii) then the amino acid sequences from the target and the template proteins are aligned. (iii) The 3D model of the target structure is created using the template protein as a reference, and (iv) finally the quality

of the new model is estimated. The creation of the new structure usually involves refinements such as clash removal and geometrical regularization of bond lengths and angles. As a rule of thumb, most attention should be devoted to steps (i), (ii), (iii) and (v), whereas global model refinement (iv) typically has a disappointing return on investment. After evaluating the quality of the model, loop steps can be incorporated to increase the accuracy of the model. For this purpose, most models go back and evaluate new structures as references or iterates the creation of the new model (Dorn et al., 2014; Webb and Sali, 2014).

2.2.3 *Ab initio* modeling

The *ab initio* or *de novo* modeling method is a template-free approach for the generation of new structures, it aims to predict the native conformation of a protein considering only the amino acid sequence. This method is used when there is no homologous structure available, or the relationship is so distant that even the threading method cannot generate reliable structures (Xu et al., 2000).

The *ab initio* method is based on principle laws of physics and chemistry, and it assumes that the native structure of the protein corresponding to the model with the global minimum free energy among all available conformations. Structural templates are not used during the process, but the structural information is used in the initial parametrization of the all-atoms potentials used in force-fields (potential energy functions). Then the method simulates the protein conformation space using the energy function, which describes the internal energy of the protein and its interactions with the selected environment. In general, the method is divided into three steps. (i) A geometric representation of the protein chain is generated. A detailed version would include all atoms and the surrounding solvent molecules (i.e., water), but this demands enormous computational effort. Therefore, normally *ab initio* folding methods use simplified geometry models in which virtual atoms represent a number of atoms in the all-atom model, which reduces the time required for the folding simulation. (ii) Potential energy functions are used in Molecular Mechanics (MM), protein design and protein structure prediction to determine the protein conformation. Bonded terms (i.e., bonds, angles and torsion) and non-bonded terms (i.e., hydrogen bonds, van der Waals forces, and dipole-

dipole interactions) constrain the lengths and angles of the molecules to find a global minimum. (iii) Molecular dynamics simulations are done to analyze the surface energy of the protein. Energies are compared using Monte Carlo simulations until equilibrium values are reached (Dorn et al., 2014; Khor et al., 2015)

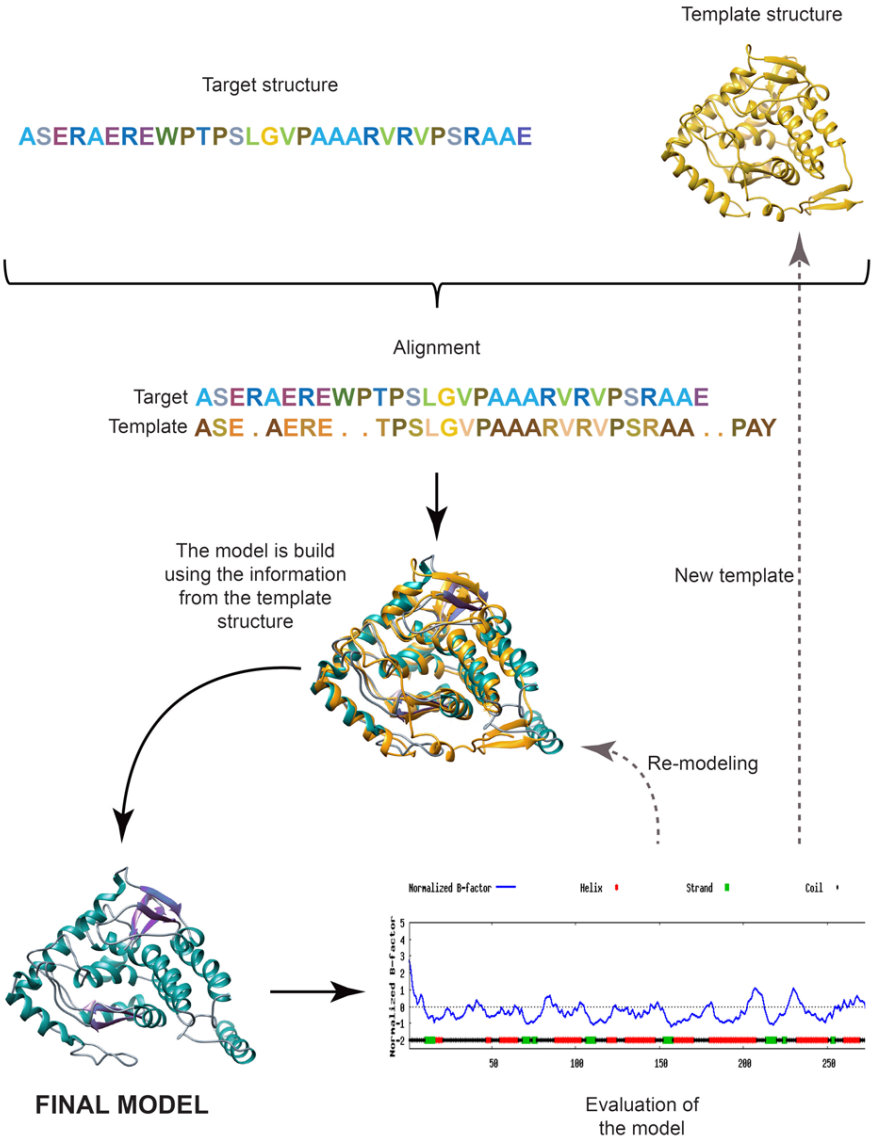

Figure 4: Flowchart of modeling assembly method. The target sequence is modeled using a template structure as a reference. The template structure is selected based on alignments. After the final model is generated, it is evaluated, and if necessary, a re-modeling step is included or the selection of a new template structure.

The major restriction of the *ab initio* method is the exhaustive analysis of all possible structures due to the computational demand. Therefore, the key point is to restrict the conformational landscape of the analysis without sacrificing the predictive capacity of the method.

Regardless of the method used, modeling of protein structure from amino acid sequences now plays a major role in structural biology. Since 1994 the scientific community has organized a biannual worldwide set of experiments called Critical Assessment of Protein Structure Prediction (CASP). The CASP experiments are designed to compare and analyze state-of-the-art methods in protein structure prediction, and other downstream processes such as protein-protein interactions, protein design, molecular docking analysis (discussed in next section), among others (Moult et al., 2016).

2.2.4 Molecular docking

The molecular docking approach can be used to model the interaction between a small molecule (ligand) and a protein (receptor) at the atomic level. This method allows the prediction of the behavior of the small molecule in the binding site of the target protein, as well as the elucidation of basic mechanistic-biochemical processes (Meng et al., 2011). The term "molecular docking" was coined in the early 1980s. Initial works did refinements in the geometry of protein-ligand interactions by optimization of the separation of the partners (Kuntz et al., 1982), but with relatively fixed orientations. Later, these relative orientations were allowed to vary but keeping the internal geometry of the receptor partner fixed. This type of modeling is often referred as rigid docking. Currently, it is possible to vary the internal geometry of the interacting partners during the formation of the complex this is known as flexible docking (de Ruyck et al., 2016).

Flexible docking has the ability to predict the conformation of a small molecule within the targeted binding site with a substantial degree of accuracy. Crucial molecular information, such as the ligand binding mode or the intramolecular interactions that stabilize the ligand-receptor complex, can be conveniently elucidated. The molecular docking algorithm executes quantitative predictions of the binding energy based on the affinity of the ligand-receptor complex and provides ranks of the docked compounds (Ferreira et al., 2015) (Fig. 5).

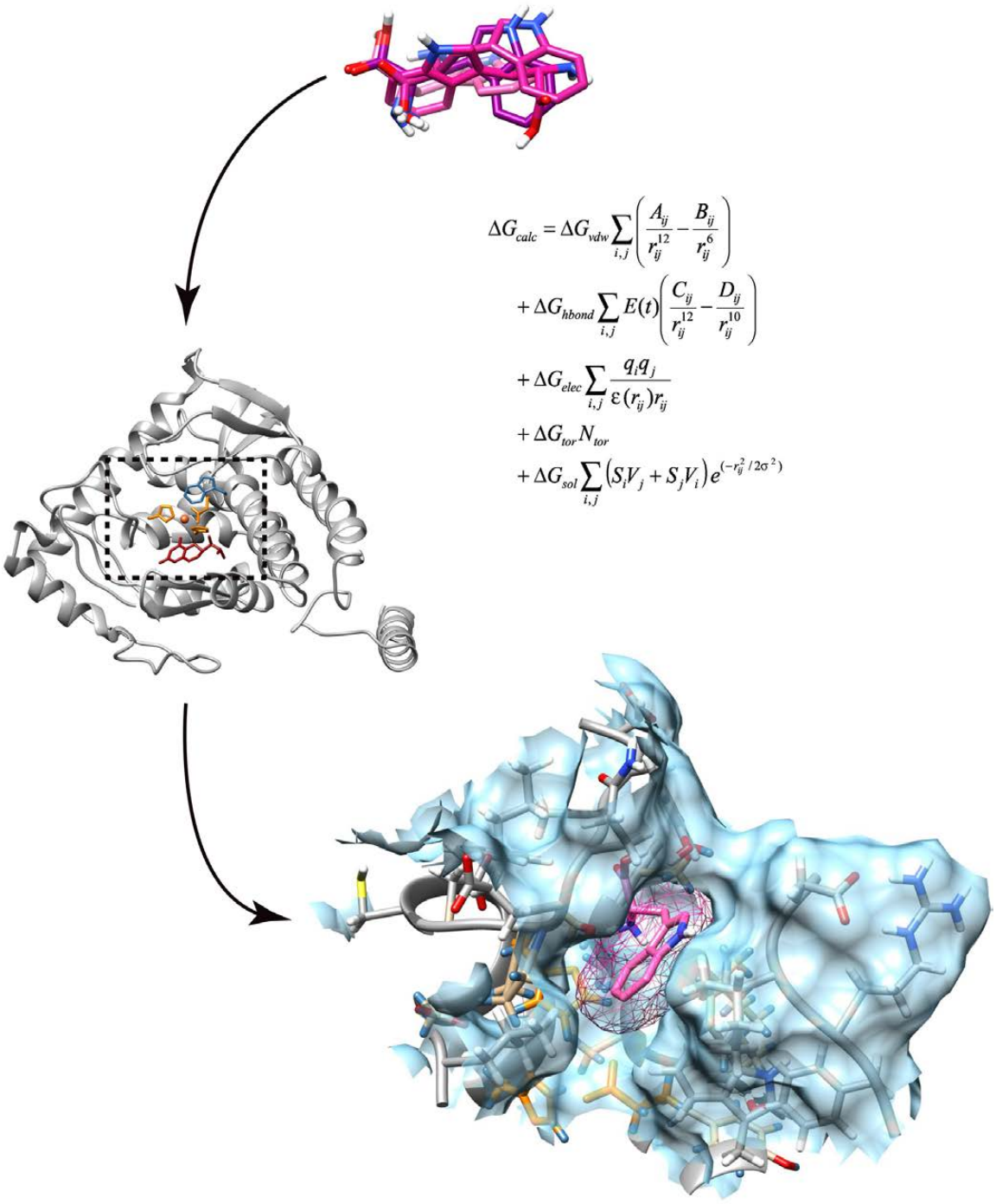

Figure 5: Schematic representation of the molecular docking process. Different conformations of a ligand are evaluated within a defined volume. The equation inserted in the figure is used to calculate the Gibbs free energy. Finally, the ligand-receptor complex is evaluated and ranked to select the most stable model.

The identification of the most likely binding conformation requires two steps. The exploration of the conformational space with various potential binding complex and the prediction of the energy associated to each conformation. This process is repeated until finding a minimum energy solution.

In the conformational search stage, structural parameters of the ligands, such as torsion, translation, and rotation are modified. The search algorithm systematically promotes slight variations in structural parameters. The method also incorporates stochastic conformational searches by randomly modifying the structural parameters of the ligand to avoid local minimum. This combination of systemic and stochastic search methods covers a broad energy landscape. The computational cost associated with this procedure is an important limitation. In every step, the binding energy of each ligand-receptor complex is estimated, and the energy variation is given by the binding constant (K_d) and the Gibbs free energy (ΔG_L). Predicted binding energy is based on physical-chemical phenomena such as intramolecular interactions, hydrophobic effect, desolvation, entropic effects (see equation in Figure 2.3). The higher number of parameters evaluated, the greater the accuracy. However, this comes at a computational cost. Ideally, efficient scoring functions should offer a balance between accuracy and speed (Ferreira et al., 2015). Finally, each conformation is ranked with a scoring function. Normally, the algorithms used to score the models are able to discriminate biologically representative solutions from a group of incorrect decoys (Ferreira et al., 2015; Meng et al., 2011; Taylor et al., 2002).

2.2.5 Computational protein design

One of the most challenging tasks in protein engineering is the design and creation of new enzymes that meet the special needs of bioreactions or metabolic pathways. This often requires catalysts with improved rates and selectivity, that can work under harsh conditions, are tolerant to changes (temperature, pH, concentrations), capable of catalyzing a broad range of substrates, to produce maximal amounts of products, and so on. Therefore, when designing a bioprocess, often native (wild-type) enzymes are not sufficient, rather it is necessary to modify them using protein engineering techniques to change their physicochemical and functional properties (Barrozo et al., 2012; Khoury et al., 2014; Yu et al., 2015).

Different strategies can be used to guide or help the development of new properties in a protein. Conventionally, directed evolution approaches or rational design (also a mixture of both named as semi-rational design) have been used to create and identify novel proteins with desired characteristics (for more detail see Section 2.3.1). However, recent developments in computational chemistry and biology have incorporated *in silico* analysis into protein design. Currently, it is widely accepted that computers have taken a prominent role in guiding and directing experimental work (Barrozo et al., 2012; Swiderek et al., 2015).

Computational enzyme design is inextricably linked to the structure of the protein; detailed analysis of the active sites reveals that specific conformations are required to facilitate catalysis. Therefore, reliable models of the target molecules are strictly required in advance. Two distinguishable methods can be used for this aim, structure-based methods and reaction analysis-based methods. In the structure-based one, molecular mechanics force fields are applied in the active site to deduce interactions between the ligand and the receptor. On the other hand, reaction analysis-based methods also involve quantum mechanics analysis. This allows the following of the complete chemical reaction including the dynamics of the system (Barrozo et al., 2012).

The redesign of an active site focused on the optimization of non-natural substrates is based on the idea that optimization of the catalytic function would imply the stabilization of the interactions between the ligand and surrounding residues in the active site at the transition state (Marti et al., 2004). Combinatorial optimization algorithms integrate ligand docking and (re-) placement of amino acid rotamers libraries to identify sequences that form complementary ligand-receptors surfaces (Martí et al., 2008). The prediction of these residues corresponds to the theoretical permutations needed to achieve a novel feature. Many examples of naturally occurring proteins have been used as starting material to redesign the activity. For instance, this strategy has been used to design variants of chorismate mutase from *E.coli* to improve its catalytic activity (Lassila et al., 2005). Computational design approaches may accelerate the creation of novel enzymes by allowing the construction of focused "smart" libraries for experimental verification.

2.3 Metabolic engineering and synthetic biology for the development of novel strains

Metabolic engineering is broadly defined as the group of methods and concepts used to improve or design cells based on the analysis and modifications of metabolic pathways or networks; typically one specific target is the goal of these modifications. As such, it aims at the engineering (i.e., design, construction, and optimization) of both, native and non-natural, routes for the production of desired molecules. This last task is intimately related to the field of synthetic biology which intends to design and construct new biological parts and systems that do not exist in nature (e.g., genetic control systems, metabolic pathways, chromosomes, cells). At this point, both fields overlap in their interest in pathway engineering. Nevertheless, it is distinguishable that the overall cell and pathway performance is the ultimate goal of metabolic engineering and for this purpose, it can use the construction of synthetic genetic networks or circuits. Meanwhile, these synthetic constructs can be seen as the final product of synthetic biology (Nielsen et al., 2014; Stephanopoulos, 2012).

In order to avoid metabolic burden, engineered pathways should have a balance between protein expression and activity, availability of the precursors and cofactors, toxic intermediates and end-products (Du et al., 2011). Several strategies at transcription level such as plasmid copy number, promoter engineering, intragenic regions engineering, ribosome binding site engineering and codon optimization, have been developed to address these problems (Eriksen et al., 2014). Nevertheless, these strategies cannot overcome limitations inherent of the enzymes as the generation of side-reactions or low activity, so protein engineering is often indispensable when optimizing a pathway. Thus, when generating industrial strains for the production of (un-) natural value-added compounds, protein engineering seems unavoidable.

2.3.1 Protein engineering methods

Protein engineering involves the modification of proteins at an amino acid level to alter the function of the proteins. Different characteristics can be selected for improvement, such as activity, specificity/selectivity, solubility, stability, product or substrate inhibition, among others. Several approaches

are used to tailor enzymes, and these are mainly grouped into three categories: directed evolution, rational design, and a third one consisting a combinatorial method of the first two which is normally called semi-rational design (Eriksen et al., 2014).

Directed evolution aims to mimic Darwinian evolution under a controlled environment in the laboratory. The enzymes can be further optimized by iterative rounds of evolution, starting with the creation of genetic diversity and followed by a selection or screening step. The most common methods used for the generation of diversity include error-prone PCR, DNA shuffling, chemical mutagenesis, use of a mutator strain, among others. To identify the beneficial mutations a mixture of selection-screening methods has been developed which can be grouped as colorimetric and fluorescent assays, or growth assays.

Directed evolution does not require previous knowledge of the structure or mechanism of the enzyme. It has the advantage that the method introduces mutations randomly in the protein, which increases the landscape of the diversity. This last point is strongly related to the size of the library, which normally is the major disadvantage of this method. Large libraries should be analyzed in order to find an enzyme with proper characteristics (Eriksen et al., 2014).

The second approach, rational design, is a knowledge-driven process. Specific structural and sequence information is needed to predict amino acids mutations that affect the properties of the enzyme. Residues within the active site of the 3D crystal structure can be investigated to direct the structure-function relationship. Deep understanding of the catalytic mechanism is desired to apply this method. The combination of all this information is used to limit the size of the libraries since often a small number of residues or amino acids mutations are selected. Therefore screening is not a limiting step when using rational design.

Currently, it is difficult to distinguish protein engineering studies using solely a directed evolution approach or rational design. Nowadays, researchers combine these techniques, in the first step hotspots (or regions) are selected, and a targeted saturation mutagenesis approach is then used. This is a powerful method since it can reduce the size of the library to be screened. These intelligent libraries rely on the ability to identify key residues with critical effect

on the structure-function relationship, and saturation mutagenesis maximizes the diversity landscape and the probability of finding positive results (Quin and Schmidt-Dannert, 2011).

2.3.2 Library design for protein engineering

Protein engineering has proven to be an effective method for manipulating and tailoring of biocatalysts. Directed evolution and rational design are commonly used for the construction of the libraries and phenotype selection during the engineering process. The construction of the library is a critical step that requires the combination of a method for the creation diversity with the proper screening system. Optimal libraries should be complex enough to contain rare beneficial mutations, and it should encode mostly functional properly folded proteins. Also, genotype duplication should also be low in the library, and the mutational spectrum of the library should be adjustable to populate desired amino acids substitutions (Wong et al., 2007).

Diversity creation methods can be divided into three categories. (i) Random mutagenesis mainly includes error-prone PCR, mutator strains and the use of chemical mutagens (and derivate methods from these). An ideal method should cover all kind of nucleotide substitutions equally and achieve three consecutive substitutions to target all amino acid changes. Nevertheless, all reported methods to date have a strong bias toward some few substitutions, and they normally fail due to two consecutive substitutions. (ii) Focused mutagenesis used to improve proteins after a beneficial mutation has been identified with random mutagenesis, or when residues have been specifically selected due to the understanding of the structure-function relationship of the protein. In this case, it is possible to apply saturation mutagenesis to one position (or neighboring positions). Most of the current methods are based on PCR reaction, and the variability is incorporated in the plasmids. Different randomization schemes can be used to reduce the size of the library or direct the properties of the library using certain combination of codons such as, VRK (hydrophilicity), NVC (hydrophobicity), NDT (small and balanced), or 22c-trick (all amino acids minimal redundancy) (Kille et al., 2013). Currently, with the combination of different PCR reaction and the synthesis of DNA fragments, it is possible to generate libraries targeting several residues. Nowadays, the limit in the number of residues is given by the screening capac-

ity and not by technical limitations during the generation of the library. (iii) DNA shuffling is used to create libraries via DNA recombination. This method requires several gene sequences encoding proteins with similar function. The new variants are generated using *in vitro* recombination to rearrange the DNA fragments randomly (Firth and Patrick, 2005; Sheppard et al., 2014).

Regardless of the method chosen to create the library, the larger the library, the higher the likelihood of discovering a novel variant. Also, the generation of large libraries is related to laborious and time-consuming procedures. Thus, the benefits and drawbacks associated with the library size should be carefully balanced. For this purpose, two metrics are often used, (i) the probability that all possible variants are present in the library (probability of full coverage), and (ii) the expected percentage of all possible variants that are represented in the library (expected coverage). The combination of these metrics can be used to estimate the sampling size necessary to have a statistical coverage of the library. This step is essential for the selection of the screening method, which should be coupled with the size library. If the screening method fails to meet the requirements of the library an enormous effort must be made to screen the library; in the opposite case, undersampling would drastically decrease the probability of finding the novel variants (Khor et al., 2015; Lutz and Patrick, 2004).

2.3.3 Biomolecular sensors

As mentioned above, metabolic engineering envisions the generation of a diversity fine chemical and pharmaceutical compounds from renewable sources using cell factories. The plethora of molecules that can be produced also requires efficient screening methods to select individual mutants carrying the desired phenotype. Early developed analytical methods were based on enzymatic assay read with spectroscopy. However, this has a limited throughput. To bypass this obstacle, genetically-encoded synthetic- fluorescent biomolecular sensors have provided means to monitor *in vivo* cellular metabolism in a noninvasive fashion in the native cellular environment and with high spatial-temporal resolution. Molecular biosensors also offer the opportunity to do high-throughput screening and selection based on fluorescence and cell survival.

Nature has developed two different types of biosensors. Transcription factors are proteins that regulate the expression of a gene in response to changes in the environment. Qualitative or quantitative analysis of a metabolite can be driven by hacking specific transcription system and coupling it to reporter genes. These sensors exhibit low orthogonality and background noise. The second class of biosensor comprises the riboswitches. These are regulatory elements in the mRNA that can selectively bind to a ligand and changes its structure. Therefore, the ligand can directly repress or activate their cognate genes at both transcriptional and translational levels (Garst et al., 2011; Nudler and Mironov, 2004). In contrast to the transcription factors biosensors, riboswitches are faster as the RNA has already been transcribed. Also, they do not rely on protein-protein or protein-metabolite interaction. In recent years, riboswitches have been extensively used in bacterial systems for many protein engineering or metabolic engineering purposes (Fang et al., 2016; Zhou and Zeng, 2015).

In *E. coli* the *tna* operon is controlled by a feedback loop that requires tryptophan. The operon consists of a leader regulatory region followed by the structural genes *tnaA* and *tnaB*. The leader sequence in the *tna* operon contains a 24-residue peptide, *tnaC*, followed by several Rho-dependent transcription termination sites. In the absence of tryptophan, tnaC termination dissociates the ribosome. This enables the terminator factor Rho to bind to the mRNA and moves forward following the polymerase. Rho catches up the RNA polymerase and unwinds the DNA-RNA hybrid in the transcription bubble, which terminates the transcription. On the other hand, when tryptophan is present the ribosome is not dissociated. Rho-dependent transcription termination sites are blocked, and the transcription and subsequent translation of *TnaA* and *TnaB* occur (Yanofsky, 2007) (Fig. 6).

Fang et al., (2016) developed a tryptophan biosensor based on the *tnaC* leader sequence, they couple the promoter sequence with a green fluorescent protein to quantify intracellular tryptophan concentration. In this study, deoxy violacein optimization was divided into two modules with tryptophan as a key intermediate. In the first step, fluorescent-ON cells were selected as tryptophan producers; and in the second step, fluorescent-OFF cells expressing the deoxy violacein synthetic pathway were recognized as tryptophan consumers, therefore cells capable of synthesizing the desired compound. Similar approach in which the substrate consumption is used to track a reaction has

been used or proposed by others, but mainly using sensors that are linked to the central metabolism (Lehning et al., 2017; Staiano et al., 2005).

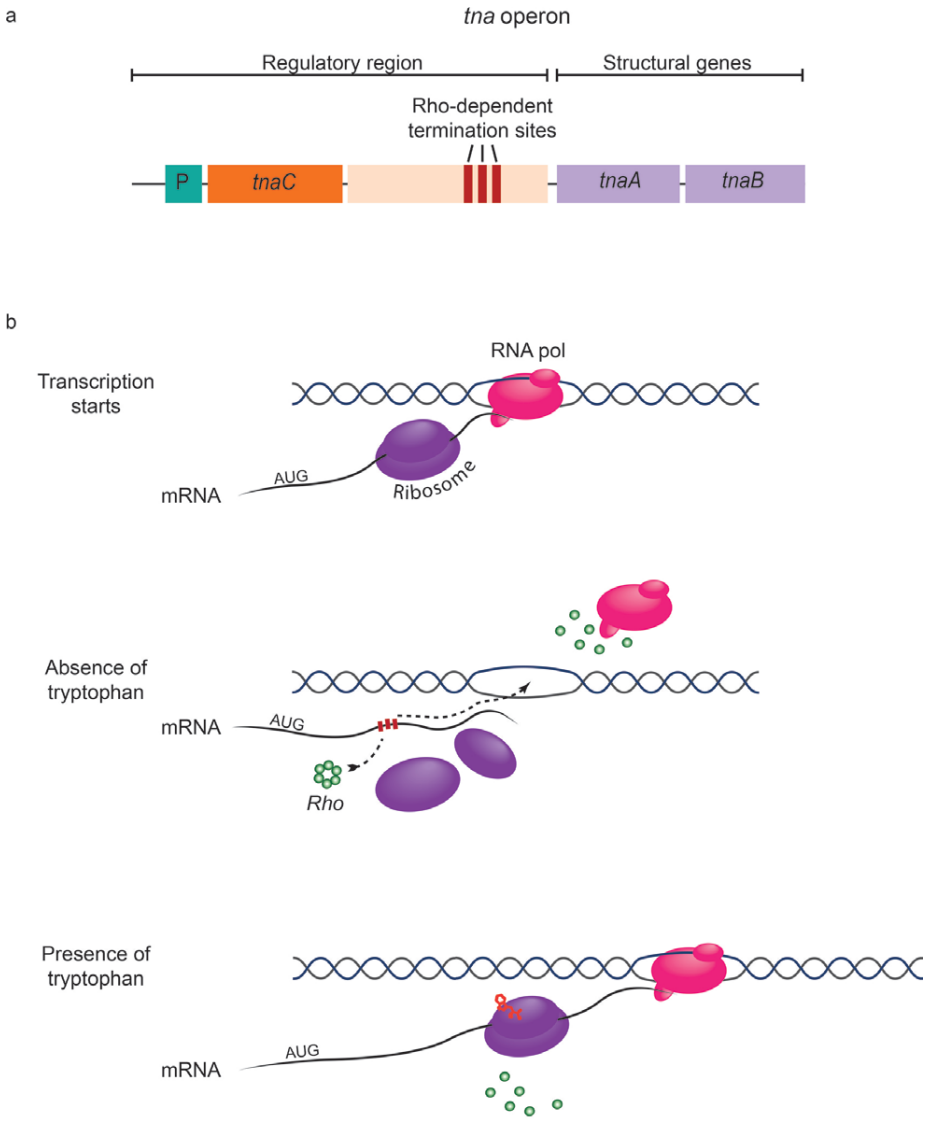

Figure 6: The *tna* operon from *Escherichia coli* and its transcriptional regulation. (a) The *tna* operon is composed of a regulatory region with a promoter, a *tnaC* leader transcript segment, and Rho-dependent transcription sites. *tnaA* and *tnaB* conform the structural region. (b) RNA polymerase transcribes the initial region of the operon, and the ribosome binds to the nascent tna mRNA. In the absence of tryptophan the ribosome dissociates after the synthesis of TnaC, the Rho factor binds to the mRNA and terminates the polymerization by unwinding the RNA-DNA complex. In the presence of tryptophan the ribosome stalls in the Rho-dependent transcription sites. The Rho factor cannot bind, and RNA polymerase transcribes the structural genes.

2.4 Tryptophan biosynthesis and metabolic pathways in *E. coli*

Different works have been done to drive the carbon flux into the aromatic amino acid pathway, which finally leads to the production of tyrosine, phenylalanine, and tryptophan. A general scheme with the tryptophan metabolic pathway and its regulation is present in Fig. 7. The first step of the biosynthesis is the condensation of phosphoenolpyruvate (PEP), which is produced during glycolysis, and erythrose 4-phosphate (E4P), an intermediate of the pentose phosphate pathway (PP pathway). This reaction is catalyzed by the 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase, and there are three isoenzymes which are specifically feedback inhibited by the three final products of the pathway (Wallace and Pittard, 1969; Zurawski et al., 1981). Moreover, DAHP synthase is also regulated at a transcriptional level by the tyrosine- and tryptophan- transcriptional repressor (TyrR and TrpR) (Brown, 1968; Davies et al., 1985).

Shikimate kinase, expressed by *aroK* and *aroL*, seems to be a second important regulation point within the pathway. *AroK* expression is constitutive. Meanwhile, *aroL* is controlled by TryR and TrpR. Nevertheless, the physiological role of AroL is preponderant when compared to AroK, since the affinity of the first one is 100-times higher (DeFeyter and Pittard, 1986; Ely and Pittard, 1979; Lawley and Pittard, 1994)(Fig. 7).

Chorismate is the bifurcation point of the pathway. Chorismate mutase/prephenate dehydratase (encoded by *pheA* and *tyrA*) catalyzes the synthesis of prephenate which is a precursor for the synthesis of phenylalanine and tyrosine. In the tryptophan branch, the tryptophan operon (termed *trp* operon) encodes a cluster of genes responsible for the conversion of chorismate to tryptophan (Umbarger, 1978).

The *trp* operon is highly regulated at different levels. Traditionally transcriptional repression, attenuation, and feedback inhibition have been recognized to control the tryptophan production pathway (Santillan and Mackey, 2001). Lately, a fourth regulatory element was discovered in Zeng‘s group: feedforward inhibition (Chen, 2017).

.

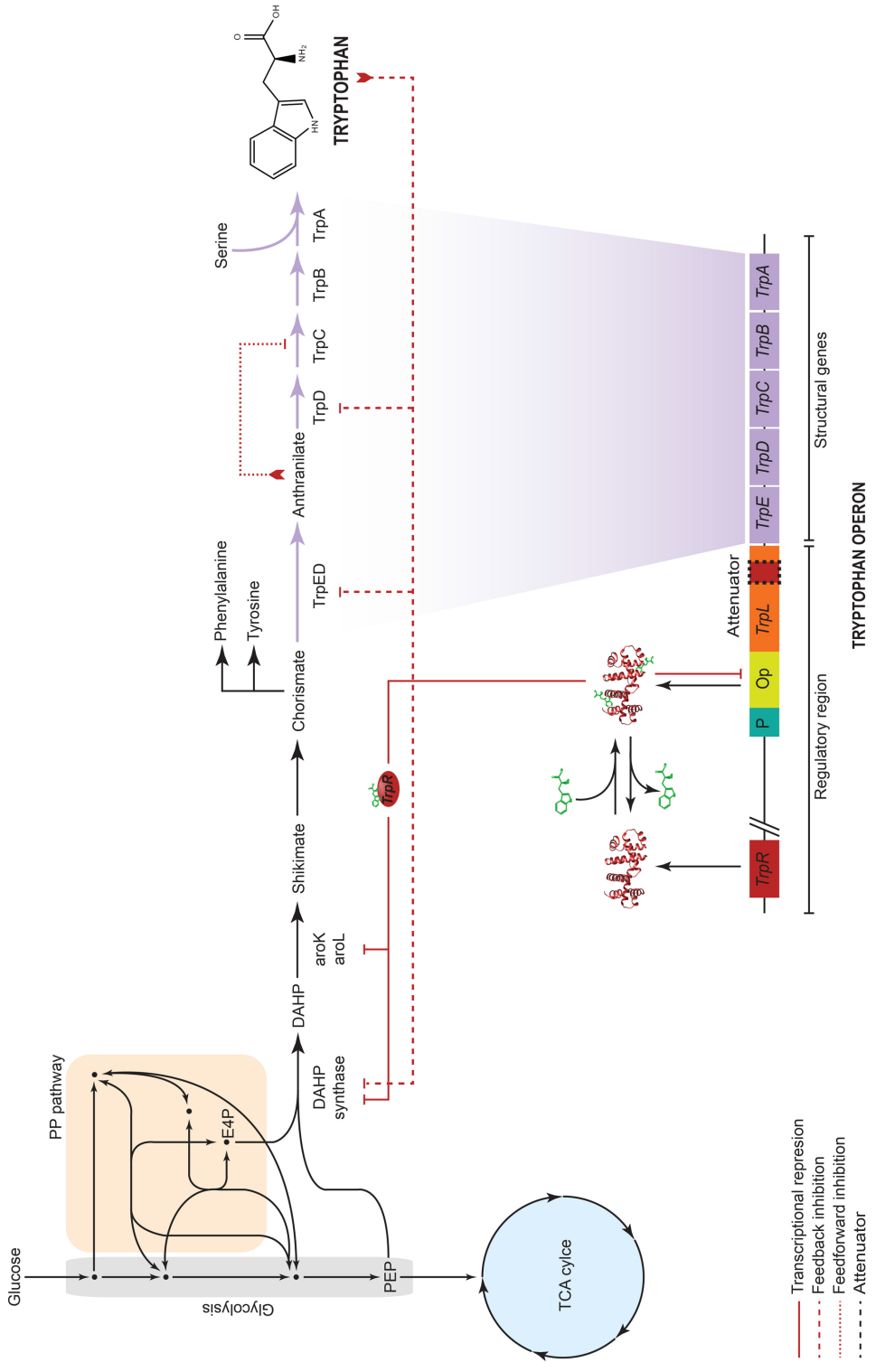

Figure 7: Tryptophan metabolism pathway and regulation. Erythrose 4-phosphate (E4P) and phosphoenolpyruvate (PE) are condensed by DAHP synthase at the beginning of the shared pathway with aromatic amino acids, which bifurcates in chorismate. Tryptophan biosynthesis from chorismate is organized in a single operon which is repressed by TrpR and regulated by an attenuator. Tryptophan has a feedback inhibition effect over different enzymes in the pathway. Anthranilate forward inhibits TrpC (Chen. 2017).

TrpR mediates the repression at a transcriptional level in response to the intracellular concentration of tryptophan. When tryptophan is present in excess, it interacts with TrpR, and together they bind to the operator sequence which overlaps with the promoter, preventing the transcription of the gene (Rose et al., 1973). The second mechanism that regulates the *trp* operon consists of an attenuator structure within a nucleotide leader region (*trpL*). Four sequences are distinguishable in the attenuator and the pairing of *sequence 3* and *4* acts as a transcription terminator. *Sequence 1* contains two contiguous tryptophan codons that are translated into a leader peptide and if tryptophan availability is high translation proceeds quickly and the ribosome occupies *sequence 2*. In this case, sequence 3 and 4 pair, form the attenuator structure and transcription ends. On the other hand, when tryptophan concentrations are low, the RNA polymerase stalls at the two-tryptophan position because the availability of the amino acid is less. *Sequence 2* is also complementary to *sequence 3* and they can pair together, disabling the formation of the attenuator structure (*sequence 3* and *4*) in this case the structural genes are then translated (Lee and Yanofsky, 1977; Zurawski et al., 1978)(Fig. 8). Besides the transcriptional regulation of the The trp operon leader region (trpL) and its function. operon, anthranilate synthase (trpE) is subject to feedback inhibition by the end-product of the pathway (Pabst et al., 1973). A novel way of regulation of the *trp* operon was recently described by Chen (2017), anthranilate forward inhibits indole glycerol phosphate synthase (trpC) which is downstream within the tryptophan pathway.

Metabolic engineering has improved the bioproduction of tryptophan significantly in *E. coli* during the last two decades. Different strategies have been developed to deregulate the common part of the aromatic amino acid pathway and to promote the tryptophan branch over the phenylalanine and tyrosine.

Overproducer strains contain different strategies to bypass the tight regulation. The common aromatic amino acid pathway, or the chorismate pathway, as mentioned before is regulated at a transcriptional and enzymatic level. Several DAHP synthase feedback-resistant have been developed to address the latter issue (Ger et al., 1994; Hu et al., 2003; Liao et al., 2001; Lin et al., 2012). At a transcriptional level, the elimination of tryR and trpR have proven to increase the carbon flow into the chorismate pathway. PEP and E_4P are the precursor molecules condensed by DAHP, so increasing the supply of this molecules increase the flux into the pathway, common strategies for

this purpose is to overexpress *ppsA* and *tktA*, which increase the availability of PEP and E4P, respectively (Ikeda, 2006; Rodriguez et al., 2014). A reduction in the amount of acetate has proven to be beneficial for the production of tryptophan. There are two explanations for this: first the reduction of acetate in the media which reduces growth, and second an increase of available PEP (Zhao et al., 2016).

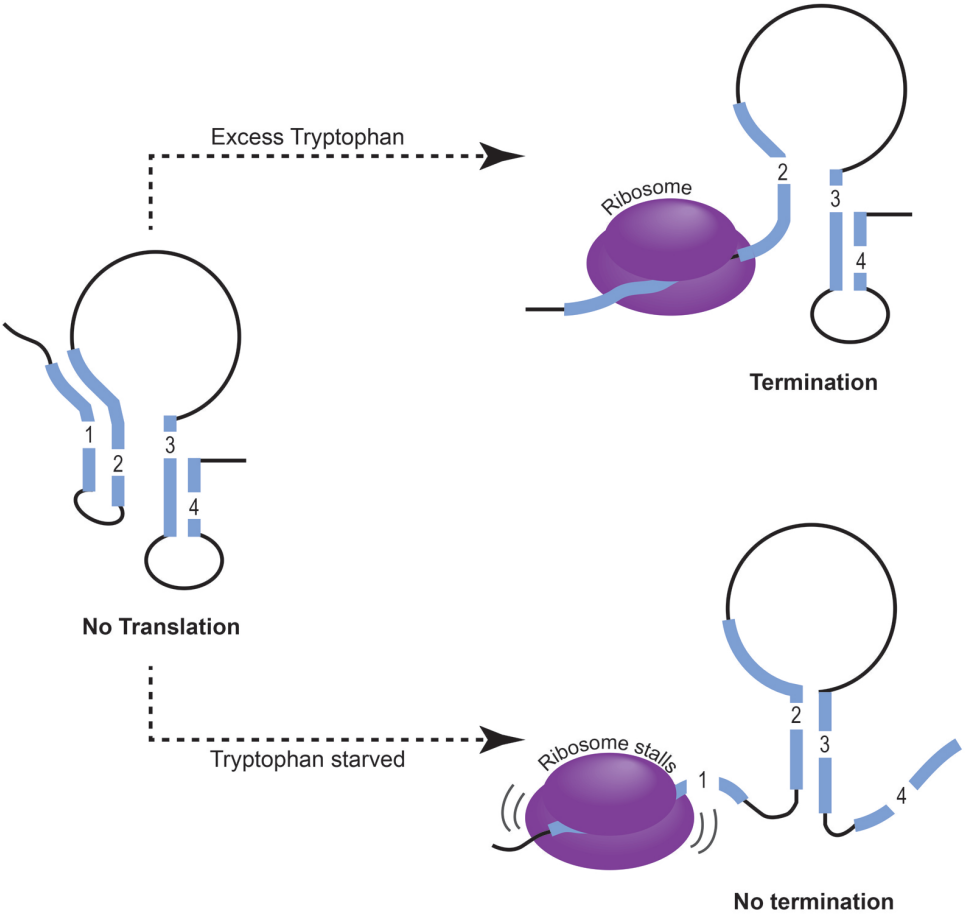

Figure 8: The *trp* operon leader region (*trpL*) and its function. The initial 141 nt of the *trp* operon can form alternative RNA hairpin structures. Whenever tryptophan level is high, the ribosome moves rapidly from region 1 to 2, and a terminator structure (3:4) is formed. This results in transcription termination. When tryptophan concentration is low the ribosome stalls in region 1, an anti-terminator structure (2:3) is formed, preventing the formation of the terminator structure.

.

Important strategies have been developed within the tryptophan branch, including the desensitization transcriptional regulation of the *trp* operon (i.e., elimination of the repression, as well the disruption of the attenuator), the removal of the feedback inhibition, and overexpression of the complete pathway. The combination of the three of them has shown significant improvement in the tryptophan production (Aiba et al., 1980; Chan et al., 1993; Dodge and Gerstner, 2002; Shen et al., 2012; Chen and Zeng, 2017). Furthermore, serine is consumed during the catalysis of tryptophan. The incorporation of the feedback-resistant *serA* gene has proven to increase final tryptophan yield (Dodge and Gerstner, 2002; Wang et al., 2013; Chen and Zeng, 2017).

A couple of extra strategies, which are downstream from the tryptophan production, have been used to increase final concentration. First, tryptophan degradation pathway has been blocked by knocking-out *tryptophanaseA* (*tnaA*) gene. Second tryptophan and aromatic amino acid transporters have been manipulated or eliminated from the producer strains (Aiba et al., 1980; Chan et al., 1993; Dodge and Gerstner, 2002; Shen et al., 2012; Chen and Zeng, 2017).

Reported metabolic engineered strains have a combination of the strategies mentioned above (Gu et al., 2013; Zhao et al., 2012, 2011; Chen and Zeng, 2017). The most recently reported strain S028 (Chen and Zeng, 2017) presents the highest titer among those rational design strains at the moment.

S028 combines several such as the overexpression the chorismate pathway, the addition of a feedback-resistant DAHP synthase, the *trp* operon was also overexpressed and trpE was substituted for a feedback-resistant variant, transcriptional repression and attenuation were eliminated. Serine availability was increased by the incorporation and overexpression of the *serA* gene. Tryptophan degradation pathway was blocked, and tryptophan uptake system was disrupted by eliminating *tnaB* and *mtr*. Finally, *E. coli* strain S028 was able to efficiently produce 34-40 g/L of tryptophan with an overall yield of 0.15 g tryptophan/glucose and a productivity of 0.60 g/L/h in fed-batch fermentation.

2.4.1 Synthetic pathways and metabolic engineering for the production of tryptophan derivatives

Different tryptophan (trp) derivatives have been produced using biotechnological methods. Perhaps the earliest report on the production of a trp derivate by extending the tryptophan pathway in *E. coli* is the production of indigo in the early 1980's (Ensley et al., 1983). Naphthalene dioxygenase was introduced in *E. coli* HB101 and combined with the endogenous activity of tryptophanase A the strain was able to synthesize de blue dye in the presence of tryptophan or indole. Later this same group reported the metabolic manipulation of a strain and the synthesis of indigo from glucose (Murdock et al., 1993).

The biosynthesis of plant hormone, indole-3-acetic acid (IAA), have also been reported using *E. coli* as cell factory. Aminotransferase (*aspC*), indole-3-pyruvic acid decarboxylase (*ipdC*) and indole-3-acetic acid dehydrogenase (*iad1*) genes were transformed in *E. coli*. The resulting strain was capable of producing IAA when tryptophan was supplied in the media (Romasi and Lee, 2013).

Tryptophan derivatives with pharmaceutical interest have also been produced in *E. coli*. Indole pyruvic acid (IPA) is used as a drug of the nervous system. Its chemical synthesis is costly and complicated, but recently Zhu et al. (2017) presented a biotechnological alternative for efficient biosynthesis of IPA. Tryptophan aminotransferase from *Aspergillus nidulans* was expressed in *E. coli* and the resulting strain was able to convert tryptophan and phenyl pyruvic acid into indole-3-pyruvic acid and phenylalanine. A second example of production of high-valued molecules with pharmaceutical interest is the biosynthesis of violacein and deoxy violacein. These compounds possess antibiotic, antiviral and antitumoral activity. A system-wide metabolic approach was applied to *E. coli* to target their production. The *vioABCE* cluster from *Chromobacterium violaceum* and the *vioD* gene from *Janthinobacterium lividum* were expressed in a strain with an improved serine, chorismate, tryptophan and pentose-phosphate pathway. The producer strain was able to accumulate 710 mg/L, and after crystallization, violacein was recovered with a 99.8 % purity (Rodrigues et al., 2013).

In the case of 5HTP and serotonin, production in *E. coli* has also been reported. For 5HTP, phenylalanine hydroxylase from *C. violaceum* was engineered to shift the preference of the substrate from phenylalanine to tryptophan (Kino et al., 2009). This enzyme was introduced in a strain with a cofactor regeneration pathway, and when tryptophan was supplied to the media 2.5 mM (550 mg/L) of 5HTP was synthesized (Hara and Kino, 2013). In an independent work Lin et al. (2014), analyzed five phenylalanine hydroxylases (PAH) from different bacteria (*Pseudomonas aeruginosa, P. putida, P. fluorescens, Ralstonia eutropha H16, and Xanthomonas campestris*). PAH from *X. campestris* was selected as starting material for protein engineering. The enzyme was incorporated in a tryptophan producer strain harboring a cofactor regeneration pathway, and 152 mg/L of 5HTP were produced from glucose. A different approach was described by Sun et al. (2015). Here salicylate 5-hydroxylase was used to convert anthranilate to 5-hydroxy anthranilate and using a two-stage strategy they were able to produce 5-HTP from glucose at a final concentration of 98 mg/L (Sun et al., 2015).

Serotonin production in *E. coli* has been achieved in media supplied with necessary precursors (tryptamine or 5HTP). Park et al. (2011) co-expressed tryptophan decarboxylase from *Catharanthus roseus* and rice tryptamine hydroxylase (T5H) gene in *E. coli*, and produced 24 mg/L of serotonin when the strain was grown in medium supplied with 404 mg/L of tryptophan. Hydroxylase activity was a clear bottleneck in the pathway since serotonin concentration did not increase as its intermediate increase in the media. In a second paper, this same group presented a strategy to overcome this problem. They identified a NADPH-cytochrome P450 reductase (OsCPR2) as a potential electron donor to the T5H. On the other hand, when coexpressed in *E. coli* and supplied with 1 mM tryptamine, they were able to produce (250 mg/l) (1.4 mM) of serotonin (Park et al., 2013).

5HTP has also been used as a precursor for the biosynthesis of serotonin in *E. coli*. A recombinant strain harboring tryptophan decarboxylase from rice was grown in medium supplied with 3 mM 5HTP (660 mg/L), 35 mg/L of serotonin were produced after a 24 h period (Park et al., 2008).

3 Materials and Methods

3.1 Chemicals

Analytical grade or molecular biology grade chemicals were purchased from Sigma-Aldrich (München, Germany), or from Carl Roth GmbH (Karlsruhe, Germany). Enzymes were obtained from Takara Bio Europa S.A.S. (Saint-Germain-en-Laye, France), or Thermo Fisher Scientific GmbH (Karlsruhe, Germany). DNA preparation kits were purchased from Macherey Nagel GmbH (Düren, Germany), and site-directed mutagenesis kits were obtained from Agilent Technologies (Karlsruhe, Germany). Bacterial strains were purchased from New England Biolabs GmbH (Frankfurt am Main, Germany), and the Leibniz-Institut DSMZ GmbH (Braunschweig, Germany). The plasmids used in this project order from Merck Chemicals GmbH (Darmstadt, Germany) or were requested to the corresponding author.

3.2 Computational analysis

3.2.1 Data collection, phylogenetic and functional divergence analysis

To identify members of the AAAH protein family, the keyword "aromatic amino acid hydroxylase" was used as a query to search against the NCBI protein reference sequence database. Hypothetical and predicted sequences were eliminated, as well as redundant entries. All amino acid sequences were examined for potential the AAAH domain for their authenticity using CD-search analysis (Marchler-Bauer et al., 2015).

Amino acid sequences of AAAH genes were aligned using MUSCLE (Edgar, 2004) or ClustalW (Thompson et al. 1994) with default parameters, refinements within the alignments were done manually. Phylogenetic analyses employing the neighbor-joining (NJ) method with 10 000 bootstrap values were done in MEGA v7 (Kumar et al., 2016). In parallel, maximum likelihood (ML) analysis was also performed using ProTest v2.4 (Abascal et al., 2005), the best-fit model predicted rooted the ML tree. This analysis was conducted in MEGA v7 using a bootstrap of 1 000 to assess the reliability of interior branches.

Functional divergence type-I (FD-I) were calculated using the method suggested by Gu et al. (2001), this estimates the level of divergence among sequences and identifies relevant amino acid residues within subfamilies. The analysis was carried out with Diverge v2.0. The method estimates the probability of changes in the site-specific shift of evolutionary rate or the site-specific shift of amino acid properties after the emergence of two paralogous sequences. Site-specific (functional diverge type-II – FD-II) analysis was used to predict amino acids residues that were crucial for functional divergence.

3.2.2 Modeling of tertiary structures and docking analysis

Tertiary structures of bacterial AAAH were generated with the I-TASSER server (Roy et al., 2010) and compared with homologous validated protein structures. A rotamer library was used during the structure determination (Shapovalov and Dunbrack, 2011), and side chains were refined using the Protein Data Bank (PDB - http://www.rcsb.org) entries 1MMK and 3E2T as guides. Molecular docking analysis of tryptophan and phenylalanine using the generated AAAH (wild-type and variants) were performed using AutoDock Vina v1.1.2 (Trott and Olson, 2010). The ligands were manually placed in the active site using the substrate position in other AAAHs crystal structures as a guide, for these analyses we used a grid box of 15 x 15 x 15 Å and the level of exhaustiveness set to 10. Molecular graphics and analysis of all the structures were performed with the UCSF Chimera package (Pettersen et al., 2004).

3.3 Molecular cloning

Chromosomal DNA from *C. taiwanensis* was prepared using NucleoSpin Tissue (Macherey-Nagel, Germany) according to the manufacturer‘s instructions. The *aaah* gene from *C. taiwanensis* was amplified by PCR using the following primers (5‘ to 3‘): *CtAAAH*-for, TATGGATCCATGCGCATTATATGCAT GATCTATCGAAATC and *CtAAAH*-rev, ATCAAGCTTTCAGATGTCTT CGGTATCGGCC. The generated fragment was cloned into the *Bam*HI and *Hind*III sites of the pACYCDuet-1 vector to yield the plasmid p*CtAAAH*.

Site-directed mutagenesis (SDM) experiments were carried out using QuickChange (SDM) kit (Stratagene, Santa Clara, USA). The resulting plasmids were transformed separately into NEB10-beta strains by $CaCl_2$ method (Sambrook and Russell, 2000). All mutations were verified by DNA sequencing.

3.3.1 Bacterial strains and plasmids

Strains and plasmids used in this study are tabulated in Table 1 and 2, respectively. *E. coli* NEB10-beta (New England Biolabs, Frankfurt) was used for general cloning purposes. The strains used for the production of tryptophan metabolites were derivate from the *E. coli* strain S028 (Chen and Zeng, 2017). The tryptophan repressor (*trpR*) gene was eliminated from strain S028 by homologous recombination using the approach previously described by Datsenko and Wanner (2000). An overnight culture of the strain S028 carrying pKD46 was grown at 30 °C in 10 mL of LB media supplied with 1 mM arabinose until an OD_{600} of 0.6. Afterward, the culture was chilled on ice for 10 min, cells were washed three times with a cold 10 % glycerol solution and finally resuspended in 400 μL of 10 % glycerol. 100 ng of linear DNA (FRT - kanamycin resistance - FRT gene flanked by *trpR* regions) was mixed with 200 μL of electrocompetent cells. After electroporation (0.2 cm cuvette, 2.5 kV, 25 μF and 200 Ω), cells were mixed with 1 mL SOC and incubated at 30 °C for 1 h before plating, from here strain TrpD was generated.

Table 1: Strains used in this study

Strain	**Genotype / description**	**Source / reference**
Cupriavidus taiwanensis	DSM No.: 17343, type strain	DSMZ
Escherichia coli NEB10-beta	*Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- ϕ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL* (Str^R) *rph spoT1 Δ(mrr-hsdRMS-mcrBC)*	New England Biolabs

continues...

... continues

E. coli BL21(DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI:: PlacUV5::T7 gene1) i21 Δnin5*	New England Biolabs
S028	W3110 ΔlacU169 gal490CI857 Δ(cro-bioA) rpsL (StrR) ΔaroFΔaroG ΔmtrΔtnaA ΔtnaB ΔaroH::$PJ_{23119-rpsL-tac}$-($aroG_{S180F}$-$serA_{H344A/N364A}$) P_{trc}-$trpE_{S40F}$D CBA	Chen and Zeng, 2016
TrpD	S028 Δ*TrpR*	This study
TrpD-Pl	S028 Δ*TrpR* carrying pBbE1k-2.2	This study
TrpD-Gi	S028 ΔTrpR::PDC DHPR	This study

Table 2: Plasmids used in this study

Plasmids	**Genotype / description**	**Source / reference**
pACYCDuet-1	Expression vector with IPTG inducible T7 promoter	Novagen
p*Ct*AAAH	pACYCDuet-1 encoding *AAAH* from *C. taiwanensis*	This study
p*Ct*AAAH-(L113Y)	p*Ct*AAAH derivative with L113Y mutation	This study
p*Ct*AAAH-(W192F)	p*Ct*AAAH derivative with W192F mutation	This study
p*Ct*AAAH-(Y222H)	p*Ct*AAAH derivative with Y222H mutation	This study
p*Ct*AAAH-(S223C)	p*Ct*AAAH derivative with S223C mutation	This study
p*Ct*AAAH-(P229A)	p*Ct*AAAH derivative with P229A mutation	This study

continues...

... continues

p*Ct*AAAH-(Y244C)	p*Ct*AAAH derivative with Y224C mutation	This study
pKD46	Temperature sensitive plasmid (repA101ts) encoding lambda()-derived Red recombination system (γ, β, *exo*)	Datsenko and Wanner, 2000
pBbE1k-2.2	pBbE1k with *DHPR* and *PCD* genes	Satoh et al., 2012
pBAD	Expression vector with arabinose inducible P_{BAD} promoter	Thermo Fischer Scientific
pACPJ23	pACYCDuet-1.2 derivative with constitutive P_{J23110} promoter	This study
pACPtrc	pACYCDuet-1.2 derivative with constitutive P_{trc} promoter	This study
pACPJ23-CtAAAH (W192F)	pACPJ23 with *CtAAAH*(W192)	This study
pACPtrc-CtAAAH (W192F)	pACPtrc with *CtAAAH*(W192)	This study
pCOLADuet-1-GSTΔ37T5H+TDC	pCOLADuet-1 with *TDC* and *T5H* fused to GST	Park et al., 2011
pCOLADuet:FL-OsCPR2	pCOLADuet-1 with *OsCPR2*	Park et al., 2013
pCOLADuet-1-*TDC*	pCOLADuet-1-*GSTΔ37T5H+TDC* derivative without the *T5H gene*	This study
pCOLADuet-1-GSTΔ37T5H-OsPCR2	pCOLADuet-1- with genes from pCOLADuet-1-*GSTΔ37T5H+TDC* and pCOLADuet:FL-*OsCPR2*	This study

continues...

... continues

pCOLADuet-1-GSTΔ37T5H-OsPCR2	pCOLADuet-1- with genes from pCOLADuet-1-*GSTΔ37T5H+TDC* and pCOLADuet:FL-*OsCPR2*	This study
pCOLAPJ23-*TDC*	pCOLADuet-1.2 derivative with constitutive P_{J23110} promoter and *TDC*	This study
pBAD-*TDC*	pBAD with *TDC*	This study
pSentrp	Plasmid encoding tryptophan intracellular concentration sensor	Fang et al., 2016
pSentrp(-LVA)	pSentrp derivative with a short peptidase sequence fused in the C-terminal of the GFP	This study

Afterward, human pterin-4 alpha-carbinolamine dehydratase (*PCD*) and dihydropteridine reductase (*DHPR*) genes were incorporated into the new strain via plasmid incorporation (pBbE1k-2.2) or genome integration, resulting in the strains TrpD-Pl and TrpD-Gi respectively. The newly generated strains contain necessary genes for regeneration of the pterin cofactor involved in the tryptophan hydroxylation reaction. The kanamycin gene was eliminated by FRT recombination in the case of the TrpD-Pl strain, or swapped for the bicistronic gene *PCD-DHPR* in the TrpD-Gi strain. Plasmid pKD46 was eliminated at 37 °C. Colonies with correct insertion were verified by PCR (see Section 3.3.2) and DNA sequencing. The first step of tryptophan/5HTP degradation was disrupted in *E. coli* BL21(DE3) strain by eliminating the tryptophanase gene (*tnaA*), BL21(DE3) *ΔtnaA* was generated using the same method described above. This strain was used for protein evolution, screening, and bioconversion from 5HTP to serotonin.

Plasmids pACPJ23 and pACPtrc, with different promoter regions, were generated from pACYCDuet-1.2. Briefly, pACYCDuet-1.2, other than the T7 protomer, was amplified by PCR with appropriate primers. The synthetic sequences containing the P_{J23110} (BBa_J23110, http://part.igem.org/) and P_{trc} (Brosius et al., 1985) promoters were used to circularize the amplicon using In-Fusion HD Cloning Kit (Clonetech, Saint-Germain-en-Laye), thus generating

pACPJ23 and pACPtrc. Aromatic amino acid hydroxylase from *Cupriavidus taiwanensis* (W192F) was subcloned into the newly generated plasmids using the conventional digestion and ligation methods. Tryptophan decarboxylase gene plus vector was amplified from pCOLADuet-$GST\Delta_3 7T_5H$+*TDC* (Park et al., 2011) and the P_{J23110} promoter was added by In-Fusion, creating pCOLAPJ23-*TDC*. TDC was also inserted into the pBAD vector following the same procedures described above. The GFP protein associated with the intracellular tryptophan sensor (pSentrp developed by Fang et al., 2016) was modified by the addition of a short peptide sequence (-LVA) to the C-terminal end of the protein (Andersen et al., 1998). All plasmids with the suffix ".2" denote plasmids in which the *lacI* genes have been deleted.

3.3.2 Colony PCR

Colony PCR was used to determine the presence of the desired DNA fragments within the plasmids or the genome. Any colony which gives a PCR result of expected size would be further analyzed by DNA sequencing to confirm the integrity of the DNA construction. Colony PCR was executed as follows: 1X Maxima Hot Start Green PCR Master Mix, 0.3 μM of each primer and water up to a final volume of 15 μL were mixed. Single colonies were diluted in 2 μL of LB media. Afterward, 1 μL was added to the PCR, and 1 μL was transferred to a fresh LB-agar plate with proper antibiotics as a backup. PCR reaction was carried out following the manufacturer's recommendation taking into account the size of the expected fragment and the melting temperature of the primers.

3.3.3 Primers

Primers used for this work are listed in Table 3.3. Primers are categorized according to their propose into plasmid design, colony PCR, homologous recombination (Hom Rec), directed site mutagenesis or smart library. All primers were designed using SnapGene^{TM} 1.1.3v, and molecules were synthesized by Thermo Fischer Scientific (Darmstadt, Germany).

Table 3: Primers used in this study

Primers	**Sequence (5‘ to 3‘)**	**Purpose**
lacI_for	ctcgcatgcactctgcgacatcgtataacgttact ggttt	Plasmid design
lacI_for	tatgcatgcgattcattaatgcagctggcacgaca	Plasmid design
CtAAAH_for	atgggcagcagccatca	Plasmid design
P-J23110_rev	ctagtaggtttcctgtgtgactctagtagctagcatt gtacctaggactgagctagccgtaaaatttcctaat gcaggagtcgcataag	Plasmid design
P-J23110-operator(lac)-rev	ggaattgtgagcggataacaattcccctgtagaaa taattttgtttaactttaataaggagatataccatg ggcagcagccatca	Plasmid design
P-trc_rev	atgaaggattatgtacatcaccatcatcaccatag ccaggat	Plasmid design
P-trc-operator(lac)rev	atgattaattgtcaaatttcctaatgcaggagtcg cataag	Plasmid design
trpR_for	cgctacaccagcggtaaggagatcc	Colony PCR
trpR_rev	aggctgttagatccacgcgtaatcgtc	Colony PCR
Cofactor1_for	atatgctatcgtactctttagcgagtacaaccgggg gaggcattttgctttgtgtaggctggagctgcttc	Hom Rec
Cofactor2_for	gctaattcccatacggttctggcaaatattctgaaa tgag	Hom Rec
Cofactor1_rev	ttgccagaaccgtatgggaattagccatggtccat atgaa	Hom Rec
Cofactor2_rev	attacgggtatttgtaggacggataaggcgttcacg ccgcatccggcattcggtgcacgactagtagagag cgttcaccgacaaac	Hom Rec
araB_for	cgccacttcacgcatgttatcg	Colony PCR
araB_rev	cggtcggcagacaaattctcg	Colony PCR
araB-KO_for	ttatagagtcgcaacggcctgggcagcctgtgccgg ggcggaagttggaagtgtaggctggagctgcttc	Hom Rec
araB-KO_for	atggcgattgcaattggcctcgattttggcagtgatt ctgtgcgagctttatgggaattagccatggtcc	Hom Rec
Y048F_for	caccagcatcgaccacgc	Direct Site Mut
Y048F_rev	aagcggtgcaccggct	Direct Site Mut
L109Y_for	gtgccggacgaggtcttctt	Direct Site Mut
L109Y_rev	atacccgggcacggcca	Direct Site Mut
L113Y_for	ggccgtgcccgggtatgtgccggacgagg	Direct Site Mut
L113Y_rev	cctcgtccggcacatacccgggcacggcc	Direct Site Mut

W192F_for	ctgtcgcgcctgtacttctacacggtcgagtttg	Direct Site Mut
W192F_for	caaactcgaccgtgtagaagtacaggcgcgacag	Direct Site Mut
Y222H_for	ggatcgattccccctggctc	Direct Site Mut
Y222H_rev	acagcctggactcggcc	Direct Site Mut
L224C_for	ggaatcgatctacagctgcgactcggccag	Direct Site Mut
L224C_rev	tgggctggccgagtcgcagctgtagatcgattcc	Direct Site Mut
tio-Vector-Bottom	t*a*c*c*a*g*t*a*c*g*g*cgcgacagcatgt	Smart library
tio-Vector-Top	g*c*c*t*g*g*a*c*t*c*g*gccagcc	Smart library
Top Linker Fragment	cacggtcgagtttggcctgatccgcacggagcagg gactgcgcatctacggcgccggcatcgtctc-gagccaggggaatcgatctacagcctggactcg	Smart library
Fragment Phe197- AHN	gcgcagtccctgctccgtgcggatcaggccahn ctcgaccgtgtaccagtacagg	Smart library
Fragment Phe197- CDB	gcgcagtccctgctccgtgcggatcaggcccdb ctcgaccgtgtaccagtacagg	Smart library
Fragment Phe197- CCA	gcgcagtccctgctccgtgcggatcaggcccca ctcgaccgtgtaccagtacagg	Smart library
Fragment Glu219- AHN	tgtagatcgaahnccctggctcgagacgatgcc ggcgccgtagat	Smart library
Fragment Glu219- CDB	tgtagatcgacdbcccctggctcgagacgatgcc ggcgccgtagat	Smart library
Fragment Glu219- CCA	tgtagatcgaccacccctggctcgagacgatgcc ggcgccgtagat	Smart library

n* correspond to nucleotides in which sulfur substitutes for one of the oxygen in the phosphodiester bonds between the nucleotides.

3.4 Semi-rational creation of library and screening of tryptophan consumers with a fluorescent sensor

Two independent protein libraries were created using a site-saturated mutagenesis approach in the residues Phe197 and Glu219. These residues were selected due that they are in the binding pocket of the enzyme and play an important role defining the shape and volume near the aromatic ring of the substrate. The libraries were generated according to the following procedure: DNA with part of the *CtAAAH-W192F* gene and the plasmid pACYCDuet-

1.2 were amplified with phosphorothioate primers using Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific, Darmstadt) following the manufacturer's recommendations. The fragment was treated with *Dpn*I enzyme and further diluted to 0.01 pmol/μL. Cleavage was done according to Blanusa et al. (2010): 8.4 μL of PCR, 1 μL of buffer (0.5 M Tris-HCl, pH 9.0), 0.4 μL of iodine solution (100 mM iodine in ethanol), and 5 min incubation at 70 °C. In parallel, four oligonucleotides degenerated in the selected residues (Kille et al., 2013) were synthesized (one plus strand and one negative strand per site) and phosphorylated with a T_4 Polynucleotide Kinase (ThermoFisher Scientific, Darmstadt). All oligos were designed in such a way that these would have sticky ends complementary to the cleavage fragment described above. DNA hybridization was achieved by mixing cleaved vector with 1 μL of the synthetic oligos (2.5 pmol/μL), to keep a ratio of 1:3 (vector:insert). After 5 min at RT, the DNA complex was directly transformed into electrocompetent BL21(DE3) *ΔtnaA* cells generated by conventional method described in Section 3.3.1. Cells were plated in LB-agar and grew overnight. Afterward, individual colonies were transferred to M9-agar plates supplied with 1 mM tryptophan and proper antibiotic. Colonies showing no fluorescence were transferred to M9-agar plates with 2 mM tryptophan. Successive 1 mM tryptophan steps were made until there was no evident change compared with the previous agar plate (Fig. 9).

Cells with less fluorescence were selected for DNA sequencing, and best mutations were combined using QuickChange site-directed mutagenesis kit (Agilent, Santa Clara). Protein structures were predicted and analyzed described in as described in Section 3.2. Protein activity was evaluated according to Kino et al. (2009) (see Section 3.7), and kinetic parameters were calculated using a Lineweaver-Burk plot.

3.5 Growth and fermentation media

LB medium

Lysogenic broth (LB medium) was routinely used for *E. coli* cultivations. LB medium contained (per liter): 10 g tryptone, 5 g yeast extract and 10 g NaCl. LB solid plate was prepared by the addition of 15 g/L agar. If necessary, pH was adjusted to 7.0 with 5 M NaOH. Sterilization was performed at 121 °C for 20 min.

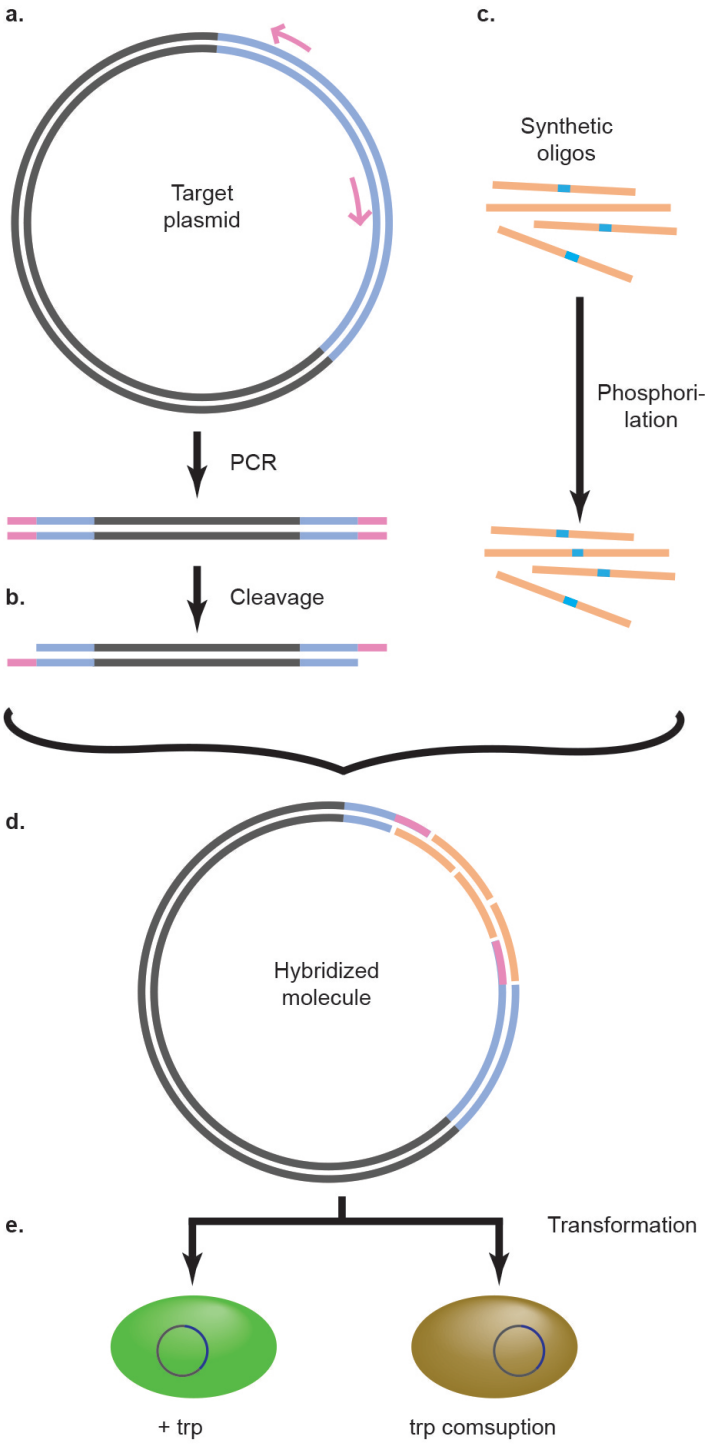

Figure 9: Library construction and screening strategies. Plasmid and part of the *Ct*AAAH-F gene were amplified with phosphorothioate primers and cleavage leaving sticky ends. The amplicon was hybridized with synthetic phosphorylated oligos and transformed into an *E. coli* strain which contains an intracellular tryptophan fluorescent sensor. Tryptophan consumers can be distinguished by the lack or diminished formation of fluorescence.

When needed, appropriate antibiotics were added to the medium before usage at the following concentrations: ampicillin 100 mg/L, chloramphenicol 34 mg/L, and kanamycin 30 mg/L. When the media was prepared with two antibiotics, the concentration was reduced by half.

SOC medium

SOC (Super Optimal Broth with Catabolite repression) medium was used for *E. coli* regeneration after heat shock transformation or electroporation. SOC medium contains (per liter): 20 g tryptone, 5 g yeast extract, 2.47 g/L $MgSO_4$, 0.584 g NaCl and 0.186 g KCl. All components rather than $MgSO_4$ and glucose were dissolved and sterilized at 121°C for 20 min. Glucose was autoclaved separately, and $MgSO_4$ was sterilized by filtration (0,22 μm). All components were mixed, and volume was adjusted to 1.0 L.

Seed medium

Seed medium contained (per liter): glucose (30 g), $MgSO_4$ $7H_2O$ (0.5 g), KH_2PO_4 (2 g), $(NH_4)_2SO_4$ (4 g), yeast extract (5 g), monosodium citrate dihydrate (2 g), biotin (0.1 mg), DL-calcium pantothenate (0.5 mg), ascorbic acid (176 mg) and 10 mL of 100x stock trace elements. The stock trace elements solution was composed of (per liter in 0.1 N HCl): Na_2MoO_4 $2H_2O$ (2.5 g), $AlCl_3$ $6H_2O$ (2.5 g), $FeSO_4$ $7H_2O$ (10 g), $CoCl_2$ $6H_2O$ (1.75 g), $CaCl_2$ $2H_2O$ (10 g), $ZnSO_4$ $7H_2O$ (0.5 g), $CuCl_2$ $2H_2O$ (0.25 g), H_3BO_3 (0.125 g), and Na_2MoO_4 $2H_2O$ (0.5 g). The pH was adjusted at 6.7 using a 5M KOH solution. When needed, appropriate antibiotics were added to the medium before usage at the following concentrations: ampicillin 100 mg/L, chloramphenicol 34 mg/L, and kanamycin 30 mg/L. When the media was prepared with two antibiotics, the concentration was reduced by half.

Fermentation medium

Fermentation medium was the same as Seed Medium, except for 1 g/L of yeast extract instead of 5 g/L, and if needed 1 g/L of tryptophan. The pH was adjusted at 6.7 using a 5 M KOH solution. When needed, appropriate antibiotics were added to the medium before usage at the following concentrations: ampicillin 100 mg/L, chloramphenicol 34 mg/L, and kanamycin 30 mg/L. When the media was prepared with two antibiotics, the concentration was reduced by half.

Feed medium

Feed medium was the same as fermentation medium, except for 600 g/L of glucose instead of 12 g/L, and if needed 1 g/L of tryptophan. The pH was adjusted at 6.7 using a 5 M KOH solution. When needed, appropriate antibiotics were added to the medium before usage at the following concentrations: ampicillin 100 mg/L, chloramphenicol 34 mg/L, and kanamycin 30 mg/L. When the media was prepared with two antibiotics, the concentration was reduced by half.

M9 mineral medium plates with trace elements and vitamins

The M9 medium was used for screening purposes. Medium contained: 33.7 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 8.55 mM NaCl, 9.35 Mm NH_4Cl, 4% glucose, 1 mM $MgSO_4$, 0.3 mM $CaCl_2$, 1 μg/L biotin, 1 μg/L thiamine and 1 X of trace elements. All salts, other than $MgSO_4$ and $CaCl_2$, were autoclaved together. Glucose, magnesium, and calcium solutions were autoclaved separately. Vitamins and trace elements were sterilized by filtration through a 0.22 μm filter. 100X traces elements are described in the "seed medium" section. 15 g/L of agar was added to the medium when preparing plates.

3.6 *In vivo* enzyme assays

The *in vivo* assays were carried out to evaluate the activity of *Ct*AAAH (wild types and variants) towards phenylalanine and tryptophan. *E. coli* BL21(DE3) Δ*tnaA* harboring different plasmids were cultivated in LB medium containing 50 mg/L ampicillin. Isopropyl-ß-D-thiogalactopyranoside (IPTG) to a final concentration of 0.4 mM was added to the culture when OD_{600} reached 0.4. Cells were then grown at 30 °C with constant agitation at 220 rpm. After a 16 hour induction period, cells were harvested by centrifugation at 3 300 x *g* and washed twice with 100 mM HEPES-NaOH buffer (pH 8.0).

5HTP and tyrosine were produced in a reaction mixture containing 100 mM HEPES-NaOH buffer (pH 8.0), 4 mM L-tryptophan or L-phenylalanine, 4 mM BH_4, 50 mM D-glucose, 0.1 mM $FeSO_4$, 1% (v/v) Triton X-100 and *E. coli* whole cells (OD_{600} between 15-20) in a total volume of 500 μL. The reaction was performed for 30 min at 30 °C. After the reaction, cells were removed by

centrifugation at 12 000 x *g* for 10 min at 4 °C, the supernatant was filtrated and stored at 4 °C before HPLC analysis.

3.7 *In vitro* enzymatic characterization

3.7.1 Enzyme overexpression in *E. coli* and purification

Enzymes were overexpressed in BL21(DE3) cells (New England Biolabs, Frankfurt am Main, Germany) using plasmids derived from the pDUET collection (Merck Chemicals GmbH, Darmstadt, Germany). Recombinant cells were grown overnight in LB media supplied with proper antibiotics at 37 °C with agitation. The next day cells were transferred to 30 mL of fresh media, the culture was diluted 1/100 and kept at 30 °C with constant agitation until it reached an OD_{600} of 0.5 - 0.7. Gene expression was induced with IPTG with a final concentration of 0.4 mM. Cells were kept at 30 °C with agitation for 18-20 h. Afterward, cells were harvest and washed twice in Binding Buffer (20 mM sodium phosphate, 0.5 M NaCl, 20 mM imidazole, pH 7.4). Final resuspension was carried out in $1/10^{th}$ of the original volume. Cells were disrupted in 2 mL tubes using a cell homogenizer with lysine beads – matrix B (MP Biomedicals Germany), 4 cycles of 40 sec at 6 m/s were used to disrupt the cells, samples were kept on ice for 5 min in-between each cycle. Cell debris was eliminated by centrifugation at 14 000 x *g* for 20 min.

Proteins were purified using Ni^{2+} -NTA columns (GE Healthcare Bio-Science, New Jersey, USA) following the manufacturer recommendations. Centrifugal filter units with a cut-off of 10 kDa were used to reduce the imidazole concentration. Pure proteins were resuspended in 20 mM phosphate buffer (pH 7.4) containing 300 mM NaCl. Finally, purity of the enzymes was checked by SDS-PAGE following conventional protocols (Sambrook and Russell, 2000), and proteins were quantified using the Bradford assay (Bradford, 1976) with commercially available dye reagents (Bio-Rad, Herculas, USA).

3.7.2 Enzyme assay

Protein hydroxylase activity was determined according to Kino et al., (2009). Briefly, activity was measured at 30 °C in a final volume of 500 μL, containing variable concentrations of L-tryptophan (1-10 mM) or L-phenylalanine (0.05-

1 mM), DMPH4 (1 mM), 500 U catalase, 0.1 mM $FeSO_4$, 5 mM dithiothreitol and 50 mM HEPES-NaOH buffer (pH 7.5).

The steady-state kinetic parameters were obtained by measuring the activity with various concentrations of L-tryptophan (1-10 mM), L-phenylalanine (0.05-1 mM) and DMPH4 (1 mM). The reaction was initiated by adding the enzyme and performed at 30 °C for 30 min. Afterwards, the reaction was terminated by the addition of 50% (v/v) methanol. 5HTP or L-tyrosine were measured using HPLC.

3.8 Production of 5HTP and serotonin by fermentation

Batch and fed-batch fermentations were done according to Chen and Zeng (2016), with some modifications. Briefly, cells were cultured overnight in LB medium at 37 °C. The preculture was inoculated into 10 mL of seed medium with an initial OD_{600} = 0.1. The seed culture grew for 10 h at 30°C. Afterward, it was inoculated in 50 mL of fermentation medium to an initial OD_{600} = 0.1 and grown at 30 °C. When batch fermentations were done in shaking flasks, 30 g/L of $CaCO_3$ were added to the fermentation medium as a pH controller. Fed-batch fermentations were carried at 30 °C in 1.5 L jar fermenters (DASGIP, Jülich) with an initial volume of 500 mL. The pH was maintained at 6.7 by automatic addition of 25% NH_4OH and 3 M H_3PO_4. The dissolved oxygen was set at 30 % of air saturation varying in subsequent order the agitation speed, the oxygen content of the gas inlet and aeration rate. Glucose concentration was controlled by supplying a feeding solution with 60 % glucose during the fermentation with a flexible feed rate.

If necessary, the culture was harvested after fermentation. Cells were separated by centrifugation and subsequent filtration. The supernatant was stored at -80 °C until further use.

3.9 Conversion of 5HTP produced by fermentation to serotonin

To produce serotonin, an overnight culture of BL21(DE3) *ΔtnaA* strain harboring the plasmid pCOLADUET-*TCD* was inoculated in LB media. Precul-

ture and seed culture were prepared as described above. Based on growing experiments, fermentation media and 5HTP-containing supernatant were mixed (4:1) and inoculated with the seed culture. Temperature (30 °C), pH 6.7 and dissolved oxygen (30 %) were controlled and kept constant during the fermentation. Induction of tryptophan decarboxylase was carried out by the addition of 0.1 mM IPTG when OD_{600} was 10. A similar procedure was used when *tdc* expression was induced with arabinose using plasmid pBAD-TDC.

3.10 Analytical methods

For fast tryptophan determinations, a spectrophotometric method (Nagaraja et al., 2003) was used. For each reaction, 120 μL of 2% PPDD (p-phenylenediamine, PPDD was dissolved in a 5 % v/v HCl solution), 80 μL of 0.5 % $NaNO_2$, and 40 μL of 3 % H_3NSO_3 were mixed. Afterward, 100 μL of the samples were transferred to the mixture and 420 μL of 1:1 water:H_2SO_4 were added. The whole setup was carried out in ice, and a mixture step was done after every addition. After 30 min, the samples were measured at 520 nm and compared against a standard curve.

5HTP was quantified with a modified Gibbs assay (Quintana et al., 1997). 100 μL of 500 mM borate-NaOH buffer (pH 9) was mixed with 100 μL of the supernatant, followed by the addition of 4 μL of 0.5 % (w/v) 2,6-dichloroquinone-4-chloroimide (Gibbs‘ reagent) in ethanol solution. After incubation at RT for 30 min, absorbance at 580 was measured for color production.

Aromatic amino acids, tryptamine, 5HTP, and serotonin were analyzed using HPLC with a modified protocol described by da Luz et al. (2014). Proteins were precipitated by conventional TCA method. Afterward, they were filtrated. Measurements were done on an Ultimate-3000 HPLC system (ThermoFisher, Darmstadt) with a binary gradient where eluent “A” was 140 mM sodium acetate with 0.1% v/v ACN and eluent “B” was 60% v/v ACN. The gradient was as follows: at 0 min 0% B, at 1 min 3% B, at 25 min 8% B, at 60 min 31.5% B, and remained constant from until 70 min at 100% B. All gradient changes were linear between the points given above. Separation was done using a Kinetex RP column (2.6 μm, C18, 100 x 4.6 mm, Phenomenex, Aschaffenburg) at 45 °C, and an injection volume of 10 μL was used.

4 Design of a serotonin synthetic pathway and proof of the concept

In this section, two possible synthetic routes for the production of serotonin are explored. Pros and cons of each pathway are compared. Results of bioconversion assays using key enzymes and intermediary molecules (tryptamine and 5HTP) are presented, as well as computational analysis of the selected pathway. Finally, a synthetic serotonin pathway from tryptophan in *E. coli* is defined.

4.1 Introduction

Currently, most fine chemicals or raw materials needed in industries are synthesized from limited natural- or nonrenewable- resources. These could be produced from simple, renewable, inexpensive starting materials using metabolic engineering and synthetic biology (Keasling, 2010). For many years, natural pathways have been the blueprints for the production of important chemicals via fermentation. Regulations at different levels often limit the production and protein engineering strategies, enzyme overexpression or the reconstitution of the pathways in heterologous hosts are some of many approaches used to circumvent this problem. When the natural pathway does not exist, or limitations at different levels prevent the production of the metabolite, it is necessary to go beyond by exploring the chemistry and synthetic capabilities of the biological systems. The *de novo* design of synthetic biological pathways by the novel combination of enzymes can overcome this problem (Zhang et al., 2008).

The design and assembly of a synthetic pathway are often limited by the identification of the enzymes capable of catalyzing the reactions of interest. Restricting the knowledge to reported enzymatic reaction limit the possibilities of the novel pathways, putative reaction developed by protein engineering should be considered to broaden the catalytic scope of the pathway. Enzyme promiscuity could facilitate the assembling of new pathways. Just as evolution recruits existing enzymes to improve or expand biological pathways, synthetic biology, and protein engineering can be used to identify enzymes with the potential to catalyze the desired reaction and modify them for specific purposes.

Strategies that incorporate directed evolution using a semi-rational and *in silico* approach reduce the size of the libraries that should be analyzed. It is possible to select and identify enzymes able to catalyze the desired reaction when these are coupled with medium- or high-throughput screening methods (Section 5 and 6) (Chou and Keasling, 2012).

After defining a novel pathway, and before it can build in the laboratory and integrated into a microorganism, it should be first evaluated. Although intuition and manual design can assist in the initial steps of the proposed path, these are not sufficient to guarantee the selection a feasible biotransformation in a given cell (Carbonell et al., 2014). Computational prediction tools and whole-cell bioconversion assays, not only facilitate the generation of the novel hypothesis, but also the screening the most efficient pathways (Hadadi and Hatzimanikatis, 2015).

Here, exploratory results for the conceptualization of a pathway that converts tryptophan into serotonin are presented. The state of the art of known reaction, the expected protein engineering work, the efficiency of whole-cells bioconversion assays and *in silico* analysis of the pathway were taken into account for the design of a novel pathway that expands tryptophan metabolism and leads to the production of serotonin.

4.2 Results and discussion

4.2.1 Conceptual design of the pathway for the production of 5HTP and serotonin

Two possible routes lead to the production of serotonin from tryptophan: tryptophan can be hydroxylated into 5HTP and further converted into serotonin by decarboxylation. Alternatively, if tryptophan is first decarboxylated into tryptamine, this can be then converted into serotonin (Fig. 10). In both cases, the decarboxylation step can be performed by an aromatic amino acid decarboxylase (AADC). AADCs are present in animals, insects, and plants. In contrast to animal and insects AADC, which accept a broad range of aromatic amino acids. Plant AADC exhibits a narrow substrate specificity. Tryptophan decarboxylase from *Catharanthus roseus* (*Cr*TDC), can only accept tryptophan and 5HTP as natural substrates, and consequently produce tryptamine or serotonin (Noé et al., 1984).

Hydroxylation of tryptamine in *E. coli* was first shown by Park et al. (2011) using tryptamine 5-hydroxylase from rice (*Os*T5H), they reported a maximum production of 24 mg/L of serotonin when the media was supplied with 2 mM (408 mg/L) of tryptophan. Serotonin concentration did not increase with the substrate concentration which suggests that either *Os*T5H catalytic activity is low in *E. coli* or that serotonin product could have an inhibitory effect on *Os*T5H enzyme activity. Later this same group reported an increase in the enzyme activity when *Os*T5H was coexpressed with its respective NADPH-cytochrome P450 reductase (*Os*CPR) (Park et al., 2013).

Efforts have also been made to engineer *E. coli* strains capable of converting tryptophan to 5HTP. Tryptophan 5-hydroxylase (TPH), which is only present in eukaryotes, is capable of synthesizing 5HTP. Mammalian TPH has been expressed in *E. coli*, but it has low activity and poor stability when expressed in prokaryotes (Wang et al., 2002). Prokaryotic phenylalanine hydroxylases

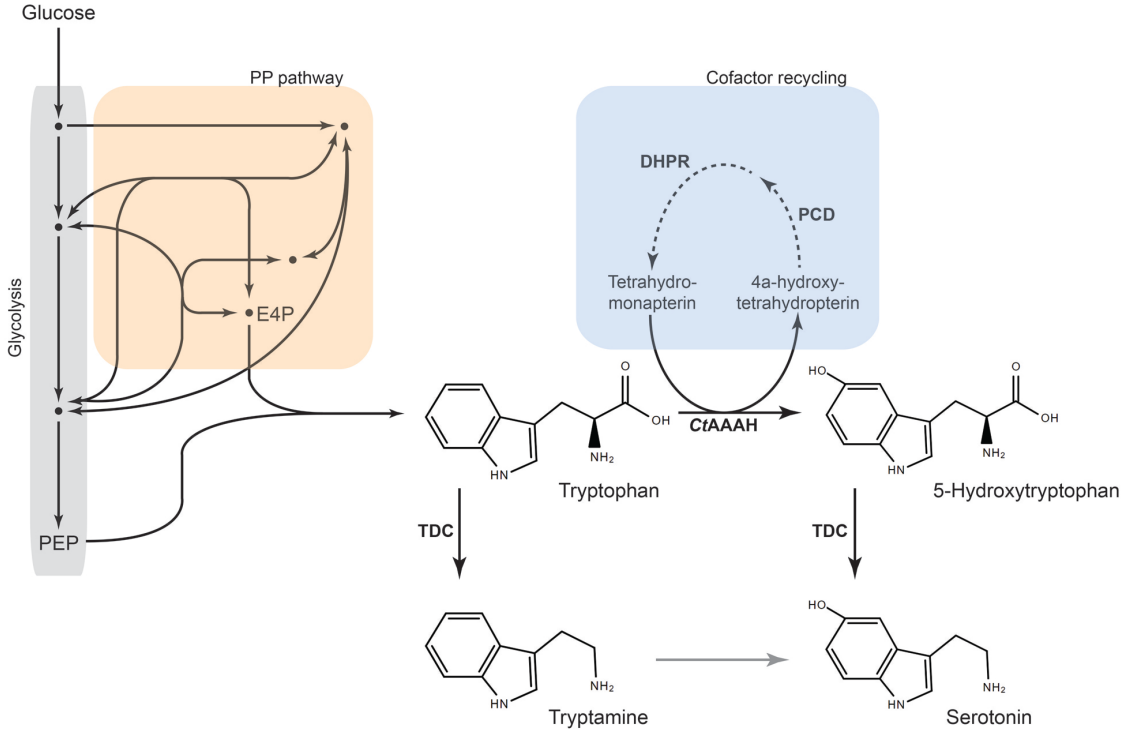

Figure 10: Novel artificial pathways for the biosynthesis of serotonin via tryptamine and 5-hydroxytryptophan in *E. coli*. *Ct*AAAH, aromatic amino acid hydroxylase from *Cupriavidus taiwanensis*; *Cr*TDC, tryptophan decarboxylase from *Catharanthus roseus*; *Os*T5H, tryptamine 5-hydroxylase from rice – *Oryza sativa*; *Os*CPR, NADPH-cytochrome P450 reductase from rice.

from different species have been engineered to change the substrate preference from phenylalanine to tryptophan (PAH) (Hara and Kino, 2013; Lin et al., 2014). Lin et al. (2014) screened out and engineered PAH from *Xanthomonas campestris,* and they were able to produce 152.9 mg/L of 5HTP from glucose.

4.2.2 Bioconversion of tryptophan for the production of serotonin production in *E. coli*: proof of the concept

The potential of tryptophan and 5HTP bioconversion by *Cr*TDC was explored. An *E. coli* culture harboring plasmid pCOLADuet-1-*TDC* was concentrated (OD_{600} 15-20) and supplied with 1 mM of respective substrates. As shown in (Fig. 11), the initial conversion rate was higher when tryptophan is used as a substrate. Nevertheless, the production efficiency was similar after 3 h of incubation at 30 °C. After 5 h incubation, there was a 93 % and 89 % bioconversion of tryptophan and 5HTP into tryptamine and serotonin, respectively. A higher affinity of *Cr*TDC toward tryptophan (K_m 0.75 mM), when compared with 5HTP (Km 1.3 mM), was previously reported, in this same study, both, tryptophan and tryptamine, were found to be inhibitors of the decarboxylation reaction (tryptamine K_i 3.1 mM) (Noé et al., 1984). Hence, TDC inhibition with 5HTP have not been reported, an inhibition assay using crude extracts and different 5HTP concentrations were conducted. A slight decrease in decarboxylation activity was observed when the 5HTP concentration was 3 mM, and at 10 mM 50 % of the activity the protein activity was detected (Fig. 11). 5HTP has been reported to have strong inhibition in decarboxylase activity of human and pig aromatic L-amino acid decarboxylase where 5HTP is the natural substrate (Bertoldi et al., 2008; Verbeek et al., 2007). On the other hand, according to the Brenda Enzyme Database (http://www.brenda-enzymes.de) tryptamine inhibits plant related TDCs (*C. roseus* and *Phalaris aquatica*), in which tryptamine is an important building block for the formation of other secondary metabolites (alkaloids) (Facchini et al., 2000).

Aromatic amino acid hydroxylase from *Cuprividus taiwanensis* (*Ct*AAAH) and tryptamine 5-hydroxylase from rice - *Oryza sativa* (*Os*T5H) with its respective NADPH-cytochrome P450 reductase (*Os*CPR) were used to compare the hydroxylase activity of tryptophan and tryptamine. *In vivo* bioconversion assays were performed with cells (OD_{600} 15-20) carrying plasmids pCOLADuet-1-*GSTΔ37T5H-OsCPR2* or p*Ct*AAAH in defined media sup-

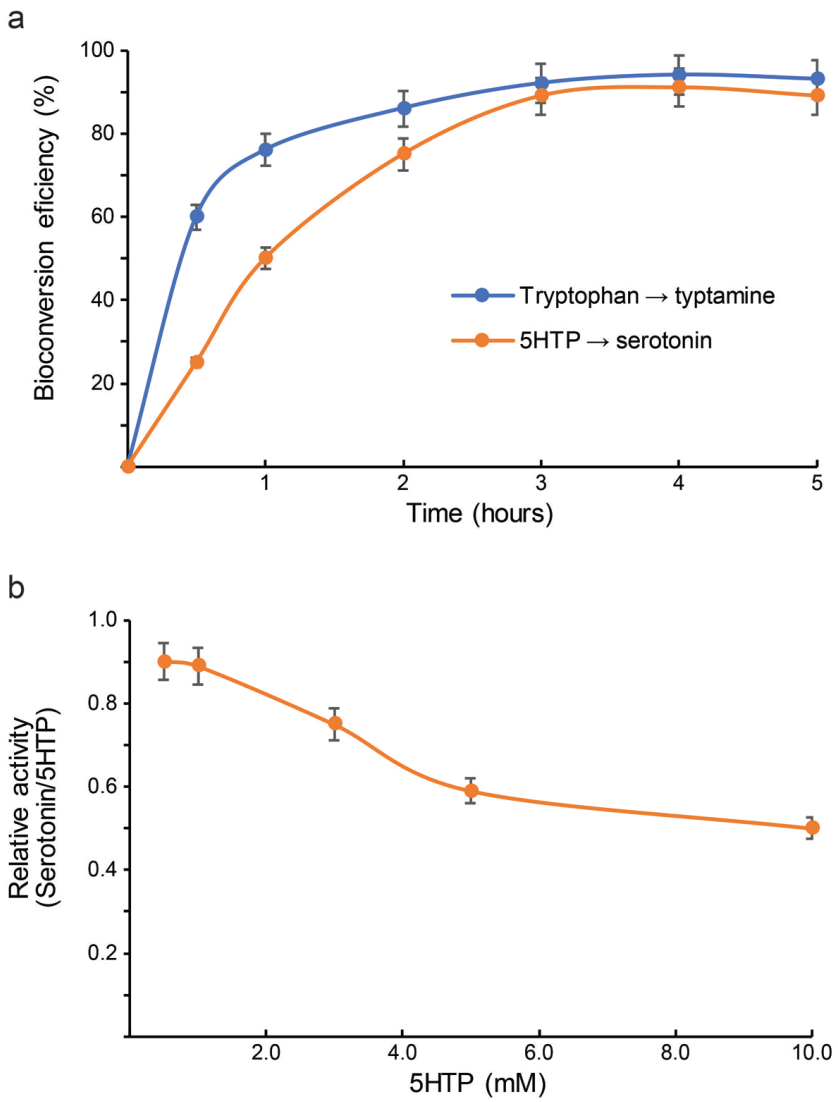

Figure 11: Tryptophan decarboxylase activity from *Catharanthus roseus* (*Cr*TDC) with tryptophan and 5HTP: a. Whole cell bioconversion efficiency of *Cr*TDC using *E. coli* (OD_{600} 15-20) in media supplied with 1 mM tryptophan or 5HTP as substrate; b. relative activity of crude extracts with *Cr*TDC protein with different 5HTP concentrations.

plied with 4 mM of tryptamine or tryptophan as substrates. Corresponding cofactors were also added to the media. After 5 h reaction around 10-12 % of the substrate was hydroxylated: 0.41 mM of 5HTP was detected in the supernatant of cells with plasmid p*Ct*AAAH, and 0.48 mM of serotonin was produced in cells supplied with tryptamine (Fig. 12). These results were expected and consistent with other values reported in the literature (Lin et al., 2014; Park et al., 2013). Despite a slightly lower conversion efficiency in cells

with plasmid p*Ct*AAAH, great potential lies in this enzyme due that to previous proteins engineering reports have been done with other AAAH (Kino et al., 2009; Lin et al., 2014). *In silico* analysis regarding this issue will be discussed in the next section.

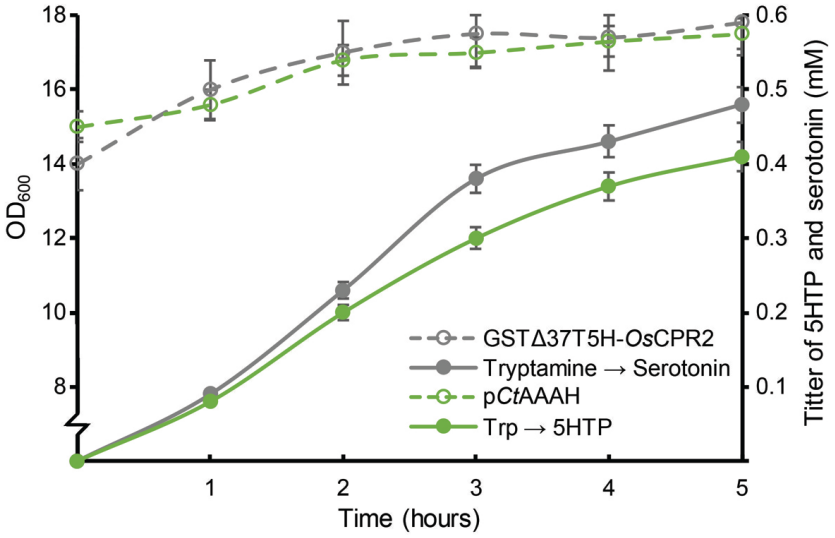

Figure 12: Whole cell bioconversion assay for the hydroxylation of tryptamine and tryptophan. 5HTP was produced using strain BL21(DE3)Δ*tnaA* harboring plasmid pCtAAAH, serotonin was produced with plasmid GSTΔ37T5H-OsCPR2. Cell growth is represented by dashed lines, product formation with solid lines.

4.2.3 *In silico* evaluation of the serotonin synthetic pathway

Computational tools can be used to evaluate predicted metabolic pathways. Proposed conversion of tryptophan into serotonin via 5HTP was evaluated *in silico* using RetroPath (Carbonell et al., 2014) in order to estimate the feasibility of the synthetic path and to determine possible bottlenecks during the production in *E. coli*.

The tool predicts four steps or reactions during the whole pathway. Gibbs‘ free energy of the system was estimated: ΔG= -26.43. This negative value suggests favorable thermodynamic conditions for the set of bioconversions. The *in*

silico analysis predicts the consumption of the endogenous cofactor tetrahydromonapterin (MH_4) by the hydroxylase enzyme. Experimental evidence related to MH_4 participating in hydroxylation reactions have been published by Lin et al. (2014) and Satoh et al. (2012). The computational analysis also predicted the consumption of a NADH during the regeneration of MH_4 and the production of one hydroxyl radical ($HO^{\cdot}$) molecule. Two hydrogen protons would be consumed during the production of serotonin, the first one is used during the regeneration of cofactor, the second in the decarboxylation step (Fig. 13).

4.3 Conclusions

In this section, two possible pathways were explored and compared for the production of serotonin in *E. coli*. Decarboxylation of tryptophan and 5HTP is possible with *Cr*TDC. However, it was previously reported that the intermediate, tryptamine, inhibits the decarboxylation reaction at 3.1 mM (Noé et al., 1984). Meanwhile, *Cr*TDC decarboxylation activity towards 5HTP maintains a 50 % of its original performance at a concentration of 10 mM 5HTP. The most challenging step for the conversion of tryptophan into serotonin is the hydroxylation since only 10 or 12 % of tryptophan or tryptamine was hydroxylated in exploratory results shown in this section.

The 5HTP pathway for the serotonin production was preferred over the tryptamine pathway. *Ct*AAAH hydroxylase activity and the potential protein engineering work that can be done on this enzyme in order to increase the preference toward tryptophan for the production of 5HTP, and the fact that *Cr*TDC can use the latter as a substrate. Further consideration should be taken into account when creating strains for 5HTP or serotonin production since the hydroxylation reaction consumes endogenous cofactor MH_4, and *E. coli* lacks a cofactor recycling pathway.

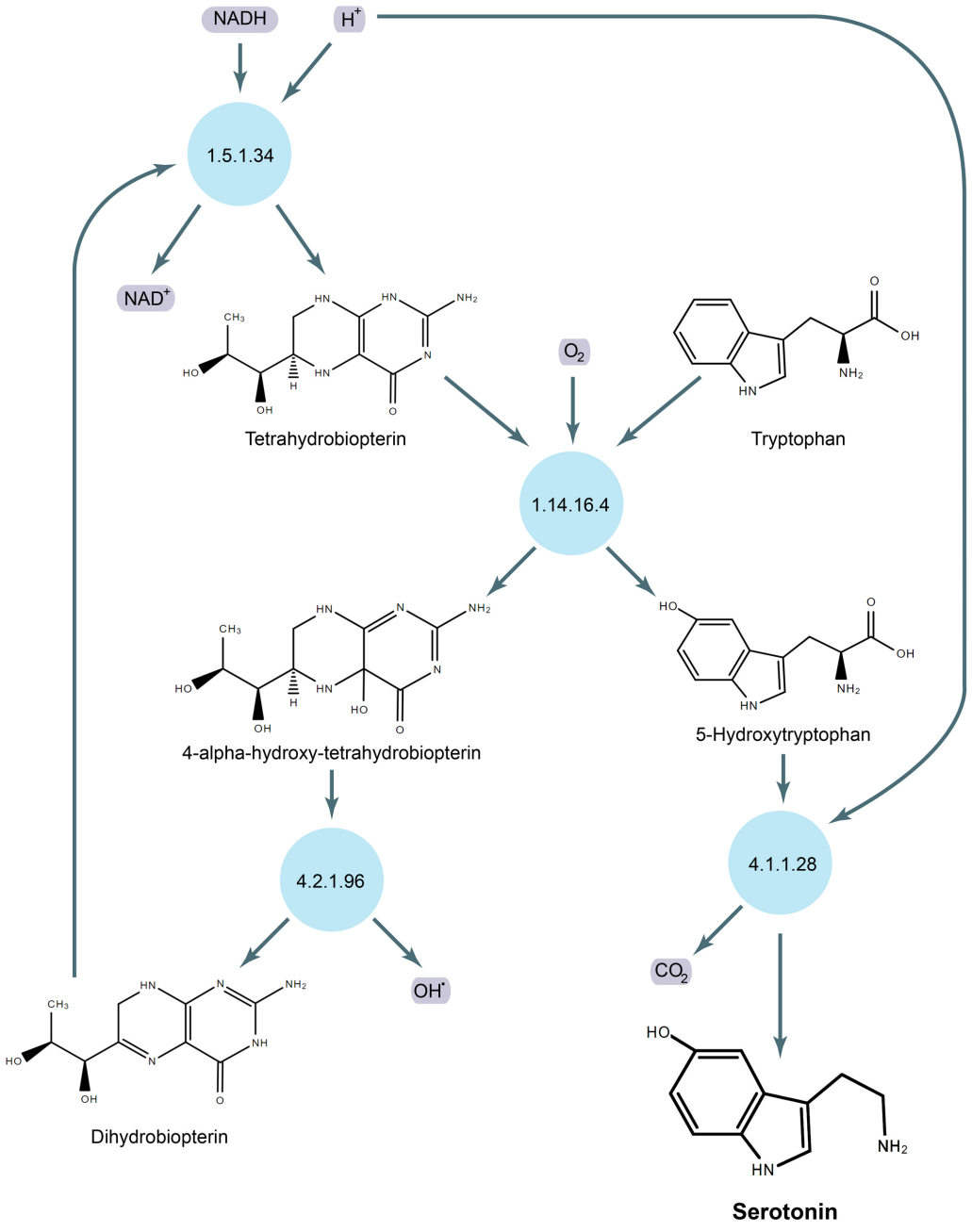

Figure 13: *In silico* evaluation of the synthetic serotonin pathway. Obtained from the RetroPath server (http://xtms.issb.genopole.fr/) (Carbonell et al., 2014). 1.5.1.34.: 6,7-dihydropteridine reductase; 1.14.16.4: tryptophan 5-monooxygenase; 4.2.1.96: 4a-hydroxy tetrahydrobiopterin dehydratase; 4.1.1.28: aromatic L-amino acid decarboxylase.

5 Protein rational design for tryptophan hydroxylation in *Escherichia coli*

Herein, phylogenetic relationships among aromatic amino acid hydroxylases (AAAH) are described, statistical differences within the sequences are evaluated to define conserved and divergent functional sites. Results from structural, modeling and docking analyses linked with substrate preference are presented. Point mutations in *Ct*AAAH were predicted to shift its substrate preference from phenylalanine into tryptophan. The results present in this chapter have been partially published (Mora-Villalobos and Zeng, 2017) as part of the requirements needed to complete the dissertation process at the TUHH.

5.1 Introduction

Protein engineering by directed evolution is a powerful technique often used to tailor enzymes with specific characteristics (e.g., improve or modify the protein activity, specificity, and stability, among others). A semi-rational approach can be used to design small- smart- libraries, for this purpose sequence, structure and functional analysis, as well as computational predictive algorithms are used to limit the amino acids diversity for protein engineering.

The first step of sequence-based analysis are MSAs and phylogenetic analyses, these provide information about conserved amino acids and the ancestral relationship among groups within homologous proteins. The comprehensive alignment of protein (super-) families leads to the identification of "hot spots" that can be related to structural or functional domains, within the functional domains conserved residues related to the activity and (enantio-) selectivity might be identified.

The correlation of specific homologous amino acid at a determined position with the enzyme function can help to point out "functional hot spots," especially if these events can be tracked down during the evolutionary history of proteins families. The variation (or conservation) of these residues gives important information that can be used to create focused and functional enriched enzyme libraries.

Critical residues can be tracked down in three-dimensional structures of the proteins. The identification of these amino acids is an important key to the design of smart-libraries since protein function is intimately linked to the structural topology. The growing number of structures available in the PDB and advances in homology modeling offer valuable assistance to locate key residues near the binding site and within the active pocket (Lutz, 2011). The described methods can also infer the stability of the protein-substrate complex, guiding the preparation of the library to a handful rational selected mutations that reduce the experimental analysis during a protein evolution evaluation.

In this section, protein engineering results for shifting the substrate preference of *Ct*AAAH from phenylalanine to tryptophan are presented. A smart-small-library was designed relying on the evolutionary, functional and structural information. Mutants were generated, and phenylalanine/tryptophan hydroxylase activity was evaluated in an *E. coli* strain which carried a cofactor regeneration pathway.

5.2 Results and discussion

5.2.1 Phylogenetic analysis of AAAH

A total of 508 AAAH protein sequences were collected after the database search and selection process. The results revealed the presence of AAAH homologs in different taxa. The mayor group corresponds to 284 eukaryotic sequences, from which 235 sequences belonged to the Metazoan taxon (Animal Kingdom). The other 49 sequences were distributed among the Fungi, Viridiplantae (mosses and green algae), and Protista groups. For the Prokaryote group, 224 sequences were retrieved from the database. These sequences were distributed among different taxa.

An NJ phylogenetic tree with the 508 sequences was generated from the amino acid sequences to explore the phylogenetic relationship among AAAH paralogous proteins. An ML tree was also constructed under the best-fit model, Jones-Taylor-Thornton (JTT), with discrete gamma distribution in three categories (gamma shape = 3,163). The tree topologies assessed by ML and NJ methods were substantially similar (Fig. 14).

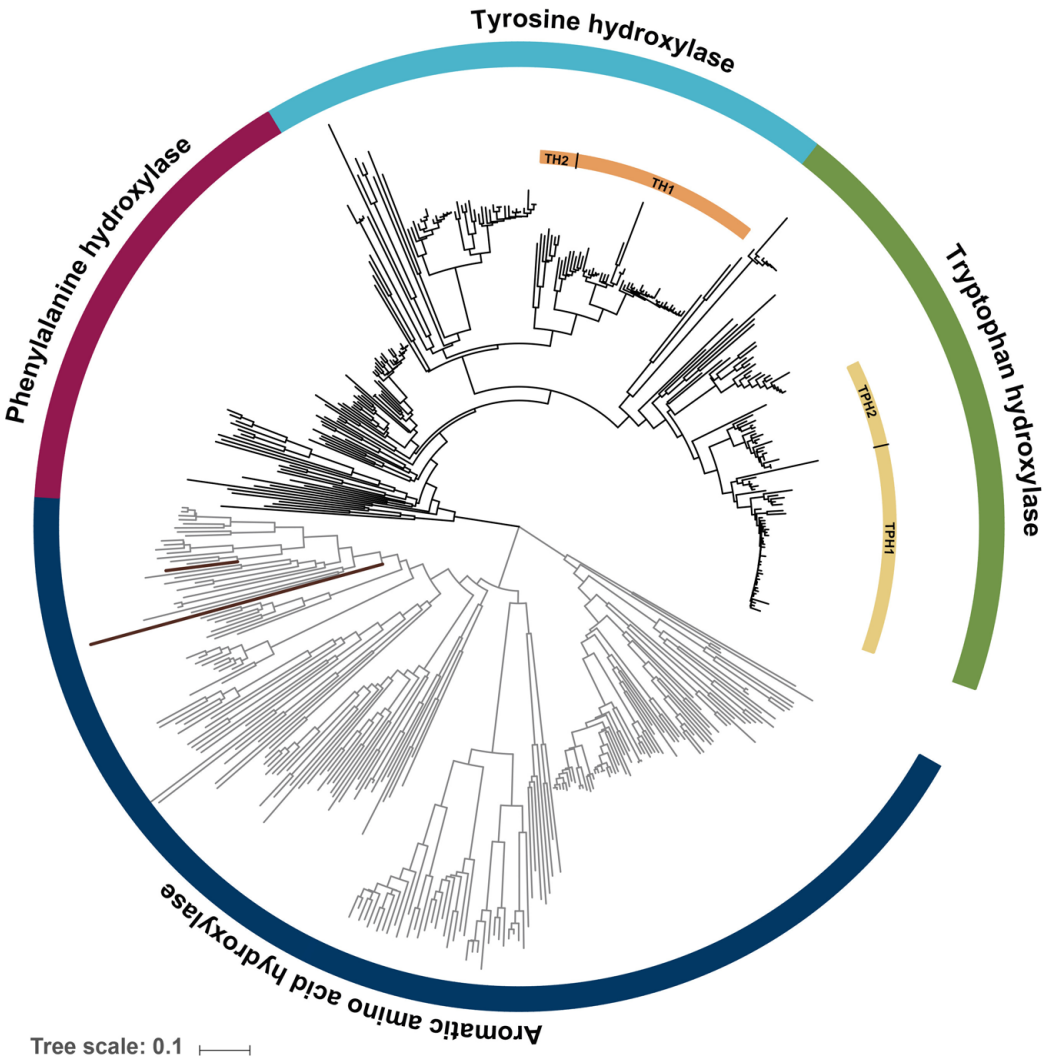

Figure 14: Phylogenetic tree of 508 aromatic amino acid hydroxylases. Branches of eukaryotes and prokaryotes are distinguishable by color, black and gray respectively. Stripes around the tree designate the type of hydroxylase which corresponds to each species. Duplication events are marked within each clade. The tree was inferred using the Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Evolutionary distances were computed using Poisson correction method. All positions with less than 2% site coverage were eliminated, a total of 495 positions were in the final dataset. The evolutionary analysis was conducted in MEGA9 (Kumar et al., 2016).

The phylogenetic analysis placed the prokaryotic AAAH at the base of the tree. As expected, our results indicate that eukaryotic AAAH derived from the prokaryotic AAAH. Later, the eukaryotic taxon divides into three distinct clusters: phenylalanine (PAH), tryptophan (TPH), and tyrosine (TH) hydroxylase. These groups were unambiguously separated from protists, fungi and lower plants (mosses/green algae), before the divergence of Choanoflagellates in eukaryotes.

Two major duplications had occurred in the TPH and TH subfamilies. The TPH duplication occurred in the vertebrate lineage, ultimately it led to the emergence of two lineages which evolved into TPH1 and TPH2. On the other hand, the duplication of the TH seems more complicated due to a duplication occurred before the divergence of jawed vertebrates, and TH2 was secondarily lost in the Eutheria taxa (placental mammals) (Fig. 14). The results are consistent with previous phylogenetic studies of AAAHs (Lin et al., 2014; McKinney et al., 2009).

5.2.2 Functional divergence analysis of PAH and TPH

Functional diverse analyses among prokaryotic and eukaryotic clusters appear to be of great interest, given that with these it is possible to identify potential sites that differ between the groups of enzymes that have phenylalanine or tryptophan preference. Such sites could be helpful in designing novel enzymes for the production of 5HTP. The FDI analysis for these clusters revealed significant θ_I value between the bacterial and eukaryotic AAAH, indicating that diversification has taken place during evolution (Table 4).

Three different groups of residues were identified in the analysis. (i) The first group represents sites that are conserved within all the clusters. Many of these amino acids are involved in the enzyme activity, e.g., conserved motifs and functionally important residues such as the 2-His-1-carboxylate facial triad which is involved in the coordination of the iron molecule in the active pocket of all hydroxylases (Erlandsen et al., 2002). (ii) A second class corresponds to the sites that are conserved within Eukaryotes and Prokaryotes taxa, but that are different between them. (iii) Finally, the third class is composed of the sites that seem to be related with specific substrate preference (Fig. 15).

The FDII analysis showed a low θ_{II} value of the TPH/PAH pair (Table 4). Since there is compiling evidence regarding of the substrate specificity of

these two groups, the FDII analysis confirms that the predicted substrate-determining residues between the latter group might have relevant functions specifically related to phenylalanine or tryptophan (activity, inhibition, activation, among others). To gain more insight into the possible function of these predicted FDII sites, the residues with higher posterior probability (Qk) values were mapped in the modeled *Ct*AAAH structure (Fig. 16).

Table 4: Functional divergence between eukaryotic TPH/PAH and prokaryotic AAAH paralogous proteins

Group 1	Group 2	Type-I		Type-II	
		θ_I[a]	Qk>0.9[b]	θ_{II}[a]	Qk>0.9[b]
AAAH	PAH	0.664 ± 0.090	37	0.374 ± 0.160	66
AAAH	TPH	0.747 ± 0.094	34	0.445 ± 0.161	78
TPH	PAH	0.201 ± 0.060	0	0.059 ± 0.095	15

[a] θ_I and θ_{II} correspond to the functional divergence Type-I and Type-II, respectively.
[b] Qk, posterior probability

By combining phylogenetic evidence with the FDI and FDII analyses, it was hypothesized that even after a long-term evolution process, prokaryotic AAAH and eukaryotic PAH still share some conserved amino acids that determine their substrate preference. Furthermore, the low divergence between animal PAH and TPH (and high sequence similarity) indicate that the substrate preference may involve only a small number of residues. Based on this hypothesis six residues were selected, and single point mutations were incorporated into the protein.

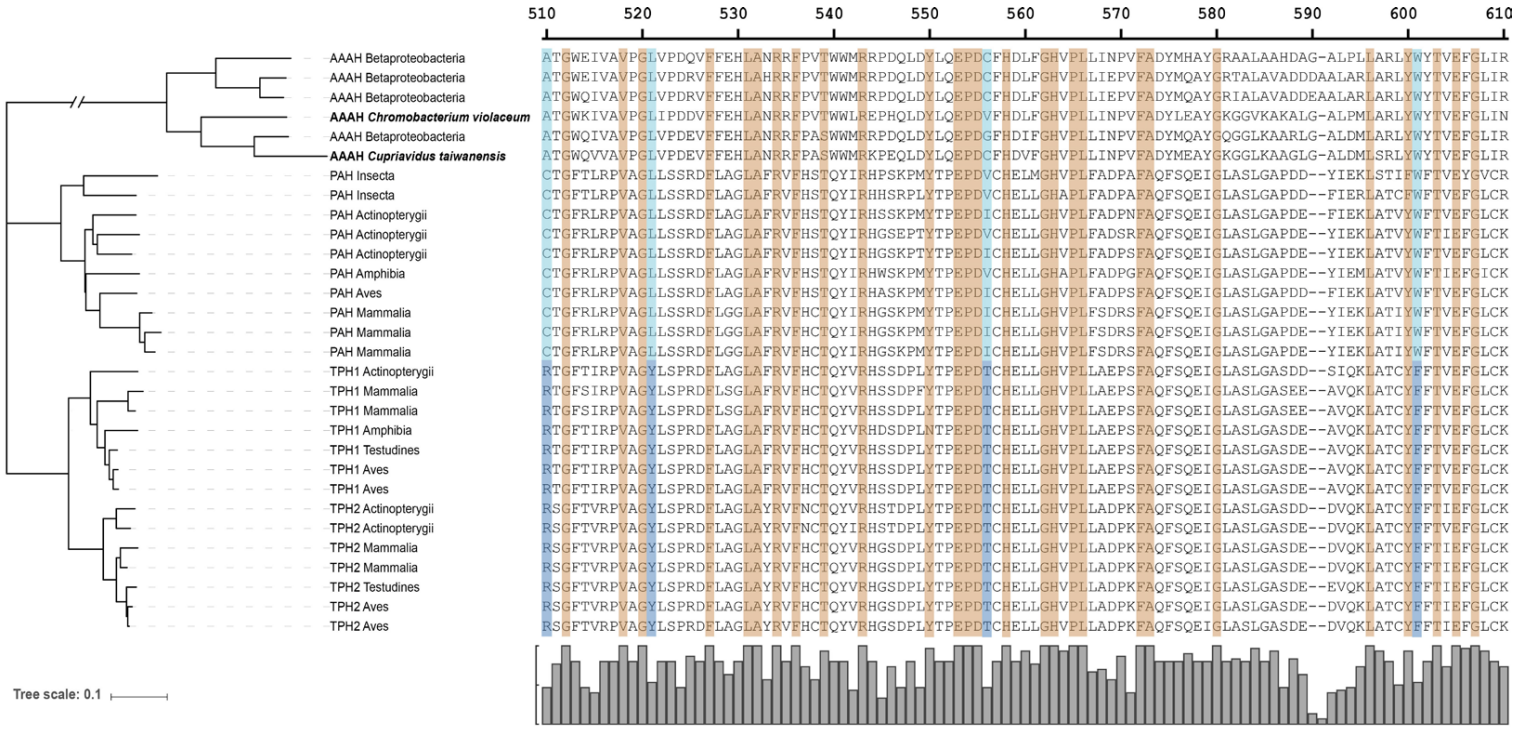

Figure 15: Protein tree and alignment for the AAAH superfamily: unrooted protein distance tree and alignment for 30 aromatic amino acid hydroxylases with phenylalanine (AAAH and PAH) or tryptophan (TPH1 and TPH2) affinity. Amino acids in the alignment are mainly present in the active pocket, amino acids colored in orange show conserve residues, amino acids highlighted with blue and light blue are strongly related to the substrate preference. A consensus graph is present below the alignment.

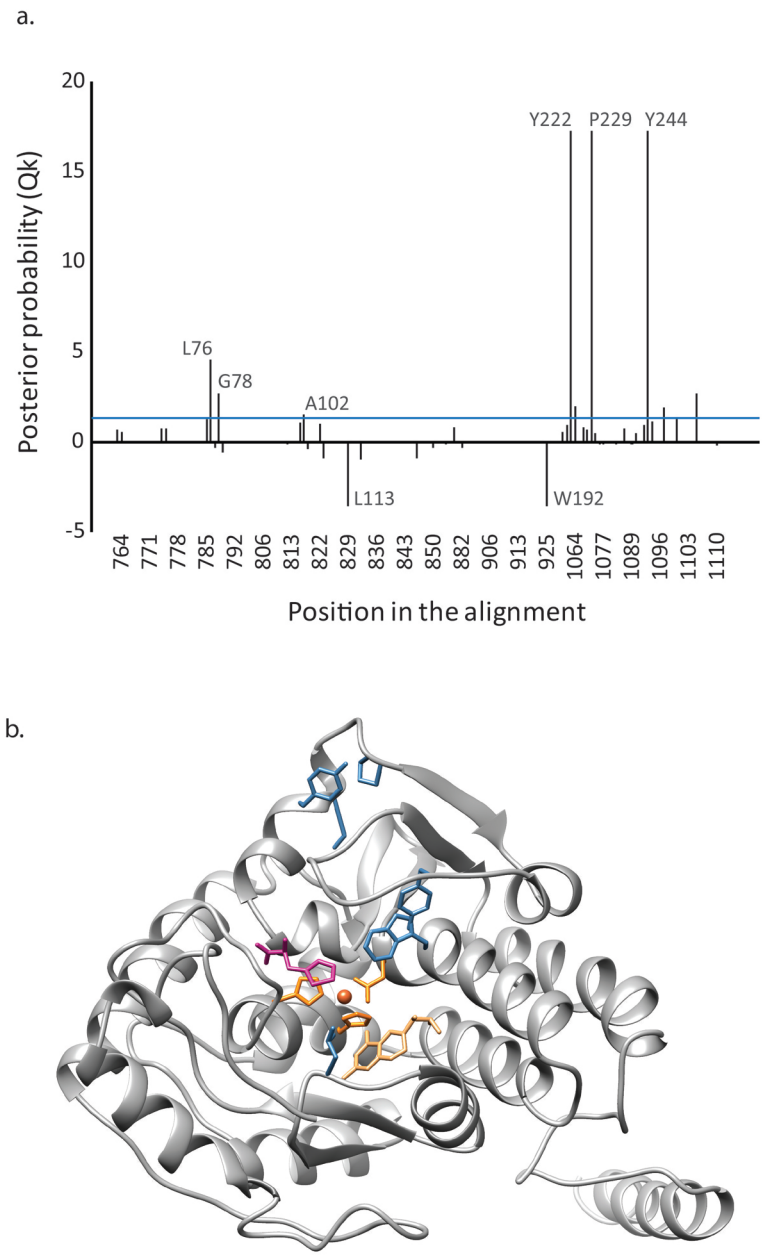

Figure 16: Prediction of substrate-determining residues based on functional divergence type-II analysis of tryptophan and phenylalanine hydroxylase, and their position in *Cupriavidus taiwanensis* modeled structure. (a) Posterior probability (Qk) of the homologous sites, a threshold line of Qk = 0.9 is indicated in the graph. (b) Predicted amino acids (blue) are mapped in the structure, phenylalanine (magenta) is present at the active site, as well as the cofactor, 2-His-1-carboxylate facial triad and the iron molecule (orange).

5.2.3 Protein modeling and substrate docking analysis

Structural information of proteins molecules is essential to annotate their biological function. Therefore, to explore the substrate-determining amino acids, the protein structure models 10 AAAHs from different bacteria were generated: *Chromobacterium violaceum, Pseudomonas aeruginosa, Xanthomonas campestris, Streptomyces kanamyceticus, Echinicola vietnamensis, Pseudoalteromonas atlantica, Pseudomonas putida, Mesorhizobium plurifarium, Archangium violaceum and Cupriavidus taiwanensis.*The crystal structure from *C. violaceum* and *P. putida* AAAH have been already determined (Ekstro et al., 2003; Erlandsen et al., 2002). These two structures, as well as others eukaryotic AAAHs were used for verification and comparison purposes. I-TASSER was used to generate the structures, the best model was selected based on the normalized Z-score, which is a consensus of alignments confidence of multiple threading programs.

Phenylalanine and tryptophan docking analyses were done using the generated structures. AAAH from *C. taiwanensis* (CtAAAH) performed better during the *in silico* tryptophan docking analysis. Hence *Ct*AAAH was selected as a prototype for further analysis. Moreover, docking results from the wild-type and variants (homologous sites that have been reported to increase the tryptophan affinity for bacterial AAAHs) showed a higher affinity of *Ct*AAAH toward tryptophan than the other structures. Relative binding energies of *Ct*AAAH for phenylalanine and tryptophan are present in Table 5.

The reliability of the structure was assessed based on the B-factor (Fig. 17). This parameter indicates the extent of the inherent thermal mobility of the residues. In the case of *Ct*AAAH structure, the core of the protein has a B-factor < 2 Å, higher values were associated with flexible loop regions. The secondary structure of the model showed high similarity with previously reported model 1LTV (Protein Data Bank) from *C. violaceum.* In both cases, 16 alpha helices and six beta strands are present in the structure (Fig. 17).

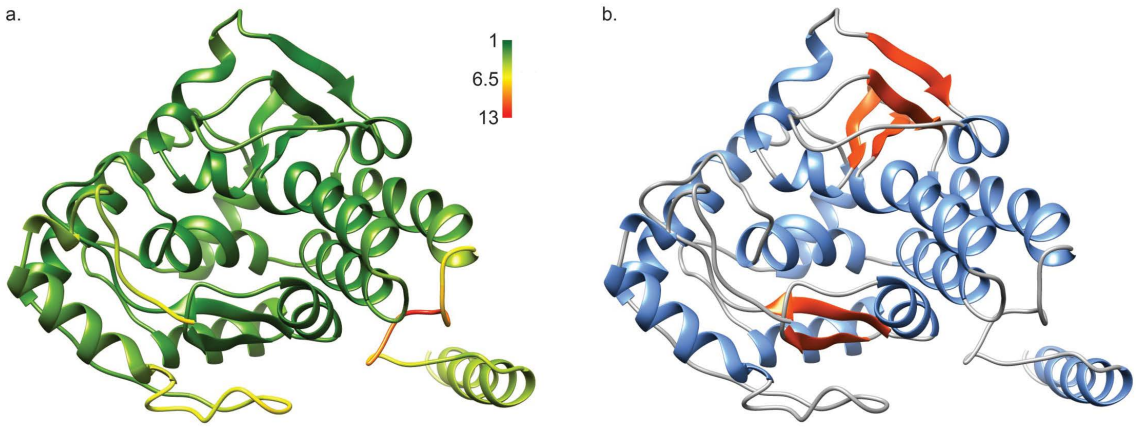

Figure 17: Distribution of the (a) B-factor and (b) secondary structures present in the *Ct*AAAH model.

Table 5: *In silico* relative docking binding energy and *in vivo* activity of *Cupriavidus taiwanensis* AAAH (wild-type and mutants) using phenylalanine or tryptophan as substrate.

AAAH	*in silico* RBE[a]			*in vivo* activity[b]		
	Phe	Trp	Preference (Phe/Trp)	Phe→Trp	Trp→5HTP	Preference (Phe/Trp)
CtAAAH (wt)	1.00	1.00	1.00	86.8 ± 5.7	10.5 ± 0.4	8.3
wt L113Y	0.83	0.84	0.99	73.8 ± 2.3	24.1 ± 0.9	3.1
wt W192F	0.75	0.86	0.87	79.1 ± 3.7	57.8 ± 1.2	1.4
wt Y222H	0.78	0.82	0.95	41.9 ± 2.9	25.0 ± 0.8	1.6
wt S223C	0.90	0.80	1.12	63.7 ± 3.1	21.6 ± 0.9	2.9
wt P229A	0.81	0.90	0.90	59.5 ± 3.5	35.4 ± 1.0	1.7
wt Y244C	0.81	0.84	0.96	48.8 ± 2.7	33.5 ± 1.2	1.5

[a] RBE, relative binding energy estimated using AutoDock Vina v1.1.2 (Trott and Olson, 2010).

[b] Activity estimated as $\mu M \cdot h^{-1} \cdot OD_{600}{}^{-1}$. All data are reported as the mean ± standard error of the mean (SEM) from three independent experiments.

5.2.4 Protein engineering for the modification of substrate preference

The design of synthetic metabolic pathways often requires a different role of the selected wild-type proteins. The *Ct*AAAH enzyme has high similarity with other hydroxylase reported to have phenylalanine preference. The alteration of the substrate specificity and the creation of a new catalyst is needed to fulfill the requirements of the selected pathway.

The *aaah* gene from *C. taiwanensis* was cloned into the low copy plasmid pACYCDuet under control of an IPTG-inducible T7 promoter. The resulting expression construction, p*Ct*AAAH, was transformed in strain BL21(DE3) Δ*tnaA*. *Tryptophanase A* was knocked out because its product encodes for tnaA which has been reported to catalyze the degradation of tryptophan and 5HTP to pyruvate, ammonia, and indole or hydroxy-indole, respectively (Gong et al., 2001; Hara and Kino, 2013).

The selectivity and activity of the proposed model were evaluated using *E. coli* strains expressing *Ct*AAAH wild-type. The *in vivo* specific activity assay results showed that the enzyme has a strong preference toward phenylalanine ($86.8 \pm 4.3\ \mu M \cdot h^{-1} \cdot {OD_{600}}^{-1}$) when compared with tryptophan ($10.5 \pm 0.4\ \mu M \cdot h^{-1} \cdot {OD_{600}}^{-1}$). However, when the tryptophan activity was compared with other bacterial wild-type AAAHs reported in the literature, *Ct*AAAH showed around three-fold higher activity toward tryptophan (*X. campestris* $= 2.91 \pm 0.21\ \mu M \cdot h^{-1} \cdot {OD_{600}}^{-1}$) (Lin et al., 2014). These findings confirm the prediction of our initial computational screening.

The FDII analysis showed 15 residues with high Qk: large Qk values indicate high evolutionary rates or physiochemical differences between homologous amino acids of two clades. These sites were tracked within the alignments, and the corresponding position of the *C. taiwanensis* sequence was compared with the sequence from the enzymes that have either phenylalanine or tryptophan preference. Six out of this 15 sites were selected because the *C. taiwanensis* amino acids were similar (belong to the same chemical group) to the sequences of the PAH cluster, and at the same time different from the TPH group (L113Y, W192F, Y222H, S223C, P229A, Y244C). Three amino acids L113, W192, and Y244, seem to be involved in the substrate and ligand stabilization and predicted mutations would be favorable for the tryptophan hydroxylation due to binding interactions and volume effect (Daubner et al., 2002) (Fig. 16).

In the modeled *Ct*AAAH-BH_4-Trp complex, the pterin ring of BH_4 -stacks with P119, the L113 residue is also involved in the binding since it is located within 3.5 Å from the cofactor. The mutation L113Y adds an extra - stacking interaction (3.2 Å) so that the enzyme-cofactor complex could be stabilized and orientated. Homologous mutations Y235L of human TPH decreased the K_m of BH_4 (McKinney et al., 2001), which in turn affected the apparent affinity for tryptophan. Similar results were reported by Kino et al. 2009

(Kino et al., 2009) using $DMPH_4$ as an analog cofactor. In the model, tryptophan is docked in the structure and stabilized by ionic interaction with S215 and R135, and through stacking interactions with H150, F197, and W192. TPHs present a phenylalanine in the W192 homologous site, in the W192F-model this position seems to be involved in the binding of the indole ring of the L-tryptophan. Also, volume-effect may play an important role in the stabilization of the substrate, since the distance between the residue and the substrate (or an analog) is between 3.8 to 4.1 Å in different reported structures. In this case, when the wild-type is docked with tryptophan the distance to the W192 residue is 3.6 Å, and it increases to 4.0 Å in the case of the W192F variant. Residue Y244 stabilizes the position W192 through a - interaction, a modification in this position could give more flexibility to the W192 residue, thus increasing the tryptophan preference due to a volume-effect as mentioned before (Fig. 18). The residues Y222, S223, and P229 are out from the binding pocket. Predicted mutations could play a role in the interaction or stabilization of the flexible loop (242-251).

Enzyme assays were done using generated mutants to investigate the performance of the predicted substrate-determining residues. As a result, the W192F mutant exhibited a 5.5-fold increase in tryptophan hydroxylation activity when compared to the wild-type. Meanwhile, the activity toward phenylalanine decreased by ∼10%. The substrate preference ratio of phenylalanine over tryptophan was shifted from 8.2 to 1.3. The other mutations, in general, presented a similar pattern, in which phenylalanine activity decreased; meanwhile, the tryptophan activity increased. It is worth noting that although a strong correlation between the *in silico* and *in vivo* Phe/Trp preference was not observed ($R^2 = 0.61$), there was correspondence among the high and the low performers regarding substrate selectivity: mutants W192F and P229A were predicted to have more preference toward tryptophan. Meanwhile, L113Y and S223C showed low Phe/Trp ratio (Table 5).

Advances have been made to predict mathematically favorable mutations for the creation of enzymes with wanted characteristics based on limited experimental data. This approach would facilitate protein engineering greatly, especially in the absence of high-throughput screening methods. The example developed here has great potential for the tailoring of enzymes with desired substrate specificity for synthetic biology and metabolic engineering. Nevertheless, final scope of this method should be prudently addressed, this is a specific example, and future experiments could bring different insights.

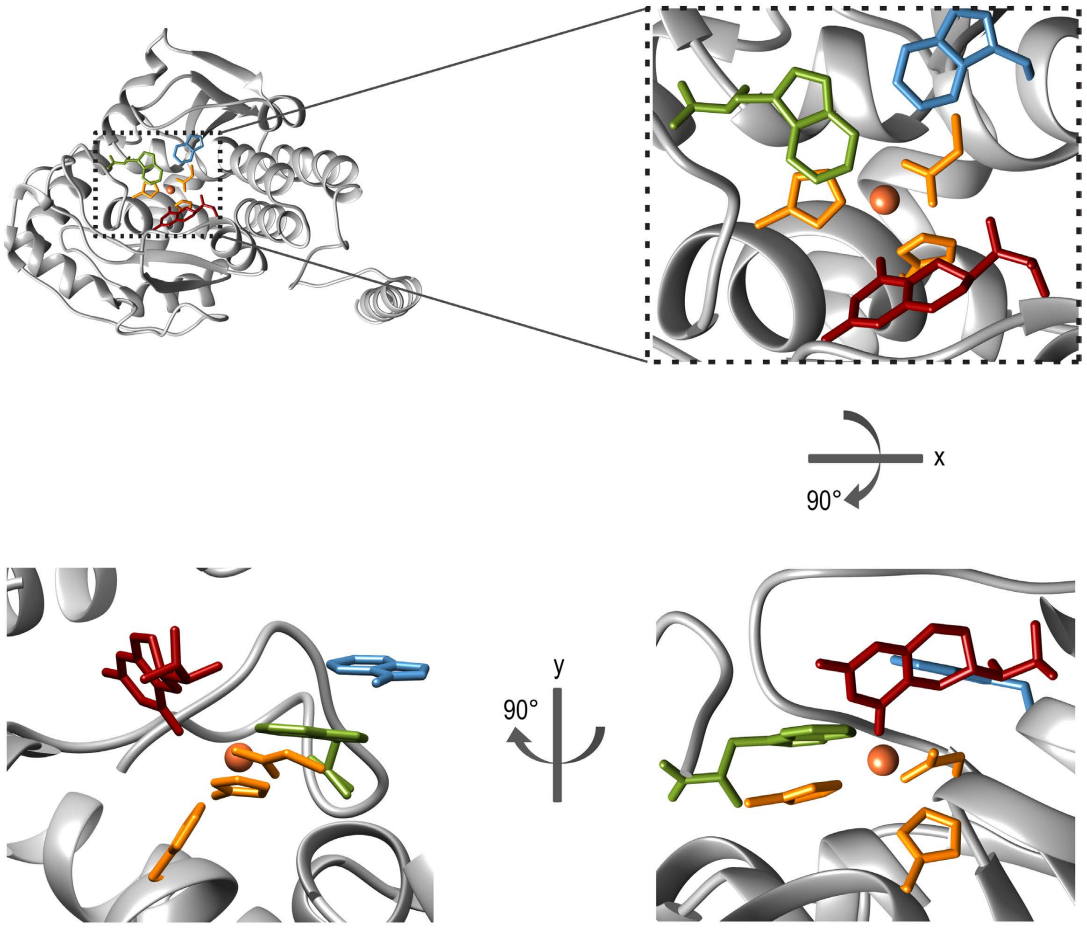

Figure 18: Docking analysis of tryptophan in *Ct*AAAH. The active site is framed in a dashed box. Docked tryptophan (green), 2-histidine-1-1carboxylate facial triad (light orange), iron molecules (dark orange), W192F mutation (blue), cofactor (red).

5.2.5 Pathway engineering for the production of 5HTP from tryptophan

Three important components are needed to expand the tryptophan metabolism in *E. coli* to produce 5HTP, it is needed a hydroxylase enzyme that recognizes tryptophan as a substrate, a pterin cofactor which is consumed during the hydroxylation reaction, and a cofactor regeneration system since *E. coli* lacks it. One of the issues of the expression of AAAHs in bacteria is the availability of the cofactor. The most common cofactor reported to participate in the hydroxylation of aromatic amino acids is tetrahydrobiopterin (BH_4). While a few bacterial taxa, such as *Cyanobacteria* and *Chlorobia*, produce BH_4, most

of the prokaryotes cannot synthesize this cofactor. As *aaah* genes are present in different groups of bacteria, they must use a different cofactor available in prokaryotes. Pribat et al. reported that tetrahydromonapterin (MH_4), which is naturally produced in *E. coli*, is the cofactor of *Pseudomonas* AAAH and can also be used in eukaryotic tyrosine hydroxylase as well. Therefore, the possibility that *Ct*AAAH could be able to use endogenous MH_4 as a cofactor was explored.

When compared *E. coli* BL21(DE3) *ΔtnaA* cultures harboring p*Ct*AAAH, wild-type, and W192F, the latter showed a stagnation during cell growth and low 5HTP production (0.21 mM), whereas the wild-type reached a maximum growth of OD_{600}= 7.6 (comparable to the control) and no 5HTP was detected. A similar effect has been observed in previous studies (Lin et al., 2014). This result could be explained because the enzyme might have been using MH_4 as a cofactor and there is no recycling system in *E. coli*. An artificial MH_4 regeneration pathway was established by transforming plasmid pBbE1k-2 into *E. coli*, this plasmid encodes for human pterin-4a-carbinolamine dehydratase (PCD) and dihydropteridine reductase (DHPR). PCD and DHPR participate in a well-known BH_4 recycling system in animals (Satoh et al., 2012). After the tryptophan hydroxylation, the pterin cofactor is converted into pterin-4a-carbinolamine, and PCD catalyzes its dehydration and produces dihydropteridine. The cycle is closed when DHPR reduces and reactivates the pterin (Fig. 19) (Davis et al., 1991).

There were no differences in the growth of BL21(DE3) Δ*tnaA* when these cells were carrying the cofactor regeneration system. As expected, when p*CtAAAH*-W192F was cotransformed with pBbE1k-2, cell viability dramatically improved, comparable with the values of the strains without the *CtAAAH*-W192F. These cells grew in media supplied with 5 mM tryptophan, the cell culture produced up to 2.5 mM 5HTP in a 24 hours cultivation period (Fig. 20), three times more than previously reported by (Lin et al., 2014). This result shows that the strain expressing the engineered *Ct*AAAH-W192F and the MH_4 regeneration pathway genes successfully produces 5HTP.

5.3 Conclusions

E. coli natural metabolic capability was expanded to produce non-canonical amino acid 5HTP using protein engineering. A computational screening ap-

proach was used for the selection of a wild-type enzyme and the prediction of key mutations. Initially, 10 structures were generated, and *Ct*AAAH was selected as the prototype enzyme to work in the laboratory.

Using sequence comparison, phylogenetic and functional divergence analyses relevant residues for phenylalanine/tryptophan substrate preference were successfully identified. All these sites were shown to increase the tryptophan preference of the enzyme at the expense of phenylalanine. The artificial pathway was introduced in *E. coli,* and L-tryptophan oxidation was achieved with mutant CtAAAH-W192F.

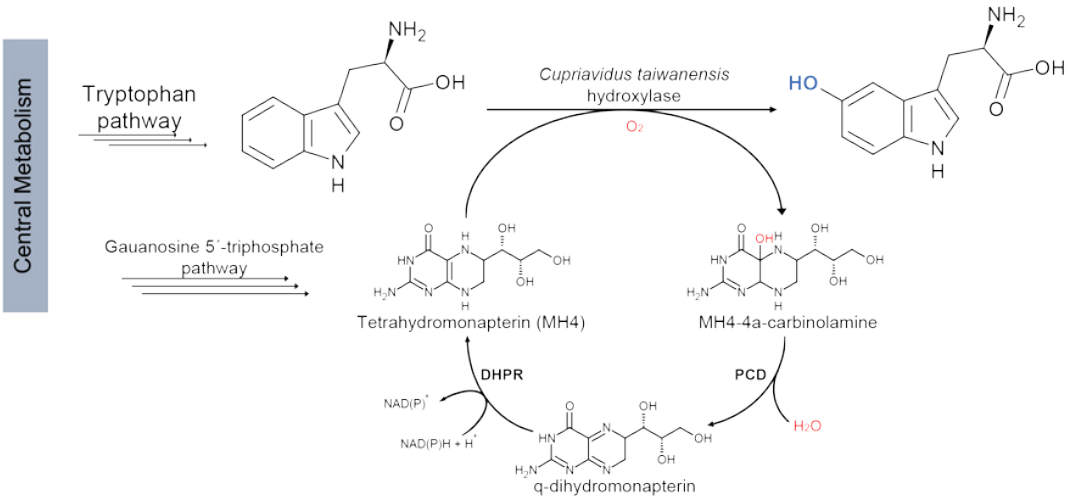

Figure 19: Synthetic pathway for the hydroxylation of tryptophan to 5-hydroxytryptophan. *Ct*AAAH hydroxylates tryptophan in the presence of oxygen and tetrahydromonapterin. In this reaction, the cofactor is oxidated. Afterward, the pterin-4a-carbinolamine dehydratase (PCD) dehydrates MH-4a-carbinolamine, and dihydropteridine reductase (DHPR) reduces back the pterin cofactor.

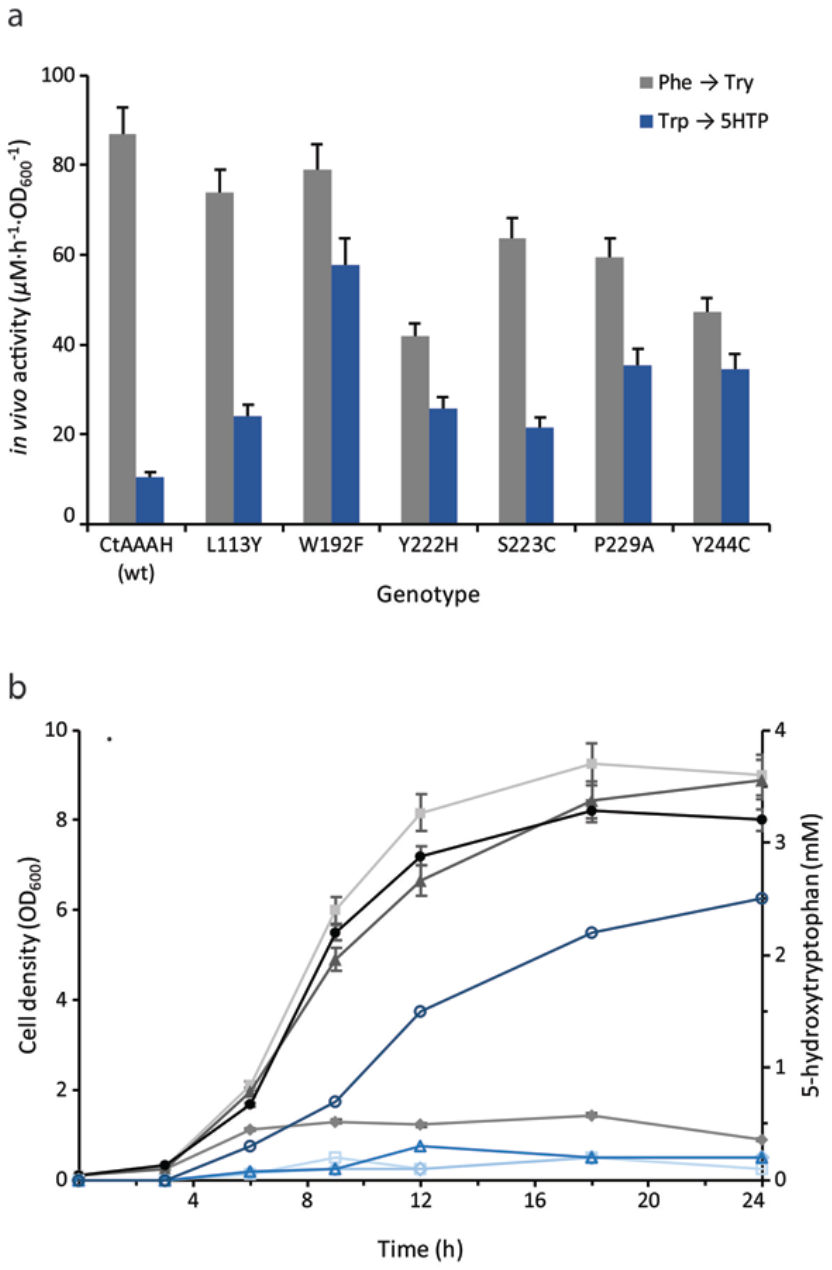

Figure 20: Modification of *Ct*AAAH via protein engineering: (a.) *in vivo* activity of wild-type (wt) *Ct*AAAH and its mutants, dark gray and blue bars indicate the activities toward phenylalanine and tryptophan, respectively. (b.) Cell growth (dark gray) and 5HTP production (blue) during fermentation by fed-batch using shaken flasks. BL21(DE3)Δ*tnaA* (squares), BL21(DE3) p*Ct*AAAH-W192F (diamonds), BL21(DE3) cofactor (triangles), BL21(DE3) cofactor p*Ct*AAAH-W192F (circles). Bars present in the graphs correspond to the standards errors means (SEM) from three independent experiments.

6 Directed evolution protein engineering and synthetic pathway for the production of 5HTP from glucose

In section five, a first round based on rational protein design was presented. This section includes results of a second iteration step in which best mutant (*Ct*AAAH-W192F) was used as the starting material for the generation of two libraries which were screened with the help of a tryptophan intracellular fluorescent sensor. The engineered enzyme was incorporated in a tryptophan producer strain, which also included a cofactor regeneration pathway.

6.1 Introduction

Compared to rational design, a key advantage of directed evolution lies in the impact of knowledge gaps (or uncertainty). To have an efficient designed protein, deep understanding of the system at different levels is required (i.e., molecular mechanism of the enzyme, relevant functional and structural amino acids, kinetics, inhibitors, molecular dynamics, among many other sources of information that can be used for this purpose) (Tizei et al., 2016). Most of the time full understanding of the biological system is not available and, in principle, a directed evolutionary approach can bypass this problem.

Direct evolution is grounded on Darwinian a principle which states that diversity within the population is necessary for selection to happen. Genetic diversity leads to phenotypic diversity, and as a counterpart, selection of specific phenotypes guides the genotype to limited number variants. Therefore, introducing diversity into a population followed by a systematically screening process entails the selection of variants with the desired activity. In practice, the introduction of genetic diversity into a population escalates the number of variants that should be sampled, and in many cases, this number can be beyond the screening capacity. Moreover, since functional variants become rarer in the population, there is a greater burden on the selection method to isolate optimal candidates.

Different methods and approaches are recommended when designing a directed evolution strategy. From an input point of view, libraries should be

as small as possible without scarifying diversity. For this purpose available knowledge on the targeted system (similar consideration used in a rational design approach, described in the section before) is used to limit and group the residues in the analysis additionally reduced codon redundancy approaches can also decrease the size of the libraries with no detriment in the diversity. The output greatly depends on the screening capacity, and this should be intimately linked to the theoretical size, diversity, and quality of the library, a minimum number of variants should be analyzed to assure a 95 % (or 99 %) coverage of the library maximizing the likelihood of success.

Bacterial colony screening is a cost-effective method to analyze a large number of variants, and when coupled with a fluorescent biosensor, this can be made in an acceptable time and cost. Intermediate sensor-assisted push-pull strategy (InterSPPS), proposed by Fang et al. (2016), uses a fluorescent sensor that responses to a molecule produced in the central metabolism (tryptophan) to target the production of a secondary metabolite. Production of tryptophan derivates can be traced by the decrease of the reporting signal.

The second iteration of *Ct*AAAH protein engineering is presented in this section. A directed evolution approach was used. Two small- rationalized-libraries were designed, and the best variants were screened out using a bacterial selection approach coupled with a tryptophan fluorescent biosensor. The best mutant from each library was each library was characterized, as well as the double mutant.

6.2 Results and discussion

6.2.1 Construction of a strain for the production of 5HTP from glucose

A pterin (cofactor) reconstitution pathway was incorporated in tryptophan producer strain S028. Two strategies were compared, the PCD and DHPR were added via plasmid (Pl) or via genome integration (Gi), whereby strains TrpD-Pl and TrpD-Gi were derived. No differences regarding growth were observed between the original strain and the newly generated strains (Fig. 21).

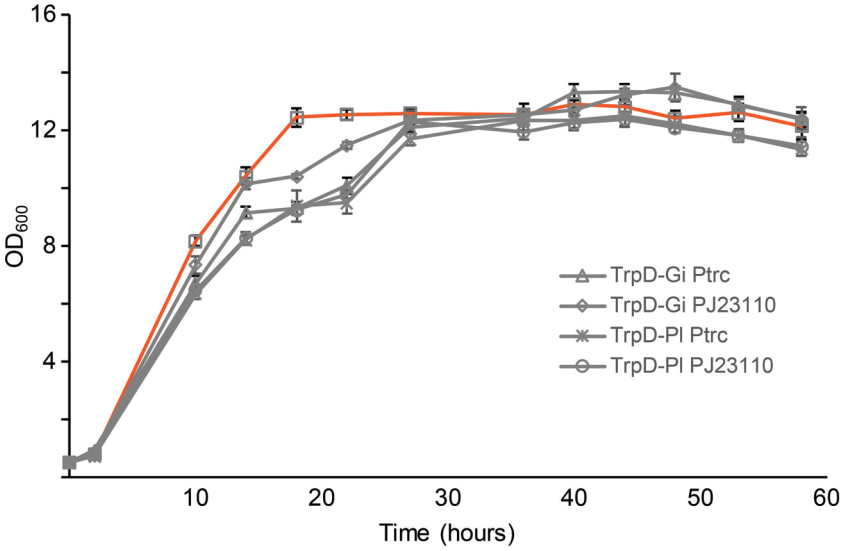

Figure 21: Cell growth of S028 derivated strains with PCD and DHPR genes incorporated via plasmid (Pl) or genome integration (Gi). Strains harbor gene CtAAAH-F with a strong promoter (Ptrc) or a medium strength promoter (PL23110). Control strain, S028, is highlighted in orange.

These strains were transformed with the plasmids pACJ23-*Ct*AAAH-F and pACPtrc-*Ct*AAAH-F. In general terms, a slight reduction in tryptophan production was observed in strain TrpD-Gi, whereas 5HTP was strongly affected when compared with TrpD-Pl (Fig. 22). In both cases, the P_{trc} promoter controls the transcription of the bicistronic mRNA, and the main differences lie in the number of copies per bacteria. The strain in which *PCD* and *DHPR* were integrated into the genome possess one copy of the genes, whereas there are around 15 to 20 copies in the strain transformed with pBbE1k-2. The low tryptophan and 5HTP production in the TrpD-Gi strains might be due to a low regeneration rate of endogenous cofactor MH_4 given that single-copy of the genes present in the strain.

Strain TrpD-Pl carrying pACJ23-CtAAAH-F produced over 2.5 g/L of tryptophan and 100 mg/L of 5HTP from glucose in shaken flasks after 60 h fermentation. The same strain with plasmid pACPtrc-*Ct*AAAH-F produced a similar amount of tryptophan, but around 25 % less 5HTP (Fig. 22). Often the expression of recombinant proteins in host cells requires a significant amount of resources causing overload in the metabolism of the host (Pasini et al., 2016). These results suggest that medium strength promoter P_{J23110} has a benefit for 5HTP if compared to the P_{trc} promoter.

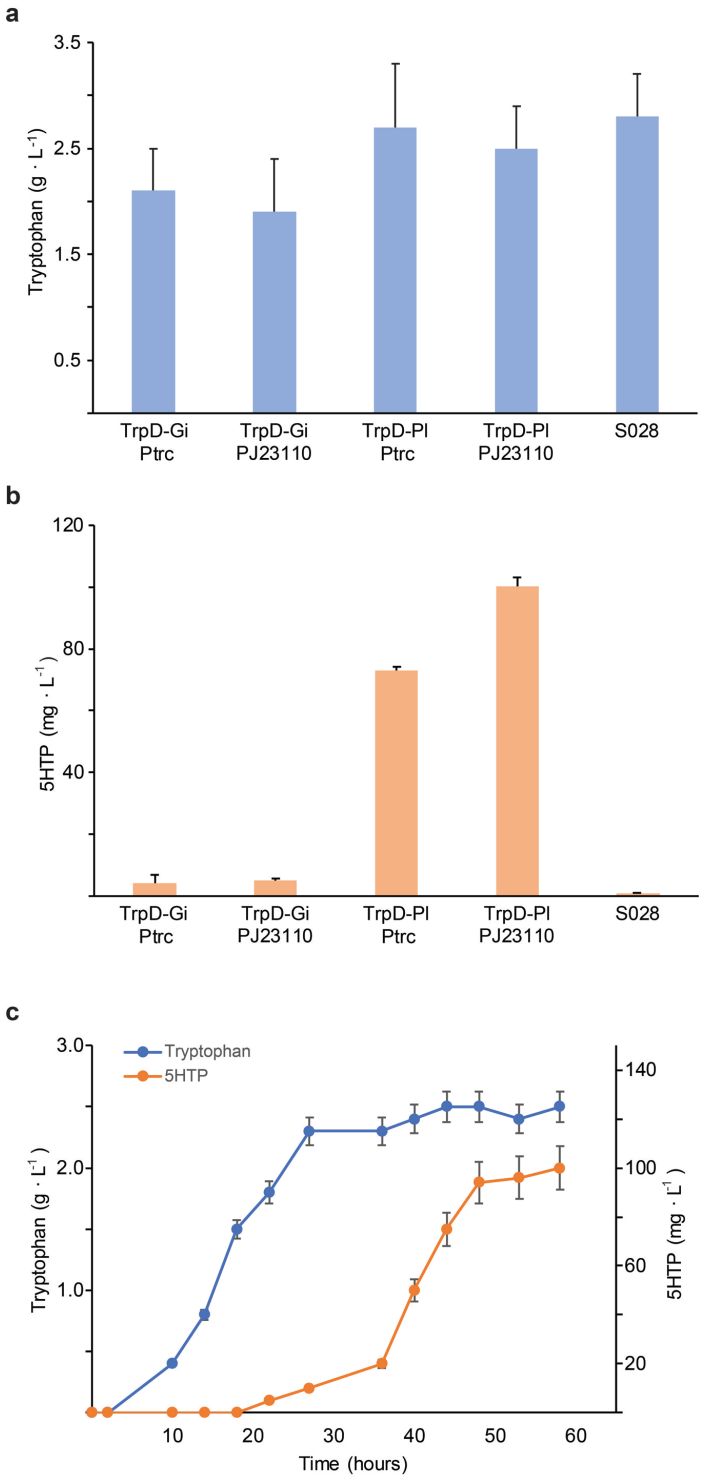

Figure 22: Tryptophan and 5HTP production in different *E. coli* strains. (a) tryptophan production, and (b) 5HTP production in strains TrpD-Gi and TrpD-Pl strains carrying the *CtAAAH-F* gene under control of the P_{trc} or P_{J23110} promoters respectively. (c) Tryptophan and 5HTP production of TrpD-Pl / pACJ23-*CtAAAH-F* over time.

Although TrpD-Pl / pACJ23-CtAAAH-F produced the highest amount of 5HTP within the set of strains, it seems that the enzyme still has a low activity due to the high amount of residual tryptophan when compared with the product. The 5HTP level did not increase with the tryptophan concentration level. Tryptophan production reached a steady state 24 h after the inoculation. Meanwhile, 5HTP production was stable after 48 h. This suggests that the enzyme still possesses a low catalytic activity. These results suggest that a second round of protein optimization (Section 6.2.3) should be done using *Ct*AAAH-F as starting material for the next iteration step. For this purpose, a tryptophan biomolecular sensor was modified and then used during the screening process (Section 6.2.2).

6.2.2 Tryptophan biomolecular sensor

Biomolecular sensors for the amino acids methionine (Mustafi et al., 2012), lysine (Binder et al., 2012), valine (Mustafi et al., 2014), and tryptophan (Fang et al., 2016) have been developed for screening purposes. The last one, the tryptophan sensor, is composed of the *tnaC* leader sequence of the *tnaCAB* operon (Fig. 6) and fused with a green fluorescent protein (GFP) (pSenTrp-GFP). Fluorescence of this sensor was measured by flow cytometry and evaluated in M9 plates. In the case of flow cytometry cells were grown in M9 media supplied tryptophan, then the cells were washed twice with fresh media and resuspend. The periodical evaluation showed the decrease of fluorescence over time. The biomolecular sensor was also evaluated in M9 agar plates supplied with 1 mM tryptophan. In this case, there were no differences in the fluorescence of cells expressing *Ct*AAAH-wt, -W192F and -Y244C.

A protease peptide signal was fused to the C-terminal of the GFP to create an unstable variant of the protein (Andersen et al., 1998). The new variant of the biomolecular sensor (pSenTrp-GFP(LVA)), presented the linear response range between 0.5 and 1.8 mM. Also, cells harboring this plasmid presented a drastic reduction in the fluorescence after tryptophan was removed from the media and colonies harboring different CtAAAH variants were clearly distinguishable in M9 plates supplied with tryptophan (Fig. 23).

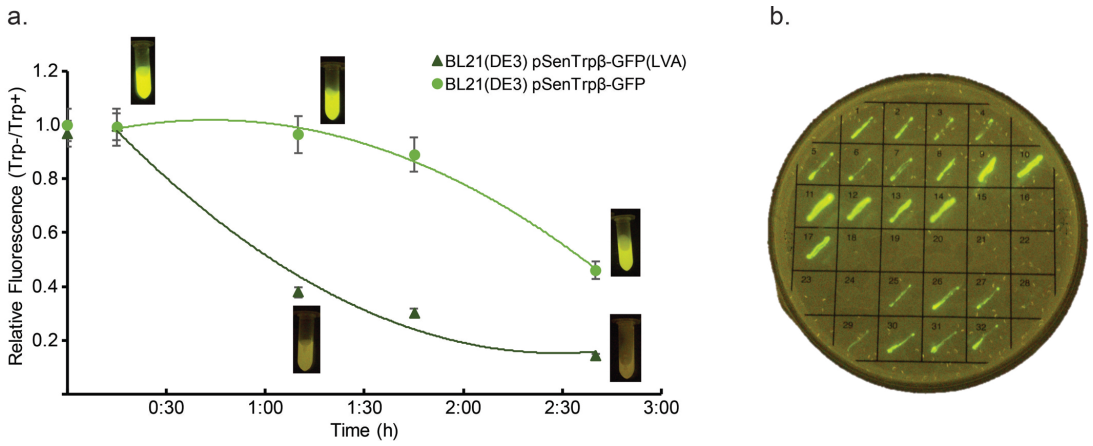

Figure 23: Performance of the tryptophan biomolecular sensor. (a) Flow cytometry measurements were done after cells were transferred in M9 media without tryptophan. (b) Fluorescence of cells harboring the pSenTrp-GFP(LVA) expressing different variants of CtAAAH (wild-type, W192F, Y224). Negative fluorescent control was also included.

6.2.3 *Ct*AAAH directed evolution

Two positions of *Ct*AAAH-F were subjected to saturation mutagenesis (SM) using the reduce codon strategy proposed by Kille et al. (2013). Residues Phe197 and Glu219 were selected for SM because these positions are part of the binding pocket of the enzyme in the region that interacts with the aromatic ring of the substrate and they are also near to the cofactor. These residues also play an important role in defining the pocket's shape and volume (Fig. 24).

Part of the gene and vector backbone from pACPJ23-*Ct*AAAH-F was amplified using phosphorothioate primers. After cleavage with an I_2/EtOH solution, a fragment with sticky ends was generated. Four synthetic oligos which include positions Phe197 and Glu219 were used to generate two independent libraries. Hybridized fragments were transformed into the strain BL21(DE3) *ΔtnaA* carrying plasmid pSenTrp(-LVA).2 which carries an intracellular tryptophan sensor. Enzymes with low activity should not consume tryptophan. Consequently, fluorescence should be high, and *vice versa*, low fluorescence means low intracellular tryptophan concentration due to the activity of the enzyme that converted it into 5HTP.

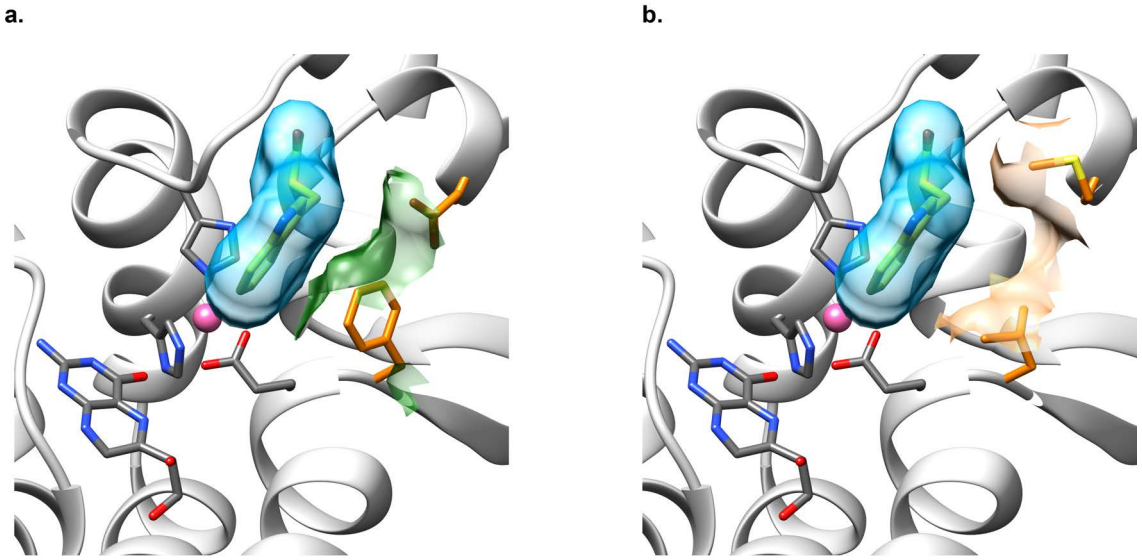

Figure 24: Protein engineering of *Ct*AAAH. (a) *Ct*AAAH-F binding pocket: the atoms of the cofactor and 2-His-1-carboxylate facial triad are present in gray, iron atom in pink, tryptophan surface is shown in blue, F197 and E219 atoms are present in orange, and their surfaces are shown in green. (b) *Ct*AAAH-LC binding pocket: as in (a), but L197 and C219 are present in orange, as well as their surfaces.

A set of 48 colonies was randomly selected. The *CtAAAH-F* gene from each colony was fully sequenced to investigate the quality of the libraries. 43 colonies were identical to the original sequence except for the saturated site. Four sequences presented *indels* near the sticky ends where the molecules hybridized and one sequence presented a point mutation out of the two categories just mentioned.

A total of 1673 colonies were screened in M9 media supplied with 1 mM L-tryptophan, 823 colonies from the Phe197-library (F197-L) and 850 from the Glu219-library (E219-L). A total of 167 colonies from the F197-L and 124 from the E219-L with low or no fluorescence were selected and transferred to a new M9-plate supplied with 2 mM L-tryptophan. Successive steps of 1 mM tryptophan were repeated three more times until four and three single colonies with low fluorescence from F197-L and E219-L were identified (Fig. 25).

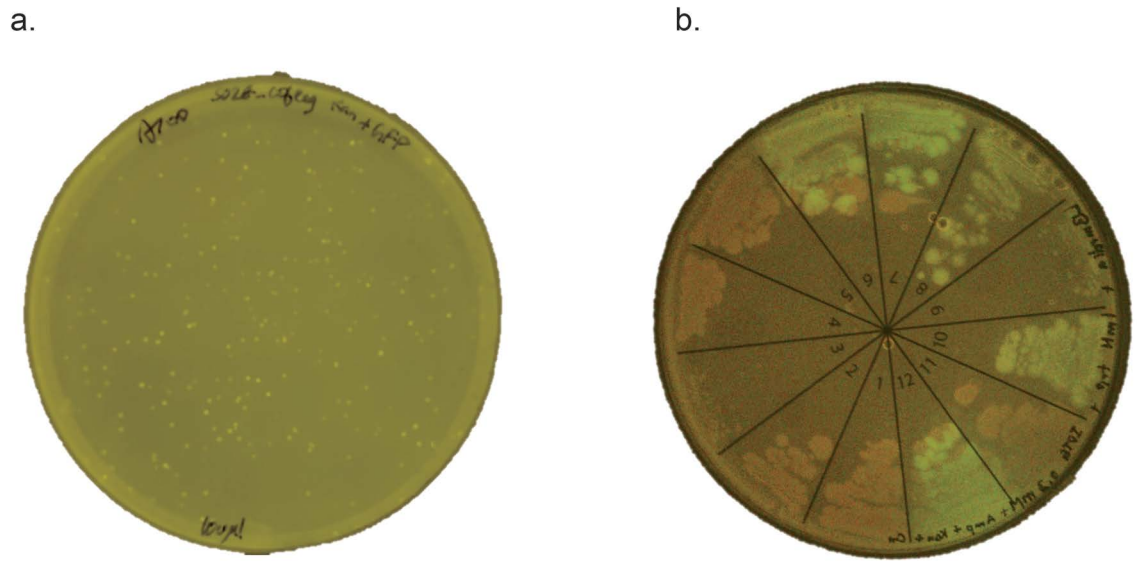

Figure 25: Screening procedure using pSenTrp-GFP(LVA). (a) Bacteria were plated in M9 plates supplied with 1 mM tryptophan, colonies with less fluorescence were transferred to a fresh M9 plate with 2 mM tryptophan. (b) Subsequent transferences with an increase in the tryptophan concentration were done until there was no distinguishable change in the fluorescent of the colonies when compared with the previous plate.

The plasmids from these colonies were extracted, and the *CtAAAH* gene was completely sequenced. Out of the F197-L, one colony had the same genotype as the wild-type (further sequencing results showed a mutation in the *tnaC* sequence which controls the GFP expression), two had a mutation that substituted phenylalanine for leucine (F197L), and one sequence had an isoleucine (F197I) in position 197. All sequences obtained from the E219-Lib presented cysteine in amino acid position 219 instead of glutamate (E219C).

Enzymatic assays with the three identified variants were done. Leu197 and Cys219, performed better in the tryptophan hydroxylation assay when compared with the *Ct*AAAH-F. Then, the double mutant *Ct*AAAH-F197L/E219C (*Ct*AAAH-LC) was created to explore the combinatorial effect of these residues (Fig. 24).

The correct size of the gene was confirmed by DNA digestion with proper restriction enzymes and by DNA sequencing. Enzymatic assays were performed to compare the kinetics of the *Ct*AAAH variants. Protein production and purity was assessed by SDS-PAGE (Fig. 26).

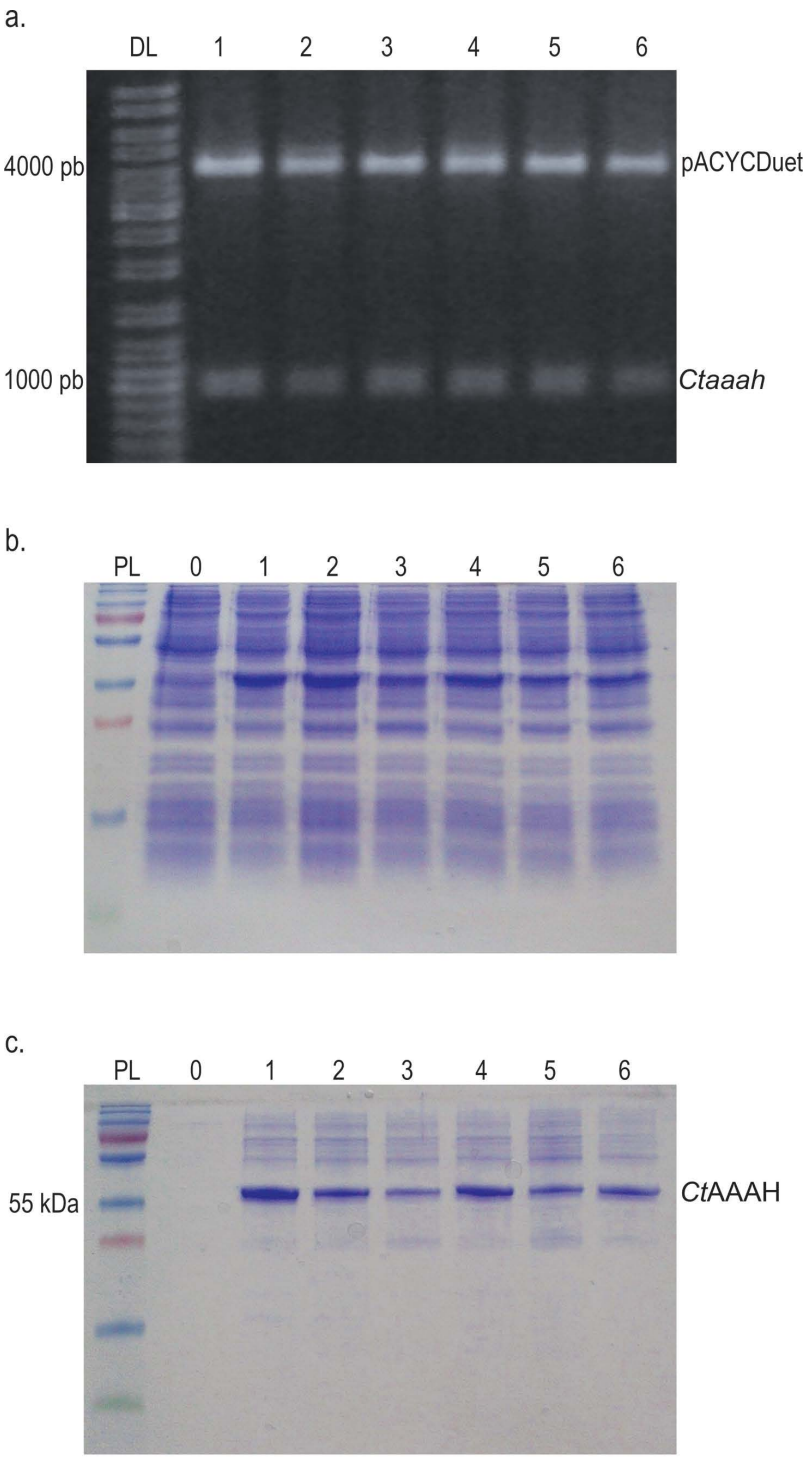

Figure 26: Molecular cloning, protein expression and purification of diferent variants of *Ct*AAAH. (a) Confirmation que the molecular cloning step by enzymatic digestion. (b) Protein expression, crude extracts. (c) Protein purification. 0, BL21(DE3) control cells; 1, *Ct*AAAH-wt: 2, *Ct*AAAH-W192F; 3, *Ct*AAAH-E219C; 4, *Ct*AAAH-F197L; 5, *Ct*AAAH-F197I; 6, *Ct*AAAH-F197L-E219C.

The double mutant showed a higher activity than the variants with single mutations. *Ct*AAAH-LC also displayed a lower K_m value (0.95 mM) and a higher reaction velocity (V_{max} = 1.9 mM $\cdot$ s^{-1}) when compared to the original *Ct*AAAH-F (Fig. 27). Iterative site mutagenesis (ISM) has proven before to be a useful approach for the improvement of the enantioselectivity, substrate acceptance or thermostability of different enzymes (Acevedo-Rocha et al., 2014; Reetz and Carballeira, 2007).

ISM was used to increase the activity. Finally, an enzyme with a lower K_m and a higher reaction velocity was selected. Certain degree of volume and shape changes in the binding pocket was predicted, this needs to be confirmed by structural determination. Nevertheless, modeling and docking analysis provide useful hints about conformational variations.

It is reasonable to assume that the size of the binding pocket changes with the substitution of phenylalanine for leucine at the position 197 and glutamate for cysteine in the residue 219. This new conformation probably stabilizes the enzyme-substrate-cofactor complex, which is important for the tryptophan hydroxylase activity.

The production of tryptophan and 5HTP was compared for TrpD-Pl strains harboring p*Ct*AAAH-F and p*Ct*AAAH-LC plasmids respectively. The growth curves were similar in both cases. After 60 h of batch-fermentation, 372.6 $\pm$ 19.7 mg/L of 5HTP were produced in the cells carrying the p*Ct*AAAH-LC plasmid, around 3.5 times higher than the cells with p*Ct*AAAH-F. In the case of the cells with p*Ct*AAAH-LC, a reduction in tryptophan production was observed. This optimized strain was used for further serotonin production (Fig. 27).

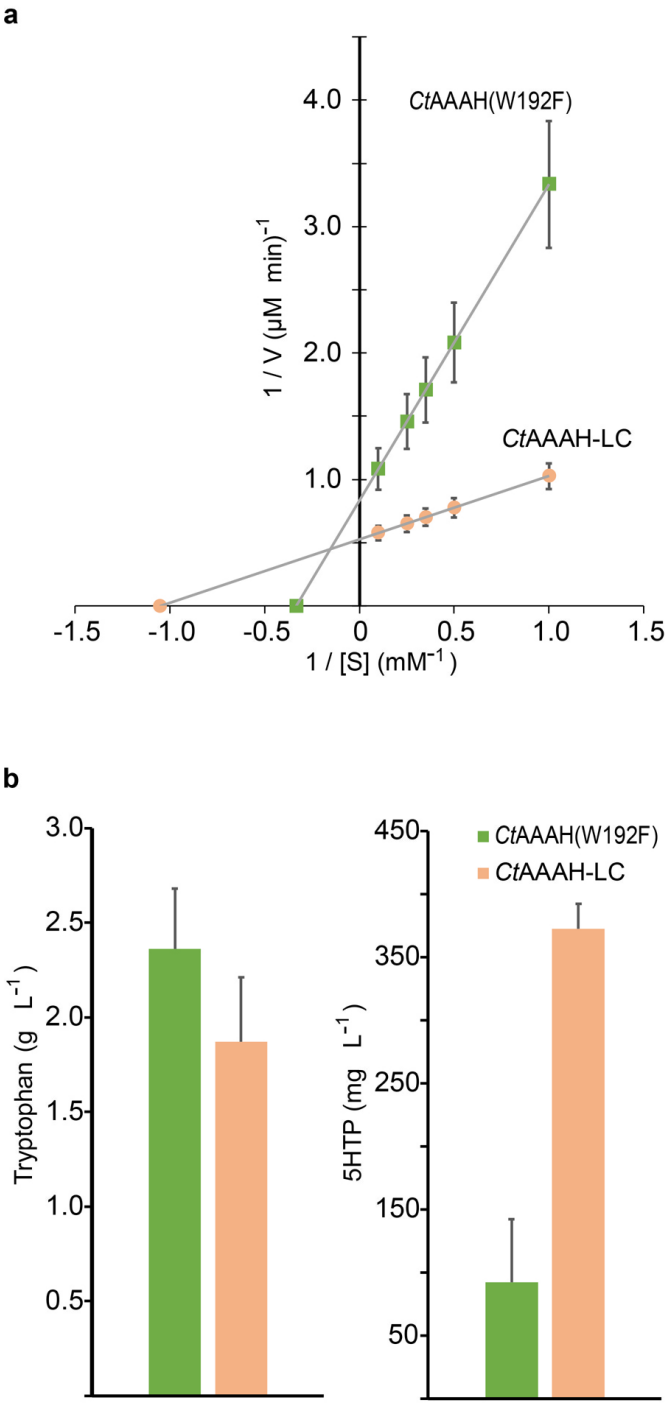

Figure 27: Enzyme kinetics of *Ct*AAAH-(W192F) and CtAAAH-LC and production of tryptophan and 5HTP. (a) Lineweaver-Burk plot of *Ct*AAAH-F and *Ct*AAAH-LC. (b) tryptophan and 5HTP production after batch fermentation.

6.3 Conclusions

Results from a second protein engineering round were presented. Two amino acids (F197 and E219) were selected due to their position within the active pocket, and two independent libraries were screen using a tryptophan biosensor. Best mutants (F197L and E219C) were analyzed and later combined in plasmid p*Ct*AAAH-LC. Variant CtAAAH-LC showed higher reaction velocity and low Km value if compared to the starting material CtAAAH-LC.

In parallel, tryptophan producer strain S028 was adapted for the production of 5THP from sugar. PCD and DHPR were incorporated in strain S028 via plasmid incorporation or genomic integration. Also, tryptophan hydroxylase expression was compared under control of a strong- and medium- promoter. Finally, results of a strain capable of synthesizing 5HTP from glucose were presented.

7 Biosynthetic pathway and processes for effective production of serotonin

This last section includes the results of an efficient process for the production of serotonin. The synthetic pathway was incorporated into the tryptophan producer strain. These results are compared with a strategy in which serotonin production is separated into two stages. A first step for the production of 5HTP and a second step that converts the later into serotonin.

7.1 Introduction

The construction of *de novo* biosynthetic pathways denotes the assembly of genes from different non-related organisms to construct new artificial pathways in a desired host. This approach enables the conceptualization and development of unprecedented biocatalysts. Once the pathway has been designed, and individual components have been shown to work independently, the system can be integrated into an existing platform for the production of the metabolite of interest. The synthesis of the compound can be achieved by two different ways. In the first case, an existing pathway from a host can be extended by connecting it to the novel pathway. In a second way, the starting material can be supplied in the media for bioconversion (Lin and Tao, 2017).

Bioconversion by whole-cell process holds promises to produce commercial-natural products. FDA and European legislation consider "natural products," compounds that are produced by biotechnological methods and are originated from natural sources (Xu et al., 2007).

The most common drawback in bioconversion includes the presence of substrate or product inhibition, the membrane acting as a mass transporter barrier o the presence of metabolic by-products due to an excess carbon flow in a pathway or to enzyme promiscuity. In many occasions, tailored enzymes and engineered pathways can cope these constraints. Nonetheless, sometimes depending on the complexity of the preferred reactions different strategies can be used to implement the fermentation process. One option to circumvent the production of by-products is to couple two or more recombinant strains. Nakagawa et al. (2016) demonstrated the total synthesis of opiates using four *E. coli* strains. In the first strain, tyrosine was produced and transformed to dopamine, while the second strain used dopamine and converted it

into tetrahydropapaveroline. In a third step, the latter was transformed into reticuline, and a final strain was responsible for the production of thebaine. Using this process, they were able to bypass oxidation and degradation of intermediates, and the negative effect of IPTG on the expression of the genes responsible for the last step.

In this last section, the tryptophan pathway was extended for the production of serotonin via 5HTP. A single culture strategy was used, serotonin and tryptamine (by-product) were detected in the supernatant. In order to boost serotonin production, the pathway was divided into two strains. A first strain responsible for tryptophan oxidation and in a second step serotonin production via decarboxylation.

7.2 Results and discussion

7.2.1 Microbial biosynthesis of 5HTP via metabolic engineering

After improving hydroxylase efficiency of the hydroxylase enzyme *Ct*AAAH. Strain TrpD-Pl harboring plasmid p*Ct*AAAH-LC was used for 5HTP production. After 52 h of fed-batch fermentation at 30 °C, the OD_{600} reached a value of 59 ± 4.7. Moreover, tryptophan production and 5HTP bioconversion were achieved, and the synthesis of both occur simultaneously, as soon as tryptophan concentration increased in the supernatant 5THP production started. Final tryptophan titer reached 24.2 ± 3.15 g/L in the media and 962 ± 58 mg/L of 5HTP (Fig. 28).

To the date, this is the highest 5HTP biosynthesis achieved by microbial fermentation from glucose. Currently, 5HTP production depends on the extraction of African plant *G. simplicifolia*. Several attempts have been made to synthesize 5HTP chemically or using enzymatic synthesis, yet its industrialization is still unsuitable. Engineered *Ct*AAAH-LC exhibits high tryptophan hydroxylation activity, this enzyme requires equimolar concentration of tryptophan and the pterin cofactor. Hence, it is expressed in tryptophan producer strain that also contains two enzymes necessary for the cofactor regeneration. After further optimization of the process, this strain could be used for industrial efficient and low-cost production of 5HTP.

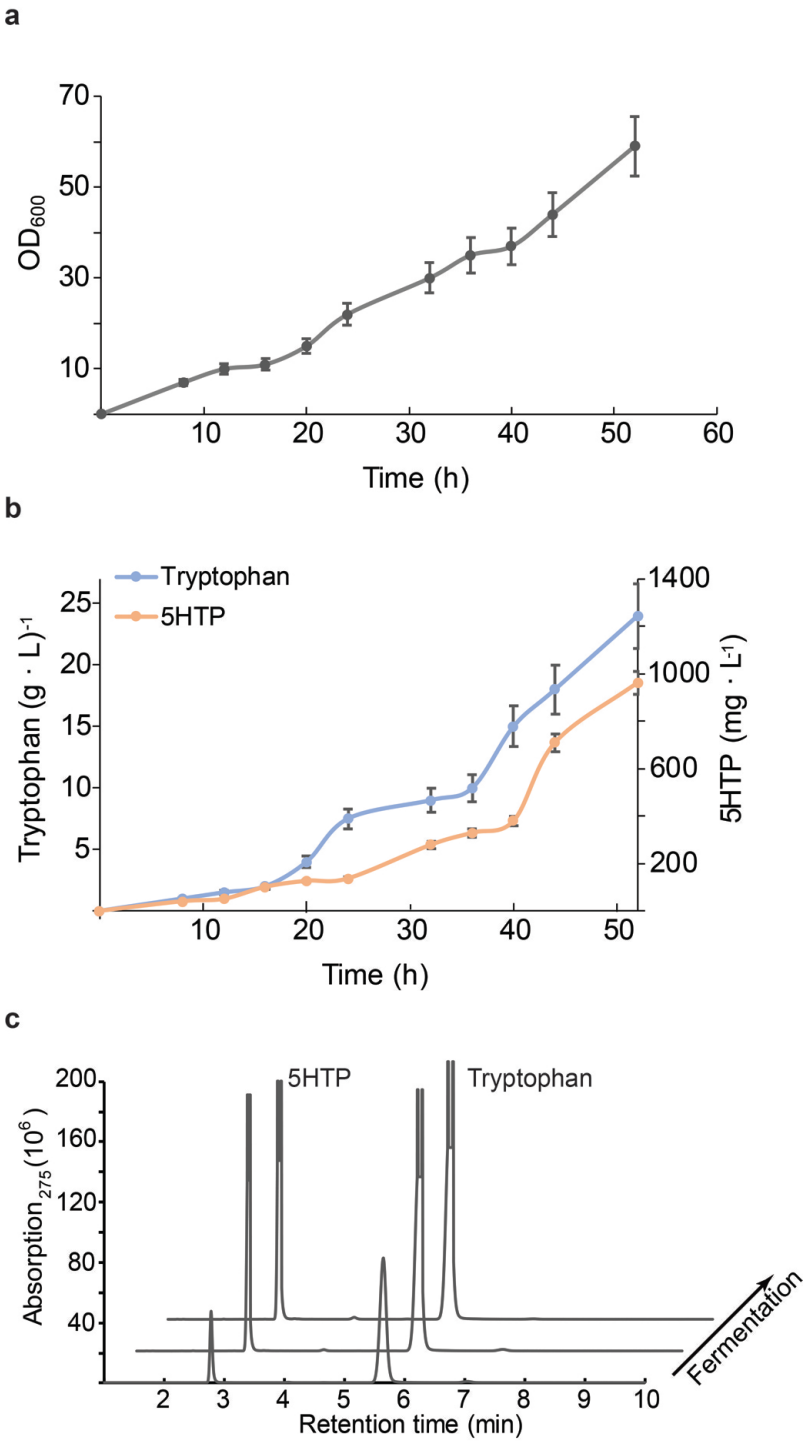

Figure 28: 5HTP production in *E. coli* from glucose in fed-batch fermentation using strain TrpD-Pl harboring plasmid pACJ23-*Ct*AAAH-LC. (a) growth curves, (b) tryptophan and 5HTP production, and (c) HPLC retention pattern of tryptophan and 5HTP

7.2.2 Serotonin production using a single culture

In a first attempt to produce serotonin from glucose, plasmid pCOLAJ23-*TDC*.2 was transformed into the 5HTP producer (strain TrpD-Pl / p*Ct*AAAH-LC). After 52 h of fed-batch fermentation the OD_{600} reached a value of 10.2 ± 0.6, and 5.34 ± 0.43 g/L of tryptophan was produced. Final 5HTP production was quite low (7.3 ± 0.6 mg/L) compared with the TrpD-Pl / pCtAAAH-LC (962 ± 58 mg/L) control. Furthermore, no serotonin was detected in the media. A high amount of tryptamine was produced (3.03 ± 0.32 g/L was present in the supernatant), which indicates a strong preference of the TDC enzyme towards tryptophan. Both tryptophan and 5HTP, are natural substrates of *C. roseus* TDC. However, apparent K_m value for tryptophan is 0.075 mM, whereas the K_m for 5HTP is 1.30 mM (Noé et al., 1984). This may explain why we detected a high concentration of tryptamine and no serotonin in the supernatant after fermentation.

Instead of using constitutive promoter P_{J23101} to regulate *TDC* expression, this gene was subcloned into a pBAD plasmid, which has an inducible expression system based on the araB promoter region. *TDC* was induced 10 h after inoculation (OD_{600} around 10). Still, 1.81 ± 0.29 g/L of tryptamine was found in the media, similar to the 1.66 ± 0.27 g/L produced in the non-induced control. 21.3 ± 3.5 mg/L 5HTP and 0.8 ± 0.2 mg/L of serotonin was detected in the supernatant after 60 h fermentation. Although the *TDC* gene was induced 10 h after inoculation, perhaps the low serotonin productivity is due to an unwanted leaky expression from the arabinose-induced promoter in early stages, this may reduce the tryptophan pool in detriment of the 5HTP pathway. Consequently, serotonin production is low.

7.2.3 Two-step fermentation strategy for the efficient production of serotonin

Decarboxylation by TDC is a key step in the synthetic pathway for serotonin production, and although this enzyme can use both 5HTP and tryptophan as a substrate, it has a high preference towards the latter one. Despite the advantages of a single strain production system, it would be difficult to integrate a tight control over the expression of the *TDC*, plus an engineered enzyme with less preference toward tryptophan, but without compromising the 5HTP

activity. To circumvent this issue, the 5HTP and serotonin production phases were separated using a two-step culture approach.

First, 5HTP was produced from glucose using TrpD-Pl / pCtAAAH-LC with a fed-batch fermentation strategy. After 60 h fermentation 23.4 ± 1.4 g/L of tryptophan and 962 ± 58 mg/L 5HTP was produced. 5HTP containing supernatant was harvested by centrifugation and filtration. In a second step serotonin production was conducted by mixing fermentation medium with the 5HTP supernatant (4:1 ratio), glucose concentration was adjusted to 30 % and pH to 6.7. The mixed medium was inoculated with the strain BL21(DE3)*ΔtnaA* harboring plasmid pCOLAJ23-*TDC*. The consumption of 5HTP and serotonin production had a strong correlation. Serotonin was continuously accumulated until 44 h when it reached a steady value. The maximum serotonin production was observed 52 h after inoculation, 154.3 ± 14.3 mg/L (Fig. 29). The initial tryptophan concentration in the medium was 5.66 ± 0.61 g/L, and after 52 h fermentation tryptophan decreased to 2.66 ± 0.54 g/L. We also detected 2.91 ± 0.46 g/L tryptamine.

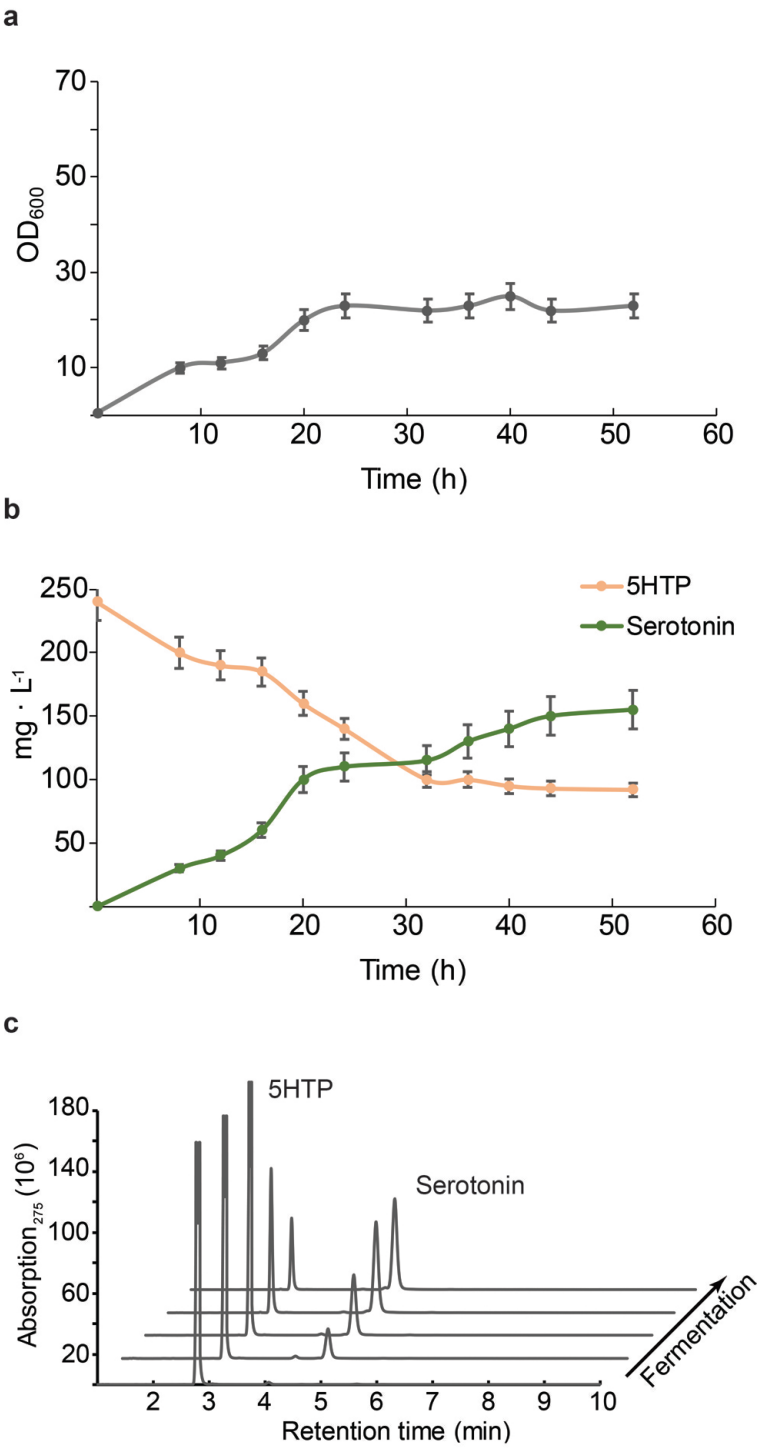

Figure 29: Serotonin bioconversion from 5HTP in *E. coli* in batch fermentation using strain BL21(DE3)*ΔtnaA* harboring plasmids pCOLAJ23-TDC. (a) growth curves, (b) 5HTP consumption and serotonin production, and (c) HPLC retention pattern of 5HTP and serotonin.

7.3 Conclusions

For many years, microbial production of industrial valuable compounds from cheap starting materials has attracted the attention. Most of the times, the synthesis of these compounds from simple carbon sources is associated with enzyme inhibition and the production of by-products due to undesired reactions. In this section, an efficient process for serotonin production was presented, in a first step 5HTP was produced from glucose and a second strain uses this as a substrate and decarboxylates it for the production of serotonin. Using this two-step production method, 154 mg/L of serotonin was produced from glucose, almost 200 times higher than the single-cell approach, in which de production of tryptamine (by-product) is high due to: 1) the similarity of the intermediates and 2) the promiscuity of TDC.

The construction of this stepwise culture method enables the optimization of individual production steps. Furthermore, the segmentation of the method facilitates the production of other tryptophan derivatives by exchanging the set of enzymes (e.g., n-hydroxycinnamoyltransferase for the production of phenylpropanoids, methyltransferases to produce other indole complex derivatives).

8 Summary

In this thesis, efforts have been made to produce 5HTP and serotonin from glucose in *E. coli*. Based on *in silico* analysis and biotransformation assays, the serotonin production pathway via 5HTP was selected over the pathway via tryptamine. Phylogenetic, structure-based and docking analyses were done to select a wildtype enzyme capable of performing tryptophan hydroxylation. Aromatic amino acid hydroxylase from *Cupriavidus taiwanensis* (*Ct*AAAH) was selected among ten other candidate due to its performance *in silico* according to molecular docking simulations with phenylalanine and tryptophan. The active pocket of the enzyme was compared using sequence, and functional divergence analysis, base on these six residues (L113Y, W192F, Y222H, S223C, P229A, Y244C) were predicted to be involved in the substrate preference. *In silico* results were confirmed experimentally by carrying out *in vivo* bioconversion studies. The method used proved to be useful for the prediction of relevant sites within the binding pocket. Nevertheless, conclusions should be addressed carefully since this is a one-case study, and further experiments should give more insight into the method.

The best mutant (*Ct*AAAH-W192F) was selected as starting material for a second round of protein engineering using a semi-rational approach. Residues Phe197 and Glu219 were selected and smart small-size libraries were designed using a reduced codon redundancy approach. These libraries were screen out using a biomolecular sensor for intracellular tryptophan concentration. The best performer of each library was combined and the double mutant, *Ct*AAAH-LC, displayed a lower K_m value (0.95 mM) and a higher reaction velocity ($V_{max} = 1.9$ mM $\cdot$ s^{-1}) than its predecessor CtAAAH-W192F. The method used in this second protein engineering step evidenced the benefits of combining rational design and directed evolution approaches. When compared with the rational design method, the total number of candidates increased drastically. Nevertheless, a fluorescent sensor was used to screen and select the colonies in which tryptophan was consumed faster. Finally, an enzyme with higher affinity and velocity was selected.

The hydroxylase enzyme, *Ct*AAAH-LC, was transformed into a tryptophan production strain which also carried a plasmid with two enzymes that participate in the tetrahydromonapeterin (MH_4) regeneration pathway. The latter molecule participates as a cofactor in the tryptophan to 5HTP conversion.

This strain can produce 5HTP from glucose at 962 ± 58 mg/L after 52 hours fermentation. This is so far the highest 5HTP concentration reported for biosynthesis based on glucose.

Finally, tryptophan decarboxylase (TDC) was incorporated in the 5HTP production strain to produce serotonin from glucose. However, serotonin was not detected, and a high amount of tryptamine was founded in the medium, which indicates a strong preference of TDC toward tryptophan. A two-step fermentation was designed to circumvent this problem, 5HTP production and serotonin conversion were separated. Using this approach 154 mg/mL of serotonin was produce using glucose as the starting material. This is the first report of its production in *E. coli*.

8.1 Outlook and future perspectives

Recent advances in metabolic engineering and biochemical pathway analysis make it possible to efficiently manipulate the biosynthetic pathways of microorganisms. This thesis presents a process for the production of serotonin via 5HTP. The major drawback of the presented scheme is the two-step fermentation approach needed to produce serotonin. As explained in Chapter 7 the process was separated due to the production of tryptamine by TDC. Two different alternatives can be used to overcome this problem and integrate the compleat pathway in one strain: (i) to increase the selectivity of TDC towards 5HTP in detriment of tryptophan activity via protein engineering as discussed in the latter Chapter, or (ii) to selectively degrade tryptamine using an enzyme such a Monoamine Oxidase (MOA) (EC 1.4.3.4). Several MOAs have been reported to degrade tryptamine and serotonin in eukaryotic organisms. Relatively little information is available regarding the bacterial degradation of tryptamine, although the oxidation of tryptamine by *Micrococcus* and *Bacillus* species has been reported (Buki et al., 1985; Leuschner et al., 1998; Nakazawa et al., 1974). In either case, protein engineering work should be done in advance due to the structural similarity of the substrates. For this purpose, the combination of the tryptophan sensor described in Chapter 6, and a recently developed 5HTP biomolecular sensor (Porter et al., 2017) could be used to facilitate the enzyme evolution and optimization of the engineered metabolic pathway.

Serotonin and 5HTP are currently produced by chemical methods or extracted from the seeds of *G. simplicifolia*. Significant benefits can be realized by switching the production that currently dependents on fossil resources to biological sources. Currently, many biotechnological alternatives have become cost competitive when compared with petroleum-based methods. Already, bio-based technologies such as enzyme catalysts are promising replacements for many industrial chemical processes. This work presented a biotechnological option for the production of serotonin and 5HTP from a simple fermentable sugar. Unlike serotonin, 5HTP could be used directly to treat different medical conditions. Serotonin can be used as the starting material for the biotechnological- or chemical- production of derivatives that could be used as antidepressant, antiviral, antibiotics, among others (Chadha and Silakari, 2017). High-value chemicals offer an opportunity to commercialize biotechnological produced molecules if quality, price and performance are equal to the reference petrochemicals. Fortunately, it seems that the number of biotech products is growing and gaining market share rapidly. This is good news for business, consumers and our environment.

9 References

Acevedo-Rocha, C.G., Hoebenreich, S., Reetz, M.T., 2014. Iterative Saturation Mutagenesis: A powerful approach to engineer proteins by systematically simulating Darwinian evolution. Methods Mol. Biol. 1179, 103–128. doi:10.1007/978-1-4939-1053-3_7

Acevedo-Rocha, C.G., Reetz, M.T., Nov, Y., 2015. Economical analysis of saturation mutagenesis experiments. Sci. Rep. 5, 10654. doi:10.1038/srep10654

Aiba, S., Tadayuki, I., Tsunekawa, H., Imanaka, T., Tsunekawa, H., 1980. Enhancement of tryptophan production by *Escherichia coli* as an application of genetic engineering. Biotechnol. Lett. 2, 525–530. doi:10.1007/BF00134901

Andersen, J.B., Sternberg, C., Poulsen, L.K., Bjorn, S.P., Givskov, M., Molin, S., 1998. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. Appl. Environ. Microbiol. 64, 2240–2246.

Anfinsen, C.B., 1973. Principles that govern the folding of protein chains. Science. 181, 223–230. doi:10.1126/science.181.4096.223

Arnold, F.H., 2001. Combinatorial and computational challenges for biocatalyst design. Nature. 409, 253–257.

Barrozo, A., Borstnar, R., Marloie, G., Kamerlin, S.C.L., 2012. Computational protein engineering: Bridging the gap between rational design and laboratory evolution, Int. J. Mol. Sci.13, 12428–12460 doi:10.3390/ijms131012428

Bertoldi, M., Cellini, B., Montioli, R., Voltattorni, C.B., 2008. Insights into the mechanism of oxidative deamination catalyzed by DOPA decarboxylase. Biochemistry 47, 7187–7195. doi:10.1021/bi800478s

Binder, S., Schendzielorz, G., Stäbler, N., Krumbach, K., Hoffmann, K., Bott, M., Eggeling, L., 2012. A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level. Genome Biol. 13, R40. doi:10.1186/gb-2012-13-5-r40

Birdsall, T.C., 1998. 5-Hydroxytryptophan: A clinically-effective serotonin precursor. Altern. Med. Rev. 3, 271–280.

Blanusa, M., Schenk, A., Sadeghi, H., Marienhagen, J., Schwaneberg, U., 2010. Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method. Anal. Biochem. 406, 141–146. doi:10.1016/j.ab.2010.07.011

Bonneau, R., Tsai, J., Ruczinski, I., Chivian, D., Rohl, C., Strauss, C.E.M., Baker, D., 2001. Rosetta in CASP4: Progress in *ab initio* protein structure prediction. Proteins Struct. Funct. Genet. 45, 119–126. doi:10.1002/prot.1170

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–254. doi:10.1016/0003-269790527-3

Brosius, J., Erfle, M., Storella, J., 1985. Spacing of the -10 and -35 regions in the tac promoter. Effect on its *in vivo* activity. J. Biol. Chem. 260, 3539–3541.

Brown, K.D., 1968. Regulation of aromatic amino acid biosynthesis *Escherichia coli* K12. Genetics 60, 31–48.

Buki, K.G., Vinh, D.Q., Horvath, I., 1985. Partial purification and some properties of tryptophan decarboxylase from a *Bacillus* strain. Acta Microbiol. Hung. 32, 65–73.

Burk, M.J., Van Dien, S., 2016. Biotechnology for chemical production: challenges and opportunities. Trends Biotechnol. 34, 187–190. doi:10.1016/j.tibtech.2015.10.007

Cameron, C., Kelly, S., Hsieh, S.-C., Murphy, M., Chen, L., Kotb, A., Peterson, J., Coyle, D., Skidmore, B., Gomes, T., Clifford, T., Wells, G., 2015. Triptans in the acute treatment of migraine: A systematic review and network meta-analysis. Headache 55, 221–235. doi:10.1111/head.12601

Cao, R., Peng, W., Wang, Z., Xu, A., 2007. Beta-carboline alkaloids: Biochemical and pharmacological functions. Curr. Med. Chem. 14, 479–500. doi:10.2174/092986707779940998

Carbonell, P., Parutto, P., Baudier, C., Junot, C., Faulon, J.L., 2014. Retropath: Automated pipeline for embedded metabolic circuits. ACS Synth. Biol. 3, 565–577. doi:10.1021/sb4001273

Chadha, N., Silakari, O., 2017. Indoles as therapeutics of interest in medicinal chemistry: Bird's eye view. Eur. J. Med. Chem. 134, 159–184. doi:10.1016/j.ejmech.2017.04.003

Chan, E.-C., Tsai, H.-L., Chen, S.-L., Mou, D.-G., 1993. Amplification of the tryptophan operon gene in *Escherichia coli* chromosome to increase L-tryptophan biosynthesis. Appl. Microbiol. Biotechnol. 40, 301–305. doi:10.1007/BF00170384

Chen, L., 2017. Rational metabolic engineering and systematic analysis of *Escherichia coli* for L-tryptophan bioproduction. Hamburg University of Technology.

Chen, L., Zeng, A.-P.P., 2017. Rational design and metabolic analysis of *Escherichia coli* for effective production of L-tryptophan at high concentration. Appl. Microbiol. Biotechnol. 101, 559–568. doi:10.1007/s00253-016-7772-5

Chen, X., Zhou, L., Tian, K., Kumar, A., Singh, S., Prior, B.A., Wang, Z., 2013. Metabolic engineering of *Escherichia coli* : A sustainable industrial platform for bio-based chemical production. Biotechnol. Adv. 31, 1200–1223. doi:10.1016/j.biotechadv.2013.02.009

Chen, Z., Wilmanns, M., Zeng, A.-P., 2010. Structural synthetic biotechnology: From molecular structure to predictable design for industrial strain development. Trends Biotechnol. 28, 534–542. doi:10.1016/j.tibtech.2010.07.004

Chen, Z., Zeng, A.-P., 2016. Protein engineering approaches to chemical biotechnology. Curr. Opin. Biotechnol. 42, 198–205. doi:10.1016/j.copbio.2016.07.007

Chen, Z., Zeng, A.-P., 2013. Protein design in systems metabolic engineering for industrial strain development. Biotechnol. J. 8, 523–533. doi:10.1002/biot.201200238

Chou, H.H., Keasling, J.D., 2012. Synthetic pathway for production of five-carbon alcohols from isopentenyl diphosphate. Appl. Environ. Microbiol. 78, 7849–7855. doi:10.1128/AEM.01175-12

Choudhuri, S., 2014. Sequence alignment and similarity searching in genomic databases, in: Bioinformatics for Beginners. Elsevier, pp. 133–155. doi:10.1016/B978-0-12-410471-6.00006-2

Christen, S., Peterhans, E., Stocker, R., 1990. Antioxidant activities of some tryptophan metabolites: Possible implication for inflammatory diseases. Proc. Natl. Acad. Sci. U.S.A. 87, 2506–2510. doi:10.1073/pnas.87.7.2506

da Luz, J.A., Hans, E., Zeng, A.-P., 2014. Automated fast filtration and on-filter quenching improve the intracellular metabolite analysis of microorganisms. Eng. Life Sci. 14, 135–142. doi:10.1002/elsc.201300099

Datsenko, K.A., Wanner, B.L., 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. U.S.A. 97, 6640–6645. doi:10.1073/pnas.120163297

Daubner, S.C., Moran, G.R., Fitzpatrick, P.F., 2002. Role of tryptophan hydroxylase phe313 in determining substrate specificity. Biochem. Biophys. Res. Commun. 292, 639–641. doi:10.1006/bbrc.2002.6719

Davidson, A.R., 2006. Multiple sequence alignment as a guideline for protein engineering strategies. Methods Mol. Biol. 340, 171–181. doi:10.1385/1-59745-116-9:171

Davies, W.D., Pittard, J., Davidson, B.E., 1985. Cloning of *aroG*, the gene coding for phospho-2-keto-3-deoxy-heptonate aldolase(phe), in *Escherichia coli* K-12, and subcloning of the *aroG* promoter and operator in a promoter-detecting plasmid. Gene 33, 323–331. doi:10.1016/0378-111990240-9

Davis, M.D., Kaufman, S., Milstien, S., 1991. Conversion of 6-substituted tetrahydropterins to 7-isomers via phenylalanine hydroxylase-generated intermediates. Proc. Natl. Acad. Sci. U.S.A.. 88, 385–389. doi:10.1073/pnas.88.2.385

de Ruyck, J., Brysbaert, G., Blossey, R., Lensink, M.F., 2016. Molecular docking as a popular tool in drug design, an *in silico* travel. Adv. Appl. Bioinforma. Chem. 9, 1–11. doi:10.2147/AABC.S105289

de Sá Alves, F.R., Barreiro, E.J., Fraga, C.A.M., Manssour Fraga, C., 2009. From nature to drug discovery: The indole scaffold as a privileged structure. Mini-Reviews Med. Chem. 9, 782–793. doi:10.2174/138955709788452649

DeFeyter, R.C., Pittard, J., 1986. Purification and properties of shikimate kinase II from *Escherichia coli* K-12. J. Bacteriol. 165, 331–333.

Do, C.B., Mahabhashyam, M.S.P., Brudno, M., Batzoglou, S., 2005. ProbCons: Probabilistic consistency-based multiple sequence alignment. Genome Res. 15, 330–340. doi:10.1101/gr.2821705

Dodge, T.C., Gerstner, J.M., 2002. Optimization of the glucose feed rate profile for the production of tryptophan from recombinant *E. coli*. J. Chem. Technol. Biotechnol. 77, 1238–1245. doi:10.1002/jctb.698

Dorn, M., e Silva, M.B., Buriol, L.S., Lamb, L.C., 2014. Three-dimensional protein structure prediction: Methods and computational strategies, Comput. Biol. Chem. 53B, 251–276. doi:10.1016/j.compbiolchem.2014.10.001

Du, J., Shao, Z., Zhao, H., 2011. Engineering microbial factories for synthesis of value-added products. J. Ind. Microbiol. Biotechnol. 38, 873–890. doi:10.1007/s10295-011-0970-3

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. doi:10.1093/nar/gkh340

Ekstro, F., Stier, G., Eaton, J.T., Sauer, U.H., Ekström, F., Stier, G., Eaton, J.T., Sauer, U.H., 2003. Crystallization and X-ray analysis of a bacterial non- haem iron-containing

phenylalanine hydroxylase from the Gram-negative opportunistic pathogen *Pseudomonas aeruginosa*. Crystallization papers. Acta Crystallogr. - Sect. D Biol. Crystallogr. 59, 1310–1312. doi:10.1107/S090744490301014X

Ely, B., Pittard, J., 1979. Aromatic amino acid biosynthesis: regulation of shikimate kinase in *Escherichia coli* K-12. J. Bacteriol. 138, 933–943.

Ensley, B.D., Ratzkin, B.J., Osslund, T.D., Simon, M.J., Wackett, L.P., Gibson, D.T., 1983. Expression of naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. Science. 222, 167–169. doi:10.1126/science.6353574

Eriksen, D.T., Lian, J., Zhao, H., 2014. Protein design for pathway engineering. J. Struct. Biol. 185, 234–242. doi:10.1016/j.jsb.2013.03.011

Erlandsen, H., Kim, J.Y., Patch, M.G., Han, A., Volner, A., Abu-Omar, M.M., Stevens, R.C., 2002. Structural comparison of bacterial and human iron-dependent phenylalanine hydroxylases: Similar fold, different stability and reaction rates. J. Mol. Biol. 320, 645–661. doi:10.1016/S0022-283600496-5

Facchini, P.J., Huber-Allanach, K.L., Tari, L.W., 2000. Plant aromatic L-amino acid decarboxylases: Evolution, biochemistry, regulation, and metabolic engineering applications. Phytochemistry 54, 121–138. doi:10.1016/S0031-942200050-9

Fang, M., Wang, T., Zhang, C., Bai, J., Zheng, X., Zhao, X., Lou, C., Xing, X.H., 2016. Intermediate-sensor assisted push-pull strategy and its application in heterologous deoxyviolacein production in *Escherichia coli*. Metab. Eng. 33, 41–51.

Ferreira, L.G., dos Santos, R.N., Oliva, G., Andricopulo, A.D., 2015. Molecular docking and structure-based drug design strategies, Molecules. 20,13384–13421. doi:10.3390/molecules 200713384

Firth, A.E., Patrick, W.M., 2005. Statistics of protein library construction. Bioinformatics 21, 3314–3315. doi:10.1093/bioinformatics/bti516

Frangatos, G., Chubb, F.L., 1959. A new synthesis of 5-hydroxytryptophan. Can. J. Chem. 37, 1374–1376.

Galperin, M.Y., Koonin, E. V, 2012. Divergence and convergence in enzyme evolution. J. Biol. Chem. 287, 21–28. doi:10.1074/jbc.R111.241976

Garst, A.D., Edwards, A.L., Batey, R.T., 2011. Riboswitches: Structures and mechanisms. Cold Spring Harb. Perspect. Biol. 3, a003533. doi:10.1101/cshperspect.a003533

Ger, Y.M., Chen, S.L., Chiang, H.J., Shiuan, D., 1994. A single Ser-180 mutation desensitizes feedback inhibition of the phenylalanine-sensitive 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthetase in *Escherichia coli*. J. Biochem. 116, 986–990.

Glasner, M.E., Gerlt, J.A., Babbitt, P.C., 2006. Evolution of enzyme superfamilies. Curr. Opin. Chem. Biol. 10, 492–497. doi:10.1016/j.cbpa.2006.08.012

Gong, F., Ito, K., Nakamura, Y., Yanofsky, C., 2001. The mechanism of tryptophan induction of tryptophanase operon expression: tryptophan inhibits release factor-mediated cleavage of TnaC-peptidyl-tRNA(Pro). Proc. Natl. Acad. Sci. U.S.A. 98, 8997–9001. doi:10.1073/pnas.171299298

Gong, F., Li, C., Xu, Y., 2013. Novel simple synthesis method of L-5-hydroxytryptophan. CN Patent App. CN 201310592943.

Greene, L.H., Lewis, T.E., Addou, S., Cuff, A., Dallman, T., Dibley, M., Redfern, O., Pearl, F., Nambudiry, R., Reid, A., Sillitoe, I., Yeats, C., Thornton, J.M., Orengo, C.A., 2007. The CATH domain structure database: New protocols and classification levels give a more comprehensive resource for exploring evolution. Nucleic Acids Res. 35, D291-7. doi:10.1093/nar/gkl959

Gu, P., Yang, F., Li, F., Liang, Q., Qi, Q., 2013. Knocking out analysis of tryptophan permeases in *Escherichia coli* for improving L-tryptophan production. Appl. Microbiol. Biotechnol. 97, 6677–6683. doi:10.1007/s00253-013-4988-5

Gu, X., 1999. Statistical methods for testing functional divergence after gene duplication. Mol. Biol. Evol. 16, 1664–1674. doi:10.1093/oxfordjournals.molbev.a026080

Gülçin, I., 2008. Measurement of antioxidant ability of melatonin and serotonin by the DMPD and CUPRAC methods as trolox equivalent. J. Enzyme Inhib. Med. Chem. 23, 871–876. doi:10.1080/14756360701626223

Hadadi, N., Hatzimanikatis, V., 2015. Design of computational retrobiosynthesis tools for the design of *de novo* synthetic pathways, Curr. Opin. Chem. Biol. 28, 99–104.

Hara, R., Kino, K., 2013. Enhanced synthesis of 5-hydroxy-l-tryptophan through tetrahydropterin regeneration. AMB Express 3, 70. doi:10.1186/2191-0855-3-70

Higgins, D.G., Sharp, P.M., 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73, 237–244. doi:10.1016/0378-111990330-7

Hogeweg, P., Hesper, B., 1984. The alignment of sets of sequences and the construction of phyletic trees: An integrated method. J. Mol. Evol. 20, 175–186. doi:10.1007/BF02257378

Hu, C., Jiang, P., Xu, J., Wu, Y., Huang, W., 2003. Mutation analysis of the feedback inhibition site of phenylalanine-sensitive 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase of *Escherichia coli*. J. Basic Microbiol. 43, 399–406. doi:10.1002/jobm.200310244

Ikeda, M., 2006. Towards bacterial strains overproducing L-tryptophan and other aromatics by metabolic engineering. Appl. Microbiol. Biotechnol. 6, 615–626. doi:10.1007/s00253-005-0252-y

Jullesson, D., David, F., Pfleger, B., Nielsen, J., 2015. Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. Biotechnol. Adv. 7, 1395–1402.

Kang, S., Kang, K., Lee, K., Back, K., Kang, A.K., Lee, A.K., Back, A.K., Kang, K., Lee, K., Back, K., 2007. Characterization of tryptamine 5-hydroxylase and serotonin synthesis in rice plants. Plant Cell Rep. 26, 2009–2015. doi:10.1007/s00299-007-0405-9

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066. doi:10.1093/nar/gkf436

Keasling, J.D., 2010. Manufacturing molecules through metabolic engineering. Science 330, 1355–1358. doi:10.1126/science.1193990

Khor, B.Y., Tye, G.J., Lim, T.S., Choong, Y.S., 2015. General overview on structure prediction of twilight-zone proteins. Theor. Biol. Med. Model. 12, 15. doi:10.1186/s12976-015-0014-1

Khoury, G.A., Smadbeck, J., Kieslich, C.A., Floudas, C.A., 2014. Protein folding and *de novo* protein design for biotechnological applications. Trends Biotechnol. 32, 99–109.

Kille, S., Acevedo-Rocha, C.G., Parra, L.P., Zhang, Z.G., Opperman, D.J., Reetz, M.T., Acevedo, J.P., 2013. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. ACS Synth. Biol. 2, 83–92. doi:10.1021/sb300037w

Kimura, M., 1968. Evolutionary rate at the molecular level. Nature. 217, 624–626. doi:10.1038/217624a0

Kino, K., Hara, R., Nozawa, A., 2009. Enhancement of l-tryptophan 5-hydroxylation activity by structure-based modification of l-phenylalanine 4-hydroxylase from *Chromobacterium violaceum*. J. Biosci. Bioeng. 108, 184–189. doi:10.1016/j.jbiosc.2009.04.002

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874. doi:10.1093/molbev/msw054

Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., Ferrin, T.E., 1982. A geometric approach to macromolecule-ligand interactions. J. Mol. Biol. 161, 269–288. doi:10.1016/0022-283690153-X

Lassila, J.K., Keeffe, J.R., Oelschlaeger, P., Mayo, S.L., 2005. Computationally designed variants of *Escherichia coli* chorismate mutase show altered catalytic activity. Protein Eng. Des. Sel. 18, 161–163. doi:10.1093/protein/gzi015

Lawley, B., Pittard, A.J., 1994. Regulation of aroL expression by TyrR protein and Trp repressor in *Escherichia coli* K-12. J. Bacteriol. 176, 6921–6930.

Lee, D., Redfern, O., Orengo, C., 2007. Predicting protein function from sequence and structure. Nat. Rev. Mol. Cell Biol. 8, 995–1005. doi:10.1038/nrm2281

Lee, F., Yanofsky, C., 1977. Transcription termination at the trp operon attenuators of *Escherichia coli* and *Salmonella typhimurium*: RNA secondary structure and regulation of termination. Proc. Natl. Acad. Sci. U.S.A. 74, 4365–9. doi:10.1073/pnas.74.10.4365

Lehning, C.E., Siedler, S., Ellabaan, M.M.H., Sommer, M.O.A., 2017. Assessing glycolytic flux alterations resulting from genetic perturbations in *E. coli* using a biosensor. Metab. Eng. 42, 194–202. doi:10.1016/j.ymben.2017.07.002

Leuschner, R.G., Heidel, M., Hammes, W.P., 1998. Histamine and tyramine degradation by food fermenting microorganisms. Int J. Food. Microbiol. 39, 1–10.

Liao, H.F., Lin, L.L., Chien, H.R., Hsu, W.H., 2001. Serine 187 is a crucial residue for allosteric regulation of *Corynebacterium glutamicum* 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase. FEMS Microbiol. Lett. 194, 59–64. doi:10.1016/S0378-109700507-3

Lin, B., Tao, Y., 2017. Whole-cell biocatalysts by design. Microb. Cell Fact. 16, 106. doi:10.1186/s12934-017-0724-7

Lin, S., Meng, X., Jiang, J., Pang, D., Jones, G., OuYang, H., Ren, L., 2012. Site-directed mutagenesis and over expression of aroG gene of *Escherichia coli* K-12. Int. J. Biol. Macromol. 51, 915–919. doi:10.1016/j.ijbiomac.2012.07.013

Lin, Y., Sun, X., Yuan, Q., Yan, Y., 2014. Engineering bacterial phenylalanine 4-hydroxylase for microbial synthesis of human neurotransmitter precursor 5-hydroxytryptophan. ACS Synth. Biol. 3, 497–505. doi:10.1021/sb5002505

Lutz, S., 2011. Beyond directed evolution - semi-rational protein engineering and design. Curr. Opin. Biotechnol. 21, 734–743. doi:10.1016/j.copbio.2010.08.011.

Lutz, S., Patrick, W.M., 2004. Novel methods for directed evolution of enzymes: Quality, not quantity. Curr. Opin. Biotechnol. 15, 291-297. doi:10.1016/j.copbio.2004.05.004

Mariani, V., Kiefer, F., Schmidt, T., Haas, J., Schwede, T., 2011. Assessment of template based protein structure predictions in CASP9. Proteins Struct. Funct. Bioinforma. 79, 37–58. doi:10.1002/prot.23177

Martí, S., Andrés, J., Moliner, V., Silla, E., Tuñón, I., Bertrán, J., 2008. Computational design of biological catalysts. Chem. Soc. Rev. 37, 2634–2643. doi:10.1039/b710705f

Marti, S., Roca, M., Andres, J., Moliner, V., Silla, E., Tunon, I., Bertran, J., 2004. Theoretical insights in enzyme catalysis. Chem. Soc. Rev. 33, 98–107. doi:10.1039/B301875J

McKinney, J.A., Turel, B., Winge, I., Knappskog, P.M., Haavik, J., 2009. Functional properties of missense variants of human tryptophan hydroxylase 2. Hum. Mutat. 30, 787–794. doi:10.1002/humu.20956

McKinney, J., Teigen, K., Frøystein, N.A., Salaün, C., Knappskog, P.M., Haavik, J., Martinez, A., 2001. Conformation of the substrate and pterin cofactor bound to human tryptophan hydroxylase. Important role of Phe313 in substrate specificity. Biochemistry 40, 15591–15601. doi:10.1021/bi015722x

Meng, X.-Y., Zhang, H.-X., Mezei, M., Cui, M., 2011. Molecular docking: a powerful approach for structure-based drug discovery. Curr. Comput. Aided. Drug Des. 7, 146–157. doi:10.1016/j.biotechadv.2011.08.021

Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T., Tramontano, A., 2016. Critical assessment of methods of protein structure prediction: Progress and new directions in round XI. Proteins Struct. Funct. Bioinforma. 84, 4–14. doi:10.1002/prot.25064

Murdock, D., Ensley, B.D., Serdar, C., Thalen, M., 1993. Construction of metabolic operons catalyzing the *de novo* biosynthesis of indigo in *Escherichia coli*. Nat. Biotechnol. 11, 381–386. doi:10.1038/nbt0393-381

Mustafi, N., Grünberger, A., Kohlheyer, D., Bott, M., Frunzke, J., 2012. The development and application of a single-cell biosensor for the detection of l-methionine and branched-chain amino acids. Metab. Eng. 14, 449–457. doi:10.1016/j.ymben.2012.02.002

Mustafi, N., Grünberger, A., Mahr, R., Helfrich, S., Nöh, K., Blombach, B., Kohlheyer, D., Frunzke, J., 2014. Application of a genetically encoded biosensor for live cell imag-

ing of L-valine production in pyruvate dehydrogenase complex-deficient *Corynebacterium glutamicum* strains. PLoS One 9, e85731. doi:10.1371/journal.pone.0085731

Nakazawa, H., Kumagai, H., Yamada, H., 1974. Constitutive aromatic L-amino acid decarboxylase from *Micrococcus percitreus*. Biochem. Biophys. Res. Commun. 61, 75–82.

Nagaraja, P., Yathirajan, H.S., Vasantha, R.A., 2003. Highly sensitive reaction of tryptophan with p-phenylenediamine. Anal. Biochem. 312, 157–161. doi:10.1016/S0003-269700504-3

Nakagawa, A., Matsumura, E., Koyanagi, T., Katayama, T., Kawano, N., Yoshimatsu, K., Yamamoto, K., Kumagai, H., Sato, F., Minami, H., 2016. Total biosynthesis of opiates by stepwise fermentation using engineered *Escherichia coli*. Nat. Commun. 7, 10390. doi:10.1038/ncomms10390

Nielsen, J., Fussenegger, M., Keasling, J., Lee, S.Y., Liao, J.C., Prather, K., Palsson, B., 2014. Engineering synergy in biotechnology. Nat. Chem. Biol. 10, 319–322. doi:10.1038/nchembio.1519

Noé, W., Mollenschott, C., Berlin, J., 1984. Tryptophan decarboxylase from *Catharanthus roseus* cell suspension cultures: Purification, molecular and kinetic data of the homogenous protein. Plant Mol. Biol. 3, 281–288. doi:10.1007/BF00017782

Notredame, C., Higgins, D.G., Heringa, J., 2000. T-coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302, 205–217. doi:10.1006/jmbi.2000.4042

Nudler, E., Mironov, A.S., 2004. The riboswitch control of bacterial metabolism. Trends Biochem. Sci. 29, 11–17. doi:10.1016/j.tibs.2003.11.004

Ovchinnikov, S., Kim, D.E., Wang, R.Y.R., Liu, Y., Dimaio, F., Baker, D., 2016. Improved *de novo* structure prediction in CASP11 by incorporating coevolution information into Rosetta. Proteins Struct. Funct. Bioinforma. 84, 67–75. doi:10.1002/prot.24974

Pabst, M., Kuhn, J., Somerville, R., 1973. Feedback regulation in the anthranilate aggregate from wild type and mutant strains of *Escherichia coli*. J. Biol. Chem. 248, 901–914.

Park, M., Kang, K., Park, S., Back, K., 2008. Conversion of 5-Hydroxytryptophan into Serotonin by Tryptophan Decarboxylase in Plants, *Escherichia coli*, and Yeast. Biosci. Biotechnol. Biochem. 72, 2456–2458. doi:10.1271/bbb.80220

Park, S., Kang, K., Lee, S.W., Ahn, M.J., Bae, J.M., Back, K., 2011. Production of serotonin by dual expression of tryptophan decarboxylase and tryptamine 5-hydroxylase in

Escherichia coli. Appl. Microbiol. Biotechnol. 89, 1387–1394. doi:10.1007/s00253-010-2994-4

Park, S., Kim, Y.S., Rupasinghe, S.G., Schuler, M.A., Back, K., 2013. Rice P450 reductases differentially affect P450-mediated metabolism in bacterial expression systems. Bioprocess Biosyst. Eng. 36, 325–331. doi:10.1007/s00449-012-0787-0

Parra, L.P., Agudo, R., Reetz, M.T., 2013. Directed evolution by using iterative saturation mutagenesis based on multiresidue sites. ChemBioChem 14, 2301–2309. doi:10.1002/cbic.201300486

Pasini, M., Fernández-Castané, A., Jaramillo, A., de Mas, C., Caminal, G., Ferrer, P., 2016. Using promoter libraries to reduce metabolic burden due to plasmid-encoded proteins in recombinant *Escherichia coli*. N. Biotechnol. 33, 78–90. doi:10.1016/j.nbt.2015.08.003

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera - A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. doi:10.1002/jcc.20084

Phillips, A., Janies, D., Wheeler, W., 2000. Multiple sequence alignment in phylogenetic analysis. Mol. Phylogenet. Evol. 16, 317–330. doi:10.1006/mpev.2000.0785

Picataggio, S., 2009. Potential impact of synthetic biology on the development of microbial systems for the production of renewable fuels and chemicals. Curr. Opin. Biotechnol. 3, 325–329. doi:10.1016/j.copbio.2009.04.003

Porter, E.B., Polaski, J.T., Morck, M.M., Batey, R.T., 2017. Recurrent RNA motifs as scaffolds for genetically encodable small molecule biosensors. Nat. Chem. Biol. 13, 295–301. doi:10.1038/nchembio.2278

Procter, J.B., Thompson, J., Letunic, I., Creevey, C., Jossinet, F., Barton, G.J., 2010. Visualization of multiple alignments, phylogenies and gene family evolution. Nat. Methods 7, S16–S25. doi:10.1038/nmeth.1434

Quin, M.B., Schmidt-Dannert, C., 2011. Engineering of biocatalysts: From evolution to creation. ACS Catal. 1, 1017–1021. doi:10.1021/cs200217t

Quintana, M.G., Didion, C., Dalton, H., 1997. Colorimetric method for a rapid detection of oxygenated aromatic biotransformation products. Biotechnology Tech. 11, 585–587. doi:10.1023/a:1018499024466

Reetz, M.T., Carballeira, J.D., 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. Nat. Protoc. 2, 891–903. doi:10.1038/nprot.2007.72

Rodrigues, A.L., Trachtmann, N., Becker, J., Lohanatha, A.F., Blotenberg, J., Bolten, C.J., Korneli, C., de Souza Lima, A.O., Porto, L.M., Sprenger, G.A., Wittmann, C., 2013. Systems metabolic engineering of *Escherichia coli* for production of the antitumor drugs violacein and deoxyviolacein. Metab. Eng. 20, 29–41. doi:10.1016/j.ymben.2013.08.004

Rodriguez, A., Martínez, J.A., Flores, N., Escalante, A., Gosset, G., Bolivar, F., Martnez, J.A., Flores, N., Escalante, A., Gosset, G., Bolivar, F., Martínez, J. a, Flores, N., Escalante, A., Gosset, G., Bolivar, F., 2014. Engineering *Escherichia coli* to overproduce aromatic amino acids and derived compounds. Microb. Cell Fact. 13, 126. doi:10.1186/s12934-014-0126-z

Romasi, E.F., Lee, J., 2013. Development of indole-3-acetic acid-producing *Escherichia coli* by functional expression of IpdC, AspC, and Iad1. J. Microbiol. Biotechnol. 23, 1726–1736. doi:10.4014/jmb.1308.08082

Rose, J.K., Squires, C.L., Yanofsky, C., Yang, H.L., Zubay, G., 1973. Regulation of *in vitro* transcription of the tryptophan operon by purified RNA polymerase in the presence of partially purified repressor and tryptophan. Nat New Biol 245, 133–137. doi:10.1038/10.1038/newbio245133a0

Roth, T., Nir, T., Zisapel, N., 2015. Prolonged release melatonin for improving sleep in totally blind subjects: A pilot placebo-controlled multicenter trial. Nat. Sci. Sleep 7, 13–23. doi:10.2147/NSS.S71838

Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. 5, 725–738. doi:10.1038/nprot.2010.5

Sadowski, M.I., Jones, D.T., 2009. The sequence–structure relationship and protein function prediction. Curr. Opin. Struct. Biol. 19, 357–362. doi:10.1016/j.sbi.2009.03.008

Sambrook, J., Russell, D.W., 2000. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press. pp 2028.

Santillan, M., Mackey, M.C., 2001. Dynamic regulation of the tryptophan operon: A modeling study and comparison with experimental data. Proc. Natl. Acad. Sci. 98, 1364–1369. doi:10.1073/pnas.98.4.1364

Satoh, Y., Tajima, K., Munekata, M., Keasling, J.D., Lee, T.S., 2012. Engineering of l-tyrosine oxidation in *Escherichia coli* and microbial production of hydroxytyrosol. Metab. Eng. 14, 603–610. doi:10.1016/j.ymben.2012.08.002

Shapovalov, M. V., Dunbrack, R.L., 2011. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. Structure 19, 844–858. doi:10.1016/j.str.2011.03.019

Shen, T., Liu, Q., Xie, X., Xu, Q., Chen, N., 2012. Improved production of tryptophan in genetically engineered *Escherichia coli* with TktA and PpsA overexpression. J. Biomed. Biotechnol. 2012, 605219.doi:10.1155/2012/605219

Sheppard, D.W., Lipkin, M.J., Harris, C.J., Catana, C., Stouten, P.F.W., 2014. Strategies for small molecule library design. Curr. Pharm. Des. 20, 3314–3322. doi:10.2174/13816128113199990602

Staiano, M., Bazzicalupo, P., Rossi, M., D'Auria, S., 2005. Glucose biosensors as models for the development of advanced protein-based biosensors. Mol. Biosyst. 1, 354–362. doi:10.1039/B513385H

Stephanopoulos, G., 2012. Synthetic biology and metabolic engineering. ACS Synth. Biol. 1, 514–525. doi:10.1021/sb300094q

Sun, X., Lin, Y., Yuan, Q., Yan, Y., 2015. Precursor-directed biosynthesis of 5-Hydroxytryptophan using metabolically engineered *E. coli*. ACS Synth. Biol. 4, 554–558. doi:10.1021/sb500303q

Swiderek, K., Tuñón, I., Moliner, V., Bertran, J., 2015. Computational strategies for the design of new enzymatic functions. Arch. Biochem. Biophys. 582, 68–79. doi:10.1016/j.abb.2015.03.013

Tang, W.L., Zhao, H., 2009. Industrial biotechnology: Tools and applications. Biotechnol. J. 4, 1725–1739. doi:10.1002/biot.200900127

Taylor, R.D., Jewsbury, P.J., Essex, J.W., 2002. A review of protein-small molecule docking methods. J. Comput. Aided. Mol. Des. 16, 151–166. doi:10.1023/A:1020155510718

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680. doi:10.1093/nar/22.22.4673

Tizei, P.A.G.A.G., Csibra, E., Torres, L., Pinheiro, V.B.B., 2016. Selection platforms for directed evolution in synthetic biology. Biochem. Soc. Trans. 44LP, 1165–1175. doi:10.1042/BST20160076

Todd, A.E., Orengo, C.A., Thornton, J.M., 2001. Evolution of function in protein superfamilies, from a structural perspective. J. Mol. Biol. 307, 1113–1143. doi:http://dx.doi.org/10.1006/jmbi.2001.4513

Tokuriki, N., Tawfik, D.S., 2009. Protein dynamism and evolvability. Science. 324, 203–207.

Tramontano, A., Büssow, K., 2006. Protein structure prediction: Concepts and applications. Anal. Bioanal. Chem. 386, 1579–1580. doi:10.1007/s00216-006-0812-8

Tribe, D.E., Pittard, J., 1979. Hyperproduction of tryptophan by *Escherichia coli*: genetic manipulation of the pathways leading to tryptophan formation. Appl. Environ. Microbiol. 38, 181–190.

Trott, O., Olson, A.J., 2010. 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. 31, 455–461. doi:10.1002/jcc.21334

Turner, E.H., Loftis, J.M., Blackwell, A.D., 2006. Serotonin a la carte: Supplementation with the serotonin precursor 5-hydroxytryptophan. Pharmacol. Ther. 109, 325–338. doi:10.1016/j.pharmthera.2005.06.004

Umbarger, H.E., 1978. Amino acid biosynthesis and its regulation. Annu. Rev. Biochem. 47, 532–606. doi:10.1146/annurev.bi.47.070178.002533

Verbeek, M.M., Geurtz, P.B.H., Willemsen, M.A.A.P., Wevers, R.A., 2007. Aromatic l-amino acid decarboxylase enzyme activity in deficient patients and heterozygotes. Mol. Genet. Metab. 90, 363–369. doi:10.1016/j.ymgme.2006.12.001

Wallace, B.J., Pittard, J., 1969. Regulation of 3-deoxy-D-arabino-heptulosonic 7-phosphate acid synthetase activity in relation to the synthesis of the aromatic vitamins in *Escherichia coli* K-12. J. Bacteriol. 99, 707–712.

Wang, J.J., Cheng, L.K., Wang, J.J., Liu, Q., Shen, T., Chen, N., 2013. Genetic engineering of *Escherichia coli* to enhance production of L-tryptophan. Appl. Microbiol. Biotechnol. 97, 7587–7596. doi:10.1007/s00253-013-5026-3

Wang, L., Erlandsen, H., Haavik, J., Knappskog, P.M., Stevens, R.C., 2002. Three-dimensional structure of human tryptophan hydroxylase and its implications for the biosynthesis of the neurotransmitters serotonin and melatonin. Biochemistry 41, 12569–12574. doi:10.1021/bi026561f

Webb, B., Sali, A., 2014. Protein structure modeling with MODELLER. Methods Mol. Biol. 1137, 1–15. doi:10.1007/978-1-4939-0366-5_1

Wong, T.S., Roccatano, D., Schwaneberg, U., 2007. Steering directed protein evolution: Strategies to manage combinatorial complexity of mutant libraries. Environ. Microbiol. 9, 2645–2659 doi:10.1111/j.1462-2920.2007.01411.x

Xu, D., Xu, Y., Uberbacher, E.C., 2000. Computational tools for protein modeling. Curr. Protein Pept. Sci. 1, 1–21. doi:10.2174/1389203003381469

Xu, P., Hua, D., Ma, C., 2007. Microbial transformation of propenylbenzenes for natural flavour production. Trends Biotechnol. 25, 571–576. doi:10.1016/j.tibtech.2007.08.011

Yanofsky, C., 2007. RNA-based regulation of genes of tryptophan synthesis and degradation, in bacteria. RNA 13, 1141–1154. doi:10.1261/rna.620507

Yu, K., Liu, C., Kim, B.-G., Lee, D.-Y., 2015. Synthetic fusion protein design and applications. Biotechnol. Adv. 33, 155–164. doi:10.1016/j.biotechadv.2014.11.005

Zhang, K., Sawaya, M.R., Eisenberg, D.S., Liao, J.C., 2008. Expanding metabolism for biosynthesis of nonnatural alcohols. Proc. Natl. Acad. Sci. U.S.A. 105, 20653–20658. doi:10.1073/pnas.0807157106

Zhang, W., Yang, J., He, B., Walker, S.E., Zhang, H., Govindarajoo, B., Virtanen, J., Xue, Z., Shen, H. Bin, Zhang, Y., 2015. Integration of QUARK and I-TASSER for *ab initio* Protein Structure Prediction in CASP11. Proteins Struct. Funct. Bioinforma. 84, 76–86. doi:10.1002/prot.24930

Zhao, C., Cheng, L.K., Wang, J., Shen, Z., Chen, N., 2016. Impact of deletion of the genes encoding acetate kinase on production of L-tryptophan by *Escherichia coli*. Ann. Microbiol. 66, 261–269. doi:10.1007/s13213-015-1103-4

Zhao, Z., Chen, S., Wu, D., Wu, J., Chen, J., 2012. Effect of gene knockouts of l-tryptophan uptake system on the production of l-tryptophan in *Escherichia coli*. Process Biochem. 47, 340–344. doi:10.1016/j.procbio.2011.11.009

Zhao, Z.J., Zou, C., Zhu, Y.X., Dai, J., Chen, S., Wu, D., Wu, J., Chen, J., 2011. Development of l-tryptophan production strains by defined genetic modification in *Escherichia coli*. J. Ind. Microbiol. Biotechnol. 38, 1921–1929. doi:10.1007/s10295-011-0978-8

Zhou, L.B., Zeng, A.P., 2015. Engineering a lysine-ON riboswitch for metabolic control of lysine production in *Corynebacterium glutamicum*. ACS Synth. Biol. 4, 1335–1340. doi:10.1021/acssynbio.5b00075

Zhu, Y., Hua, Y., Zhang, B., Sun, L., Li, W., Kong, X., Hong, J., 2017. Metabolic engineering of indole pyruvic acid biosynthesis in *Escherichia coli* with tdiD. Microb. Cell Fact. 16, 2. doi:10.1186/s12934-016-0620-6

Zurawski, G., Elseviers, D., Stauffer, G. V, Yanofsky, C., 1978. Translational control of transcription termination at the attenuator of the *Escherichia coli* tryptophan operon. Proc. Natl. Acad. Sci. U. S. A. 75, 5988–5992. doi:10.1073/pnas.75.12.5988

Zurawski, G., Gunsalus, R.P., Brown, K.D., Yanofsky, C., 1981. Structure and regulation of aroH, the structural gene for the tryptophan-repressible 3-deoxy-d-arabino-heptulosonic acid-7-phosphate synthetase of *Escherichia coli*. J. Mol. Biol. 145, 47–73. doi:10.1016/0022-28360334-X

Curriculum Vitae

Surname, first name:	Mora Villalobos, Jose Aníbal
Date of birth:	30. October 1981
Place of birth:	Viçosa, Minas Gerais, Brazil
Nationality:	Costa Rican
Home address:	Hannoversche str. 137a
	37077 Göttingen, Germany

Education:

07.2013 – present	Graduate student in the Institute of Bioprocess and Biosystems Engineering (IBB) at the Technical University of Hamburg (TUHH)
	Thesis: "Protein engineering and synthetic pathways in *Escherichia coli* for effective production of 5-hydroxytryptophan and serotonin."
01.2004 – 12.2006	Master of Science. Master in Biology with emphasis in Genetics and Molecular Biology
	University of Costa Rica, Costa Rica
01.2000 – 12.2003	Bachelor in Biology
	University of Costa Rica, Costa Rica
01. 1995 – 12.1999	Highschool Diplom in Medium Education
	Saint John High School, Costa Rica

Professional experience:

07.2009 – 06.2013	Associate researcher – Molecular Biology National Center of Biotechnological Innovation (CENIBiot)
01.2007 – 06.2009	Researcher – R&D Commercial Distributor Agrotico S.A.
01.2004 – 12.2006	Laboratory Assistant (Scholar) Cell and Molecular Biology Research Center (CIBCM-UCR)
01.2003 – 12-2003	Lecture Assistant (Scholar) Genetics for agronomist and Genetics for microbiology

Publications:

Mora-Villalobos, J.A. and A.P. Zeng. 2017. Synthetic pathways and processes for the effective production of 5-hydroxytryptophan and serotonin from glucose in *Escherichia coli*. J. Biol. Eng. Accepted

Mora-Villalobos, J.A. and A.P. Zeng. 2017. Protein and pathway engineering for the biosynthesis of 5-hydroxytryptophan in *Escherichia coli*. Eng. Life Sci. 17, 892-899.

Avandeno R., E. Garcia Diaz, M. Valdez Melara, N. Chavez Solano, J.A. Mora-Villalobos, F. Aguilar Cascante, B. Williamson Benavides and L.Y. Solis Ramos. 2015. Genetic diversity analysis of *Jatropha* species from Costa Rica using AFLP markers. American Journal of Plant Sciences, 6:2426- 2438.

Funding:

06.2013 – 03.2017	German Academic Exchange Service (Deutscher Akademischer Austauschdienst – DAAD). Research grant for doctoral candidates and young academics and scientists.
09.2011	Cochran Fellowship Program (CFP) for short-term training opportunities. Workshop at Michigan State University, USA.
05.2006	Generation Challenge Program (GCP) for general meeting research platform for the exchange of young scientists. Workshop at Tropical Agriculture Research Center (CIAT), Colombia